GEOGRAPHIC BASE FILES:
Applications in the Integration and Extraction of Data from Diverse Sources

Patrick E. Mantey, Eric D. Carlson, IBM Research Laboratory, San Jose, California

Abstract
This paper addresses the development of integrated municipal data bases, with consideration given to political realities and to the sources of data now available in municipalities. First, the potential users and potential uses of a municipal data base are discussed, and an information system which would serve these users and uses is considered. Next, the "current" status of data bases in municipalities is reviewed and it is concluded that there is a large quantity of data available in many municipalities, but that integrated data bases for supporting an information system are not usually a reality. The problem of building an integrated data base from the variety of data sources presented by local agencies is then addressed. A central ingredient for an integrated data base is the Geographic Base File (GBF) which permits the construction of extracted data files from these multiple sources to support information system applications. The concept of extraction, for the integration of diverse source files via geographic references, is developed (and a prototype implementation is described in the Appendix). Using the GBF, and source data from various municipal functions, extracted data bases can be rapidly built to serve a variety of applications of an information system in the decision-making situations of municipalities.

I. APPLICATIONS OF A MUNICIPAL DATA BASE

Municipal governments are, in essence, created to deliver services to a geographical area. There is an unusual variety, in comparison to private industry, of services offered. Local government is often structured (or fractured) along functional lines into special districts, as well as by geography. Such structuring has precluded concentration of power, but it has increased the complexity of planning, resource allocation, or management.

Many of the problems in municipal government require decisions which are not routine. Rather, they require the professional insight and judgment of human decision makers who consider the specific conditions of each problem. Ideally, this insight and judgment would be aided and guided by appropriate information derived from a comprehensive data base. This is the objective of a municipal information system: to facilitate effective analysis and solution of specific problems by supporting human decision makers with data resources and analysis functions in readily usable form. Because a municipality provides services to a geographical area, much of the data relevant to decision making or problem solving in local agencies will have geographical attributes, and can be given spatial interpretation via maps. A key attribute of a municipal information system is the capability for displaying information in the form of maps. Another requirement for such systems to be effective is that they support ready use by decision makers who know very little about computers. The system must help the decision makers develop their precise objectives or decision criteria for solving a problem. The solution process in such decision making requires exploratory analysis, selection of relevant data, and meaningful data presentation in an interactive environment. A system which provides such capabilities, called GADS (Geo-data Analysis and Display System) has been developed and evaluated in several applications, such as police manpower allocation and analysis of urban development policies [1-4]. The evaluations of GADS indicate that interactive analysis and display systems have a great potential in the operations, management, and planning of municipalities.

As an example, consider a municipality which maintains a computerized property information file (via the tax assessor function). Such a file would have data on each parcel, possibly including:

- address
- owner
- zoning
- improvements
    - date constructed
    - type construction
    - size (area)
    - current use
- area
- centroid
- assessed value

If this data were accessed via an interactive information system, a decision maker could readily obtain, for example, the address and assessed value of all residences constructed between 1960 and 1962 and having floor area between 1600 and 1800 square feet, on lots with 6800 to 7200 square feet. If the user of the system were a real-estate appraiser, this information in tabular form might be of value in determining if a particular home is fairly appraised (Figure 1). A histogram giving the distribution of assessed value of these homes would provide additional insight

(Figure 2) and a map relating the average value of such homes in the city's planning areas to the city-wide average (Figure 3) would provide the appraiser with information in a spatial framework.

If the user of the information system were an assessor concerned with determining the neighborhoods which could be considered equivalent for purposes of computer-aided appraisal, additional data would be required. If recent sales records are the basis for calibrating the assessment model, it may be found that the assessor's data alone cannot be used to model variations in selling price of houses fitting the description above. Showing on a map the mean and variance of the selling price of such houses by neighborhood will offer the assessor a visual means for examining the quality of the assessment model. The display may cause the assessor to consider other factors to explain the variations; e.g. crime rate, level of public facilities and services (such as the influence of an adjacent regional park [5]) or the influence of other near-by land uses.

In making decisions related to residential zonings, pertinent questions and information displays would relate to adjacent land use, the effect that the proposed development would have on the mix of housing stock available in the community, or to the effect these new residents would make on the per-capita park acerage in the area. The school officials (and in some areas the local zoning authority) need to evaluate the impact such a development will have on the existing school facilities. Each of these questions, and many others, are of interest to different decision makers involved in the area of community development.

An example of the use of a municipal data base in resource allocation could relate to finding the best approach to the reduction of burglaries in a residential neighborhood. One group may advocate better street lighting. If the queries relating to these burglaries show they are day-time crimes, another approach probably offers greater promise. Certainly the level of police patrols would be examined. If the area is found to have a high percentage of two wage-earner families, and the majority of burglaries are during the week, then some suspicion will fall on the children. If a large number of burglaries occur on school days, it may appear that the school-age children are not the perpetrators. However, if the data base contains school information so that it can be determined that a school adjacent to the neighborhood has flexible scheduling and that school children are free to roam about off school grounds, the source of the trouble may have been identified.

As a last example, consider the application for a building permit for an automobile service station. With the recent wave of service station abandonments, many municipalities are reluctant to grant such permits. Similarly, financial institutions are probably wary about loaning money for such ventures. For all concerned, and for the public good, a careful analysis of such a proposal is required. Factors of interest would include location and number of existing stations, traffic access and traffic patterns at the proposed site, and an estimate of the automobile ownership and disposable income in the surrounding areas.

In these examples it has been assumed that the information system has access to a comprehensive municipal data base. However, such data bases are a rarity today. In the next section the current status of municipal data bases is discussed, and in Section III an approach toward the provision of integrated data is offered.

II. CURRENT STATUS OF MUNICIPAL DATA BASES

The importance of a comprehensive integrated data base to support decision-making in municipal government has been widely recognized. There have been several different approaches taken to the development of comprehensive municipal data bases.

One approach, which was popular in the 1960's, was the development of a comprehensive "data bank". This data bank was usually generated as a special collection or census, and often was carried out and funded as part of a comprehensive transportation and land use planning study. As a part of such studies, detailed field surveys of land use (at the parcel level) were conducted, and survey data also was gathered on employment, income distribution, and commercial activity (Figure 4). The data acquisition consumed a major portion of the study resources, and did not provide any information relating to many municipal services (e.g. public safety). Although accurate data was often gathered in the development of these data banks, their value was short-lived because they were at best a snap-shot of the state of a very dynamic system, and no means were provided for updating and extending these data banks. At about the same time, the computer was becoming a tool in the operations of local governments. The application of computers by municipalities beginning in the 1960's can be characterized as a function-by-function approach, with data processing introduced into tasks involving high-volume routine transactions. The computer is generally utilized in those functions which have previously been computerized in private industry: payroll, accounting, billing, budget status reporting, personnel records, etc. Also, the 1960's saw wide-spread use of computers in the processes associated with elections and in operations of law enforcement agencies. Usually these applications were isolated from each other, and no attempts were made to make this information available for use by other municipal functions.

The real property assessment function of local government in the late 1960's began to recognize the potential of computer applications. The court decisions in various locales requiring property to be appraised at current market value placed a significantly increased work load on assessment officials. In numerous areas, the assessor has turned to computer-aided appraisal to meet these demands. If a model is to be built and calibrated to reliably estimate fair-market value of residential properties, a comprehensive real property data base is a must. These data bases were constructed, often by computerizing the data contained on the assessor's file cards which were maintained on each parcel. Some property data bases, besides the usual references to assessor map, book and page and to situs address, also added geographic data such as a "centroid". Computerized real property systems required the field acquisition of very comprehensive data. A work sheet for field surveys by appraisers in a California county is shown in Figure 5. Note the detailed land and building attributes used in this system. There was a significant increase in the amount of data used and required by the introduction of this computerized appraisal system, but much of the data on this sheet would not change from year-to-year, and the appraiser in the field would only need to correct those data items which had changed.

An additional requirement for computer-aided appraisal is data relating to current selling price of residential properties. This data provides the calibration information for the regression models, and is available, depending upon the state or local laws, from the registrar of deeds, from the collector of transfer taxes, or from title companies. Assessors in some locales obtain this data by questionnaires which buyers are required by law to complete and return (Figure 6). These sources and/or others also can be used to obtain financial data (e.g. mortgage terms) for the property. Clearly this data base is an integral part of a municipal data base for applications such as illustrated by the examples in Section I.

Another approach related to the development of municipal data bases is characterized by the USAC projects [6], particularly those of Charlotte, North Carolina, and Wichita Falls, Texas. These cities were funded by the Federal USAC project to build Integrated Municipal Information Systems, (IMIS). The concepts of IMIS are [6]:

"(1) Integrated data processing systems should inter-relate municipal processes.
(2) A fundamental analysis of municipal operations and identification of related data processing components is a precondition to the effective use of computers.
(3) A systems approach is required throughout the development process.
(4) The automation of municipal operations must exploit the full range of computer technology.
(5) Automation of routine municipal processes is a fundamental condition to the realization of an IMIS.
(6) An IMIS views the municipality as a basic building block for intergovernmental information systems.
(7) Municipal information systems are by-products of computer-driven, operations-based systems.

(8) Adequately designed data processing systems can be transferred from one municipality to another.

