COMPATIBILITY OF URBAN LAND UNIT DATA

Robert C. Colwell, Department of Housing and Urban Development

This decade is witnessing a major revolution in data handling as a result of rapid technological progress in the capabilities of computers. Installations of electronic data processing equipment in private industry are being used for a growing list of chores ranging from inventory control to stock market transactions and from airline reservations to material processing equipment. Wherever a job involves assimilation of large quantities of data, it seems that electronic equipment can be devised to do that job faster, more accurately and cheaper. The computer is making significant contributions to the increase in productivity of the U. S. economy.

In Government, computers have found a major place. The Bureau of the Census has long been a pioneer in developing new data processing equipment, and Univac I used in the 1950 Census of Population and Housing was one of the early Federal installations. Today, I suspect that there is hardly a Department or agency of the Federal Government that does not make some use of automatic data processing equipment, directly or indirectly. The Comptroller General recently reported that "the Federal Government is spending directly and indirectly this year nearly $3 billion for the acquisition and use of computer equipment, compared to only minute outlays 15 years ago."

State and local governments are also proceeding to install ADP equipment, finding that burgeoning files of data can be maintained and retrieved more effectively and less costly via the electronic route. Just as electrical bookkeeping equipment began replacing manual records several decades ago, so the computer is replacing EAM in the current decade. The pace is fast and the opportunities are great. Like other technological revolutions, there can be no turning back to older methods. Our economy and our governmental operations have been restructured around the computer. We would drown in a depthless sea of data if these machines ever went on a strike.

The major focus of this paper is a narrow one, but perhaps an important one for the future application of electronic equipment to geographically identifiable data in governmental activities. The Department of Housing and Urban Development, as well as the Bureau of the Census, has a deep involvement in this problem. Increasingly, State and local governments are sharing a similar concern.

The element of geographic location is common to the files in nearly every program that HUD administers. Most of the substantive programs involve land parcels, either separately or in contiguous groupings. Policy analyses and program execution in urban land-use and land-related activities rely on statistical studies of the Department's records as well as on urban and metropolitan universe data, ranging in scale from a block to an SMSA, to the total U.S.A.

Let me illustrate with a few examples:

1. City A proposes a new urban renewal project involving 100 acres. Are the boundaries of the plan properly located for the elimination of slums and blight and the creation of a well-planned neighborhood after renewal? What proportions of clearance and rehabilitation can be expected? Reference to Census block data can throw considerable light on these questions.

2. The XYZ Development Company seeks a commitment to insure mortgages on a 200-house subdivision at prices ranging from $18,000 to $23,000. Is the subdivision too large, and are these prices proper for the proposed location? The records of recent market activity in other subdivisions that FHA has insured will be useful in judging the prospective market absorption, and Census data will help to identify directions of growth in various housing price-groups.

3. Long-range fiscal plans for senior-citizen housing are being formulated. What levels of Federal support are appropriate, and what income groups need to be served? Here, a combination of statistics need to be analyzed, relating demographic and income distribution data to records of available housing and its utilization.

4. A mass transit grant for city K is being considered. Is the proposed corridor the best to serve the most pressing commuter needs? At what level of service will saturation be reached in this corridor? In this case, block and tract data on population and housing characteristics, labor force data, and journey-to-work patterns should be quite helpful.

The list of problems and the variations from place to place are literally never ending. While decisions rest on judgements, the accuracy of the decision is almost invariably better if it rests on empirical data rather than intuitive opinions formed from memory. These informational needs are currently served by three principal data collecting activities. A summary statement followed by a fuller discussion of each may be helpful.

Major Sources of Land-Oriented Data

First, the decennial census of population and housing provides the basic source of information about the housing stock and its occupants, and about the labor force and its characteristics. The 1970 information is expected to cover more items of pertinent information than any prior census, and to be available for more land units than ever before. The wealth of data provided by a decennial census helps to measure the past accomplishments of programs, and by reflecting national needs, furnishes guidance for re-orientation of older programs and the need for new ones. This is the grist for solid analysis on a national scale, as well as for smaller areal units down to the tract and block.

