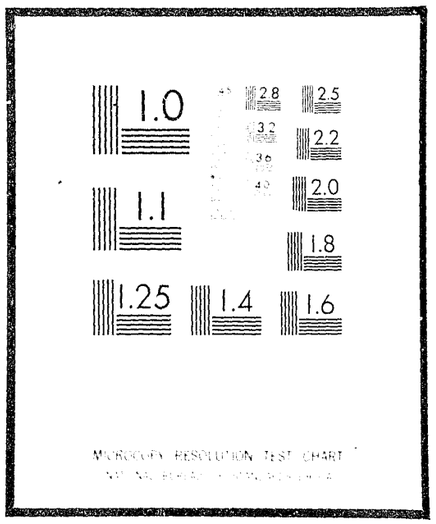


Geographic Base (DIME) System— A Local Program

Computerized Geographic Coding
Series GE 60 No. 6

Conference proceedings

November 18 and 19, 1974
Columbus, Ohio

Issued April 1975

U. S. DEPARTMENT OF COMMERCE

James L. Pate, Assistant Secretary
for Economic Affairs

Social and Economic Statistics Administration
Edward D. Failor, Administrator

BUREAU OF THE CENSUS
Vincent P. Barabba, Director

BUREAU OF THE CENSUS

Vincent P. Barabba, Director
Robert L. Hagan, Deputy Director
Alva L. Finkner, Associate Director for Statistical Standards and Methodology

Geography Division
Morton A. Meyer, Chief

ACKNOWLEDGMENTS

This report was prepared under the supervision of Jacob Silver, Assistant Division Chief for Operations, Geography Division. Ann Liddle and Peggy Heekin were instrumental in organizing the conference and in compiling and editing these proceedings.

The Bureau of the Census is extremely grateful to Kline Roberts, President, Columbus Area Chamber of Commerce, and William C. Habig, Executive Director, Mid-Ohio Regional Planning Commission, for cosponsoring the conference and for their fine cooperation. We wish to give special thanks to James R. Thomas of the Columbus Area Chamber of Commerce, and Eileen Flowers and Ronald Henderson of the Mid-Ohio Regional Planning Commission for their efforts in making the conference a success.

For **Library of Congress Cataloging in Publication Data,** see inside back cover.

For sale by the
Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402
Price $1.90

PREFACE

This report presents the proceedings of the sixth in a series of conferences devoted to the Census Bureau's Geographic Base (DIME) System. This conference was held in Columbus, Ohio, on November 18-19, 1974. The papers presented and the resulting discussions centered on the theme "Geographic Base (DIME) System—A Local Program." The question and answer sessions have been edited for conciseness of presentation.

The purpose of this series of conferences is twofold: first, to improve communication among those agencies and organizations which are assisting the Census Bureau in maintaining the file for their respective areas; and second, to provide a vehicle for the mutual exchange of information and experience in the management and use of the files.

Copies of the first five conference proceedings—

1. U.S. Bureau of the Census, *Use of Address Coding Guides in Geographic Coding—Case Studies,* Report GE 60 No. 1, Washington, D.C. 1971 ($1.15)

2. U.S, Bureau of the Census, *Geographic Base Files—Plans, Progress, Prospects,* Report GE 60 No. 2, Washington, D.C. 1971 ($1.40)

3. U.S. Bureau of the Census, *Geographic Base File System—Uses, Maintenance, Problem Solving,* Report GE 60 No. 3, Washington, D.C. ($1.65)

4. U.S. Bureau of the Census, *Geographic Base File System—Establishing a Continuing Program,* Report GE 60 No. 4, Washington, D.C. 1973 ($1.50)

5. U.S. Bureau of the Census, *Geographic Base (DIME) File System—A Forward Look,* Report GE 60 No. 5, Washington, D.C. 1974 ($2.00)

can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, or from any Department of Commerce district office.

Participants

Dr. Robert T. Aangeenbrug
Resident Scholar
Office of the Associate Director for Statistical Standards and Methodology
Bureau of the Census
Washington, D.C.

Mr. Harvey N. Albond, AIP
Director of Planning and Development
City of Niagara Falls
Niagara Falls, New York

Ms. Martha Alexander
Acting Director
Long Range Planning Division
Urban County Planning Commission
Lexington, Kentucky

Mr. Vincent P. Barabba
Director
Bureau of the Census
Washington, D.C.

Mr. Kenneth E. Bare
Acting Manager
Information Service
Tampa Bay Regional Planning Council
St. Petersburg, Florida

Mr. Tom Beerman
Battelle Memorial Institute
Columbus, Ohio

Mr. Forrest K. Betche
Associate Planner
Lima-Allen County Regional Planning Commission
Lima, Ohio

Ms. Susan A. Brayford
Research Support Specialist
Research Division
International Association of Chiefs of Police
Washington, D.C.

Mr. Frederick R. Broome
Special Assistant to the Division Chief
Geography Division
Bureau of the Census
Washington, D.C.

Mr. Don Brown, Jr.
Franklin County Hall of Justice
Columbus, Ohio

Mr. Donald Bufkin
Program Manager
Tucson Area Transportation Planning Agency
Tucson, Arizona

Mr. George Butzer
Columbus Division of Traffic Engineering
Columbus, Ohio

Mr. Roy K. Campbell
Vice President
Director of Research and Planning
Manufacturers Hanover Trust Company
New York, New York

Mr. Larry Carbaugh
Acting Chief, Users' Services Staff
Data User Services Division
Bureau of the Census
Washington, D.C.

Mr. Jeff Connell
Planner
Lorain County Regional Planning Commission
Elyria, Ohio

Mr. Aurel Damian
Columbus Division of Engineering and Construction
Columbus, Ohio

Mr. Jim Davis
Community Planner
Columbus Area Office
Department of Housing and Urban Development
Columbus, Ohio

Dr. Joseph W. Duncan
Deputy Associate Director for Statistical Policy
Office of Management and Budget
Washington, D.C.

Mr. John Evans
Research Associate and Project Coordinator
Center for Urban Studies
University of Akron
Akron, Ohio

Dr. Alva L. Finkner
Associate Director for Statistical Standards and Methodology
Bureau of the Census
Washington, D.C.

Ms. Eileen Flowers
Mid-Ohio Regional Planning Commission
Columbus, Ohio

Mr. James Gannon
Area Development Planner
Bureau of Urban Systems Planning
Ohio State Department of Highways
Columbus, Ohio

Mr. Forrest B. Garner
Programming Officer
Huntington National Bank
Columbus, Ohio

Mr. Paul Gatsch
Columbus and Franklin County Criminal Justice Coordinating Council
Columbus, Ohio

Ms. Paula Gavulic
Genesee County Metropolitan Planning Commission
Flint, Michigan

Mr. Charles Gebhardt
Transportation Engineer
Ohio Department of Transportation
Columbus, Ohio

Mr. Bob Gebhart
Columbus Public Schools
Columbus, Ohio

Mr. Ralph Giambra
Dataware
City of Niagara Falls
Niagara Falls, New York

Mr. Jim Giffin
Manager
Polimetrics Lab Special Projects
Ohio State University
Columbus, Ohio

Mr. Ken Graska
Mid-Ohio Regional Planning Commission
Columbus, Ohio

Mr. Charles Groves
Administrator
Bureau of Transportation Technical Services
Ohio State Department of Transportation
Columbus, Ohio

Mr. William C. Habig
Executive Director
Mid-Ohio Regional Planning Commission
Columbus, Ohio

Mr. Joseph Hadley, Jr.
Planner
Tri-County Regional Planning Commission
Akron, Ohio

Mr. Roger L. Haiflich
Software Analyst
Allen County Data Processing
Fort Wayne, Indiana

Mr. Richard E. Hartgen
Geographic Planning Specialist
Geography Division
Bureau of the Census
Washington, D.C.

Ms. Peggy Heekin
Geographer
Geography Division
Bureau of the Census
Washington, D.C.

Mr. Ronald Henderson
Mid-Ohio Regional Planning Commission
Columbus, Ohio

Ms. Francine Hochman
Montgomery County Planning Commission
Dayton, Ohio

Mr. Mike Hogan
Assistant Manager
Marketing Research Department
Huntington National Bank
Columbus, Ohio

Mr. John Holl
Chief Planner
Lorain County Regional Planning Commission
Elyria, Ohio

Mr. N. Jack Huddle
Director
Department of Development
City of Columbus
Columbus, Ohio

Mr. Thomas E. James
Director
Benchmark Data Center
The Academy for Contemporary Problems
Columbus, Ohio

Mr. Robert Jilla
Transportation Planner
Brooke-Hancock-Jefferson Metropolitan
Planning Commission
Steubenville, Ohio

Dr. Robert E. Johnson
Columbus Health Department
Columbus, Ohio

Ms. Ann Jones
Assistant Housing Administrator
Columbus Division of Planning
Columbus, Ohio

Mr. James A.F. Kelly
Assistant Director, Research Division
International Association of Chiefs of Police
Washington, D.C.

Mr. Edward L. Kutevac
Director
Trumbull County Planning Commission
Warren, Ohio

Mr. Santo J. LaTores
Information Division
Systems Research and Analysis Directorate
Transportation Systems Center
Department of Transportation
Cambridge, Massachusetts

Mr. Robert Layton
Northeast Ohio Areawide Coordinating Agency
Cleveland, Ohio

Mr. Dale Little
Columbus Division of Police
Columbus, Ohio

Mr. Richard Luckay
Assistant General Manager
W.A. Storing Company
Columbus, Ohio

Mr. John McClain
Task Force Coordinator
Ohio State Department of Transportation
New Philadelphia, Ohio

Mr. Robert McWilliam
Director
Detroit Regional Office
Bureau of the Census
Detroit, Michigan

Mr. Anthony J. Manch
Liaison Engineer
District Six
Ohio Department of Transportation
Delaware, Ohio

Dr. Edward Menge, Jr.
Manager
Behavioral Science Laboratory
Ohio State University
Columbus, Ohio

Mr. Morton A. Meyer
Chief
Geography Division
Bureau of the Census
Washington, D.C.

Mr. John D. Meyers
Transportation Analyst
Niagara Frontier Transportation Committee
Buffalo, New York

Mr. Bruce Miller
Columbus Division of Planning
Columbus, Ohio

Mr. Neil K. Mills
Battalion Chief
Columbus Fire Department
Columbus, Ohio

Ms. Rebecca Minshall
City of Columbus
Columbus, Ohio

Mr. Tom Moody
Mayor
City of Columbus
Columbus, Ohio

Mr. Harold Nisselson
Chief Mathematical and Statistical Advisor
Bureau of the Census
Washington, D.C.

Ms. Sally Odegaard
Director
Ohio Government Information Center
Columbus, Ohio

Mr. Bob Orr
Columbus Division of Planning
Columbus, Ohio

Mr. Jack Pascoli
Planning Engineer
Urban Planning Division
West Virginia State Department of Highways
Charleston, West Virginia

Mr. Allen L. Pearson, II
Research Division
International Association of Chiefs of Police
Washington, D.C.

Mr. L.J. Pendleton
Director, Market Research
The Ohio National Bank
Columbus, Ohio

Mr. Alan Pisarski
Special Assistant for Information Policy
Office of the Secretary
Department of Transportation
Washington, D.C.

Mr. Michael Plett
Systems Analyst
Ohio-Kentucky-Indiana Regional Council of
Governments
Cincinnati, Ohio

Ms. Lois Preston
Department of Development
City of Columbus
Columbus, Ohio

Mr. Charles F. Pscholka
Richland County Regional Planning Commission
Mansfield, Ohio

Dr. Gerald F. Pyle
Associate Director
Center for Urban Studies
University of Akron
Akron, Ohio

Mr. Kline Roberts
President
Columbus Area Chamber of Commerce
Columbus, Ohio

Mr. Edward Roehrenbeck
Ohio National Bank
Columbus, Ohio

Mr. John E. Rowe
Director
USES
Indianapolis, Indiana

Mr. Sheldon Rubin
Assistant Chief
Demographic Census Staff
Bureau of the Census
Washington, D.C.

Mr. Elias Samaan
Executive Director
Northeastern Indiana Regional Coordinating Council
Fort Wayne, Indiana

Mr. Anthony J. Savoia
Columbus Department of Data Processing
Columbus, Ohio

Mr. Fred Sheldon
Mahoning County Planning Commission
Youngstown, Ohio

Mr. Jacob Silver
Assistant Division Chief for Operations
Geography Division
Bureau of the Census
Washington, D.C.

Dr. James Stevens
Project Director
Law Enforcement Geo-processing Systems
Geography Division
Bureau of the Census
Washington, D.C.

Mr. Herman Stofle
Planning Director
County of Franklin
Columbus, Ohio

Mr. W.A. Storing
President
W.A. Storing Company
Columbus, Ohio

Mr. Michael Stover
Hillsborough County Planning Commission
Tampa, Florida

Mr. Robert M. Stutz
Building Industry Association
Columbus, Ohio

Dr. David C. Sweet
Director
Department of Economic and Community Development
State of Ohio
Columbus, Ohio

Mr. James R. Thomas
Vice President, Research
Columbus Area Chamber of Commerce
Columbus, Ohio

Mr. Clyde R. Tipton, Jr.
Vice President, Communications
Battelle Memorial Institute
Columbus, Ohio

Mr. Mitchell Tucker, Jr.
Director
Management Planning and Systems Department
City of Tulsa
Tulsa, Oklahoma

Mr. Larry Wackerly
Technical Director
Stark County Area Transportation Study
Canton, Ohio

Mr. Jacob J. Wang
District Representative
Stark County Area Transportation Study
Canton, Ohio

Mr. C. Weate
Battalion Chief
Columbus Fire Department
Columbus, Ohio

Mr. Charles Weber
Planner
Ohio Department of Natural Resources
Columbus, Ohio

Mr. Charles B. Wenner
Columbus Safety Department
Columbus, Ohio

Mr. Robert Wexler
Transportation Analyst
Niagara Frontier Transportation Committee
Buffalo, New York

Ms. Gayle D. Wiley
Research Analyst
Miami Valley Regional Planning Commission
Dayton, Ohio

Mr. Forrest B. Williams
Battelle Memorial Institute
Columbus, Ohio

Mr. Robert Zack
Supervisor of Metropolitan Planning
Bureau of Urban Systems Planning
Ohio State Department of Highways
Columbus, Ohio

Agenda

November 18, 1974

MORNING SESSION

AFTERNOON SESSION

EVENING SESSION

November 19, 1974

MORNING SESSION

Chairman's Introduction

MR. ALAN PISARSKI

Good morning. This is the Geographic Base (DIME) System Conference—A Local Program, a conference co-sponsored by the Columbus Area Chamber of Commerce, the Mid-Ohio Regional Planning Commission and the Bureau of the Census. My name is Alan Pisarski, and I will be serving as the chairman of this conference today and tomorrow. I think I should point out that the focus, the emphasis, of this conference is on the local character of a successful program of systems technology.