(9) The integrated approach to municipal systems development must proceed on the basis of a plan within which incremental installation may be achieved in accordance with the priorities and resources of any particular city."

The USAC efforts involve city governments, and were the consequence of studies such as the IBM/New Haven project [7] and the USC/Burbank project [8]. These groups sought to develop a methodology, via a "systems approach", for the application of computers by municipalities [6] but did not result in system implementation of integrated municipal information systems.

The USAC approach wisely focused on operational sources to provide the current data required for municipal decision-making. In the implementations, which are still in progress, the cities have concentrated on building up operational uses of computers, and on implementing these applications on a central computer under an integrated data-base management system. The value of computers to these operational functions has been confirmed [9] but the applications of IMIS in the areas of management decision making are still to be demonstrated. One of the difficulties that must be overcome in providing a comprehensive municipal data base, (even in cities with a fully integrated and operational IMIS constructed according to the USAC philosophy), is that complete integration, where all municipal functions use the same computer and data base management system, is not a likely prospect with current local governmental structure and with the limited resources of local governments. For example the data pertinent to decision making in a city may be gathered by another agency, such as the tax assessor (and conversely), and may reside on different computers, under different data management schemes and in different file formats. In addition, problems of data security, compatibility of files, and high processing costs may make complete integration unrealistic for many municipalities.

Special data collections, such as the U.S. Census, and data available from state sources, must also be readily incorporated into a municipal data base. With census data gathered according to blocks, block groups and census tracts, with assessors property data coded according to assessor map, book and page, with public works data in state-plane coordinates, and school data gathered by school attendance area, the building and maintenance of a truly integrated municipal data base presents a formidable task.

III. APPROACHES PERMITTING DEVELOPMENT OF INTEGRATED FILES

A completely integrated data base would have all data relating to any functions of a municipality residing on the same computer system, under the same data management system and organized and indexed to facilitate correlation. This ideal is not attainable, given present organizational structures and computing capabilities in most municipalities. However, if the computerized files of various municipal functions are "properly structured", it will be possible to achieve the same benefits as if there existed a completely integrated municipal data base. In addition, such an approach will not require re-implementation of current applications, but rather leaves the application data base in the control of the function responsible for its primary maintenance and use.

The approach taken is to make use of data, when data files are "properly structured", to develop effectively the results as if there existed a completely integrated data base without requiring that complete integration take place. This should not be construed as an argument <u>against</u> integration. If integration is politically and technically possible, provides required data security, and is economically attractive, it should be implemented. Even with an integrated data base, there will always be decisions which require different groupings of data than those supported by the integrated data base. (There will also remain, in practice, data sources which cannot be integrated.) So, the problem of providing a comprehensive data base from multiple data sources is unavoidable and is not completely solved by an "integrated" data base.

The "proper structuring" required to make data integration possible can be illustrated by example. If one is interested in information about burglaries, and wishes to relate this information to neighborhood conditions, data sources could include police dispatch files, criminal justice arrest files, census data and tax assessor files. Suppose the police dispatch data is used for burglary incidence, and that such data is available in terms of police beats, e.g. the number of burglaries in each beat for each day. If one wishes to use census data for socio-economic information, and if the census tracts and beats have few common boundaries, no small area information is obtainable relating these data sources. Alternatively, if the police dispatch data is captured by the street address of the call, and if a directory exists for the city which will permit identification of the census tract for each street address, then burglaries and socio-economic data can be related at the census tract level.

"Proper structuring" of data files only has meaning with respect to potential uses of the data (i.e. data files are not an end in themselves). If the objective is to offer data to support decision making in a wide range of problem areas, then the data files must be as detailed as possible, within the constraints of economics, privacy and security. The detailed data can then make possible the development of the widest variety of data subsets and aggregations, and is more likely to permit development of the required set of integrated data for a particular decision-making context. An additional requirement is the existence of data elements in each file which will facilitate relating the data to that from different files. (In this paper, geographical references will be singled out as data elements serving this function in municipal files. Common references to account numbers, project numbers or personnel identifiers are other examples of data elements permitting the relating of data from different source files.) A set of files will be called "properly structured" if they contain information permitting the relating of data from different source files so that integrated subsets of data at the appropriate level of detail can be developed to support the requirements of problem solvers.

Because municipal government is a service delivery function, mutual references to geography can often be used to relate data from the diverse files available in municipalities. A powerful file in facilitating the relating of data, based on these common geographical references, is a Geographic Base File (GBF). Functionally, the GBF contains data to support the relating of data from other files to geographical location and also the display of this data on a map. The creation of a GBF for a municipality is a key requirement in the development of a municipal data base from source files. Several different approaches have been taken.

The simplest GBF is a file sometimes called a Property Location Index (PLI) which contains a list of the valid addresses in the municipality and an x,y coordinate for each. This approach is the one used in Lane County, Oregon, and by the Assessor in Santa Clara County , California. To make this more useful, a list of public place and street intersections and their x,y coordinates is appended. With such a GBF it is then possible to automatically convert addresses (in the police call file for example) to x,y coordinates. If the GBF also contains the police beat, census tract, and municipality for each address, then it is very simple computationally to count the number of calls in each beat. Evaluation of calls by census tract would also permit consideration of socio-economic data with the crime data (of course, police officers could also encode calls by beat and census tract, but this approach is liable to significant errors and seems to be a poor use of police manpower).

The most detailed GBF's contain digitized land parcel boundaries, easement locations, building outlines, utility placements, and even topographic information, along with street address information on all parcels and names of all public lands and buildings. This GBF is at the level of detail of surveyor's data, and is suitable for engineering applications and detailed map building. Ottawa, Canada's National Capital Commission [11] has pioneered in the development of this kind of GBF.

The most common GBF at the present time is the result of work by the U.S. Census Bureau in conjunction with the 1970 Census. Using the Metropolitan Map Series, a massive feature labeling and digitization was performed for 200 major metropolitan areas in the United States. The resulting computerized maps were called the DIME (Dual Independent Map Encoding) files [12]. Each entry (record) in the DIME file represents a line segment (a portion of a street segment, railroad, creek, city limit, etc.). Figure 7 shows a sample record and the map data from which the record is derived. The segment has a "From" node and a "To" node, as well as a Left and Right side. Thus, each entry describes its two ends and its two sides. The description of the nodes includes x,y coordinates, produced by the Census Bureau's map digitization. The nodes are labelled as falling on a particular map of the series, and given a sequential number within map and census tract. The other data on each entry is feature identification for the segment and its sides. Each feature is identified by a prefix, name, suffix, type (e.g., North Army Southwest Street); only name and type are required (e.g., Coyote Creek). The description of a side actually describes the adjacent land. The census tract number, the block number, and the place (city) code are included for each side. High and low address ranges for the street segment sides are also given. The records are ordered by feature name and low address (features without addresses have no secondary ordering). Administrative overlays (e.g. beats, census tracts) can be readily defined in terms of segments of this file. Used in combination with "point-in polygon routines", these computerized overlays facilitate development of counts of events in areas of any specified overlay map.

As in the development of any large machine readable file, high startup costs, data errors, and poor standardization have hindered development of GBF's. But the key problem in the development and use of a GBF is editing (corrections and additions).* Because of the startup cost, accuracy, and standardization problems, editing is a key aspect of development. It is particularly important to verify the topological and coordinate accuracy of the file. Even if there were no developmental problems, "geographic" changes, such as new streets or changing area boundaries, make file editing essential to a useful GBF.

The Census Bureau and related efforts have produced programs for off-line creation and batch editing of a DIME file [12-13]. These programs require a digitizer for data entry and take large amounts of computer and clerical time for editing. Although the procedures were used to create 200 GBF's, there has been little editing, and hence little use, of these files. Some cities (e.g., [14]) have developed their own GBF's similar to DIME. These efforts are also characterized by the use of a digitizer and batch computer programs for file creation, and by cumbersome file editing procedures. There have been a few efforts to develop on-line digitization systems, (e.g., [15], and there are experimental systems which could support on-line digitization with visual feedback [16-17]. Yet, none of these systems provide all of the capabilities required for effective GBF creation and editing.

---

*The Census Bureau uses the word "editing" to mean topological verification, and uses "update" for what we define as editing in this paper.

Conclusions drawn from the IBM study [10] regarding the requirements for interactive GBF editing and maintenance were:

1. There must be a capability for projecting hard copy maps and/or photographs onto the display screen. It must be possible to select arbitrary (contiguous) sections of the maps, and to produce a range of scales.
2. The display system must be able to handle multiple, non-rectangular geographic coordinate systems.
3. The display system must be able to produce both text and lines, with at least three colors for lines (in order to be able to distinguish two maps).
4. The display system must enable selection of any addressable point on the screen, whether or not anything is displayed at that point.
5. The creation and editing functions must include: digitization of base and overlay maps; labeling of points, lines, and polygons in the maps; moving and deleting points and lines; display of any section of the maps, and of specific points, lines and polygons; and checking for topological accuracy.

IV. DATA EXTRACTION

A. Philosophy and Operation

In the previous section, the combination of a GBF and properly structured source files containing geographic references were identified as the basis for offering a problem solver an effectively integrated data base to support decision making. Recent studies of interactive information systems applications in the solution of unstructured problems [4,18,19] have identified the need for reduced subsets of data for supporting the problem solving. Data reduction is required because:

a. the potentially useful data base will be much larger than the data actually used,
b. the user will want access to varying levels of detail in the data base,
c. the relevant subset of data will vary during the problem-solving process,
d. some data (e.g. census and event data such as police calls) may not be compatible at the detail level of the data captured in the source files.

