INTRODUCTION

The Need

The rapid growth that has occurred in this country since World War II has led to a genuine concern for the urban decision-making process. It has been recognized that better data are needed in order to make appropriate decisions in the changing urban environment. As a result, a wealth of information has been gathered about the urban area. Almost every facet of urban life has been examined. Data have been collected on many aspects, from the health of the individual to the number of trips he makes. Until recently, most of this information has been gathered by a specific agency or committee for its own use. Individual definitions for data characteristics have prevailed. Definitions of land use density, for example, can range from trips generated per acre to people per square foot of floor space. The definition has depended upon the information user. The result of this multiplicity of data definitions and uses is a hodgepodge of data, collected many times by many agencies without knowledge of each other's efforts (1).*

Public agencies gather much of their data for normal operations; these data are potentially very useful to other agencies at little additional cost, if only common definitions and parameters could be established. Without these common definitions and parameters each agency sees the urban environment only in the perspective of its narrowly defined information requirements (1).

Each urban area has developed a multiplicity of plans to channel the growth of the area in a manner that is deemed best for the community as a whole. Up to the present, just as the different agencies collect and use their own data, so do the various urban studies. While a recreational or school plan may use much of the data collected in a transportation study, the difference in data definitions and aggregation units makes the information nearly useless for any study other than the one for which the data was collected.

* Numbers in parenthesis refer to numbers in list of references.
Many segments of the urban environment desire information on the community. The present means of getting this information is to go out into the urban area and collect the data directly. Any company, organization, or group that now wants data on the community must collect the information itself or accept the narrow definitions of data now established by existing governmental groups or studies that have collected data for their own purposes.

In order to facilitate the use of data by other than the primary information receiver, a set of universally compatible definitions is needed. This set of definitions is not impossible to develop if one attempts to direct the collection of "pure" data. The term "pure" simply means that the information should not be aggregated before collection. For instance, when square feet of space is collected, it should be recorded as square feet, not square feet per some other dimension. For example, square feet per employee may be useful to an industry, but square feet of building and number of employees is much more useful for planning purposes while still serving the purpose originally intended.

The data system that is described below seeks to develop a tool for urban decision-making that utilizes data from many sources and makes this information available and usable by other sectors of the urban community.

The Scope

The project described involves the development of an urban data system for an area of approximately 100,000 population. The Lafayette-West Lafayette area was used to demonstrate application. The first developmental problem involves choosing the degree of sophistication needed for such a system. This involves choosing a particular level in a hierarchy of data system complexity. Once the level of sophistication has been decided, the basic data collection and aggregation module must be chosen. The data to be used must be decided along with specific definitions for each data item. How these data can be entered into the system should be developed along with updating procedures to keep information current. A logical and easily usable means of data storage and retrieval must be developed to facilitate the use of the system by a wide variety of users.

DESCRIPTION OF EDSARS

Level of Data Sophistication

The information used in the Environmental Data Storage Retrieval System (EDSARS) is taken from the level in data hierarchy of a data library. Banked data, which is another level in the information
hierarchy, are organized into machine records but need not be functional or logical in format. Raw data make up the lowest level of data sophistication. These data are not machine digestable and therefore are not usable in an organized data system. The data library information is logical and functional in format and can be updated, searched, and retrieved; these requirements are essential for any urban data system.

**Level of System Sophistication**

The three levels of system sophistication vary in the complexity of models incorporated. The first level uses no models, the second utilizes specialized models, while the third level uses simulation. EDSARS, being an attempt at developing the initial phase of an urban data system, utilizes the first level of sophistication. The system contains tabulated data but no specialized or simulation models. It is felt that the model requirements will evolve from the demands of the users on the data system. The addition of models to the system can be made within the present format; the data in the system will directly feed any models developed in the future.

The computer hardware that is incorporated also influences the level of system sophistication. EDSARS utilizes the CDC 6500 computer at Purdue University. This is a general purpose computer; the programming language used is Chippewa Fortran. The data system can be initialized and information retrieved or updated by merely submitting the correct program deck to the computer science center. The updating, retrieval or initialization will be run just as any other job that is submitted to the computer. The information for EDSARS is now stored on tape. As the system is initialized and the amount of stored information grows, the incorporation of a disk pack will become feasible. A disk pack is a mountable disk storage device that enables random access of information. This direct access feature will save valuable computer time when the system searches large quantities of data.

The decisions on the level of system sophistication were the result of many factors. Models were not incorporated into the system because of the necessity of actual data to test the validity of a model. This project outlines the initialization of EDSARS without actually inserting real data. The amount of data needed to initialize the system makes initialization another entire project of at least one to two years in duration. Once the initialization is complete, then the addition of models can be considered.