I have been associated with this program and these conferences for several years. It seems terribly inappropriate with that theme that a Federal employee, such as myself, is up here serving as the chairman. So perhaps I should rapidly attempt to establish my credentials as a spokesman regarding local needs. Number one I was not born a Federal employee. I am convinced that some people were. None in this room, of course. My own professional experience in geographic systems began as the Chief of Coding and Special Projects at the Tri-State Transportation Commission, which is a regional and land use transportation planning agency for the New York City metropolitan area. It is a local program, but I guess in terms of local programs it certainly had a very large locale; we had an 8,000 square mile study area and about 17 million people in the region.

At that time, about 1962, we were preparing coding maps, coding guides, and automated coding systems for coding on the order of a million addresses for the large scale survey programs that were underway. At that time there were no Federal programs in geocoding, as such. There was no such thing as Federal assistance. In fact, a great deal of our time at the local level was spent attempting to convince the Federal people that we had a problem. There was a general problem of cities trying to get the Federal government concerned and interested in local geocoding programs and activities.

I think the roles today are to some extent reversed. Federal people are generating the geocoding tools. They are out promoting local interest and use and providing technical assistance. But I think although the roles have changed somewhat, the prime fact that an effective program is still a local program remains with us. Designing, developing, utilizing, and maintaining a comprehensive area systems technology is something that can only be generated and sustained at the local level.

The Department of Transportation has always operated from this belief, and it has been the basis of its pragmatic approach. We have supported geographic systems for many years, both financially and methodologically, but always with this emphasis on local initiative. Our view is based on the very unique character of cities—the special requirements of each city—the kind of focus that says that each city is going to have its own special applications and needs. The Department will be ready to support those needs, but the Department itself cannot legislate active "use" of such systems.

I know that the Mayor of Columbus, sitting here, is wondering what this is all about. I can only say that at the Federal level it has been my experience that attempting to convince the decision makers in Washington about the importance of this work is probably just as difficult as it is to convince local officials of what it is that we are all attempting to do. When one sits down with an Assistant Secretary and says, "We need two or three million dollars to make this program work throughout the country," he generally looks at you in astonishment. You are in the position that I have often described of telling him that you have found a solution to something that he did not know was a problem. And when you say, "I have discovered a marvelous way of handling the geographic description of addresses," he smiles and says, "Oh, is that a problem?" It takes a while to tell him what a terrible situation it is and then tell him that you have found the solution. I think overall, with both the local and Federal initiative, we have been successful in both identifying the problem, the needs, and at least some of the solutions. In that context DOT has supported, and will continue to support, geographic systems development in partnership with local users.

Opening Remarks

MAYOR TOM MOODY

As Mayor I wish to welcome all of you to Columbus. To those of you who are Columbusites or are greatly familiar with this territory by virtue of your work, I want to extend to you on behalf of the city a continuing welcome even though some of you may wonder on occasion whether you really are welcome in town.

I cannot help but respond to some of your opening remarks, Mr. Chairman. We are rather delighted that you were not born a Federal official. Our concern is not with those who were or were not born Federal executives, however, our concern is with those who died in office and continue to serve. It is rather interesting to me also that you expressed some concern about whether the Mayor had any appreciation of what it was all about. I understand that because I know a lot of mayors, and I am not sure that many of them do know what anything is all about. However, I would like to tell you that in Columbus, Ohio, we may not know all the answers, but we think we have identified most of the problems. That is why this city has committed itself in very large substance to a solution of these problems.

The people who report to me tell me that Columbus is far ahead of any other city in the development of this geographic base program, and I am always concerned about their accuracy. If this is not true, then I would beg of you that you identify for me where we are not ahead and how we can get ahead because it is our purpose in this area to be in the forefront. We recognize the nature of this problem. We recognize the value of the solution. We recognize a number of areas where this information can be used, not only for the benefit of government but for the benefit of industry and commerce. Certainly there is no one here, so far as I know, who is opposed to knowledge. That is not always true at the municipal level, Mr. Chairman, as you have properly identified.

One of the great difficulties, however, and this is not limited to this kind of area, is that even when you have the knowledge, it is sometimes difficult to know what to do with it. If you find that to be rather exotic reasoning, think for a moment about your wives, and then to be fair to the ladies, think for a moment about your husbands. You probably know more about those folks than any other person, but what good does it do you on some occasions? Well, we hope to make intelligent use of this information; we hope to be able to plan things better. We hope to be able to plan in a number of areas far better than we presently do.

Knowledge helps us do things, but its greatest use is to impart it to others so that they will understand what we are about. Politicians and planners are both in the same category. Some of you planners may not like that, but on the other hand, I would like you to think for a moment about how we politicians feel about being in the same boat with you fellows. If we cannot communicate the proper information about data, if we cannot establish a common fact bank, then we simply cannot communicate one with the other, no matter what we say. This is why it is important that we master the subject so that we can teach at least some of it to others, so they will have an understanding of what we are attempting to do and why we are attempting to do it.

When I speak of Columbus I am not referring solely to the municipal government of Columbus, I am talking about this entire area because it is an area project. If Columbus has had any success whatsoever, and indeed I do believe it has, it is because of the fine spirit of cooperation. Mid-Ohio Regional Planning has been in the forefront. I would indicate to all of you the City of Columbus does not regard Mid-Ohio as a bad actor. We pay half the bill down there, and we are proud to pay half the bill. The Columbus Area Chamber of Commerce has been truly exemplary in providing citizen leadership to this effort, and it would be remiss for me not to take a moment to thank Mr. James Thomas. Usually when I am talking to Mr. Thomas, I am asking for something or complaining, but today, I want to say to Jim Thomas that he has done a good job in this work. We are not at all ashamed or abashed about the role the City of Columbus has played in this, either, because we feel that the City of Columbus through its staff work has made an important ongoing contribution to this effort, and financially we have supported it by providing that staff.

There are a number of persons who are not able to attend this technical meeting who wish to all of you the very best for a good experience, not only because it will contribute something to us, to make our jobs easier, perhaps, but it will make it more understandable because we have confidence in what you are doing. We are spending in one way or another a considerable sum of money to see that you can do it. We have hopes that each of you will be enriched by this conference.

You know it is traditional for a mayor to say to a visiting group of whatever nature that we hope that you spend a lot of money here. Well, I am not going to say that. What I am going to say is that I hope that you take something away from this meeting and away from Columbus which is far more than what you came here with. If you do not, then this conference has failed. I am sure that it will not fail. I am sure that wherever you came from, whatever your purpose is, you will walk away from here enriched, with the kind of enrichment and knowledge that can be spent again and again and again and leave you with a fatter purse than ever in the bank of knowledge. God bless you and have good luck with your computers today. I have planned my morning such that I am just going to watch and see what Mr. Tony Savoia and Ms. Eileen Flowers have to say, because I want to make sure we are the best and I want to make sure that if we are not, you are going to tell me.

Status of the Nationwide GBF/DIME System

MORTON A. MEYER

On behalf of the Bureau of the Census I would like to express officially our appreciation for the warm welcome and the hospitality shown by the City of Columbus, the Columbus Area Chamber of Commerce, and the Mid-Ohio Regional Planning Commission. It is great to be here. Mayor Moody's interest in, and recognition of, the value of information geared to meet the needs of local officials is the guarantee that Columbus will maintain its premier position. I would also like to say that any time Mayor Moody wishes to become a Federal employee and would like to work on the GBF/DIME, he has a job in the Geography Division of the Bureau of the Census.

I would like to introduce to you several of the various officials who are here from the Census Bureau. You will have additional opportunities to talk to them and meet the other members of the Bureau during the course of the conference. First, Dr. Alva Finker, the Associate Director for Statistical Standards and Methodology, my boss; Mr. Harold Nisselson, Chief Mathematician and Statistician for the Bureau of the Census; and Mr. Robert McWilliam, Regional Director for the Bureau of the Census for the states of Ohio and Michigan, and portions of Indiana, Kentucky, Pennsylvania, and West Virginia (the Regional Headquarters are located in Detroit, Michigan). I also am going to be very brief so that the presentations of Mid-Ohio and the Columbus Area Chamber of Commerce can take place without delay. It is true that Columbus is in the forefront of the GBF/DIME program and is one of those few cities that have just about completed the GBF/DIME correction and update processes and have available to them a geographic data base ready for use in ongoing programs.

For those of you who may not know, the Geographic Base Files were first developed for use in the 1970 Census as a way of assigning geographic classifications to respondent mailing addresses. Without it, the mail-out/mail-back self-enumeration census technique used in 1970 would not have been possible. The GBF/DIME-File became the device required to code addresses to a specific census block, tract, enumeration district, and such other areas as needed by the Census Bureau to produce statistics classified geographically.

The necessity for small area data is, of course, obvious. Statistics for the United States, or for the State of Ohio or even for the City of Columbus as a whole are not of very great use. What is important for management and analytical purposes is knowledge of the detail upon which the summary data are based. I should note also that GBF/DIME-File capabilities permit not only the classification of Federal information but also make possible the spatial classification of local address referenced data. Thus, information which was too voluminous or complex to be amenable to analysis can be readily classified geographically, mapped if desired, and easily understood.

The Census Bureau is now beginning to prepare for the 1980 census and among the very earliest operations are those of geography. The preparation of the Metropolitan Maps and Geographic Base (DIME) Files in areas where mail-out/mail-back censuses will be conducted represents one phase of this program, which in its entirety includes the preparation of enumeration maps covering the entire United States, Puerto Rico and various other territories.

The Bureau recognizes, of course, that development of the GBF/DIME-Files and maps is a major undertaking and that—although the GBF/DIME System also has important uses locally—Federal Government funding is essential if we are to develop the kinds of files needed for the 1980 census. I am, therefore, happy to report that the Congress has approved our first funding request for this purpose and within the next several months we will be able to supply modest sums to various of our participating agencies—including Columbus, Ohio.

I mentioned briefly a moment ago that the files have important local uses. I will not attempt to enumerate them here (the other conference speakers, who are actually working in the vineyard, will do it much better) but one of the reasons that this is so is that the files are public documents which are freely and readily available either from your own local area or from the Census Bureau. They contain no confidential information, just street names and address ranges, that is, the maximum range of addresses between any two

intersecting streets. They also include X-Y coordinates for street intersections (and non-street features) from which maps can be drawn automatically by computer. We make no restrictions on the use of the files, believing that maximum benefit will accrue from their widest possible dissemination. To assist local users, the Bureau now provides, as most of you already know, the procedures, methodology, and computer programs (and even the computer processing where necessary) needed to correct, update, extend, and maintain the files. As part of its long range program, the Bureau will soon provide—again at no cost to the local areas—a set of its GBF/DIME System applications packages. The estimated date of the availability of these programs will be announced shortly.

Before I close, let me add one final item of great local importance. The Bureau's present plans call for enumerating in 1980, some 90 percent of the population of the United States by mail. This means that we will be establishing GBF/DIME-Files for every metropolitan area and extending the files from the urbanized core of the SMSA (their present area of coverage) out to the SMSA boundaries. As a part of this program, in order to make the 1980 census data most useful to local areas, we also intend to provide block statistics data for the entire SMSA to enable local areas to aggregate population counts as needed to meet the requirements of the "one man-one vote" decisions.

Once again, it is a pleasure to be here in Columbus. If there is anything that the Census Bureau can do to help you in your endeavors in the area of geography and GBF/DIME-Files and Metropolitan Maps, we are at your service.

GBF: A Community Success

EILEEN FLOWERS JAMES THOMAS FORREST B. WILLIAMS

ANTHONY J. SAVOIA RONALD HENDERSON

Columbus GBF—History, Development, Maintenance

EILEEN FLOWERS

I would like to introduce you to the Columbus Geographic Base File, known as the GBF/DIME-File, by briefly going into the history, development and maintenance. I say briefly because there is no doubt that some of you here have heard about correcting and updating as often as we in Columbus have. Also, because our purpose here during these two days is to tell you about our experiences and applications, share them with you, and learn of the experiences of other users so we can really begin to take advantage of the unlimited potentials of this reference file.

To bring everyone up-to-date, I will go back to 1966 when the Bureau of the Census began thinking ahead to the 1970 Census. In the past, the decennial census had been conducted by enumerators who went from house to house asking each family in the country, one by one, all the questions concerning their family size, ages, income, and all other detailed information that had to be collected.

By using this method, certain problems were encountered. Approximately 180,000 interviewers knocking on the door of each and every household in the country took many months. When the interviewers went to the homes, usually one person was home, rather than the entire family, and only they were interviewed and answered the questions. Also, the quality of the data was sometimes diminished by the influence of the enumerators. In other words, there may have been some enumerator bits reflected in the answers written down.

To remedy these situations and to speed up the census taking process, a decision was made to establish a system of mail-out/mail-back questionnaires for gathering census data. To help make this possible, a file containing house number ranges for each block side with their associated geographic identifiers would have to be created. This was the beginning of the development of a file which later would come to be known as the Geographic Base (DIME) File.

In 1968 the Address Coding Guide (the title of the earlier system) was compiled. This guide contained records for block faces which included census tract and block numbers, street names and address ranges. The problem with this system was that it was impossible to access information for both sides of a street separately or simultaneously. Nor could one tell the block sides preceding or following the one you were looking at. Recognizing this problem, the DIME system was initiated. Each Address Coding Guide block-side record was combined into a dual record in order to describe each side of a street segment in the same segment record. Where streets intersected, node numbers were added and geographic information was recorded for both sides of the segment, such as census tract numbers, census block numbers, left and right side address ranges, ZIP codes, and minor civil division identification which are all useful location identifiers for the specific local area. These, in addition to the other geocoding identifiers in the file, are all of specific importance to the Census Bureau on a national level.

The node points were then digitized, and latitude, longitude, and State plane coordinates were established to enable us to pinpoint locations at points of intersections. Non-street features were also added to the file. Railroads, rivers, streets, political boundaries, and

the digitized node points would make it possible to draw complete maps using computer graphics systems.

The addition of ZIP code information and boundaries such as minor civil divisions further extends the use of the file for spatial analysis. For our own in-house use, we will add traffic zones for the benefit of our transportation studies and HUD division. This expanded file is what we now know as the GBF.

The development of the file did not just happen. It was only possible due to the combined efforts of many groups of people. First of all, the entire concept of the GBF/DIME was conceived by the Bureau of the Census. They are the ones who were responsbile for providing us with manuals, maps, coding sheets, edit listings, computer edit and correction programs, technical assistance, etc., etc. I guess you get the picture. For this we thank Messrs. Meyer and Silver. Their needs, combined with the foresight to design the system so that each local area could benefit by the utilization of this geocoding tool, is very important indeed.