Extraction is a process by which an integrated subset of data is developed from the source files relevant to a particular problem-solving application. Extraction thus provides the user with a capability effectively indistinguishable from a fully integrated data base, without requiring the development of such an integrated data base at the detail level of the source files, i.e. it provides a "virtually" integrated data base.

The extraction approach builds a data base subset from the source files according to a priori specifications for a particular application. Total integration of the source files, and dynamic aggregation and subsetting of the data at the time the data items are required is of course an alternative approach. This approach is not attractive in today's environment because:

a. for any application all the relevant source files would have to be on-line to support conversational interaction,
b. protection of the source files would be more difficult,
c. development of conversational information systems would require additional standardized data structures and codes for the dynamic aggregation and subsetting,
d. better conversational performance is possible when the problem solving accesses a smaller data base.

Clearly the development of a fully-integrated, on-line data base from the source files, solely for problem-solving applications, is not (currently) economical. Such an approach would also require special procedures for keeping the duplicate records current and consistent. With the extraction approach, the subset of data thought to be relevant to the particular problem is developed and made accessable to the problem-solving system in an extracted data base. The subset is an extract from the available source files at the level of detail desired by the decision maker for (that phase of) problem solving. This extracted data base may be thought of as a set of tables. Each table contains values for a set of variables extracted from the source files. For each variable there is one value in the table for each basic unit (e.g. zone, account, employee) used for the problem solving. New variables can be added directly to the extracted data base as an added column of the tables. An example of an extracted data base is shown in Figure 8, for use in crime analysis. The extracted data tables are formed from: source files containing 10 years' data on crimes, land use, and population; a special purpose map of police beat-building-blocks (basic zones); and an extraction specification for computing 20 crime categories and selecting population and number of houses by year. The result is 10 tables (one for each year) giving crime by category, population, and number of houses for each basic zone.

The extraction approach leaves control of the operational source files in the hands of the originating application. The extracted data bases are "snapshots" which are current at the time of their development. The problem solver can re-invoke the extraction process at any time to get a more current extracted data base. This process decouples the data base used in problem solving from the operational files, and assures the problem solver that the data base upon which he makes decisions is under his control. This user control of the extracted data base, and the potential performance advantages offered by access to the smaller extracted data set as compared to access to the total set of data, make the extraction approach attractive even in installations where an integrated data base exists (as with a complete IMIS as in the USAC approach described in Section II). Extraction is simplified with the existence of an integrated data base, because there are then no difficulties with file formats and data conversion.

B. Extraction System Architecture

The architecture of a municipal information system designed using the data extraction philosophy would have three major sets of programs and data bases (e.g. Figure 9). The first set would be the source data files and related programs for data entry, update, and other routine processing. These files should be "properly structured" as defined earlier in this paper. The data base management for these files may be an integrated system, such as IBM's IMS, or a more traditional system such as those provided by IBM's DOS. The second component includes accurate reference files (indices), such as the Geographic Base File, Programs for maintaining these files, and programs for providing the data extraction functions of data matching, subsetting, and aggregation. This component is the key to integrating the data base of source files. The GADS experience indicates that it is possible to develop general purpose programs for the data extraction functions. Essentially, these functions provide integration through user-invoked processing, rather than through the complicated data structures and accompanying processing overhead often found in integrated data base systems. The data extraction programs are the interface between the municipal data base and the third component of the architecture, the extracted data bases and associated decision support system. The GADS analysis and display functions are an example of a decision support system for non-programmer users. A data extraction interface can provide multiple extracted data bases for a single decision support system, or for multiple decision support systems. For example there might be decision support systems for cash management, budget preparation, urban planning, computer-assisted appraisal, crime analysis, etc., all supported by a common extraction interface. The data management techniques for the extracted data bases should be tailored for each decision support system. However, the data access techniques may be the same as those provided for the source data files.

The details of the data extraction architecture and the implementation requirements are beyond the scope of this paper and there will be installation-specific comments. (An extraction implementation is briefly described in the Appendix). There is, however, one general requirement for any data extraction system. This requirement pertains to the data aggregation functions of extraction and can be described by considering examples of the data sources encountered in municipal governments and the kinds of extracted data to be developed from these sources. Consideration here is limited to data which can be related to points or areas. Data related to networks, budget items, part numbers, etc., should be handled in an analogous fashion.

1. Compatible data

This is the easiest, and fortunately the most frequent situation, if data are captured as specified in Section III. The data in source files which can be identified with geographic points (x,y) can be directly related. If the extracted data base is to be relevant to a study of slum dwellings, for example, and if health cases, fire alarms and building code violations are all data sources which are available at the event level, (i.e. by address) then an extracted data table showing incidence of each of these events for specified address can be directly developed. Another frequently used extracted data base is the tabulation of such event data by geographical area, in terms of a specified map. (The extracted data base in Figure 8 is an example of this). Extracted data bases in such cases are obtained by matching coordinates of events to the corresponding map areas (via point-in-polygon processing of the event coordinates against the map boundaries specification). Figure 10 illustrates the extraction and aggregation to relate property (assessment) data and census data to support inquiries at compatible levels (e.g.) blocks or block groups), and further aggregation to support transit planning models.

2. Non-compatible area data

If data is available by areas in the source files, and these areas are not compatible (i.e. one map is not a subset of the other), then the extraction process is more complicated. For a chosen set of variables from the source files, there is a minimum level of aggregation at which an extracted data base is possible. For example, school attendance areas and police beats (and therefore the associated data) may only be compatible at the census tract level, i.e. they may both be (different) finer partitions of census tracts. The extraction process should alert the user to the non-compatibility and display for the user the minimum level of aggregation necessary for compatibility of the data sources of interest, in the form of a map, and permit the user to specify further aggregation from this map as desired.

If the user desires an extracted data base at a detail level finer than is compatible with the data sources given, the user must supply additional information. For example, suppose the user is studying property values vs age distribution of inhabitants, with the age data on citizens available from the census only at census tract levels of aggregation. Compatability exists at the census tract level. Any finer detailed extracted data base, at the city block level for example, could only be developed if the user is willing to make assumptions (such as homogenity of the distribution of population ages in the census tract).

V. SUMMARY AND CONCLUSIONS

The development of information for decision-making in municipalities requires integration of data from the various operational files which are generated in local government. Even when an integrated municipal data base does not exist, it is possible to develop integrated data from properly structured source files in conjunction with a well-maintained Geographic Base File. The current sources of information developed in municipalities, in particular the property data of the tax assessor function and the operating files of various service delivery functions, provide a rich source of information, augmented by special collections such as the U.S. Census.

Data Extraction is the process of developing integrated data subsets from diverse source files to support interactive problem solving. Extraction provides the interface to large data bases of source files and provides data description, subsetting and aggregation functions. Our experience with GADS has shown that data extraction is useful when the user or problem characteristics require access to varying amounts, detail, and selection of data, and conversational (rapid response) interaction with a decision support system. These characteristics are likely to be encountered when designing decision support systems for nonprogrammer, professional users working on unstructured problems. The data extraction interface matches the functional and response time requirements of interactive decision support, can be implemented on a variety of computer system configurations, and can reduce the operating costs of the decision support system.

Because data extraction operations can produce multiple extracted data bases, with different structures, a single data extraction interface can support multiple decision support systems. In addition, existing decision support systems can be supported and enhanced by data extraction without major program revisions.

APPENDIX: An Extraction System Implementation

A project in the IBM Research Division has developed an interactive Geo-data Analysis and Display System (GADS as a vehicle for studying interactive problem solving [1-4]. GADS supports nonprogrammer users solving unstructured problems where the relevant data can be related to a geographic location. Examples of the problems for which GADS has been used include: land use planning, police manpower allocation, school districting, and commercial site location. The need for data extraction was recognized during the first studies of the use of GADS. In particular, the need for data aggregated to a variety of geographic levels (e.g., block, police beat, census tract, neighborhood), and changing data needs expressed by users indicated the inadequacies of the static, special purpose data base and the one-level, integrated data base approaches.

GADS data extraction is configured essentially as shown in Figure 9. The extraction implemented in GADS is limited to compatible event data. There is a requirement that each record of each file in the large data base contain a geographic code (such as an address, x,y coordinates or block number) so that extracted data can be related to points, lines, and polygons on a map. A utility program is provided to transform geographic codes into x,y coordinates if necessary for data extraction or display. A data base developed by extraction is a table; an example is shown on Figure 8. Adding another crime type, acres of commercial land use, or re-aggregating by census tracts would take only a few minutes.

In the GADS implementation the large data base management system is a special purpose one designed to handle fixed format files with no hierarchies or repeating groups. Simultaneous access to multiple files, and shared access to single files are not supported. Multiple file extractions are handled by consecutive extractions from the individual files. Sequential and direct access I/O are provided. Character, fixed binary, packed decimal, and floating point (binary) data representations can be used. The entire large data base is stored on disk, and there is a utility for loading files from tapes.