Second, the operating records and files of HUD are a major statistical resource, partly because they show what HUD programs have

done, and where they have been used. The cumulative involvement of mortgage insurance, low-income public housing, urban renewal and college housing is quite substantial in many urban areas. Other programs of more recent vintage are already making their mark. The new Department is undertaking to create an integrated information system, pulling together the operating records of the various activities previously conducted by the five constituent agencies of HHFA. This system is needed for both internal management and for broader program analysis.

Third, since computers began to find a place in the offices of local governments, the era of municipal data banks has begun to bud and some blooms are already open. While the first vision of the potential uses of parcel-based data banks may have been seen in the decade of the 1950's, few serious efforts were undertaken until the statistical sixties arrived. Early in 1961, five cities in southwest States submitted an application for a demonstration grant under Section 314, the urban renewal demonstration program, and the Maryland National Capital Planning Commission published a research monograph financed under the urban planning assistance program, Section 701. Concurrently, transportation planners, bogged down in O and D records, also began to turn to land parcel records as a way to anticipate urban traffic rather than see what it had been. From these beginnings, the metropolitan data bank concept has taken root in various forms in a growing list of communities. Where the installation has been able to support comprehensive planning or the community renewal program, assistance from HUD has been available.

In designing automatic data processing systems, many problems have bubbled to the surface and solutions satisfactory to the users have been developed. The record of some of these has been described in the growing volume of reports and articles, but some system designers who have failed to heed the experience of others have paid the price of reliving that experience.

Unit Identification Plans

Every mass data handling system must begin with a unit identification plan. These plans can be distinguished by three major categories in a simple taxonomy.

1. Arbitrary numbers

If the unit is a parcel of land, the identification plan may be based on an arbitrary numbering system unique to the particular set of records. Some property assessment files are so numbered. But where such a unit identification plan is used, there can be no compatibility or comparability with other data files unless a secondary identification which can be recognized geographically is also added to each unit record, or unless a cross-index is compiled.

Arbitrary numbering plans may have the attributes of simplicity and economy. For instance, mortgage loans which are secured by a lien on a parcel of real estate are often identified by the lender and/or the underwriter by means of a serial number containing a prefix or a suffix, or both, which relate to geographical areas such as a State or a branch office jurisdiction. Number assignment in a serial system is simple and automatic. Such a system may also serve as an operations control and a statistical source when there are no number voids or cancellations. But serial numbers are of little use for geographical analysis unless the prefix or suffix is quite sophisticated and detailed. Nevertheless, serial number plans are very attractive to some commercial types of activity, both governmental and private, and probably predominate in the universe of unit identification plans.

2. Geo-political units

Another frequently used plan is that of identification in terms of some geo-political system. Street names and numbers come under this heading, as do census blocks and tracts, tax-roll description, postal zones, wards, precincts, town and city identification, and other such areal elements. Sometimes two or more geo-political units are identified in the heading in order to facilitate the computation of aggregates in terms of such units or to identify cognizance or responsibility. Fire, police, voter registration, insurance, school enrollment and similar data files would be expected to require geo-political identifications.

The geo-political unit was a necessity in the pre-computer era and continues to perform useful functions. Manual records and punched cards often require the notation of larger areal units in order to be economical of time and cost. These limitations are not as compelling when data files are maintained in modern electronic storage.

Boundaries, names, and numbers of geo-political units can be changed by those who create them, and are revised whenever necessary. This risk must be considered when the unit identification plan is designed, recognizing the time and cost involved in file revision as a trade-off against the convenience of ready access to large areal unit location. If the file is expected to be maintained and used for a long time, i.e., for several decades covering the life of a building or the maturity of a long-term lien, file revision may be a necessity and prove costly. But if the file is to be used for a one-time operation or have an expected life of, say, a decade or less, or if the areal unit is expected to be very stable--such as county boundaries--the risk of revision may be small.

Time series analysis based on geo-political unit aggregates also involves a hazard. Boundary

revision raises the question of whether absolute location is more significant than areal unit cognizance. For example, the granting of Statehood to Alaska and Hawaii has made time-series analysis of many U. S. statistics somewhat more difficult to compile and certainly more cumbersome to describe.

The continuous process of urban annexation, sometimes of small tracts and other times of large areas, is an ever-present hazard for the urban economist and demographer. This may be illistrated by the population and the area of the city of Los Angeles; both of which have expanded greatly in the 20th Century. For some purposes, it may be meaningful to compare the population living within the city limits in 1900 with the population inside the 1960 boundaries. In other cases, it might be necessary to know how many people in 1960 were living inside the 1900 boundaries or how many in 1900 lived within the city area of 1960.