At this point, we need to add the cooperation and support on the local level. Many people have played a major role in getting the GBF to the stage it is in today. The City of Columbus Engineering Department provides us with new site plans that contain address numbers to update our maps. Aerial photos are examined closely to verify site plans and subdivision plats and fill in gaps in the event one or two or either of these is missing or has gone astray. Address numbers of existing streets are taken from the Columbus Fire Department address books. These books include address numbers at intersections for all the streets in Columbus. The Fire Department has for some time been giving us very valuable assistance by mailing us update information as it materializes.

At the Mid-Ohio Regional Planning Commission, where the actual gathering together of all these facts and figures takes place, there are many steps to go through. Our Graphic Arts Department begins the process by drafting the new site plans onto the maps and keeping corporate boundary line changes up-to-date. Research analysts do the necessary field surveys to check out any questionable or missing information. They then coordinate and compile all of this data, and do the coding to put it all into a few lines of numbers and letters; but those few lines of numbers and letters represent a lot of work and pack a lot of information.

The next step is to send the map transparencies to the Ohio Department of Transportation Aerial Engineering Department where any new node points that we have added are digitized. All of the coded information is then keypunched and sent to the City of Columbus Data Processing Center where computer runs are made. Here again is an example of everyone being drawn into the act. FIXDIME II is the major update program that we have used so far, but for our 3-year update we found that we needed some peripheral programs. This update covering that great time span yielded approximately 16,000 update records, so a staff member at the city wrote the program to load all the cards to disk, sort them and then dump them to a mag-tape rather than have someone stand in front of a card sorter for about 4 hours.

After FIXDIME II was run and our file updated, we sorted the file by map sheet, primarily by street name, and then chained the addresses from low to high. A print program was then written to select the fields we needed to keep under surveillance and produce a listing of the file that would follow each street from beginning to end, or at least from map edge to map edge. The listing is then in the right format to aid in continued maintenance.

Another very important part of the GBF, in which there must be a cooperative effort, is finding the dollars that are necessary to fund the work that goes into the continuous process of maintaining and updating the file. Back again to 1966, when the Census Bureau implemented the Address Coding Guide, the Franklin County Regional Planning Commission, which is now our present Mid-Ohio Regional Commission, was chosen as the coordinating agency for the Columbus SMSA. We were able to obtain funding for two-thirds of the ACG project from a HUD 701 grant.

In 1970 the Address Coding Guide Improvement Program (which translated the ACG into the Geographic Base (DIME) File was implemented, and once again two-thirds funding was provided by a HUD 701 grant.

In July, 1971 Columbus was chosen as one of the five pilot cities for a GBF/DIME Maintenance Program. These cities were told that the Bureau had no funds to support local work on this project but would provide staff, computer programs, and mapping support. Columbus, as well as other cities, thus had to go elsewhere for funds to perform this function. It was estimated that it would take $12,000 to complete the correction program and a 2-year update. In our area this money was raised through local government and public and private businesses by the Columbus Area Chamber of Commerce, through a GBF Committee. Because of our need to depend on the Census Bureau for computer edits and other materials, there was a

delay in getting the original edit/error listings, and therefore, the schedule was greatly exceeded. It turned out to be a 3-year update instead of the originally planned 2-year. The total cost incurred by the Planning Commission for the correction phase alone was $17,500. This was well above the $12,000 raised.

We have recently completed the initial update phase of the program to January, 1974. We are once again in the process of bringing the file up-to-date, and have just completed coding corrections that have been brought to the surface by a topological edit (TOPOEDIT) of the segments in our file. This edit primarily checks to make sure that all segments that comprise the boundaries of a block have been coded, and certain fields, which permit the segments to be chained around the block, are coded correctly. The corrections are at the computer center now. FHWA, HUD, and county funds, as well as a minimal amount of local funds have been used to bring us to this point, and will continue to be used as long as they are available for the future work that must be done.

In addition to our incurred costs, in-kind services have been generously provided by many organizations as a result of our GBF Committee environment. The Columbus Area Chamber of Commerce has provided staff support and promotion of the use of the file; the Ohio State University has plotted the file to edit and correct coordinates; Battelle has done program testing; the Columbus Board of Education has contributed a large block of personnel hours on coding and keypunching. The City of Columbus has provided computer facilities and staff support. Without the help of these organizations and the many hours of time put in by their respective staff members, the GBF could never have reached its present state.

The next major step we have to face here in Columbus is the extension phase of the GBF. Our coding limit line now lies entirely within Franklin County. The first step will be to extend the coverage to the county line. After that, the entire SMSA, which includes five counties, will have to be coded. Work on editing the map coverage has been started, but there is still a lot of work that will have to be done.

I think that most of you are pretty aware by now, if not from what you have seen here, probably from your own experience, that building a Geographic Base (DIME) File takes a lot of thought, a lot of time, a substantial amount of money, and certainly a very large amount of dedication to the project. It has not been an easy task, but we in Columbus think that what we have to gain from the project makes it worthwhile.

I have covered a lot of time and a lot of work in just a few minutes, and because of that it probably sounds as though everything went well, but it didn't. We have to admit to a lot of discouraging moments. There were times we questioned our intelligence in taking on this project. At times we even questioned our sanity. The very thing that I would suggest to anyone just embarking on a CUE program would be to start using the file as soon as possible. It does not need to be perfect for certain applications. You will find applications are exciting and are a great help to the maintenance of the file if users feed back the discrepancies encountered. Applications are pretty limitless, and we will be telling you about some in operation and some planned.

Geographic Base File: Cornerstone of a Computer-Based Municipal Information System

ANTHONY J. SAVOIA

Over time, most American cities have undergone traumatic changes, changes brought about by technological advancements, urban migration, the inability of some city governments to systematically develop alternatives to improve performance and productivity, and a myriad of other factors.

During the 1960's the demands for vital services, such as police, fire, health care, and emergency medical services, have increased at an astronomical rate. In the past, these increased demands would have been solved by merely hiring more personnel. However, because of spiraling operational costs and the magnitude of the increased demand for services, this alternative is no longer a viable one.

To further complicate matters, public administrators are also plagued with other severe problems. Problems such as inadequate or overcrowded housing facilities, the proliferation of urban blight, inadequate public transportation, pollution of our natural resources, increased use of drugs, and the continued exodus of the white and black middle-class to the suburbs, leaving the inner city to those who can least bring about social and economic change—the elderly on fixed incomes, the lower income and the no income segments of our society.

Herein lies the bulk of our contemporary problems which experience has shown cannot be solved by the application of "hardware" type solutions.

There are some theorists who take a pessimistic view of the future of our cities and propound that they are doomed, insisting that the political system, as presently structured, cannot cope with today's urban problems. Conversely, others say that there is much hope for our cities and that the "systems approach" to problem identification and ultimate solution, used first in the aerospace industry, is a necessary methodology for the control and monitoring of our social, economic, and physical problems.

Innovative technological advancements are partially responsible for our urban problems and it will require a combination of the same type of innovative technology, dedication, high degree of cooperation between the private sector and the local, State, and Federal governments, and an unfaltering determination to solve these problems.

Local governments are charged with the responsibility of providing for the safety, health, and welfare of all of the citizens within their respective jurisdictions. In order to provide these services in the most effective and efficient manner, it is necessary for public administrators to know as many facts as possible about the people they are serving, the physical makeup of their political jurisdiction, and the resources which may be used for the delivery of these services. These facts must also be in a form in which they can be quickly retrieved, updated, modified, correlated, and assimilated so that they may be truly an effective part of the decisionmaking process.

That is to say, there is a need for a highly integrated municipal information system, a network of interdependent and intradependent service delivery subsystems. This information system, because of the large amount of data that needs to be stored, manipulated, correlated, and printed, can only be accomplished by use of man's mental extension—the computer.

If we analyze the requirements of a municipal information system, we can break them down into four basic elements:

1) description of the physical properties of the serviced area, such as streets, housing, buildings, quality of air and water, and public facilities;
2) a description of the people being served (demographic information);
3) a description of resources available for the services such as manpower, funds, and equipment; and
4) a description of the services to be delivered, fire and police protection, refuse collection, recreational services, electrical, water and sewer services, health care, and health planning services.

Until 1970, the element most essential to a municipal information system, a geographic locator file, was virtually non-existent. The Census Bureau's Address Coding Guide, forerunner of the current Geographic Base (DIME) File, did not contain the necessary data to allow its use in a municipal information system. Today, however, Columbus, Ohio has the most current geographic locator file in the nation.

The question now becomes, "So we now have a geographic locator file. Of what use is it?" The answer to that question is a difficult one, difficult in that, practically speaking, the scope of uses of a geographic locator file is unlimited. Consider the following: virtually every city department collects information that, if allied to a geographic locator file and plotted using existing computer mapping techniques, would be more comprehensive than the production of large quantities of computer printouts.

From an operational point of view, our more vital services could be greatly enhanced by developing capabilities such as computer aided dispatching of police, fire, and other emergency vehicles; computer mapping of incidence of crime for the purpose of graphically displaying trends and as an aid in law enforcement resource allocation; computer mapping of multiple data sets to arrive at various types of neighborhood profiles; computer mapping of water, sewer, and electrical networks; computer mapping of census data; and dynamic routing of vehicles. This is only a sample of what a well planned municipal information system can and should provide in an operational environment.

Also, from an urban and regional planning point of view, consider these capabilities: computer mapping of census tabulations, both housing and population; land-use inventories; computer mapping of transportation patterns and flows; computer mapping of health statistics; computer aided location of capital improvement sites, such as health care centers, police substations, fire stations, and multiservice centers.

These capabilities, and more, are possible with a geographic reference file coupled with the proper data files. These files should be a part of any good municipal information system.

Last, but not least, consider this—what one city division does affects other city divisions, because the city is a complex organization with an enormous amount of interdependent social, economic and physical functions. Therefore, every action a city official takes has resounding effects as it interacts with other functions and services performed in the city. For example, what effect would rezoning a portion of a

residential neighborhood to heavy commercial have on that neighborhood? What would it do in terms of property values and tax revenues? What would be the effect on parking, water supplies, waste disposal, fire and police protection, and traffic flow? And how will all of these items affect adjoining neighborhoods? Wouldn't it be desirable if a city administrator could take all of these facts and place them in a "crystal ball" and be able to instantaneously foresee the impact of his decision before it is implemented?

Of course it would, and there is a way—a special technique known as computer simulation. This is accomplished by feeding the computer information relevant to the problems at hand, and by altering some of the pertinent facts, one can determine the impact of several alternative plans. The computer will calculate or, if you will, predict the probable results of each plan. Furthermore, the computer can show how various city divisions must coordinate their efforts to make a proposed plan work effectively. Again, I would like to stress that the prerequisite for accomplishing any of the objectives that I have described, is a computerized, highly integrated municipal information data base of which the geographic locator is a very integral part.

Finally, the development, update, and maintenance of the Geographic Base (DIME) File for the greater Columbus area, not only demonstrates the necessity of a cooperative effort among the private sector, the Federal, State, regional, and local governments, but more important, it is proof that this concept will work. Without that cooperative effort, I am sure that we would be nowhere near our present level of Geographic Base (DIME) File development. It seems to me that it is essential that this same spirit of cooperation be applied in the solution of other contemporary difficulties.

With that cooperative effort, we have a good chance of eliminating or at least controlling today's urban problems, and through an increased planning capability, avoid some of the problems tomorrow.

Columbus GBF—A Community Success

JAMES R. THOMAS

OBJECTIVE

Columbus has been involved in the Census Bureau's GBF development program since its inception, and during the last 2½ years the Columbus Area Chamber of Commerce has played a leading role in coordinating our community effort. At the time of the introduction of the Census Bureau pilot city's program, the Chamber established a Geographic Base File Committee to bring interested parties together to achieve an updated, corrected GBF for Columbus, Ohio. Our initial objectives, to make it possible to update and correct our file, have been accomplished and this fact was reported at the Census Bureau's Boston GBF/DIME Conference in April, 1974.

EXPERIENCE

We believe that it is through usage of the GBF that the program can be kept alive, it is through a cooperative effort that a comprehensive system will evolve, and it is under the leadership of an action oriented community organization that jurisdictional conflicts may be resolved. In Columbus the Chamber of Commerce has been the community organization that has assumed the leadership role of coordinating the development of the GBF. In other communities, other organizations may be equally as effective and the organization of choice.

I would like to take the next few minutes to give you a brief overview of the community involvement that has made the Columbus GBF what it is today. The process we used is analogous to a candy recipe in many respects. Just as it is not always the ingredients that make a good candy, so the presence of equipment, funds and people will not assure the updating and correcting of a GBF. It is equally important that the ingredients be properly mixed, and held at the correct temperature.

To be specific, let us take a closer look at a recipe for divinity candy.

FIGURE 1. **Recipe for Divinity**

2 cups sugar
½ cup light corn syrup
¼ teaspoon salt
½ cup hot water
2 egg whites
1 teaspoon vanilla extract

Dissolve sugar, syrup, and salt in hot water. Cook without stirring to 248° or to firm ball stage.

Remove from heat and pour gradually over stiffly beaten egg whites beating constantly with wire whisk. Add extract and continue beating until mixture will hold its shape when dropped from a spoon. Drop by teaspoonfuls onto waxed paper.

Now let us look at how the Columbus GBF ingredients were handled. Within the Chamber structure a committee was organized to administer the program. In organizing the committee many diverse interests and skills were assembled. Our initial committee was quite small and composed of representatives from organizations from whom we obtained financial support. This included the City of Columbus, Mid-Ohio Regional Planning Commission, Columbus Public Schools, and several financial institutions. This initial committee was vitally concerned with the correction and update aspects of the GBF and major efforts were expended to that end. Just as in the candy recipe it is necessary to cook the mixture at just the right temperature, so it was necessary to allow adequate mixing and association with our committee members. To this end, the committee met frequently, both formally and informally. During our meetings, problems were ironed out before they became critical. The committee structure created a pool of technical expertise which provided a ready forum for technical discussion. The forum creates a situation in which professional discussion enhances the quality of a product through open dialogue provided to the mutual benefit of our total system.

We recently had a user request a copy of our file. This user seemed unclear about the file format, content, and how he might use the file. When he got his copy file, I was able to give him names of people who had used the specific software he intended to use and the names of others who had been using the exact equipment he had. By this process, we are able to provide information regarding the latest software developments in Columbus and the working experience of users. With the winding down of the correction and update phase, our committee efforts have been reoriented to extend use of the file in city government, Columbus Public Schools, and in the Regional Planning Agency. In addition to that effort, we have been pursuing a goal of creating interest, uses, and applications within the private sector. We feel several uses are of interest in the private sector, such as:

(1) marketing,
(2) population estimating by census tracts, and
(3) new residential construction by census tracts.