Figure 11 gives examples of the data description and subsetting capabilities. The data description implementation allows different formats to be used for the same file or the same formats to be used for different files (Figure 11a). Seven data types are allowed. The subsetting language includes constructs for: subsetting based on any arithmetic or logical combination of the items in a file, creating of new items, conditional subsetting or creation (IF, THEN, ELSE), and function calls (Figure 11b.). Results from subsetting can be displayed as lists (Figure 11c) or as locations on a map (Figure 11d). Using the display capabilities, two dimensional subsetting is possible. That is, the user can draw a polygon on the screen, and select only those elements of a file whose location is within that polygon. This facility is much more user-oriented than algebraic specifications for subsetting, and other graphic subsetting operators would be useful (e.g., display all the crimes of the same type as the one being pointed at).

The aggregation operations in the implementation are restricted to forming the extracted data base for the GADS analysis and display functions. This data base is aggregated by areas of a map. It is stored on disk, and is accessed by column name.

GADS is implemented in FORTRAN, but the data extraction components were implemented in PL/I because of its larger set of data types, and better functional suitability for the extraction tasks. The combination system runs on the IBM S/360 or S/370 series under the Time Sharing Option (TSO). The combination requires 220K bytes of main storage. Separating extraction reduces the main storage requirement to about 120K. The user terminals may be IBM 2250s or storage tube display terminals. The I/O time from the large data base, and the data rate to the terminal are the limiting facotrs in extraction response times (i.e. selection and aggregation times are negligible compared to I/O times). Although five minutes may be required to list or display an entire file, the user can see the results unfolding (e.g. the selected items are listed as they are selected). Thus users seem willing to wait during extraction. After all, the batch mode equivalent capabilities have response times of days, and manual methods have response times of weeks or months.

REFERENCES

[1] P. E. Mantey, J. L. Bennett, E. D. Carlson, Information for Problem Solving: The Development of an Interactive Geographic Information System. IEEE Int. Conf. on Communication, Vol. II. Seattle, Wash. June 1973.

[2] E. J. Cristiani, R. J. Evey, R. E. Goldman, P. E. Mantey. "An Interactive System for Aiding Evaluation of Local Government Policies," IEEE Transactions on Systems, Man & Cybernetics, Vol. SMC-3, No. 2, March 1973, pp. 141-146.

[3] E. D. Carlson, J. L. Bennett, G. M. Giddings and P. E. Mantey. "The Design and Evaluation of an Interactive Geo-data Analysis and Display System," Proceedings of the IFIP Congress 74, International Federation for Information Processing, Stockholm, August 1974. North Holland Publishing Company, Amsterdam, 1974.

[4] E. D. Carlson and J. A. Sutton, A Case Study of Non-programmer Interactive Problem-Solving, IBM Research Report, RJ 1382, IBM Research Laboratory, San Jose, California, April 1974.

[5] T. R. Hammer, R. E. Coughlin, E. T. Horn IV, "The Effect of a Large Urban Park on Real Estate Values," Journal of the American Institute of Planners, Vol. 40, No. 4, July 1974, pp. 274-277.

[6] "City Hall's Approaching Revolution in Service Delivery," Nation's Cities, January 1972.

[7] Concepts of an Urban Management Information System," a Report to the City of New Haven, Connecticut, by Advanced Systems Development Division, IBM Corporation, Yorktown, January 1967.

[8] A Municipal Information and Decision System. University of Southern California, School of Public Administration, 1968.

[9] R. L. Stickrod and L. C. Martin. Data Processing: Analysis of Costs, Benefits, and Resource Allocations. Lane County, Oregon, Management Report, February, 1973.

[10] G. M. Giddings and E. D. Carlson, An Interactive System for Creating, Editing and Displaying a Geographic Base File. IBM Research Report, IBM Research Laboratory, San Jose, California, 1973.

[11] D. C. Symons, A Parcel Geocoding System for Urban and Rural Information, Ottawa, Ontario, National Capital Commission, 1970.

[12] U. S. Bureau of the Census, Census Use Study, The DIME Geocoding System Report No. 4, Washington D.C., 1970.

[13] U. S. Bureau of the Census, Census Use Study, The DIME Editing System Washington D.C., 1970.

[14] R. Jull, Geo-modeling: A Local Approach, Eugene, Oregon, Lane Council of Governments, 1972.

[15] R. D. Hogan, Remote Graphic Terminal and Urban Geographic Information System Demonstation, Gaithersburg, Maryland, IBM Federal Systems Center, 1968.

[16] R. D. Merrill, "Representation of Contours and Regions for Efficient Computer Search," Communications of the ACM, Vol. 16, No. 2, February 1973, pp. 69-82.

[17] B. V. Saderholm, "Paper 'Keyboard' Runs Experimental IBM System," IBM Research Division Press Release, Yorktown Heights, N. Y., March 8, 1973.

[18] R. M. Cyert, H. A. Simon, and D. B. Throw. Observation of a Business Decision. Journal of Business, 29, (1956), 237-248.

[19] D. M. S. Peace and R. S. Easterby. The Evaluation of User Interaction with Computer-based Management Information Systems. Human Factors, 15, April 1973, pp. 163-177.

PARCEL NUMBER	ADDRESS	YEAR CONST	LOT SIZE	FLOOR AREA	ASSESSED VALUE
123-06-174	100 ABBYWOOD CT.	1962	7145	1675	28000
123-06-175	102 ABBYWOOD CT.	1962	6965	1650	29000
123-06-176	104 ABBYWOOD CT.	1962	6500	1600	27000
123-06-177	106 ABBYWOOD CT.	1961	7125	1650	25000
123-06-150	112 ABINATE LN	1960	7200	1800	35000
123-06-151	114 ABINATE LN.	1960	7185	1800	37000
123-06-152	116 ABINATE LN.	1960	6995	1700	30000
123-06-153	117 ABINATE LN.	1960	6955	1850	31000
123-06-154	120 ABINATE LN.	1860	7095	1900	32000
123-06-158	132 ABINATE LN.	1961	7155	1750	29000
123-06-160	111 ABINATE LN.	1960	7005	1975	34000
123-06-161	113 ABINATE LN.	1960	6885	1650	27000
123-06-162	115 ABINATE LN.	1960	7100	1750	30000
123-06-163	117 ABINATE LN.	1960	7055	1780	30000
123-06-184	31 AFTON CT.	1960	7020	1650	29000
123-06-185	33 AFTON CT.	1960	7165	1600	27000
123-06-186	35 AFTON CT.	1961	7190	1775	29000