The decision of using the CDC 6500 computer was made in light of the hardware available. Purdue University now has an IBM 7094 computer which could handle a data system such as EDSARS, how-
ever, the 7094 is a second generation computer. This type of computer is now in the process of being phased out by many organizations, being replaced by a third generation computer such as the CDC 6500. Any work done in the future on data systems will most probably be done on the more advanced type of equipment. Chippewa Fortran was used because of the author's knowledge of the language and the fact that the Fortran developed for the CDC system is quite efficient in its data-handling capabilities.

**Data Module**

The data module chosen for EDSARS is the parcel. This aggregation module seems to be almost the universal choice by existing urban data systems. The parcel provides a flexible, multi-purpose base from which to work. The data to be incorporated into an urban data system are easily keyed to the parcel. The tagging methods, which will be discussed below, work well with the parcel module. The parcel forms a very useful aggregation unit in that it is the largest common denominator that can be used to build zones. Any zone in an urban area can be represented fairly accurately, by a composite of parcels. This gives the system maximum flexibility in the designation of zones while containing a minimum number of data units.

The parcel in EDSARS is defined as all contiguous land under one ownership and one general land use. This definition closely parallels the parcel used in assessors' records. If two adjacent pieces of land are owned by the same person and used for the same purpose, they would be listed as one parcel. If two adjacent pieces of land have different uses, they would be listed as two parcels. This definition, being general, allows a certain measure of ambiguity in the designation of a parcel; the system has the facility, however, of being able to join two or more parcels into one new parcel, or break up one parcel into two or more parcels. This facility for redefining parcels allows the system to establish its own equilibrium as the data are used and reevaluated.

A special definition of the parcel is utilized when coding rights-of-way. Each street segment and utility right-of-way is coded just as any other parcel. A street or right-of-way is broken down into block-long segments if the block length is 500 feet or less; if the block length is longer than 500 feet, the block is broken down into segments of 500 feet or less. An intersection is taken as a street parcel. The parcel boundaries are defined as the right-of-way line for the street segments. An example of an area divided into parcels can be found in Figure 1.
Data Tagging Methods

EDSARS uses both the name method and location method of tagging data. The name tag utilized is the street address of the parcel. The street number, name and type (e.g., Drive, Street, Lane, etc.) are all noted in the name tag of the parcel. For rural areas, the street number is replaced by the rural route number, and the street name is replaced by “Rural Route.” The name method of tagging gives the system the facility of locating data for the user on a basis that is familiar to all segments of the urban environment. Street addresses are universally known and understood, and therefore enable all potential users to be familiar with at least one retrieval method.

Street segments are coded by the street name and the number (in hundreds) of most of the houses on the street segment. A street segment
along a street called Main Street, where house numbers go from 100-225 would be coded as 100B Main Street. This indicates that this is the one hundred block of Main Street. This gives the benefits of the name tag to street segments as well as individual parcels.

The location tag utilized by EDSARS is a rectangular grid coordinate system which is superimposed over the entire development area. The grid coordinate uses one foot as the basic unit. The parcels and street lengths are tagged by the coordinates of their approximate centroid. The actual digitizing of the coordinates is accomplished by an automatic coordinate digitizer. By utilizing a location tag, internal logic is added to the data in EDSARS. The coordinates facilitate the retrieval of data on an areal basis. Data for certain geographical segments of the development area can be directly retrieved with the use of coordinates. Density computations become immediately possible with the use of coordinates.

The utilization of rectangular grid coordinates provides another very useful capability. A zone, such as a census tract or transportation zone, can be represented by the grid coordinates of its boundary. This is accomplished by representing the zone by a series of triangles and digitizing the coordinates of the vertices. By representing zones in this way, a dictionary of zone names and grid coordinates is developed. When any information is desired on a zonal basis, the coordinates of the zone are read and each parcel is tested to establish whether or not it lies within the zone in question. The information for each parcel within the zone is then retrieved and aggregated thereby giving information on the desired zonal basis. Figure 2 shows a zone broken into triangles for coding.

In order to coordinate the actual data incorporated into the system and the tags for each parcel, a dictionary with the parcel number, street address (or block number for street segments), and grid coordinates is developed. The actual data are stored in conjunction with a parcel number. The data are related to the name and location tags through the parcel number-street address-grid coordinate dictionary. The parcel number is merely a unique number of one to six digits given to each parcel. The numbers need not be consecutive or have any logical order. The only requirement is that each parcel have one and only one unique number.

Data Dimensions

The definition of land use developed by the Metropolitan Washington Council of Governments was utilized to aid in determining the data needed to define the different areas of land use. Data were examined in the light of how well they defined:
In order to completely describe the urban environment, the information on each parcel is broken down into three categories:

1. Parcel Information—information on the parcel itself, including dimensions, restrictions, zoning, use, etc.
2. Building Information—information on each building in a parcel, including age, value, type of construction, condition, size, etc.
3. Establishment Information—specific information on each unit within a building such as a business, a dwelling unit, etc., including space use, number of employees, number of residents, age of residents, number of vehicles, etc.