Our initial committee organization began at the highest levels of our local government and business communities and continues in the same vein. We obtained financial support from our local leaders at an early stage and have attempted to involve decision-makers as well as operational specialists in our committee structure. We found on several occasions that the vested interests of some of our group began to distort our effort to achieve a comprehensive system. When this occurred, we were able to take advantage of our top level commitments to remind the participants of our goal of an area-wide system. The financial support we had obtained gave a high degree of visibility to our backing and tended to reinforce our stated purpose for obtaining a corrected, updated file. Our committee combined representatives from many community organizations and enabled us to discuss individual needs and facilitated compromise in the formulation of policy and establishment of priorities. If impasse seemed to be looming, it was possible to discuss the problems at higher levels and resolve differences before they devastated our program. We always tried to resolve differences within our committee structure, but we did not hesitate to seek outside help in times of stress.

This broad involvement has created operational communications across jurisdictions enabling us to deal with problems as they arise, thus giving us an independence so necessary for meeting the target dates in our program. The involvement at our top levels has given us access to materials and operations which might never have been possible without such involvement. As a result of these efforts, our file is current to January 1974 and new data is being supplied for input to our file on a regular basis.

CENTRAL PROBLEMS

One might expect a Chamber of Commerce to report in glowing terms how successful this program has been, but it has not been without its problems. The time required to get the program started, to make maps available, and to allocate resources for correcting and updating has resulted in the creation of our current GBF with a high degree of utility at a time when 1970 Census data is becoming outdated and of limited use on a current basis. If we expect a high degree of utility from the GBF, we must have new data to match against this file. For example, if the computer tapes for the 1972 Economic Census could be made available for a reasonable cost, there is an immediate new use for the GBF. The private sector is unwilling to pay the reported cost ($18,000) for this two year old Columbus tape. As we are trying to market our GBF among potential consumers, we find that the most aggressive organizations have already developed a system approximating the capabilities of the GBF, and they do not feel the added expense of conversion is warranted at this time. We have found retailers with their own systems and a utility with its own system. It is only if we can show ongoing uses for the long term, and the availability of current data files compatible with the GBF, that any of these organizations will be interested in the added expense of the conversions of their files.

PROSPECTS

Just as in making candy there is extensive need for stirring the mixture, if we are to get more use of our file we must create high visibility and broad interest, just combining the ingredients will not provide a good product. We feel that our prospects for local use for the immediate future are very bright and are changing the thrust of our committee structure to promote use of GBF. We are moving from a correction and operational phase to an implementation phase.

We find that the staging of this conference has provided an excitement and thrust which was difficult to kindle. The creation of the demonstrations you have seen has stimulated increased interest and we hope the demonstrations will spark the imagination of our local participants as budgets are being prepared for 1975, and that they will begin the implementation of projects using our local GBF. We expect to accelerate the local use of GBF through better understanding of its capabilities and what is involved in local applications.

A *second* thrust that we are currently pursuing is taking us to organizations that maintain current data files. When meeting with these organizations, we are requesting access to their files, and discussing how we expect to make use of information they have by utilizing the GBF and adding their data. Using this technique, we have found we are able to stimulate additional interest within local organizations to make use of their own data utilizing the GBF. Some local applications of interest to our data generators are among the following:

1) Residential customer list from a utility for use in making population projections,

2) Traffic and circulation data for downtown development and planning from Regional Planning,

3) A current housing inventory file from the building department,

4) Bus route design and transit stop locations from the bus company, and

5) Local crime reporting by tracts intending to correlate resources allocation with socioeconomic conditions and crime incidence for the police department.

Each of these organizations can supply badly needed current data for use with our GBF. It is up to us to discover other sources of data so important for planning and development in our growing community.

A *third* approach is to obtain media coverage in providing publicity for local success in using the GBF. Our first such effort related to the development of car pools during last winter's energy crisis. As a result of this publicity, we received numerous requests for information about GBF. We are currently preparing publicity for a direct-mail marketing organization and soon expect to be reporting on a health project showing incidence of disease. These are but a few examples of our ideas of how we can make better use of the GBF in Columbus, Ohio. At the beginning of this year we had some rather ambitious ideas which we have not been able to implement. We have met resistance in the private sector to sharing software development; we have noted skepticism regarding use of Census Bureau's software; we have experienced delay in obtaining some materials from the Census Bureau as they develop and debug new materials; and we sense a lack of communication between cities. There is a need to create a community of GBF organizations to provide a means to learn of working programs and developments reported to the Census Bureau. Each of these points can be overcome, but we feel they should not be a problem exclusively within the domain of Columbus, Ohio, but that each interested community should share their findings as fully as possible. Through these efforts we can all meet and benefit from the successes of each other.

SUMMARY

1. The Chamber of Commerce, a local organization, assumed the leadership role in bringing together interested parties to achieve a program of benefit to our total community.

2. We obtained the necessary capitalization to provide the seed money to initiate the program, and have attempted to develop a source of local funding to provide continuous support.

3. The committee structure has been maintained to provide equal system identity for all involved parties.

4. Individual vested interests have been suppressed for the good of the whole community by use of broad involvement in policy making situations.

5. A sharing of experiences is readily available through our committee structure.

The process we used enabled us to develop a corrected, updated file through the dedicated efforts of many people determined to have a successful GBF. Our local interest remains high and as new data becomes available, we will be using it along with our GBF to maintain our leadership role within this vital program.

Applications of the GBF in the Research Community

FORREST B. WILLIAMS

Over the past 5 years the research community, particularly contract research firms, have shown increasing interest in the Census Bureau's Geographic Base Files (GBF). In central Ohio, Battelle-Columbus Laboratories (BCL) has recently been working with the GBF, and undoubtedly Battelle's applications are typical of most of the rest of the community.

This paper describes what Battelle has been doing with the GBF. In the course of the discussion, BCL's interest in GBF will be explained and applications that Battelle has made of the GBF will be discussed. A number of these applications have involved ADMATCHing use of the GBF, which should be of some interest.

Battelle Memorial Institute was founded in 1929 under the will of Gordon Battelle as a not-for-profit research institute. Current research and development at the Columbus Laboratories covers a wide range of subjects in both hard and soft sciences.

The group within BCL which provides resources related to the census is the Battelle Census Center. The Census Center was established partly as a public service to the community so that local decisionmakers can have access to the summary tapes from the 1970 Census of Population and Housing. Most of the users of the Census Center have no other means of accessing this data set since few can afford to maintain the multireel tape files involved. In conjunction with the Ohio State University's Behavioral Sciences Lab, a considerable census tape library has been archived here in Columbus.

Battelle began to participate in the Columbus CUE program primarily because of interest in providing technical support to community projects, but also because of the requirements of in-house, soft-science research groups.* As a member of the Columbus GBF committee, BCL has provided both technical and financial support to the CUE program.

Currently, most of Battelle's applications of the GBF have utilized the address matching capability of the file. Typically, a sponsor has a data file of customers, users, or employees and their addresses, and would like to know where in the city these individuals reside. The Bureau of the Census distributes a set of computer programs named ADMATCH which is designed to perform this task. The GBF contains address ranges on both sides of any street in the urbanized area of the SMSA. Given an address from the data file, ADMATCH first locates the individual street and address range in the GBF. The program then determines, from information in the GBF, where the individual lives in terms of township, place, ZIP code, census tract, or any of a number of summary areas. The result is a file containing the original data records, each of which has been augmented by selected information from the GBF, such as the summary area where the individual lives. The researcher can then generate tabulations for these summary areas. For example, he could list the number of his customers residing in census tract 1.10, 1.20, and so on.

BCL's first application of the GBF took place last fall during the energy crisis. Battelle was interested in encouraging its employees to car pool. One way of doing this is to distribute to employees, lists of individuals who live relatively close together. Census tracts were selected as the most appropriate summary level, the BCL employee file was ADMATCHed, and employees were listed by census tract and block code.

Marketing research applications present an exciting research topic for using census data and sophisticated modeling techniques. Battelle has conducted several projects which required matching customer files to census tracts for marketing research purposes. Typically, these efforts involve the development of a model to examine the relationships between the number of customers and socioeconomic variables across census tracts. This procedure seems particularly attractive for identifying promising new branch locations for a parent firm.

Another application which was investigated involved matching utility company records to census tracts. The objective was to estimate the elasticity of demand for electricity and gas with regard to income in order to determine how price changes affect consumption.

Also under consideration is an application which would help local planners estimate population change after 1970 in various areas of the city. Building permits and utility company meter installations and removals would be ADMATCHed for this purpose.

Given BCL's experiences with ADMATCH, it seems appropriate to make some technical comments about

*It is interesting to note that Battelle was also involved in the original Census Use Study, The New Haven project, although in a minor role.

it. In order to put the remarks in perspective, one should consider BCL's data processing capabilities. Major facilities consist of large-scale Control Data computer systems: a CDC 6400 and a CDC CYBER 73-16. These machines are equipped with over two billion characters of direct access mass storage, plus some 100 communications ports to support teleprocessing and timesharing. The two systems are furnished with remote batch and interactive terminals, and a full complement of standard peripheral units. BCL also has the support of a variety of on-line graphics and plotting equipment.

These facilities give BCL the capability of processing a wide range of scientific, administrative, and other information processing tasks. This flexibility is needed in any diversified contract research organization.

Because Battelle has CDC machines, it was not possible to use the Census Bureau version of ADMATCH which is written in OS/Assembler, and so a COBOL ADMATCH from Florida State University was obtained. In this program, the logic of the matching process is the same as that presented in appendix B of the ADMATCH Users Manual published by the Bureau of the Census. In the original procedure, all records from the GBF pertaining to one street are stored in a table in memory at one time. Each address from the "customer" file on that street is compared to each entry on the table to locate the best match. This approach proved too time consuming for the volume of data required to be processed for BCL applications. The procedure was changed by sorting both the GBF street records and the records from the customer file by the sort key in table 1, and then making record-by-record comparison of the "customer" file and the GBF file.

Table 1

ZIP Code
Street Name
* Street Type Code (primary and secondary)
* Direction Code (primary and secondary)
Odd/Even Indicator
Low House Number Range
* High House Number Range

* Not included in original version of sort.

One disadvantage of our procedure is that it allows only perfect matches, and the options to permit a number of the address items not to be matched cannot be used. In table 2, the results of using the modified procedure are presented for three different cases.

Table 2

CASE #1

Data File	25,000 records
Pre-Processor Rejects, non-valid address formats	69 records
Matching Run Rejects	9,013 records

Matching run rejects were broken down as follows.

Street Type	8,330 records
House Number	557 records
Zip Code	126 records

CASE #2

Data File	225,175 records
Pre-Processor Rejects, non-valid address formats	2,466 records
Matching Run Rejects	67,402 records
Overall Match Rate	68.9 percent

Matching run rejects were broken down as follows.

Street Type	60,146 records
House Number	7,256 records

CASE #3

Data File	2,350 records
Pre-Processor Rejects non-valid address formats	165 records
Matching Run Rejects	824 records
Overall Match Rate	57.9 percent

The BCL experience has been that most ADMATCHing rejects are due to errors in the "customer" file rather than the GBF. It is interesting to note that Case #3, which has the lowest match rate in table 2, is the BCL employee file. Business communications at Battelle are distributed by interoffice mail, but the address file is used to mail a number of other less important items to the employee's home. One wonders what the match rate would be if employees' payroll checks were mailed to their homes.

The match rate in itself may not be important to the client. For example, some projects dealt with a sample of the total customer files from which statistics on the customer population were generated. In other applications the match rate was important. In these cases BCL first ADMATCHed using the efficient BCL-modified procedure, and then the rejects were ADMATCHed using the slower, but more flexible, Census Bureau procedure.

Other applications of the GBF are being considered. Battelle has facilities for interactive graphics, including direct view storage tube terminals, and a CDC CYBER Graphics Terminal. Obviously, BCL is eager to begin working with the topological part of the GBF with these graphics devices.

The unusual flexibility of the GBF provides an organized approach to any information system which requires a locational component. Currently, in a project co-sponsored by the State of Kentucky, BCL is investigating the possiblity of using the GBF as the basis for land record information systems. Another project under investigation would utilize the GBF in a fire-dispatcher information system. Such a system would provide information on special conditions at a given street address of which firemen should be aware. For example, a house may have an invalid living in an upstairs bedroom, or an excessive amount of paint may be stored in a garage, etc. A final application being investigated involves using the GBF to locate all census summary areas with a given distance of a street address and then printing summary statistics for those areas.

This paper has described Battelle's experience with the GBF and briefly discussed projects envisioned in the future. Battelle is excited about applications of the GBF and about the CUE program here in Columbus. The real challenge of this technology would appear to be its successful application and BCL hopes to contribute to that effort.

Cooperation—Working Together for Mutual Benefit

RONALD HENDERSON

In the papers presented here today, we have been familiarized with the history, development, and applications of the Columbus Geographic Base File. We have seen how the GBF has expanded from its original purpose as an improved data file gathering technique for the Bureau of the Census to a reference file containing valuable data being utilized in various ways by a host of public and private organizations.

Originally, the GBF was to be a modification of the Census Bureau's Address Coding Guide and was designed by the Census Bureau primarily for purposes of improved and more efficient data gathering and retrieval. The GBF was implemented for the Columbus SMSA in 1970. Since that time, however, the GBF has become not only a useful tool for the Census Bureau, but its valuable potential has also been recognized by a number of interested groups within the public and private sectors. In the public sector, the City of Columbus is applying the GBF to a number of city functions. In the private sector, organizations such as the Columbus Area Chamber of Commerce and financial and marketing institutions are all interested in the GBF because of its great potential as an effective instrument providing valuable information necessary for important business decisions. The research community, as well, has shown an equal interest in the GBF and its applications, as demonstrated by the work of Battelle-Columbus Laboratories. Because of its unusual flexibility and increasing interest in its applications, the GBF and its uses have grown considerably since being introduced 5 years ago.

The purpose of this paper is to discuss one of the key elements which has made the GBF a success in Columbus. That key element is cooperation—cooperation in funding, cooperation in research, and cooperation in application. Because of the many groups and organizations involved in the GBF process, cooperation is vital in all phases of its operation from initial implementation, through continuous maintenance, to innovative application.

All of those interested in and involved with the GBF share a common need. Simply stated, this common need is an easily accessible reference file of geographic information, which is what the GBF actually is. When

used in conjunction with any one of an assortment of other files, the GBF becomes an invaluable tool for public and private organizations in making marketing decisions, in observing and analyzing population and housing trends, in aiding the dispatching of service vehicles, especially emergency vehicles, just to name a few of the wide range of uses. The need for a geographic reference file such as the GBF is universal, even though its uses and applications vary from user to user.