Figure 1. Tabular display of assessment data

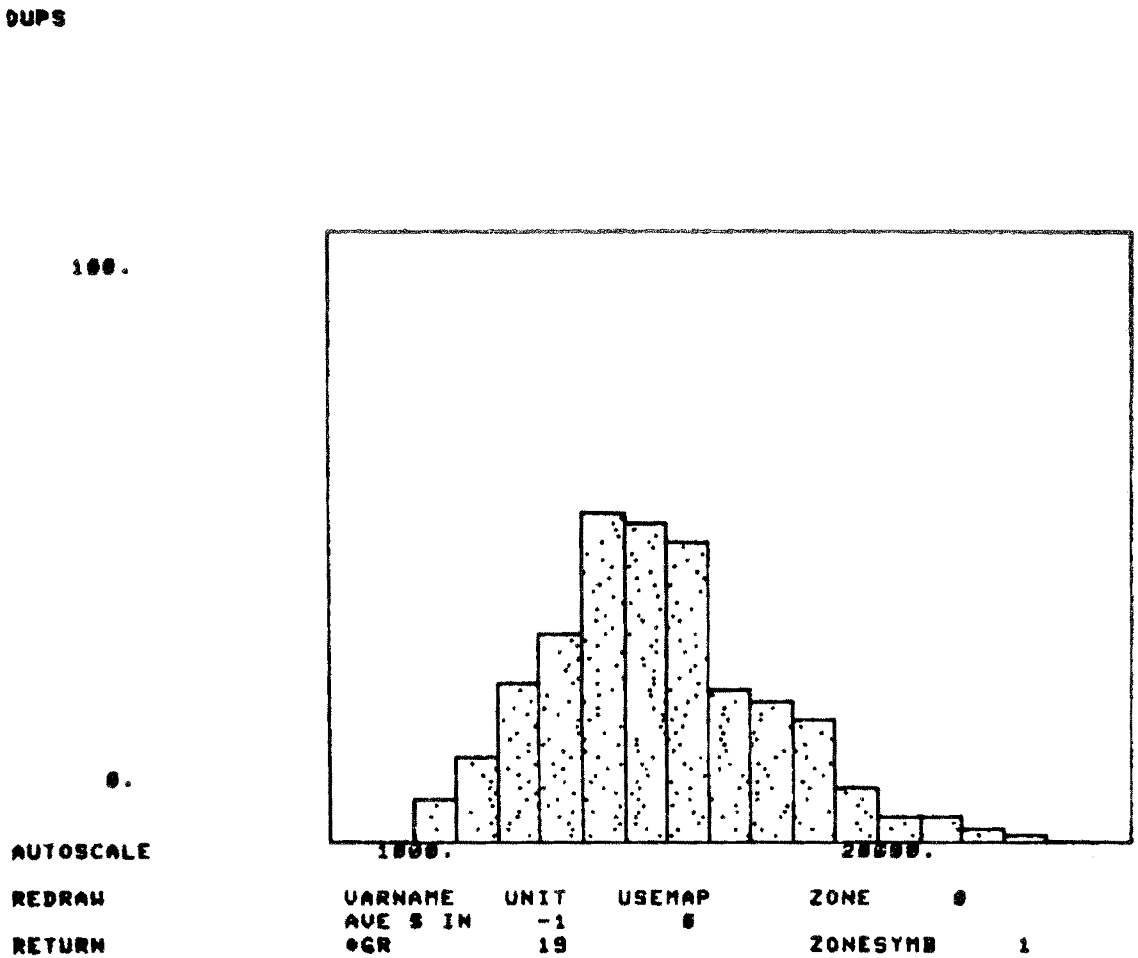


Figure 2. Histogram display of housing values

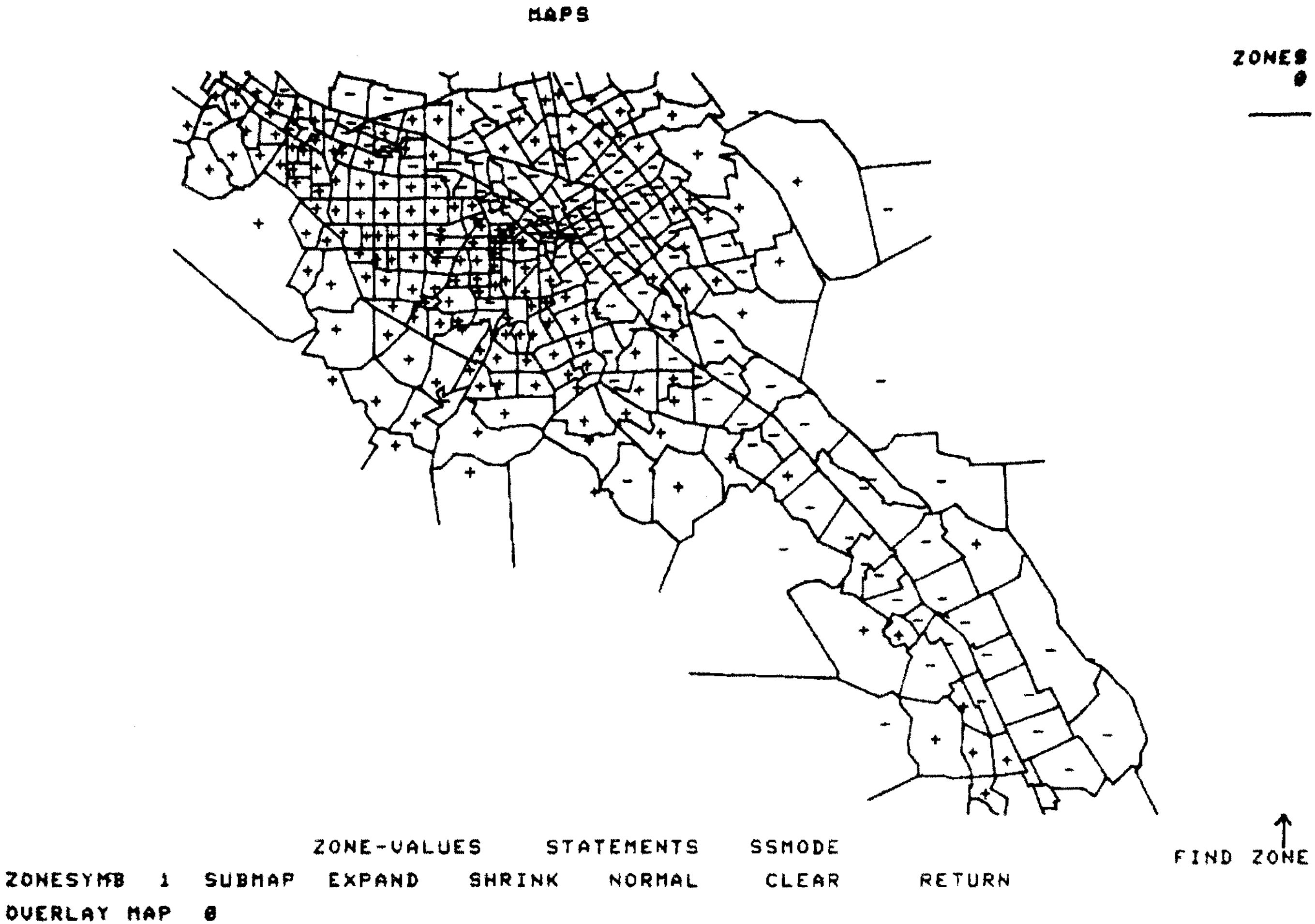


Figure 3a. Map display of relative housing values

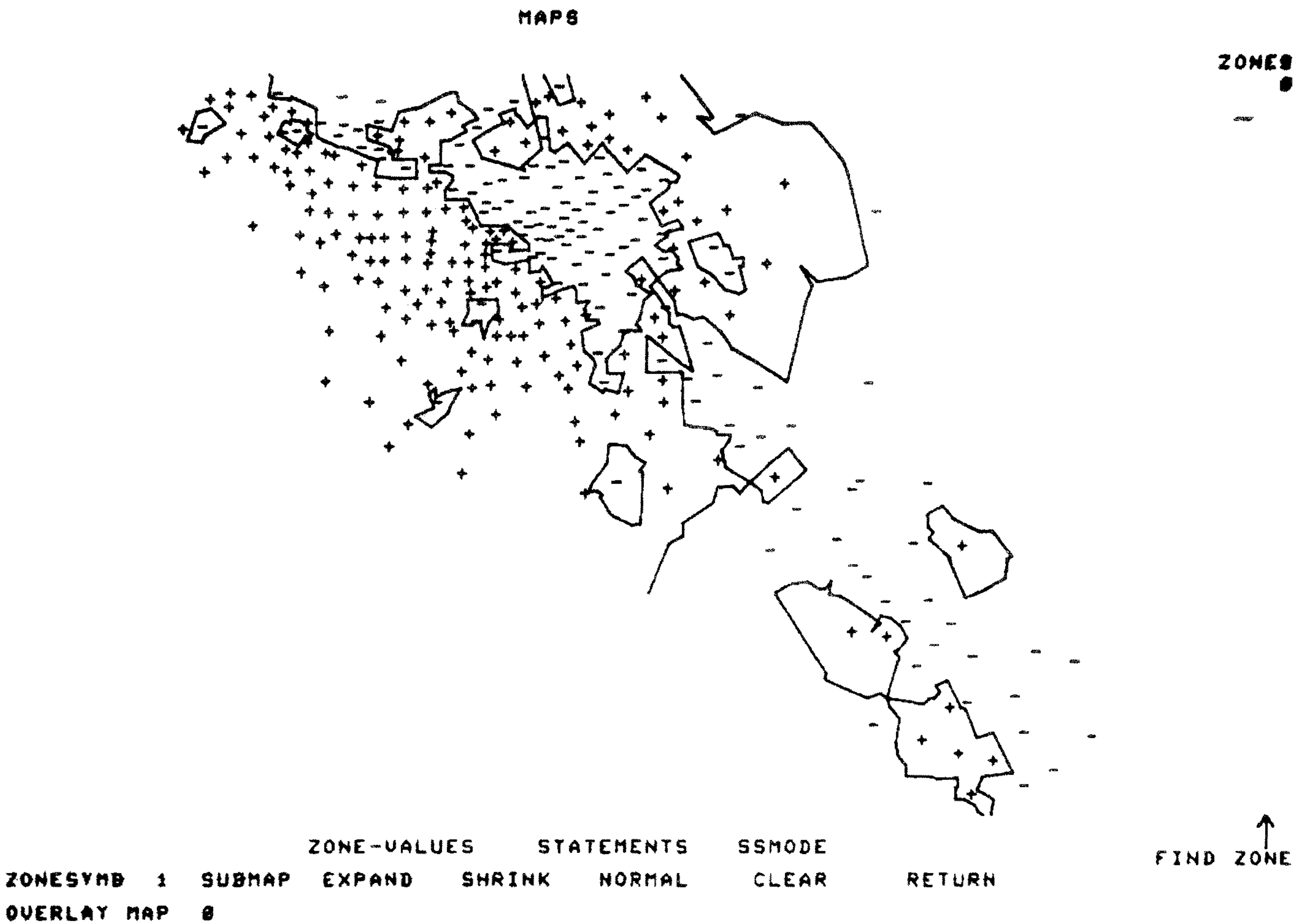


Figure 3b. Simplified display of housing values

AVE $ IN	AVERAGE INCOME PER HOUSEHOLD
ACCESS $	ACCESSABILITY TO DISPOSABLE INCOME
ACCESS	ACCESSABILITY TO EMPLOYMENT
AVLAND-S	ACRES AVAILABLE FOR SINGLE FAMILY DEVELOPMENT
AVLAND-M	ACRES AVAILABLE FOR MULTIPLE FAMILY DEVELOPMENT
AVLAND-C	ACRES AVAILABLE FOR COMMERCIAL DEVELOPMENT
AVLAND-I	ACRES AVAILABLE FOR INDUSTRIAL DEVELOPMENT
GROW NXS	GROWTH FACTOR IN SINGLE FAMILY
GROW NXM	GROWTH FACTOR IN MULTIPLE FAMILY
ODWU/A-S	NO SINGLE FAMILY DWELLING UNITS PER RESIDENTIAL
ODWU/A-M	NO MULTIPLE FAMILY DWELLING UNITS PER RESIDENTI
ODWU/A-T	TOTAL NO DWELLING UNITS PER ACRE OF RESIDENTIAL
EMP-MFG	NO. OF EMPLOYEES WORKING IN MANUFACTURING
EMP-WHOL	NO. OF EMPLOYEES WORKING IN WHOLESALE AND TRUCK
EMP-COMM	NO. OF EMPLOYEES WORKING IN COMMERCIAL(RETAIL)
EMP-TC&U	NO. OF EMPLOYEES WORKING IN TRANS, COMMUN,AND U
EMP-GOVT	NO. OF EMPLOYEES WORKING IN GOVERNMENT
EMP-TOTL	TOTAL NO OF EMPLOYEES
EMPDEN-C	NO. COMMERCIAL EMPLOYEES PER ACRE OF COMMERCIAL
HHOLD0-6	NO. OF HOUSEHOLDS WITH INCOME: 0- 6000
HHLD6-10	NO. OF HOUSEHOLDS WITH INCOME: 6000-10000
HHLD1015	NO. OF HOUSEHOLDS WITH INCOME:10000-15000
HHLD15+	NO. OF HOUSEHOLDS WITH INCOME:15000+
ISHOPCTR	ZONES OF ANTICIPATED SHOPPING CENTERS
IBAY	ZONES TREATED AS BAYLANDS
OCDWUN-S	NO. OF EXISTING SINGLE FAMILY DWELLING UNITS
OCDWUN-M	NO. OF MULTIPLE FAMILY DWELLING UNITS
OCDWUN-T	TOTAL NO. OF SINGLE AND MULTIPLE FAMILY DWELLIN
OCLAND-S	ACRES OCCUPIED BY SINGLE FAMILY DWELLINGS
OCLAND-M	ACRES OCCUPIED BY MULTIPLE FAMILY DWELLINGS
OCLAND-C	ACRES OCCUPIED BY COMMERCIAL DEVELOPMENT
OCLAND-I	ACRES OCCUPIED BY INDUSTRIAL DEVELOPMENT
PRODEN-S	PROJECTED DENSITY FOR SINGLE FAMILY DEVELOPMENT
PRODEN-M	PROJECTED DENSITY FOR MULTIPLE-FAMILY DEVELOPME
POPUL-S	TOTAL POPULATION IN SINGLE FAMILY DWELLINGS
POPUL-M	TOTAL POPULATION IN MULTIPLE FAMILY DWELLINGS
POPUL-T	TOTAL POPULATION
POP/HH-S	POPULATION PER HOUSEHOLD FOR SINGLE FAMILY DWEL
POP/HH-M	POPULATION PER HOUSEHOLD FOR MULTIPLE FAMILY DW
POP/HH-T	POPULATION PER HOUSEHOLD FOR ALL DWELLINGS
RES-LAND	RESERVED LAND-NOT AVAILABLE FOR CURRENT DEVELOP
SLOPE	PERCENT OF LAND WITH LESS THAN 10% SLOPE
TSKTR	CHANGE FACTOR IN INTRAZONAL TRAVEL TIME
DISP $IN	DISPOSABLE INCOME PER HOUSEHOLD
SEWER	SEWER SERVICE DISTRICT
FLOOD	FLOOD CONTROL AREA

Figure 4. Land use/transportation planning data base [2]

RESIDENTIAL FACT SHEET

DISTRICT (AC16) ______ NEIGHBORHOOD (AC17) ______ APN (AA01) ______

TRACT (AC06) ______ LOT ______ BLOCK ______ SITUS (AA10) ______ SHEET ______ OF ______ SHEETS

RECORD DATA			LAND ATTRIBUTES			BUILDING DATA			COST DATA		
App'l. Date	AC10		Width Ft.	AI01		Year Built	AJ01		Quality Class	AC18	
Appraiser No.	AB02		Depth Ft.	AI02		Eff. Year	AC19		Total Living Area	AC05	
Site Use Code	AB01		Sq. Ft. (Actual)	AC04		REL.	AJ02		1st Flr. Area	AK01	
Zone	AC01		Sq. Ft. (Useable)	AI04		Dining Room	AJ03		2nd Flr. Area	AK02	
Total Prop. Val.	AE01		Typical	AI07	$N^1\ Y^2$	Fam.-Den-Rumpus	AJ04		2nd Flr. Factor	AK03	
Land Value	AB04		Irregular	AI08	$N^1\ Y^2$	No. of Bedrooms	AJ05		3rd Flr. Area	AK04	
Imp. Value	AB06		Cul-De-Sac	AI09	$N^1\ Y^2$	No. of Baths	AJ06		3rd Flr. Factor	AK05	