3. Grid systems

The third method of notation that may be used in a unit identification plan involving geographic location is a grid system with a permanent spatial anchor. Undoubtedly, the most widely used grid locational system is that of latitude and longitude. This method differs from the other two systems in that the intersection of grid coordinates identifies a single point, whereas arbitrary numbers and geo-political units usually refer to a land parcel or area.

If the grid scale is small enough, a single point within a parcel may provide sufficient identification, or several points may be needed to describe boundary lines. Grid point assignments to land parcels involve problems and costs, but once made are unchanging except where parcels are subdivided or merged.

Such plans overcome many of the shortcomings of other methods of identification and are adaptable to computer operations. If used in conjunction with geo-political systems, either as primary or secondary identifiers, grid coordinates provide a ready means of file correction if adjustments are made in either the boundaries or the numbering plan of geo-political units.

We hear that there are some local land parcel data files that employ grid systems other than latitude and longitude. Where these are private installations serving a special purpose, such as an electric power company, an arbitrary grid system scaled for its particular needs may be more suitable and easier to use. In such cases, comparability with other data files may be of no concern. However, where public purposes are involved and the uses of the file may be varied, it would seem that the unit identification plan would be most useful if it carried grid intersects, together with whatever other locational descriptions and unit classes might be helpful.

Recognition of Needs for Comparability

Machine processed files require compatibility if comparative studies are to be made, or if new elements are to be added from other files. We find a growing recognition of this need, as computer applications increase in number and complexity, and as equipment becomes available to do larger and more difficult jobs.

1. Census plans

The Bureau of the Census has stated that it proposes to identify latitude and longitude in degree decimal terms in connection with block-face addresses being developed for the 1970 Census of Population and Housing. Land parcels in Census tracts and in about 1½ million blocks will be so identified. The Census Bureau will discuss this procedure in a paper to be presented this afternoon. This procedure has far-reaching consequences, and opens possibilities for research with 1970 data that have heretofore been impossible.

2. HUD information plans

The new Department of Housing and Urban Development is undertaking to construct an information system to record and describe the activity in each of the many programs it administers. This system will encompass the millions of land parcels securing FHA-insured loans; the manifold characteristics both before and after execution of nearly 2,000 urban renewal projects; low-rent public housing projects going back to 1937; and scores of other programs ranging from college housing loans to open space grants from mass transit loans to code enforcement grants and public facility loans, not to mention FNMA's portfolios.

All of these records are now in various kinds of filing systems, ranging from dockets to manual card records to magnetic tape. Each individual case has some kind of a serial number and some form of geographic location reference, and all have urban land characteristics. The translation of these records into a national information system so that HUD can readily determine what it is doing--and has done--in any and every urban sector is a tremendous undertaking and will be a lengthy and detailed process. But it is essential in order to carry out the broad urban responsibilities which the Congress and the President have now assigned to the Department. The ultimate benefits will far outweigh the initial investment, not only in making program administration more effective but also more efficient.

While HUD has not yet reached the stage of planning the details of its overall information system, it may ultimately adopt the same latitude and longitude land parcel descriptions that the Bureau of the Census plans to use in 1970. These would be carried alongside the case, project, or loan numbers and various geo-political unit identifications. One of the more obvious analytical gains would

be the opportunity to aggregate HUD records into Census blocks and tracts for the study of program activity in relation to small-scale universes.

3. Metropolitan data banks

During recent years, HUD and its predecessor agency have given some financial support to the development of metropolitan data banks based on land parcels. This support has come from the urban renewal demonstration program (Section 314), the urban planning assistance program (Section 701), and the community renewal program. Meanwhile, other data banks organized on land parcels have come into being without HUD assistance.

These installations have interesting possibilities both as tools in comprehensive planning and as aids to efficient local governmental operations.

There are now enough of these local data banks in operation to permit a study of their various features as well as the uses to which they have been put. Such a survey is about to be undertaken and is expected to be completed before the end of the current fiscal year. We hope that this appraisal will help the Department in shaping its policies regarding future support of these efforts.