Although there is a common need among interested organizations in acquiring information contained in the GBF, it is important to realize that no one organization possesses all the facilities and resources to develop, correct, update, and extend the GBF. A few basic necessities include solid financial support, trained personnel, and a substantial amount of time. Since the GBF is a computer based information file, adequate computer facilities are mandatory. Because of the nature of the various elements required to operate and maintain a successful GBF, it undoubtedly must be a cooperative effort.

The Columbus GBF relies on cooperation among six basic components: a support group, decisionmakers, a coordinator, a GBF technician, a computer specialist, and a group of analysts. These components are the backbone of the Columbus GBF, and do not necessarily apply to GBF's in other areas since every area has different needs and constraints.

The support group is the largest and most influential component. It is composed of a number of organizations sharing and cooperating in providing essential resources, facilities, and services. The largest member of this group is the Bureau of the Census. They have provided us with such things as extensive instructions, program software, and initial tapes. GBF would not be as successful as it is today without the guidance and assistance of the Census Bureau. The Columbus Area Chamber of Commerce has provided the environment and facilities for meetings and has served as coordinator and promoter on the local level.

The Mid-Ohio Regional Planning Commission has served as an operation facilitator, providing mainly technical skills and services. It should be noted that as developer of the GBF in Columbus, MORPC did not have and still does not have an in-house computer system. As coordinator and promoter of the GBF, the Chamber of Commerce also does not have a computer. The computer processing has been performed by many organizations in the Columbus area, but mostly by the City of Columbus Data Processing Center.

MORPC also serves as a data collector by obtaining needed information from Franklin County and the numerous jurisdictions within the Columbus GBF area. The Ohio Department of Transportation has supplied digitizing of node dots. Computer graphics provided by the Instruction and Research Computer Center of the Ohio State University serve as quality control in checking and correcting coordinates in the file. Several departments from the City of Columbus have contributed substantially to the GBF cause. The City Data Processing Center has allowed us the use of their facilities and expertise in making update runs for the GBF. The City Engineering Department serves as an update source by continuously providing site plans and plats with addresses for new developments.

Address ranges for existing streets are made available through the Columbus Fire Department. The Columbus Board of Education has contributed many staff hours for coding and keypunching. By integrating the GBF into their line of work, the W.A. Storing Company, a local mailing service, serves as an additional quality control check. They have been able to identify ZIP code errors and provide necessary corrections. Battelle-Columbus Laboratories has been instrumental in the GBF work by supplying research expertise and advisory service in fitting ADMATCH and other Census programs to our facilities. In working with several different applications using the GBF, Battelle has also become an important source of technical information. Finally, several financial institutions were responsible for a sizeable share of the initial GBF funding.

From the various contributors in the support group are molded the other basic components. Decisionmakers are necessary in determining the status and fate of the Columbus GBF project. A coordinator is required to organize the activities of a GBF technician and a computer specialist. The GBF technician must have a thorough understanding of the design and structure of the GBF. Because of the vast amount of computer related work involved in the GBF, the knowledge and skills of a computer specialist are essential. Lastly, there are the many analysts and technicians who are responsible for the surveying, drafting, coding, keypunching, and various other duties which are so vital to the success of the GBF.

While I have mentioned six rather general and basic components of the GBF, there is a seventh which may be included. This additional component is feedback. In order that improvements and refinements may be made in the GBF, feedback is essential at all levels, especially from those who use the GBF in practical applications. Battelle-Columbus Laboratories and W.A. Storing Company, through their applications of the GBF, have been able to provide valuable feedback in the form of suggestions for the use of the GBF and the correction of erroneous data in the files.

It should be stated again that the components described here have been the vital elements for a successful GBF in Columbus, and these components may differ from city to city. Our purpose is to illustrate what elements are involved in the Columbus GBF and to emphasize that it would not be at all possible without the cooperation of all those involved.

Based on the current status of the GBF and Columbus experience with it, the question we are presently faced with is: "How do we move ahead?" Answering the question is difficult, but there are several factors which we might review and study if we are to continue our progress and success with the GBF: 1) a better public understanding and increased public interest; 2) continued regular maintenance of the file; 3) continued sharing of ideas; 4) monitoring of GBF activities; and 5) continued research and development of applications.

First of all, in our work with the development, maintenance and application of the GBF, we have found that with the seemingly unlimited number of uses which the GBF has to offer for many public and private organizations, there seem to be too few people and organizations contributing to a project which has such large scale potential. It is hoped that more public interest in the GBF may be stimulated through an improved understanding of the GBF and its applications. A better understanding and more public interest are important for many reasons, but primarily for the initial and continuous support. Funding, by far, is the major support mechanism. Were it not for the money raised by the Columbus Area Chamber of Commerce through local government and private businesses in the initial stages of the GBF, the Columbus GBF would not be what it is today. Although more and more money is coming from government rather than private sources, funding, whether public or private, is still the prime element in the support of a successful GBF.

Secondly, it is suggested that continued maintenance of the GBF be provided on a regular basis so that data contained in the file and inferences made from that data may be made more accurate for the users. A regularly maintained GBF means cleaner data for more accurate applications. Consideration of local priorities and resources will determine whether maintenance or application of the GBF is more beneficial for a particular area.

Thirdly, it is important that there be some mechanism available to allow for the free flow and sharing of ideas among those organizations involved with and interested in the GBF. In all phases of the GBF there are many processes and techniques utilized. In many cases there are unique conditions which are encountered and there are undoubtedly many mistakes and problems which arise. The purpose of sharing ideas and experiences at all levels is to broaden the scope of knowledge in the field and to encourage the development of new techniques and applications.

A fourth factor which is closely related to the sharing of ideas is the monitoring of activities. Since more and more government agencies and private organizations are getting involved with the GBF, it is possible that much of the information on the subject may be repetitive in nature and may receive unnecessary attention. With the monitoring of activities on both a local and national scale, the unnecessary duplication of processes, techniques and general information could be minimized.

On the local level, this is already being done with metropolitan and regional agencies directing and correlating GBF activities. On the national level, it is suggested that the Bureau of the Census expand its role as director of GBF development to coordinator of national GBF activities. As a super-coordinator, the Census Bureau could gather and coordinate data from all GBF's in the country so that the information from local GBF's could be made more readily available on a national scale. Acting as a national coordinator, the Census Bureau could also be instrumental in receiving and disseminating feedback from the local areas which could be useful in guiding and improving GBF operations throughout the country.

Lastly, it is hoped that research and development of the GBF, especially its applications, will be continued. Without practical applications, the GBF is a useless file. If the GBF is to be used as the valuable tool it was intended to be, it is necessary that continuous development and advances be made in applications of the GBF.

After looking at various ways of how we might move ahead with the GBF, the question that concerns us now is, "Why continue the GBF in Columbus?" So far, we have spent over $70,000 in public and private funds and thousands of staff hours in tedious work. Along with this has gone much thought and dedication. With this much time, money, and effort already invested, abandonment of the project would be inconceivable, especially since the Columbus GBF is

quite sound. Since the GBF has been recognized as a valuable tool, many applications of it are either being considered or are already in use in the Columbus area. Many new procedures and techniques have been established for maintenance of the file, and at this time, maintenance of the Columbus GBF is progressing quite well.

The future of the GBF in Columbus is encouraging. Increased funding and expanded services are anticipated. Many innovative applications of the GBF are being considered. Extension of the GBF to include the total five-county SMSA is not too far off. With an almost unlimited number of uses, the future holds much in store for the development, maintenance, and application of the GBF. If the Geographic Base File is to continue to be a valuable reference file for public and private organizations in the Columbus area, a cooperative effort is essential in its funding, research, and application. Without the involvement of the various organizations in the Columbus area, we would not be one of the national leaders in GBF development. Only through total cooperation in all phases may mutual benefits be obtained from the use of the Geographic Base File.

Question Period

Mr. Alan Pisarski—I want to thank our panel this morning for what has been a beautiful presentation—well integrated, good graphics, very well structured. If the overall GBF program in Columbus is as well organized as this presentation has been, I can see why the Mayor feels that it is the best in the country. I also think Ms. Flowers' opening remarks make a very important point and that is that while the technology is complex, and the logistics are massive, many, many of the problems are in fact institutional, and to organize the cluster of interests and focus people to work in the same direction is really a major part of the whole undertaking.

Dr. Robert Aangeenbrug—I would like to ask a question of Ms. Flowers: You talked about $17,500 in terms of cost and then you alluded to additional in-kind services from about four or five organizations. Without putting you on the spot, I would like you to perhaps huddle with those involved and give us some approximate additional information about the cost or value of these in-kind services. I am rather concerned that we document the real cost of doing this prior to 1980 in a place that does not have Battelle, Ohio State University, and a vigorous Chamber of Commerce in a supportive environment such as you seem to have here. I think it would be useful to have that.

Ms. Eileen Flowers—I was wondering if what Mr. Thomas presented in his paper answered your question? You will have to remember too that we are one of the Census Bureau's pilot cities, and as such, we did a lot of trial and error work that is not shown in the results. A lot of those in-kind services would not be necessary for everybody else. We feed back our information and experiences to Washington, and all other areas learn by some of our stumbling blocks. I do not think I can put a dollar value on it right at the moment. As an example, for the Update phase, just to get us out of a time bind, the Board of Education loaned us a coder for a short time.

Dr. Robert Aangeenbrug—Yes, For example, I just heard the figure $70,000 mentioned. I also heard in the previous presentation, $17,500 mentioned. I think we might try to talk about what the $70,000 was and what the $17,500 was.

Ms. Eileen Flowers—The $17,500 was the amount that it actually took for one phase of the corrections operation; and what we first thought was going to be a 2-year update, ran into a 3-year update. Now, that was one period of time. This $70,000 figure would essentially be the cost starting from the very beginning and our first involvement in the Address Coding Guide Program for the 1970 Census and before we become involved in the CUE program.

Mr. Anthony Savoia—This is not just computer; this is total cost. It is everybody's time and effort. The computer, the whole bit.

Ms. Eileen Flowers—This is what I did not know, myself. Mr. Savoia's estimate is $100,000 to $150,000, but then, that includes "extra" clerical hours without pay and computer time that is no doubt recorded at City Hall but was not passed on to us at MORPC.

Mr. Anthony Savoia—There is one point that I would like to make about the $100,000 to $150,000 figures. The $50,000 variance is quite a range. That $50,000 is the value we have put on people's time that is not recorded. For instance, Ms. Flowers, a very modest person, has put a lot of effort on her own time into this program, and there is no way of estimating what the cost is. We have lost track of that type of time and effort.

Mr. Alan Pisarski—It is a fascinating number, and it is better to have a number that you can look at and evaluate than to say it is intangible, which begins to sound infinite, if not absolutely priceless.

Ms. Sally Odegaard—Mr. Thomas, nothing has been mentioned about confidentiality, and it certainly is a national concern at the present time. You mentioned that you approach organizations that keep current data files for your ADMATCHING activities. How do you deal with the problem of confidentiality, since these presumably have names and addresses in them?

Mr. James Thomas—We deal with each of these situations on an individual basis. At no point in time do we add the names and addresses to any file which is made available as part of our GBF operation. The GBF itself does not contain any confidential information. The GBF tape is a tool that permits the aggregation of data. For example, a customer listing or a utility listing (data file), the GBF, the ADMATCH routine, and summary programing together would permit geographic aggregation of the data file. It would be this aggregate number which could be released and would eliminate the problems of confidentiality.

By aggregating the number of people in a census tract, for example, rather than the specific names and addresses, the only information we have would be total numbers related to individual tract areas or block areas. I do not believe there is any difficulty with that kind of information and its confidentiality.

We anticipate going to the organization and working with them, in their own facility, they doing the matching or aggregating to achieve that high degree of matching within their own file. We will never actually have their files except in a "sanitized" version which should not create any problems.

Mr. Alan Pisarski—Let me comment on that from the Federal perspective because it is a fair question. It is one that the Department of Transportation, and a lot of other Federal agencies are quite concerned with. I feel that Geographic Base Files are actually very supportive of programs and laws that attempt to avoid disclosure and protect privacy.

Certainly, the file in itself does not contain anything of disclosure concern. But more importantly it is a tool by which the kinds of items of information that might be subject to disclosure are submerged, and permits the summarization and suppression of that kind of information, specifically names and addresses, and gets it out of manipulated data sources. I see it is an aggregating tool, rather than a disaggregating tool; and therefore supportive of the kinds of programs that try to protect privacy.

Mr. Ronald Henderson—The question of data security is an extremely interesting one. There are organizations, mostly government, that have requirements for reporting detailed data, for instance, the Public Health Service. In my experience with government and security-type applications, I find a lot of strange things going on. I recall when I was at the City of Columbus working on an automated program for Civil Service—the government instructs that you are not able to ask an applicant their race, and certain other kinds of characteristics that somehow might be discriminating. But in order to get Federal funding, you have to report detailed statistics for individuals, such as race, sex, age etc. Many government agencies and private agencies are in a real bind because when they take an application they have to guess and "quietly" record race, etc. The whole object of the GBF is to make the files so readily available that every organization can have their own copy of the GBF file, do their own kinds of security matching, and for any kind of data that would be released, they can strip off the sensitive information.

Mr. Michael Plett—Some peripheral discussion was given by the presentors about extending the Geographic Base File to the limits of the SMSA, which involves four additional counties, aside from Franklin County. I would like comments on how they intend to perform this activity, particularly with the cost that may be involved, and also with the idea in mind that the people who supported research activities inside of an urbanized area, such as Franklin County, may be adverse to supporting those same kinds of activities in rural counties. Also, how would the composition of your support change in any such extension?

Mr. James Thomas—We have been grappling with that particular problem for some time. We feel that there are organizations that are already working in many of those areas that can contribute much of the data needed. We have also discovered that some of the people who are reluctant to support the program because of the limited area currently covered would probably be quite excited as soon as we start extending the GBF beyond the boundaries of Franklin County. There are many organizations in the private sector whose customer coverage goes far beyond the range of our existing GBF. For example, our telephone company covers some 8 or 10 counties (I do not know the exact number), and their interest in one county is

rather limited. Similarly with our electric utility, they cover some 22 counties in Ohio, and any time we approach them and say, "How about funding our one-county operation?", they say, "When you get something that is a little more comprehensive, we would be much more interested in it." There are those organizations that serve one small area, and they are supporting our program currently. We anticipate, therefore, that some of the organizations that have not been interested will be more approachable as we expand the operation.

Mr. Alan Pisarski—I might say that I very much support your comments. The Department of Transportation, for instance, requires urban transportation studies to, in effect, work at the boundary of a problem, rather than within any confined set of geography, and so the Department's view would be that any Geographic Base File program should operate over the whole SMSA structure; and in fact, origin-destination surveys, etc., are always conducted in that kind of frame. I think HUD, EPA, and other Federal agencies would have similar requirements in their programs.