New Lot Value	AE02		Non-Thru-St.	AI10	$N^1\ Y^2$	Util. Rooms	AJ07		Bsmt. Area	AK06	
SALES DATA			St.-Frontage	AI11	$N^1\ Y^2$	Total Rooms	AC21		Bsmt. Factor	AK07	
Rec'd. Date	AA13		Corner	AI12	$N^1\ Y^2$	Funct. Plan	AJ08	$P^1\ A^2\ G^3$	Addn. Area	AK08	
Sales Price	AC12		Alley	AI13	$N^1\ Y^2$	Condition	AJ09	$P^1\ A^2\ G^3$	Addn. Factor	AK09	
Confirmed	AE03	$N^1\ Y^2$	Util. U./G.	AI14	$N^1\ Y^2$	Workmanship	AJ10	$P^1\ A^2\ G^3$	Gar. Area	AK10	
TEMPORARY VALUE			Curbs & Gutters	AI15	$N^1\ Y^2$	Stg. Space	AJ11	$P^1\ A^2\ G^3$	Carport Area	AK11	
Part Compl.	AC20	$N^1\ Y^2$	Sidewalks	AI16	$N^1\ Y^2$	Heating (Ducted)	AJ12	$N^1\ Y^2$	Carport Factor	AK12	
Board Action	AF02	$N^1\ Y^2$	St. Lights	AI17	$N^1\ Y^2$	Cooling (Ducted)	AJ13	$N^1\ Y^2$	Pch. Area	AK13	
Other	AF03	$N^1\ Y^2$	Common Green	AI18	$N^1\ Y^2$	Gar.	AJ14	$N^1\ Y^2$	Pch. Factor	AK14	
IMMEDIATE AREA			Common Rec.	AI19	$N^1\ Y^2$	Gar. Conv.	AJ15	$N^1\ Y^2$	Fireplace Cost	AK15	
Mkt. Demand	AG01	$P^1\ A^2\ G^3$	H. & B. Use	AI20	$N^1\ Y^2$	Carport	AJ16	$N^1\ Y^2$	A. C. Cost	AK16	
Trend	AG02	$P^1\ A^2\ G^3$	Sewer	AI21	$N^1\ Y^2$	Patio	AJ17	$N^1\ Y^2$	Patio Area	AK17	
Trans.	AG03	$D^1\ S^2\ G^3$	View	AI22	$N^1\ Y^2$	Decking	AJ18	$N^1\ Y^2$	Patio Factor	AK18	
Res. Area	AG05	$N^1\ Y^2$	View Qual.	AI23	$P^1\ A^2\ G^3$	Pool	AJ19	$N^1\ Y^2$	Pool Area	AK19	
Single Fam.	AG06	$N^1\ Y^2$	Traffic Flow	AI24	$H^1\ A^2\ L^3$	Fence	AJ20	$N^1\ Y^2$	Pool Extras	AK20	
TOTAL PROPERTY			Water Front	AI25	$N^1\ Y^2$	Guest House	AJ21	$N^1\ Y^2$	Misc. Cost	AK21	
Arch. Attr.	AH01	$P^1\ A^2\ G^3$	Beach Front	AI26	$N^1\ Y^2$	Fireplace	AJ22	$N^1\ Y^2$	Misc. Struct. Cost	AK22	
Prop. Improved	AH02	$N^1\ Y^2$	Docking Rights	AI27	$N^1\ Y^2$	Built-Ins	AJ23	$N^1\ Y^2$	TOPOGRAPHY		
Prop. Lot Util.	AH03	$N^1\ Y^2$	Horses	AI28	$N^1\ Y^2$	Stru. Failure	AJ24	$L^1\ M^2\ H^3$	Grade	AL01	$B^1\ A^2\ E^3$
Landscape	AH04	$P^1\ A^2\ G^3$				Tile Roof	AJ25	$N^1\ Y^2$	Bank	AL02	$N^1\ Y^2$
Nuisance Infl.	AH05	$N^1\ Y^2$	FIELD USE ONLY - DO NOT ENCODE						Slope	AL03	$N^1\ Y^2$
Condominium	AH06	$N^1\ Y^2$	% Base Lot	A130					Other	AL04	$N^1\ Y^2$

AO-226 Ventura County Assessor

Figure 5. Field work sheet for appraisal data

Office of the County Assessor
201 County Administration Building
70 West Hedding Street
San Jose, California 95114
299-3941 Area Code 408

County of Santa Clara
California

Date Recorded ____________

Recorder's Deed #____________

Property Description #____________

1. Our records indicate you purchased this property.

   What was the full price? ____________

   a. Amount of cash down payment: $____________

   b. Please enter details concerning any balance:

      (1) 1st Deed of Trust $____________ at ____________% interest
      Duration of Loan________years

      (2) 2nd Deed of Trust $____________ at ____________% interest
      Duration of Loan________years

   c. Was a trade involved? Yes__________ No__________

   d. Outstanding Improvement Bonds against property $____________

   e. Did price include personal property? Yes________ No________
      If yes, please estimate value of such property $____________

   f. If this is income property, please enter the monthly gross income as if it were 100% occupied $____________

2. Remarks: Please enter on the reverse side or by attachment any information you feel may help us to make a fair appraisal of your property.