Fig. 2. Zone Divided into Triangles for Data Coding
Data Items

With the tagging methods established, the module chosen, and the data categorized within the parcel, it now becomes possible to choose the actual data items to be incorporated into the system. The methodology for this is three-fold:
1. Examine data collected by existing urban data banks.
2. Establish the data needed for all urban planning functions.
3. Query local planners as to data needed for operations and planning.

Once a complete list of data is collected by the above means, a thorough evaluation of each data item is needed to establish its usefulness, its cost of collection, and its ability to be updated.

To determine the types of data collected for existing data banks, a study of literature on existing and planned urban data banks was made. From this literature, a list was made of data collected. These data were then categorized under parcel, building or establishment information. Data needed for all phases of urban planning were then examined. A case study was made of the Louisville Metropolitan Area Transportation and Development Study to determine the data collected for this project (3-16). The type of data needed for a comprehensive environmental health plan was determined (17). General notes on the data collected for other urban studies, including a recreational plan, school plan, urban renewal plan, major thoroughfare plan, transportation plan, land use plan, and an economic development plan, were studied for the data needed (18). All listed data items were then appraised by local planners. The data incorporated, listed by general category, are as follows:

I. Parcel Information
   1. Land use
   2. Ownership
   3. Frontage
   4. Area
   5. Year of subdivision
   6. Assessed value of land
   7. Easement
   8. Landmark
   9. Neighborhood characteristics
   10. Land appearance
   11. Number of structures
   12. Year of zoning change
   13. Zoning
   14. Zone change request number
   15. Variance number
   16. Comprehensive plan use
   17. Utilities
   18. Parking spaces
   19. Loading area
   20. Assessed value of improvements
   21. Total assessed value
   22. Sale date
   23. Sale price
   24. Nuisances
(The following data are collected for street segment parcels)

25. Intersection
26. Length of segment
27. Right-of-way width
28. Pavement width
29. Functional class
30. Structural composition
31. Percent grade
32. Average daily traffic
33. Number of accidents
34. Traffic control signs and signals
35. Speed limit
36. Curb parking regulations
37. Curb type
38. Sidewalks
39. Number of lanes
40. Loading zone
41. Bus route
42. School route
43. Access control
44. Condition

II. Building Information

1. Year built
2. Type of construction
3. Type of structure
4. Building condition
5. Year of latest building permit
6. Cumulative cost of building permits
7. Number of floors
8. Total floor area
9. Basement
10. First floor area
11. Number of dwelling units
12. Building setback
13. Required setback
14. Rehabilitation cost
15. Type of building code violation
16. Number of building code violations
17. Number of establishments

III. Establishment Information

1. Space use
2. Total number of employees
3. Peak shift number of employees
4. Optimum total employment
5. Number of rooms for rent
6. Number of residents by sex and age group
7. Family income
8. Vehicles owned
9. Police calls
10. Fire calls
11. Welfare payment
12. Number of communicable diseases
13. Type of communicable diseases
14. Rent

Each piece of data that was entered into the system was judged to be important to the planning community, able to be updated, and relatively easy to collect. Data that were too expensive to collect or not updatable were not incorporated into the system.
CONCLUSIONS

The following conclusions about EDSARS and its potential can be made:

1. EDSARS should facilitate efficient and economical handling of planning data for an area of about 100,000 population.

2. The utilization of a general purpose computer and general purpose programming languages should make EDSARS available to most metropolitan areas in the United States.

3. The concept of a unified data system is the most important contribution of EDSARS.

4. The data proposed for EDSARS is the most usable and easily obtainable information available to the urban area.

5. The incorporation of a flexible method of representing zones by their location is essential to an efficient urban data system such as EDSARS.

6. The information for an urban data system should be in three separate files so that one file can be updated and improved without disturbing the other files.

7. Zone names and boundary locations should comprise one file; parcel numbers, parcel location, and street address should compromise another file, and the third file should be made up of general data.

8. The best unit for data collection is the parcel.

9. The data system should be flexible so that improvements can be made as use and technology increase.

10. The streets and rights-of-way should be represented as special parcels in order to insure full territorial and information coverage.

11. All data incorporated should be potentially useful and updatable.

12. Utilization of applicable theory and practical experience of existing data systems are needed to develop a useful, efficient and improved data system.

The concepts that are represented by these conclusions when tied together into an urban data system such as EDSARS give the planning community and the urban environment as a whole a flexible and useful tool which should make more information available to more people at a much lower cost and with much less effort.

The details of the EDSARS system, including initialization, update and retrieval programs, codes utilized, and complete descriptions are available in a Purdue University Joint Highway Research Project Information Report No. 5, March 1968. Copies are available at the cost of reproduction.
This project was funded by the Public Health Service through the Environment Health Institute of Purdue University.

LIST OF REFERENCES
18. *Planning Methodology and Techniques*, class notes for CE 697, Prof. W. L. Grecco, Purdue University, 1966.