Mr. Morton Meyer—I would like to add several observations to Mr. Pisarski's point that the GBF program should cover as wide an area as possible. This has been recognized by the State Highway Department of Massachusetts which, in cooperation with the Census Bureau, has embarked upon a program to develop a GBF (DIME) System covering the entire State of Massachusetts. The State of Rhode Island is developing a similar statewide file which they plan to use for the location and spatial analysis of traffic accidents.

Mr. Alan Pisarski—I think we should make the point that around the world there are several countries that are, in effect, building Geographic Base Files for the country. I think many of us in Washington, D.C. have been visited by representatives of these countries. We in DOT are also doing national geographic base file building. Not exactly block GBF/DIME-Files for the country, but analogous programs in systems. For instance, we do have a DIME type of file for all the county boundaries of the United States. So, I think there is a hierarchial set of requirements in scale here that is appropriate.

Dr. Joseph Duncan—I would like to pick up on Mr. Meyer's point, in another direction. It seemed to me that one of the themes of the panel discussion this morning was the value and need for local technical understanding in the manipulation of the data and clean-up of the data files. Near the end Mr. Henderson talked about the need for deeper involvement by the Census Bureau, in terms of local overall national coordination, a point which could be in conflict with local technical involvement and discussion. I would be interested if any of the panel members have any specific suggestions for the Census Bureau beyond the activities which they are presently doing in terms of the types of publications that were passed out at this meeting, this type of conference, and so forth. If there are any specific suggestions for the Census Bureau, I know it would be very helpful.

Mr. Alan Pisarski—Since Dr. Duncan is from OMB, I am sure the Census Bureau will be fascinated with the answer.

Mr. Ronald Henderson—What can they do? The major thing, which was another theme of the panel today, is that they could give us more money. That would be of great help. That is just a general feeling, and not a complaint. We see, as we are moving from the update and maintenance phase into the application of the GBF, that we have spent a lot of money developing and correcting the files; and I am sure anyone here can tell you that we probably made more mistakes than we did things right. One of the things that we identified as being very important is that we communicate with each other. I think that the kinds of things that we have learned here in working with the GBF should be communicated to everyone else who has been working on the GBF file.

Mid-Ohio Regional Planning has been visited by a number of other Regional Planning Agencies and we have tried to give them a rundown of our experiences. But every agency, the city, private organizations, the people who have been contributing their time, and I might say a lot of their own personal time, to get the file to this state have budgets; they have projects that they have to work on. They do not have time to respond to everyone coming in.

I would suggest that there should be some forum through which there would be increased communication to GBF users. We could supply information to the Census Bureau so the Census Bureau could disseminate the information. By information, I mean real feedback on particular problems that we have encountered. I would suggest that would be one great aid in helping everyone in the country; that is, have the same kind of feedback, even though it may not apply especially to their organization. I think this would be important, and I think the Census Bureau would be the best forum for that.

Mr. Alan Pisarski—I can never let one of these conferences go by without making the statement that, historically, the prime source of funding in the DIME program has been the Department of Transportation, the Department of Housing and Urban Development, and to some extent, HEW and LEAA, and it is really only recently that the Census Bureau has begun to put extensive amounts of funding into it. I think a lot of the support structure should not come from the notion of maintenance and the description of this whole activity in and for itself, but rather in relationship to the functional programs that are going on where this tool can be highly applicable. If considered in this perspective, the significant support will come from that direction as opposed to a kind of the tool-in-itself development.

Mr. Morton Meyer—We have heard the suggestion, we take it to heart and with your help and support, we will see what can be done to establish a clearing house as recommended.

New Programs in State and Local Statistics

JOSEPH W. DUNCAN

INTRODUCTION

I am especially pleased to return to Columbus to participate in this conference on the theme "Geographic Base (DIME) System—A Local Program." Several years ago, when I was responsible for the Census Summary Tape Processing Center of Battelle Memorial Institute, I had the opportunity to help raise funds for the first Columbus effort to update maps and create the local address coding guide. It is very satisfying to note the progress and achievement of the sponsoring organizations since that date.

Now that my responsibilities and activities have shifted to the governmental side, it is with great enthusiasm that I speak with you today about new programs in State and local statistics. After 13 years of daily experience as a user of State and local statistics, while working with private business and State and local governments, it is now my privilege to help create a national program of improving these important statistics.

I plan to discuss two specific areas. The first deals with several immediate activities which are underway and will shortly affect the quality of national data on local areas. Second, I want to outline a long-range program which is being initiated by the Statistical Policy Division of the Office of Management and Budget in order to extend these developments to many other areas.

CURRENT IMPROVEMENTS

Work is presently underway to improve two key series of great interest to local planners and policymakers. These efforts relate to intercensal population estimates and estimates of local unemployment rates.

In order to meet the needs of making general revenue sharing allocations to some 39,000 local areas and jurisdictions, the U.S. Bureau of the Census has been using data collected from 1972 income tax forms (Form 1040) to develop estimates of migration patterns between 1970 and 1973. These data on migration, when linked with vital statistics on births and deaths, are expected to yield useful population estimates for individual small-area jurisdictions.

The estimating work is nearing completion and it is anticipated that results will be published by February of next year. We hope that these estimates will be widely used and accepted. However, we know that considerable controversy surrounds them. In addition to concerns about the technical problems associated with the fact that nearly 30 percent of the tax return information provided was inadequate[1] there are strong feelings about the estimated undercount which occurred in 1970 (as well as in previous years).

The Bureau of the Census estimates that between 3.1 million and 5.8 million persons were not counted by the 1970 Census. The Bureau's preferred estimate is 5.3 million, but I should underscore the fact that the *exact* scope and profile of the undercounted are really not known.

Several public interest groups have argued that the central cities were adversely affected since 1.8 million or 35.3 percent of the 5.3 million undercount is estimated to represent Negroes. Approximately 58 percent of the 1970 Negro population resided in central cities.

Unfortunately, not only is the exact size of the undercount unknown, but the distribution in geographic terms is even less well understood. While it may be technically feasible to make some reallocation of the undercount to States (using demographic subgroups to estimate appropriate shares), there is no

[1] It should be noted that the Census Bureau was able to use address data to supplement the Form 1040 information provided. Thus, over 90 percent of the returns were coded to subcounty areas; 97.1 percent were coded to county; and 99 percent were coded correctly to State levels.

acceptable method for equitably estimating all 39,000 jurisdictions and for assuring that the adjustment will, in fact, be more accurate than the 1970 estimates and subsequent adjustments.

In view of this situation, my office has drafted two directives that require the use of the official census intercensal estimates and officially recognize that there is no reliable adjustment for the undercount. We are turning our attention to efforts to obtain an even more accurate count in 1980.

In the area of local unemployment estimates, two recent developments are of interest. Following passage of the Comprehensive Employment and Training Act last year, the Bureau of Labor Statistics (BLS) was given responsibility for setting standards and procedures for making local area unemployment estimates. They developed procedures which moved from the establishment focus of employer reports filed under unemployment compensation programs to a household-based concept more consistent with national employment/unemployment estimates. Thus, the focus of local unemployment estimates shifted from place of work to place of residence.

The Current Population Survey (CPS), which is the basis for national employment/unemployment rate estimates, is based upon an interview of approximately 45,000 households. It provides specific estimates for 19 of the largest States. Recently, the Census Bureau has developed ratio estimating techniques to provide estimates for five more States and in July 1974, OMB authorized design of two sample supplementation programs to provide estimates for all 50 States and perhaps up to 100 SMSA's. At this point, it appears that the State supplementation will proceed with estimates becoming available over the next 18 months. Further, BLS has revised the local area estimating procedure to assure that the technique used for all local labor market areas is internally consistent. The new procedure will be effective on January 1, 1975.

These two developments—definition of the base on a household concept and expansion of the CPS—will have the result of dramatically improving our understanding of labor market developments throughout the Nation.

The growing demands for improved State and local data are evident to Federal decisionmakers and to State and local decisionmakers. The responsibility for developing such data is a shared responsibility and, therefore, it is essential that the planning of new programs and the setting of priorities result from a shared planning effort. Our intent in the Statistical Policy Division of OMB is to pursue the following three stages:

Stage 1. Through a set of regional consultations, we plan to obtain a sense of the needs of regional statisticians. Further, in these meetings we are seeking suggestions concerning the process for joint planning which can be instituted by States and our office in relation to finding ways to meet these needs.

Stage 2. On the basis of these suggestions and the ideas of major Federal statistical agencies, we then intend to establish an appropriate and continuing Federal/State planning relationship.

Stage 3. Finally, through that relationship and our normal role of coordination, we plan to work with statistical agencies in both the Federal Government and in the States to assure implementation of efforts which will assure the long-term development of effective statistical programs.

In the balance of this review, I would like to briefly explore each of those stages with you.

SOME FACTORS IN HOW WE REACHED THE PRESENT STATE OF AFFAIRS

I think it is helpful to reflect at the outset of these remarks on a couple of elements of the historical concern with Federal/State relations in statistical policy. The predecessor agency to the Statistical Policy Division (earlier within the Bureau of the Budget and now in the Office of Management and Budget) was the Central Statistical Board. In its fourth annual report, covering the period July 1, 1937 to June 30, 1938, the Board summarized its chief activities during fiscal year 1938 in an eight-paragraph summary. One paragraph reads as follows:

> Another notable development in which the Board took a strong and friendly interest concerned the statistical activities of State governments. State requirements are an important factor in the task of coordinating statistical inquiries in this country. During the past year two States, Wisconsin and Delaware, established agencies with functions similar to those of the Central Statistical Board. The Statistical Committee of the Wisconsin Executive Council especially made extensive use of the reports of the Board in the formulation of its programs and procedures. The Board has expressed officially to the agencies of both States its desire to develop effective continuing cooperation.[2]

In the text of the annual report, the report stated, "The Board believes that every encouragement should be given to the development in each State of an agency responsible for coordinating the statistical activities of the State government. The Board welcomes the establishment of these two agencies and offers them its full cooperation."[3]

The U.S. Department of Agriculture and the U.S. Department of Labor have operated Federal-State cooperative statistical programs for many years. Recently, new programs have been created by specialized agencies such as the Law Enforcement Assistance Administration. In addition, major Federal-State cooperative programs affecting education statistics and health statistics have been started recently by the Department of Health, Education, and Welfare. For example, the National Center for Education Statistics has developed a program to establish a "common core" of basic data on educational manpower, facilities, and programs.

The Congress has recently passed several major pieces of legislation that require detailed and accurate local data. The most widely recognized piece of legislation in this area is the General Revenue Sharing Act with its requirements for detailed data on over 39,000 jurisdictions. More recently, the Comprehensive Employment and Training Act (CETA) legislation, which was signed in December 1973, specified data requirements for prime sponsor areas which, generally speaking, are equivalent to local labor markets. Current legislation associated with elementary and secondary education includes proposals for several specific programs of State and local data, including data at the State level on the total number of children aged 5-17 inclusive, from families below the poverty level.

These accelerating demands of the early 1970's reinforce the needs which were so clearly expressed in the late 1930's.

FISCAL YEAR 1975 STATISTICAL POLICY PLANS

For the Federal Government to encourage open discussion of Federal, State, and local statistical requirements may invite skepticism—either with respect to the real intent of the Government or with respect to the capability of the Statistical Policy Division to achieve major new programs; especially in a period of fiscal restraint with its obvious implications for budget reductions (rather than major new statistical program initiatives).

I assure you that it is not our intent to create "rising expectations" for quick solutions or for major injections of Federal funding into State statistical programs. Rather, it is our conviction that the development of Federal-State statistical programs is a *long-range* process and that many current programs can make an even more effective contribution to improving the statistical base for State and local decisions. Our primary purpose, therefore, is to establish a process for planning and priority determination as we deal with the current needs in the perspective of longer term requirements.

As I commented earlier, national legislation in the past two years has created a potential demand for State and local area statistics which exceed current production capabilities of the Federal statistical system. To create the full capability required would be extremely costly. In addition to its direct expense, the development of a large scale Federal effort would further aggravate the imbalance which presently exists between the statistical resources now available to State and local governments in relation to Federal statistical programs. The demands for consistent national data may merit this imbalance in certain cases such as the decennial census. However, it is clear, on the face of the current situation, that development of State and local data systems as the sole responsibility of the Federal Government would be contrary to both Federal interest and policy—which in recent years have emphasized the need for increasing State and local government *decisionmaking capabilities.* Consequently, at a minimum, it is essential that whatever programs are established, that the efforts be undertaken on a joint basis.

In the past, the needs for State and local area data have not been comprehensively addressed. While many needs were articulated by most governments, few governments were in agreement concerning priorities. Apart from a wide interest related to the need for a mid-decade census, a uniform expression of requirements has been absent. Nevertheless, in the Federal statistical agencies and OMB we have been forced to make decisions concerning State and local data in the absence of a clear consensus on immediate needs.

In contrast to a decade ago, the current rate of improvement of Federal data and in the development

[2] Fourth Annual Report of the Central Statistical Board, July 1, 1937, to June 30, 1938, page 2.
[3] Ibid, page 9.

of new statistical series is quite impressive. However, as real as these accomplishments may be, this level of effort does not cover all of the statistical requirements implicitly or explicitly embodied in revenue sharing programs, the New Federalism, and other steps being taken to return the power and capability of decision-making to State and local governments.

If the Federal Government is to expand the production of State and local area data, and if it is to do so without sacrifice of vital programs or without a quantum increase in its own operations, then serious exploration must be made of the desirability and feasibility of increased Federal-State cooperation in the production of data. Close attention must be given to State capabilities, their data needs, and their resolve to find solutions to issues such as 1) statistical standards and 2) the development of means which will guarantee public confidentiality and industry nondisclosure.

We are well aware of the difficulties and complexities that will be encountered in the interagency and intergovernmental planning process we are proposing. We do not presume to solve these matters this year or the next, but we do believe that specific, interim objectives can be achieved.

We have some notions as to some interim objectives. Federal experience with cooperative programs provides a baseline of departure. However, we know that we do not adequately understand the needs of States and localities. Moreover, we are equally aware that many States do not understand the constraints which affect the Federal statistical system. Therefore, we will be seeking the views of the States in regard to our plans and incorporating their requirements insofar as possible.

To begin to move ahead requires the development, or institutionalization, if you will, of means of communication and the surfacing of a workable consensus concerning what is needed and what can be achieved.

RECENT DEVELOPMENT AT THE STATE LEVEL

Earlier this year (March 21-22), the University of North Carolina, Carolina Populations Center, sponsored a conference labeled the "North Carolina Demographic Data Workshop." The perspective of State-level statisticians was reflected in Dr. Harry Rosenberg's comments that "... the history of the Federal statistical system with the exception of the constitutionally-mandated decennial censuses, is essentially episodic in character, a history whose chapters are built around programs inspired by crises and public sector intervention....at the State level....the statistical system reflects the episodic history of the Federal level, exacerbated in its fragmentation by the admixture of special State programs instituted with their own data requirements, but often handicapped by underinvestments in information as well as by the absence of a coordinating office such as the Division of Statistical Policy in the Office of Management and Budget, even though the need at the State level may be relatively greater."