3. If you would like mail concerning this property sent to a different address from the one above, please indicate below.*

4. If there are questions regarding this questionnaire please contact the Assessment Standards Division at 299-3941.

5. See Reverse Side.

____
Signature of Owner

____
Telephone Number

____
Date

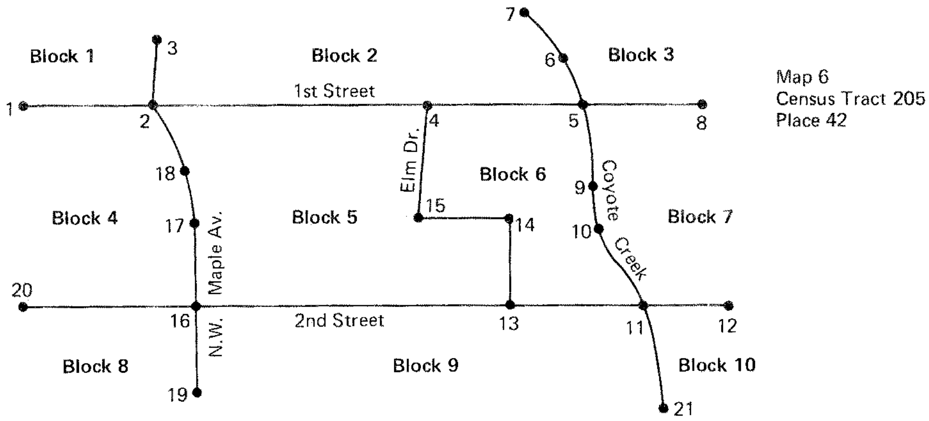
*____________________
Address

____
City, State, Zip

7047 REV. 11/71

Figure 6. Assessor's questionnaire for financial data

Source Map



Number of Segments

Maple Av.	5	
Elm Dr.	3	
Coyote Cr.	6	= 22 segments (records)
1st St.	4	
2nd St.	4	

Contents of a Sample Record

Contents

Maple Av. N.W.

2	1530000	31000	6
18	1530100	30100	6
205	205 42 42	5	4
100	198		
101	199		

Legend

Name Type Suffix

From node number, x, y, map number
To node number, x, y, map number
Left tract, right tract, left place,
right place, left block, right block
Left low address, high address
Right low address, high address

Figure 7. 'DIME' geographic base file structure

Year 10 . . .

Year 2

Year 1 Zone	Crimes #1	Crimes . . . #2	Houses . . .	Population
1	3	15	50	146
2	17	6	53	203
3	4	4	41	192
⋮				
N	26	9	15	47

Figure 8. Example of tables in extracted data base

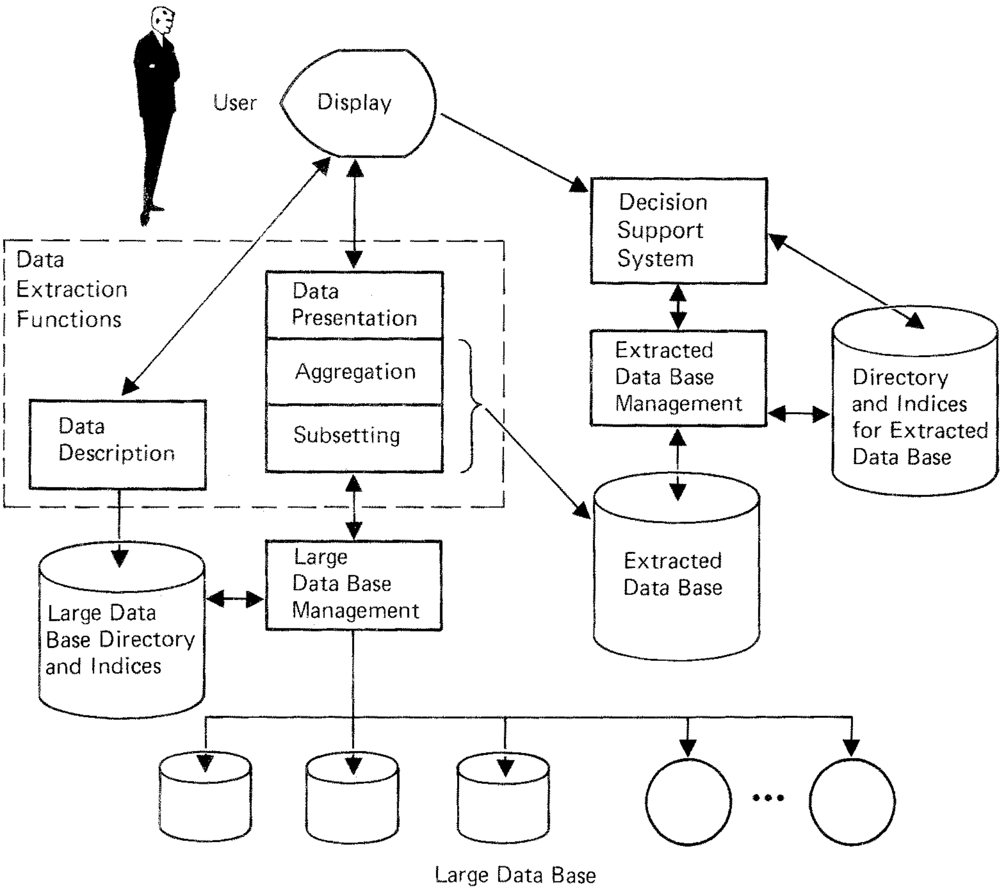


Figure 9. Interactive data extraction and problem solving system

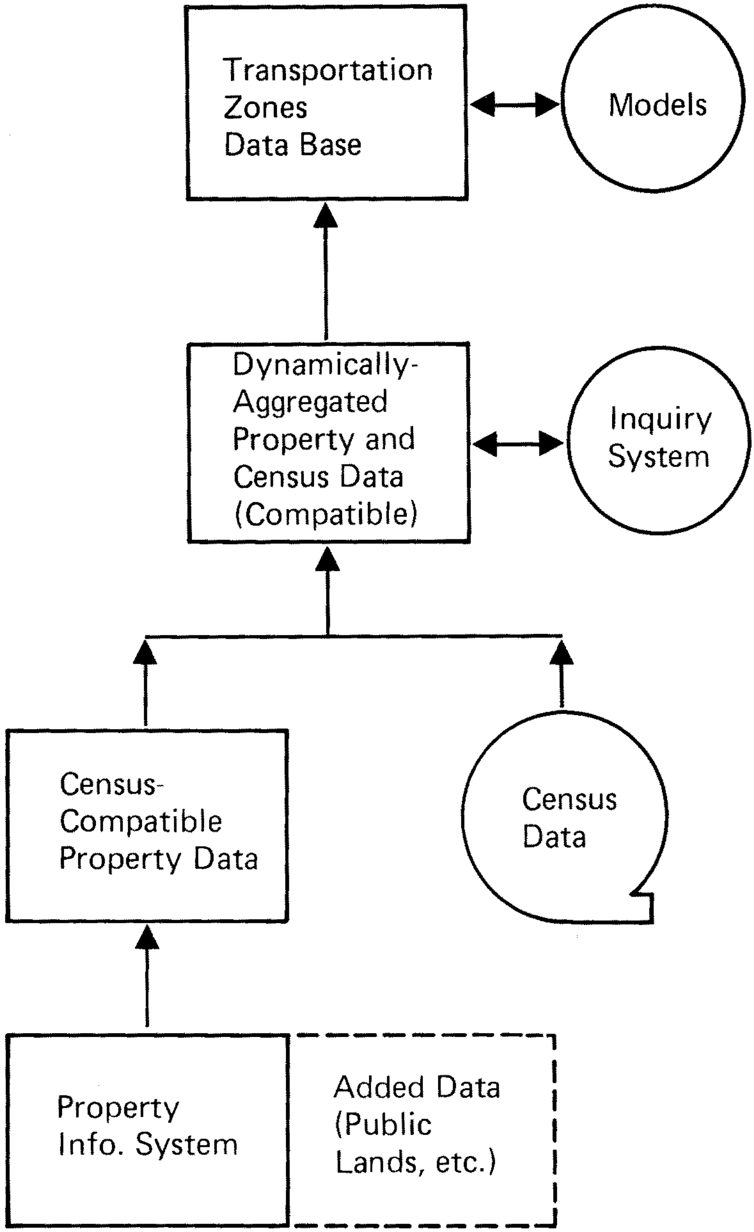


Figure 10. Extraction and aggregation relating property and census data

FUNCTIONS AVAILABLE FOR EXTRACTING
DATA FROM EVENT FILE

CREATE AND EDIT

DATA FORMAT DESCRIPTIONS

SELECTION SPECIFICATIONS

BUILD-UNIT SPECIFICATIONS

VIEW EVENTS ON MAP

LIST EVENTS

BUILD GADS UNITS

RETURN (TO GADS MAIN MENU)