Without seeking to direct the form and structure of local proposals, it would seem reasonable that HUD would probably encourage local agencies to incorporate the same system of land parcel identification that both Census and the Department are expecting to use, viz., a grid coordinate system mathematically translatable into latitude and longitude or into some other plan of coordinates. Inclusion of grid location reference points in local data banks would open up compatibility with Census block and tract data, giving local government the same kind of opportunities that HUD would gain in its own information system.

The Problem of Disclosure

Serious concern is being expressed in various quarters about the potential invasion of privacy that is threatened by computer-based data. This threat arises partly because the capacity of electronic storage makes possible the combining of many data series that previously had to be maintained in separate files under separate jurisdictions. It has been pointed out that the picture described by the totality of several data series presented in concert is different from the view of each by itself. Even though each item could be publicly researched, the mosaic made by the total is one that only a computer could readily produce.

This contention is probably not debatable, for it is a method of analysis that has been successfully used in both financial credit review and in military intelligence to name only two examples. It is important, therefore, that the structure of data files be carefully examined to anticipate whether control of the file by persons who were not benevolent in their motives could be harmful or capricious. It may be helpful to examine the policies of the three principal collectors and repositories of geographically identified data that have been discussed in considering safeguards against the risk of invasion of privacy.

First, the Bureau of the Census operates under very strict laws and regulations regarding disclosure. Its publications and releases are required to show only aggregates which are sufficiently large to preclude identification of individual components. The apprehensions expressed about the invasion of privacy do not relate to Census releases, and the Bureau has never been viewed as a threat because of its laws and its vigorous administration of them. The activities of the Census Bureau relative to this issue are twofold: (a) block and tract data that can be used as benchmarks for other studies; (b) the techniques being developed for translating city street addresses to grid coordinates. Both of these are tools of policy analysis and program planning, but hold no threat to privacy.

Second, some of the programs of HUD deal directly with individuals and private industry, viz., FHA and FNMA, while others deal largely with various kinds of public agencies and local governments. In all of its nonpublic actions, the Department carefully guards the privy relation with its program participants. Its files are treated as confidential as a physician's clinical records of his patients. Where public bodies participate in the Department's program, their actions and the details of the project are usually public information in the offices of the municipality. Public projects often involve open public hearings to assure full disclosure and consideration of local decisions. These actions hold no threat to personal privacy.

Third, local data banks differ from the Federal activities discussed above. They can, and some do, contain a wide range of information, varying from land parcel items to facts about the people who live or work at various addresses. It is the latter type of data that offers the threat to personal privacy.

The recording laws of each of our States require public recording of deeds and liens to establish seniority of interest. This is the essence of land tenure policy in the United States, and our mortgage and real estate industries have developed around this principle. Other local public land records relating to assessments, taxes, zoning, building permits and similar activities are available to anyone who takes the trouble to inspect them. In fact, it is difficult to identify any type of land parcel information in local governments that is not available to the public, except, of course, preliminary planning of future public land acquisition for such things as rights-of-way, public buildings, open space, etc.

In the case of social and demographic data, the picture is different; the threat of

invasion of privacy is a potential reality. It would be helpful in the administration of many local social, educational and welfare activities to be able to make small-scale geographical studies of the incidence of crime, truancy, disease, and arson to name only a few items. Also, geographical patterns of income distribution, educational levels and other economic and social data have a use in capital budgeting and many elements of comprehensive planning. But as long as there is public concern about the threat of disclosure of personal information, we should expect to find public resistance to its inclusion in data bank storage.

Possibly there is a way out of this dilemma by following the practices of the Census Bureau. For practical purposes, the needs of local planning and analysis would probably be served just as well by aggregates for small areal units as by maintaining data about persons on an address basis. If the inputs to data banks from files in other municipal departments were limited in detail to a geographic level that protected individuals against possible adverse disclosure, the logic supporting the critics of this analytic tool would vanish.

However, the maintenance of public confidence rests more on faith in the integrity of those intrusted with data than it does on logic. A single indiscretion or an innocent error might destroy the confidence built up over many years in many separate cities. Hence, it will be well for those who design and those who operate local data banks containing personal, social and economic information, to examine the structure of inputs and storage carefully lest outputs impinge on personal rights. Failure to do as much could jeopardize the longer range use of modern equipment that promises help with some of the most difficult kinds of public decisions.