The South Carolina Budget and Control Board has recently reorganized the Division of Research and Statistical Services to acknowledge its direct responsibilities:

> To review and evaluate continuously the research and statistics operation of the several departments, institutions, and agencies of the State government insofar as they pertain to the production and projection of economic, population, labor force, and employment data with the view to coordinating such activities and avoiding duplication of effort. (Section 99 of the 1974-75 Appropriations Act provides that the Budgeting Control Board has full power and authority to make surveys, studies, and examinations of department, institutions, and agencies of this State, etc.)

Further, the Division maintains a program of statistical standardization and coordination for the purpose of reducing the reporting burden on business and the public, analyzes and makes projections of the State's economy, and provides economic and statistical consulting services by contract to the agencies of State and local government.

These two developments illustrate the growing attention to statistical planning and coordination which is emerging at the State level.

CONCLUSION

At the outset I indicated that we are undertaking three activities:

Stage 1. We are holding a set of regional consultations to obtain a sense of the needs which regional statisticians outline (meetings in Boston and Seattle occurred in August). In these sessions we seek to obtain suggestions concerning the process for joint planning which can be instituted in relation to these needs.

Stage 2. On the basis of these suggestions and the ideas of major Federal statistical agencies, we then intend to establish an appropriate Federal-State planning relationship.

Stage 3. Finally, through that relationship and our normal role of coordination, we plan to work with statistical agencies in both the Federal Government and in the States to assure the long-term development of effective statistical programs.

The Statistical Policy Division of OMB is looking forward to working with you in the development of an effective program for State and local statistics improvement in the next several years.

The achievements of local programs such as the Columbus effort in the development and utilization of the Geographic Base File will make an important contribution to this entire effort.

Micrographics: A New Approach to Cartography at the Census Bureau

FREDERICK R. BROOME

ABSTRACT

The production of choropleth maps for publication can be a tedious and expensive task, especially when the mapping involves presentation of data for the 3,000 plus counties of the United States. The micrographics mapping system developed at the Bureau of the Census overcomes these problems and produces quality graphics rapidly and at a fraction of the cost of coventional manual methods. This method demonstrates the efficiency of a total system approach where advanced techniques are balanced with conventional procedures.

INTRODUCTION

Several factors have influenced recent developments in thematic cartography. Among the most significant is the application of automation to the production of maps. Not all map-making steps, however, lend themselves equally to automation. This paper describes a total system approach to automated mapping as used by the Bureau of the Census to produce the choropleth maps for the **1969 Census of Agriculture Graphic Summary.** The system is a blend of conventional techniques and a computer-output-microfilm (COM) automatic drafting operation.

ORIGIN OF THE SYSTEM

The techniques of computer output to microfilm are not new to the Bureau of the Census. A computer output to microfilm unit has been in use since the early sixties to supplement line printers. However, except for "forms flash," the COM unit used provided for alphanumeric output only and had no graphics capability. In spite of this, some attempts were made to output graphics; but the equipment limitations were too severe.

In December 1972, the Census Bureau's Agriculture Division approached the Geography Division about the possibility of using automation to generate maps for a planned Agriculture Census Atlas. At that time the Bureau's "in-house" capability consisted of a limited set of programs capable of generating maps on a Geosapce plotter, a drum pen plotter, and line printers. After a review of the publication requirements, the line printer was selected for producing the maps. In addition, a commonly known software package, SYMAP, was available for use.

The Agriculture Atlas, or **Graphic Summary** as it was finally designated, would have both dot distribution maps and choropleth maps. In-house capability did not exist for authomatically generating dot maps. However, there was an excellent program at the University of Wisconsin which not only distributed the dots randomly within predetermined boundaries but also employed a land use geographic filter to assure logical dot placement. The Bureau, therefore, decided to contract for the production of the dot maps but to utilize in-house capability for the preparation of the choropleth maps. This paper is limited to the evolution of the choropleth mapping system from the line printer and SYMAP to the operation of the final system.

Following prescribed procedures, agriculture data was fed into the SYMAP program along with a digital description of the U.S. county boundaries. The program shaded each county area according to the computed class for that county. Shading was done using the line printer by overstriking various symbols. The resulting U.S. maps were made up of several strips of computer paper, each 16 inches wide, taped together to form individual maps about four feet wide by three feet high. A film positive of the U.S. showing state outlines was placed over each. Then the maps were photographically reduced to publication size (figure 1). This product was considered to be comparable to the choropleth maps manually produced for the **1964 Census of Agriculture Graphic Summary** (figure 2).

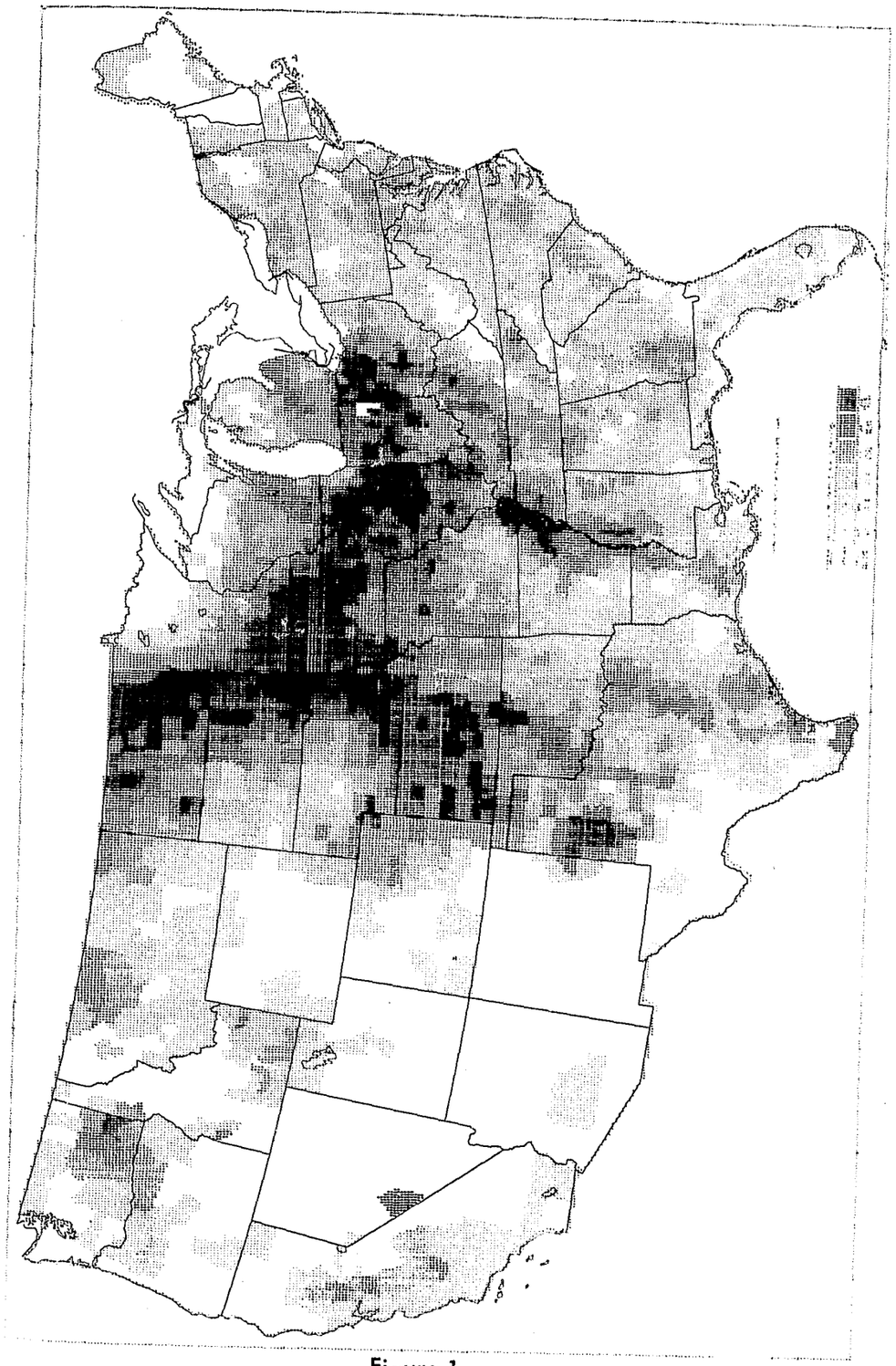


Figure 1

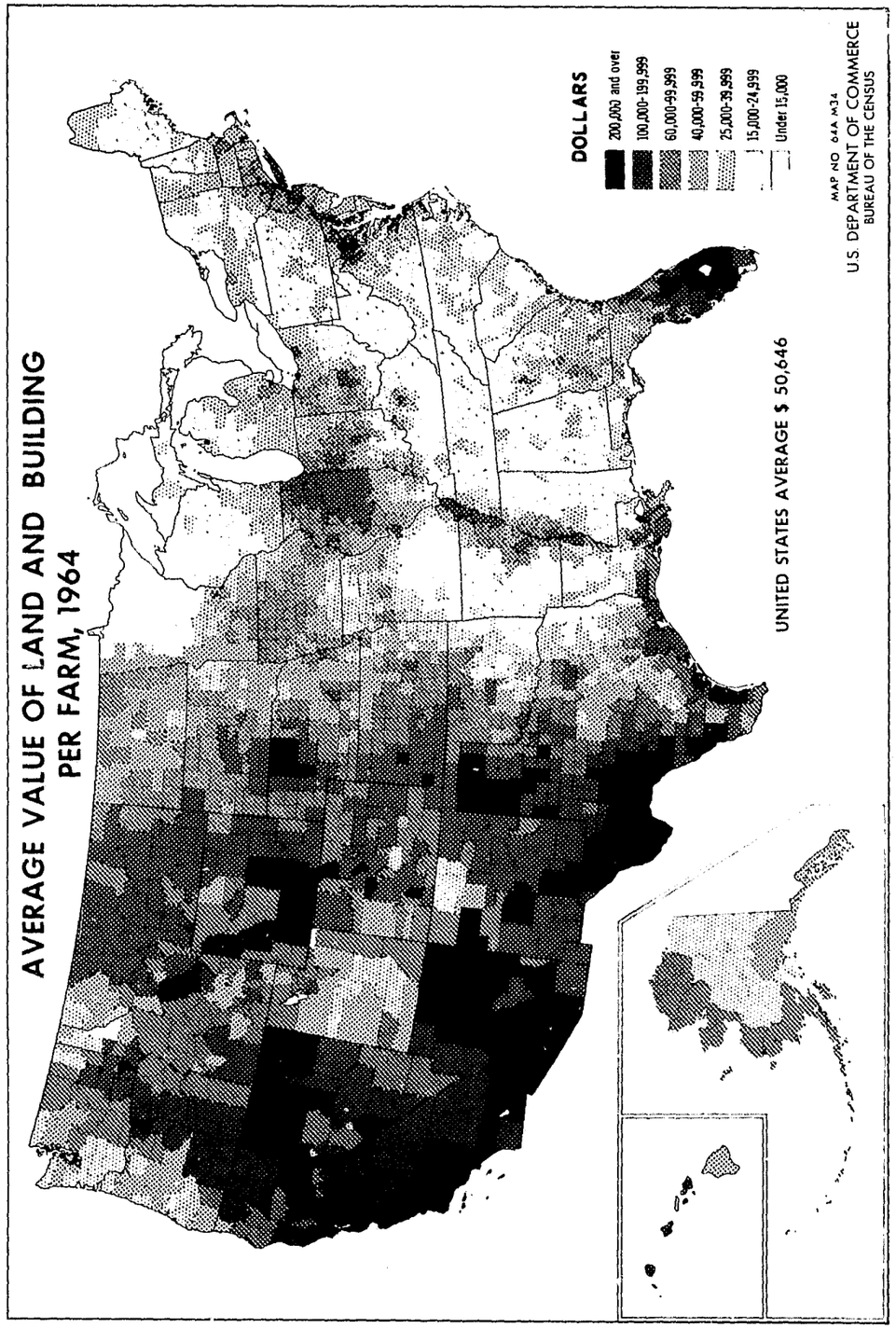


Figure 2

However, during this same time period, the Geography Division was studying the feasibility of using a COM unit belonging to the National Oceanic & Atmospheric Administration for making maps for the Bureau's Urban Atlas Project. This particular COM was able to draw lines (vectors) in addition to alphanumeric characters and had been used quite successfully by NOAA to draw graphs and to prepare data tables in support of their geodetic data publication program. Initial tests proved that high quality graphics could be generated by the COM unit. Since this procedure also avoided the operation of mosaicing and splicing line printer paper, it was decided to prepare SYMAP outputs on microfilm.

Computer programs were written which took the output commands that normally would have gone to the line printer from the SYMAP program and reformatted them to meet the requirements of the NOAA COM microfilm plotter. This permitted an entire U.S. map to be printed on a single frame of microfilm. Shading on the microfilm was still accomplished by overwriting character combinations. While this improved the quality of the shading because the variations in a mechanical printhead were eliminated, the cost per map actually increased. The SYMAP program still had to be run to create the map. Then a reformatting program had to be executed, followed by the COM run. Under the Bureau's charging procedure at that time, the SYMAP maps were costing about $280 each for computer processing alone. This was quite acceptable, however, since it was far less expensive than conventional manual techniques. A similar map produced conventionally would have cost $1,000 or more.

It was soon realized that the COM unit had a capacity for producing much higher quality graphics than those obtainable from merely emulating a line printer. The whole approach was reevaluated. It was quickly apparent that the best quality shading would result by using the computer to produce "open window" negatives and then having shading produced through screens applied conventionally in the photographic laboratory. This procedure also yielded several other benefits that will be discussed later.

A conventional choropleth map making operation usually consists of the following general steps: 1) Extract the data from the data file; 2) Compute the class intervals for each area; 3) Prepare a base map of the areas to be mapped; 4) Produce one open window negative for each class interval; 5) Photographically composite and screen the negatives into press ready copy; 6) Make a proof map.

Computer algorithms were developed to cause the COM unit to produce the open window negatives. The algorithms were incorporated into a series of programs which are the automated equivalent of steps one through four above. An example of the first microplotted test is shown in figure 3. The computer program and plotter run to produce the microfilm for figure 3 took 2½ minutes and cost only $80. This was a significant reduction in cost coupled with a significant increase in quality over the line printer map. As an example of the versatility of this system, the photographic laboratory using the same negatives also produced a color version of the map (not shown here).