Figure 11. Extraction system functions

```
                    CREATE AND EDIT DATA DESCRIPTIONS

EMP 92    IBM EMPLOYEE FILE WITH X,Y
MATCH CH 13 1      MATCH KEY FROM UNIMATCH
STNO CH 5          STREET NUMBER
BLANK1 CH 1
STNAME CH 16       STREET NAME
STTYPE CH 2        STREET TYPE
CITY CH 10         NAME
ZIP CI 5           CODE
BLANK2 CH 1
X CI 7             COORDINATE
BLANK3 CH 1
Y CI 6             COORDINATE
_
_
_
_
_
_
_
_
_                           SCRATCHPAD




INPUT MODE         GET PAGE    6      RESTORE     DELETE LINE
REPLACE MODE =ON   VIEW TITLES                    INSERT BLANK LINE
CHECK              SAVE PAGE ___                  COPY LINE DOWN
RETURN             REDRAW          ERASE             PRINT
```

Figure 11a. Data description function

```
                    CREATE AND EDIT SELECTIONS

EMP
_
_
_
IF X<=0 OR X>2000000 THEN X=0 ;

IF Y<=0 OR Y>1000000 THEN Y=0;

SELECT WHERE ZIP = 95125
_
_
_
_
_
_
_
_
_
_                              SCRATCHPAD





INPUT MODE    =ON   GET PAGE     15        RESTORE      DELETE LINE
REPLACE MODE        VIEW TITLES                         INSERT BLANK LINE
CHECK               SAVE PAGE    15                     COPY LINE DOWN
RETURN              REDRAW               ERASE             PRINT
```

Figure 11b. Data selection function

```
MTØØØØØ2      ØØ1ØØ   GLENEYRIE            SANJOSE   95125   1591723   29Ø3Ø4
MTØØØØØ4      Ø1245   GLENEYRIE         AV SANJOSE   95125   159Ø839   297561
MTØØØØØ4      Ø16Ø8   GLENFIELD         DR SANJOSE   95125   1592561   2857ØØ
MTØØØØØ4      Ø1631   GLENFIELD         DR SANJOSE   95125   1592377   285847
MTØØØØØ4      Ø1161   GLENN             AV SANJOSE   95125   1591593   29632Ø
MTØØØØØ4      Ø1148   GLENN             AV SANJOSE   95125   159175Ø   296376
MTØØØØØ4      Ø1487   GLENPINE          DR SANJOSE   95125                  Ø
             Ø
MTØØØØØ4      Ø1538   GLENPINE          DR SANJOSE   95125                  Ø
             Ø
MTØØØØØ4      Ø1499   GLENPINE          DR SANJOSE   95125                  Ø
             Ø
MTØØØØØ4      Ø1754   GLENUNA           AV SANJOSE   95125   159614Ø   293698
MTØØØØØ2      Ø193Ø   GLENUNA              SANJOSE   95125   1596628   292822
MTØØØØØ4      Ø1742   GLENUNA           AV SANJOSE   95125   159614Ø   293698
MTØØØØØ2      Ø1348   GLENWOOD             SANJOSE   95125   1592291   293314
MTØØØØØ4      Ø1551   GRACE             AV SANJOSE   95125   1587514   293Ø85
MTØØØØØ4      Ø1482   GRACE             AV SANJOSE   95125   1587719   2941Ø2
MTØØØØØ4      Ø1623   GRACE             AV SANJOSE   95125   1587514   293Ø85
MTØØØØØ4      Ø15ØØ   GRACE             AV SANJOSE   95125   15876Ø7   293232
MTØØØ2Ø4      Ø155Ø   GUADALAJARA       AV SANJOSE   95125   1592633   2673Ø4
MTØØØØØ4      Ø146Ø   HAMILTON          WY SANJOSE   95125   15911Ø9   292471
MTØØØØØ4      Ø188Ø   HARMIL            WY SANJOSE   95125   159629Ø   293944
MTØØØØØ4      Ø1Ø24   HAZELWOOD         AV SANJOSE   95125   1596Ø56   292681
MTØØØØØ4      Ø1479   HERVEY            LN SANJOSE   95125   1596632   297ØØ4
MTØØØØØ4      Ø1514   HERVEY            LN SANJOSE   95125   159734Ø   2963Ø5
MTØØØØØ4      Ø15Ø2   HICKS             AV SANJOSE   95125   1591625   293923
MTØØØØØ4      Ø2243   HICKS             AV SANJOSE   95125   1592833   289482
GRADDRESS ØØ  Ø1656   HILLSDALE         AV SANJOSE   95125                  Ø
             Ø
MTØØØØØ4      ØØ947   HUMMINGBIRD       DR SANJOSE   95125   1599186   2895Ø7
MTØØØØØ4      ØØ91Ø   HUMMINGBIRD       DR SANJOSE   95125   1599225   289352
MTØØØØØ4      ØØ87Ø   HUMMINGBIRD       DR SANJOSE   95125   1599225   289352
MTØØØØØ4      ØØ91Ø   HUMMINGBIRD       DR SANJOSE   95125   1599225   289352
MTØØØØØ4      Ø1674   HUSTED            AV SANJOSE   95125   1591335   285412
MTØØØØØ4      Ø1436   HUSTED            AV SANJOSE   95125   15945Ø1   287659

 LISTING: FILE = EMP        SELECTION = EMP               RETURN
 USING P1P2: NO                                           PAGE FORWARD
                                                              PAGE    5
```

Figure 11c. Listing of data selected

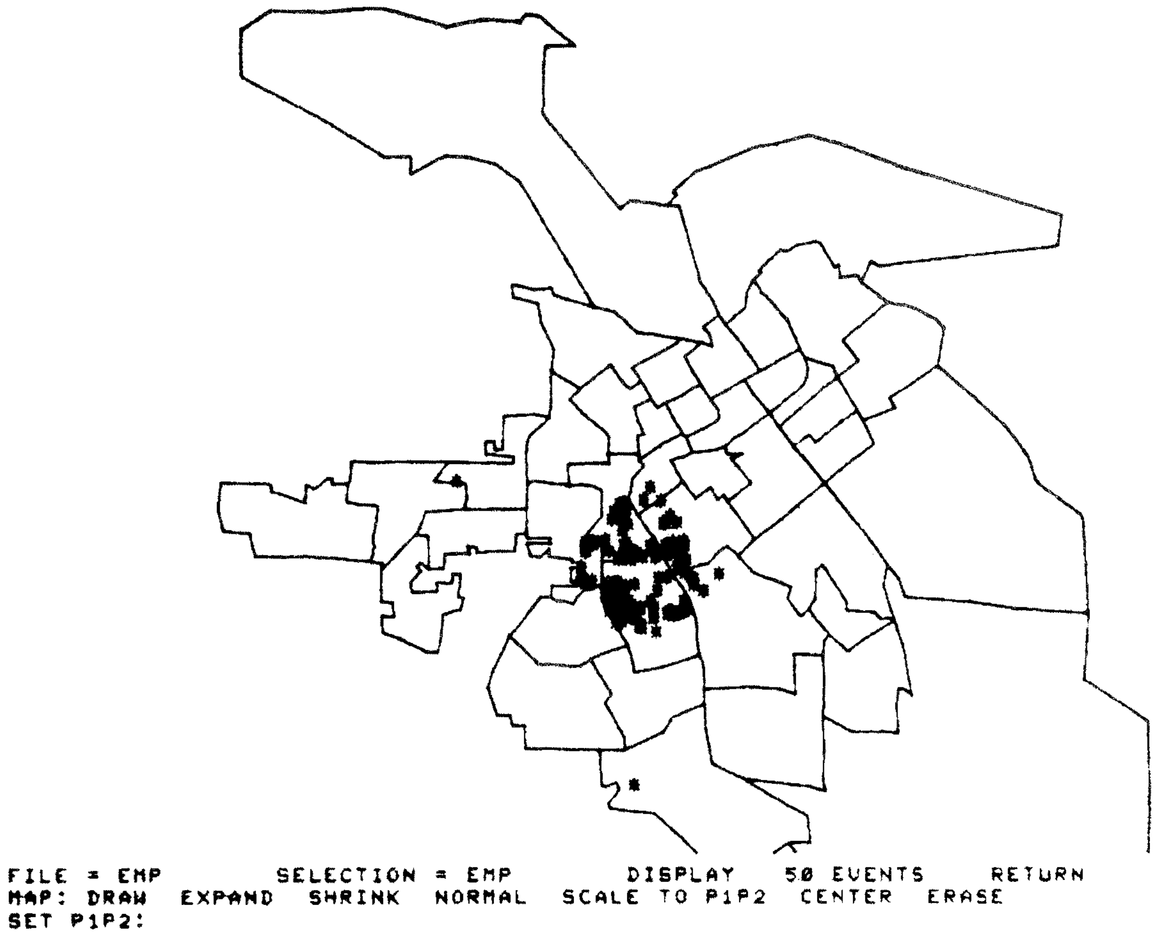


Figure 11d. Map showing location of selected data