The line printer map, the black and white micrographics produced map, and the color micrographics produced map were submitted to the Agriculture Division within four months after the start of the program. After reviewing budgetary considerations, they decided to use the micrographics system to produce the 69 choropleth maps needed for the **Graphic Summary** and to print them in color. Final changes were made to the computer programs to consolidate and streamline the process and production work began immediately thereafter.[1]

Between March, 1973 when the go-ahead decision was made and the following June, the final screened and composited negatives for all 69 maps to be published were shipped to the printer.

THE MICROGRAPHICS MAPPING SYSTEM

The micrographics system developed by the Bureau differs from conventional cartography in the techniques used but not in the steps taken. A diagram of the overall system is shown in figure 4. For purposes of illustration, the production of a typical map will be followed through the system in detail.

DATA EXTRACTION

This program performs both the data extraction operation and the computation of the class intervals. The parameters of the class intervals are usually provided to the Geography Division by the sponsor of the map. Under the direction of control cards, the program extracts from the data file the variable specified to be mapped, and the geographic codes for each state and county. The class for the county is computed and a record containing the state code, the county code, and the class is output for each county. The output file then becomes input to the matching program described later.

[1] Most of these final changes were made by Richard Buhrman who also produced the agriculture data file used in the project.

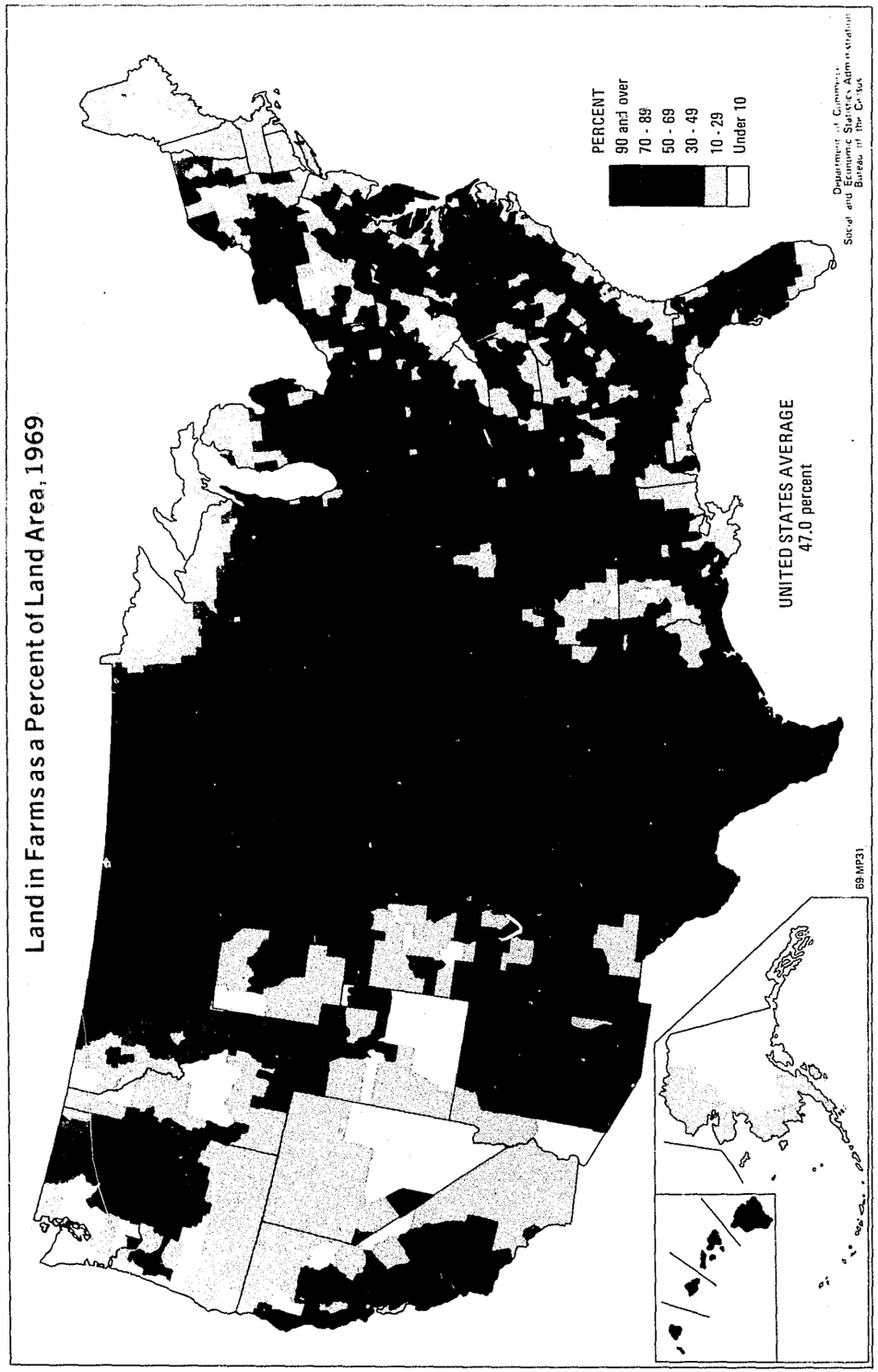
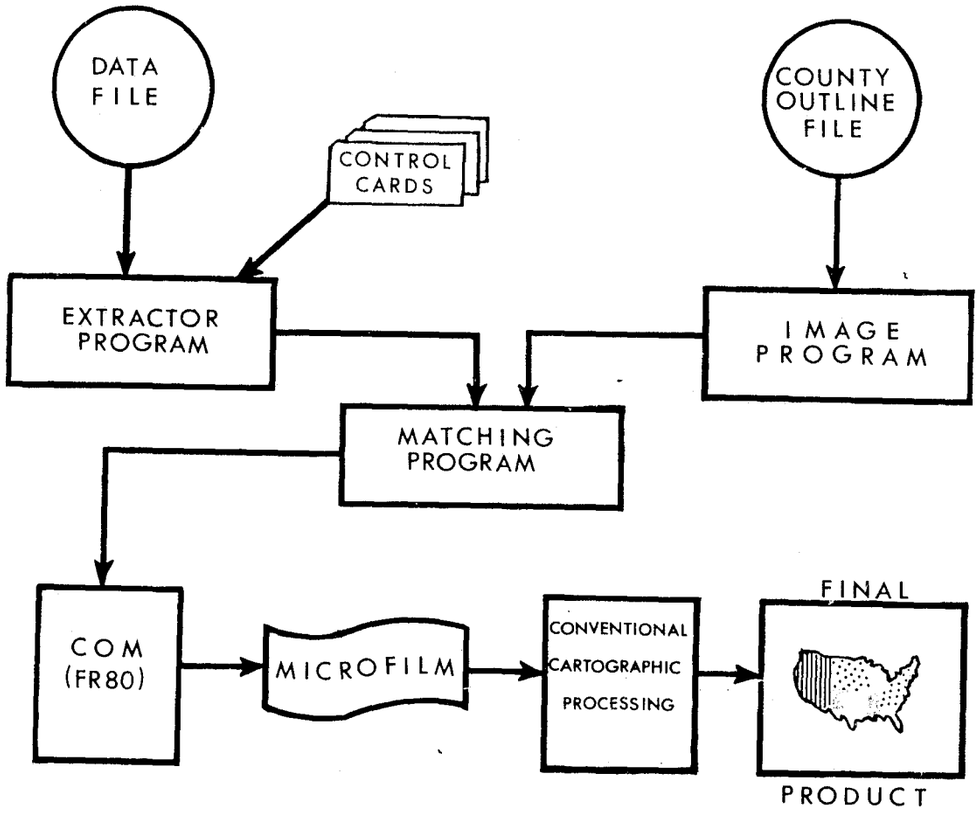


Figure 3

FIGURE 4. **Micrographic Mapping System**



OUTLINE FILE PROCESSING

The heart of the automated mapping system is the program which converts the digital descriptions of the county boundaries into images for the open window negative. A digital county boundary file is merely a computer file containing for each county: a state code, a county code, and the coordinates for a series of points around the boundary. Since not all points along the boundary are included, the boundary is actually approximated by a series of straight line segments. (figure 5.A). The coordinates may be in any given system as long as all counties are relative to the same coordinate origin. The file used was based on the county boundary file originally digitized for the Bureau of Public Roads (now part of Federal Highway Administration, Department of Transportation.)

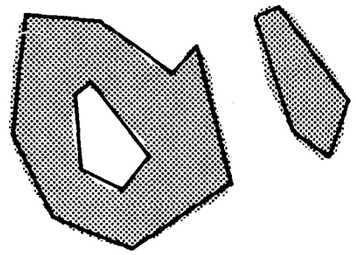
The outline coordinates are read and transformed one county at a time into the size and position that the county will appear on the microfilm frame. The transformation is based on the beam width and beam spacing characteristics of the NOAA COM unit. When the smallest beam width is used and 4,095 points are addressed along each axis, the beam just touches the adjacent beam at this spacing[2] Thus, a whole film area may be "painted" by stepping one unit at a time along the Y axis and sweeping the beam from side to side along the X axis (figure 5.B).

The transformation of the boundary coordinates into the much smaller scale microfilm image may cause two adjacent boundary points to have the same computed microfilm coordinate position. Duplicate coordinates, if any, are removed to speed up the processing. At the same time, maximum and minimum trans-

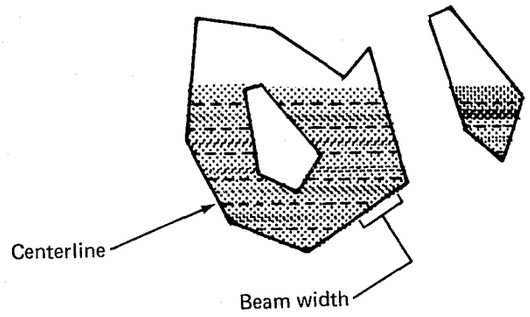
[2]The NOAA COM is capable of addressing 16, 383 points along each axis. However, since the beam width would still be the same, the extra resolution was considered unnecessary.

Figure 5

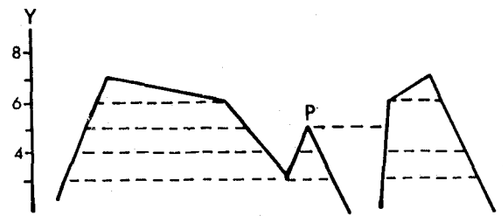
A. Example of county with interior hole and island. Shaded area represents actual county shape. Coordinates for digital file are read at endpoints of the straight lines that approximate the boundary.



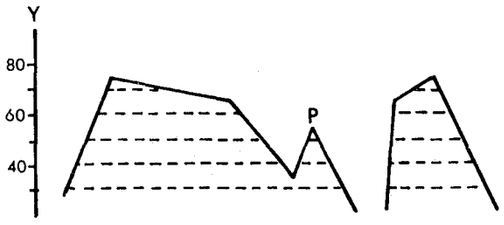
B. County outline showing some paint lines. Dashed lines represent computed centerlines of paint lines. Shaded area represents width of beam. Note slight overlap of beam sweep.



C. Example of what can happen during sweepline computation when program logic does not test for peak or pit along outline. Note the erroneous paint line to island and resulting absence of line inside island.



D. When the endpoints of the bounding lines are multiplied by 10 and have 5 added, they are offset so that none of the endpoints fall on a potential paint line.



formed Y values are determined. The difference between the maximum and minimum Y for each county gives the number of beam sweeps needed to "paint" the image of that county onto the microfilm.

Starting at the minimum Y, the X intersection with the county boundary segments are computed for each Y increment from the minimum to the maximum. This is nothing more than computing the intersection between two lines, a boundary segment and a line parallel to the X axis at a given Y.

Three techniques are employed here to save computation time. First, the slope and Y-intercepts are calculated once for each boundary segment and are stored in another array for later use in computing X-intercepts. Almost all segments will be intercepted at least once and most, more times. Thus, computation time is saved. Second, all transformed boundary coordinates are stored as integers. Doing so permits more rapid searches of the boundary for intercepts since integer operations in a computer are generally quicker than real number operations. Third, the problem of determining whether a point is the beginning or end of a paint line can require some rather complex and slow logic. For example, point P in figure 5.C is at a peak. If this is used as a start point for a paint line, a line would be drawn from there to the island. To avoid this problem of determining whether or not each point is a true start point or just an inflection point, the whole transformed county start point or just an inflection point, the whole transformed county boundary is mutiplied by 10 in the X and Y direction and 5 is added to each value, i.e., mapped into a finer grid and offset by 5. (figure 5.D)

Then the segment points will never be on a Y paint line because the computation for paint line values are stepped by 10 in the Y direction starting at 0. Since there is no ambiguity about intersections, the computation is faster.

The output of the image program is a record for each county giving the state code, the county code, and a series of coordinates for the paint lines that make up the county image. Each line is defined by a Y value and a beginning and ending X value. If the paint line is broken into several segments as in figure 5.B where the lines is interrupted by a hole and a skip to the island, a start and stop X is output with a Y for each portion.

MATCHING THE FILES

The third program "tags" the county image with its computed class by matching state and county codes. The output files from the two previous operations are input to this matching program. When a match is found, the image record is tagged with the class interval for that county and output to an intermediate file. The program then sorts this intermediate file by class and outputs the images by class to tape.

COMPUTER OUTPUT TO MICROFILM

The output tape from the above match operation is read into the COM unit. The commands on the tape direct the plotter to draw the open window negatives, one for each class. All the county images for one class are drawn on the same microfilm frame. The output of the COM unit is 35mm microfilm. A typical series of frames are shown in figure 6.

CONVENTIONAL PROCESSING

The microfilm is delivered to the photographic laboratory where it is enlarged and the final maps are produced by following conventional cartographic procedures (figure 7). First, each frame is enlarged into a film positive at publication size. The individual positives for each enlargement (i.e., class interval) are registered to the others and to a positive outline base map. (These positives are punched as a group to hold this registry.) Lettering, as required, is manually applied to the base map. Next, film negatives are prepared by contacting the positives. The negatives are then screened and composited to meet the needs of the final product. Three composite negatives were prepared for each of the agriculture maps, one for each color, red, yellow, and black.

It should be noted that the COM unit is quite capable of plotting all the lettering as well as shading in the counties in up to 64 levels of gray, or of plotting "dots" in the open windows to produce screens directly. However, the "total system approach" directed that these capabilities not be used. The computer algorithms for aesthetic placement of type are extemely complex, particularly when the title lengths, legend and text wording vary from map to map. The work of human draftsmen proved to be more economical and pleasing to the eye than automated type placement techniques. Although the 64 levels of gray were not available to us through the plotter software package we had to use; even if the levels had been available, the final product would have to be screened like a continuous tone photograph before it could be printed. The computer generated dot screen would have had several disadvantages. First, the computation time and, therefore, cost would have been excessive.

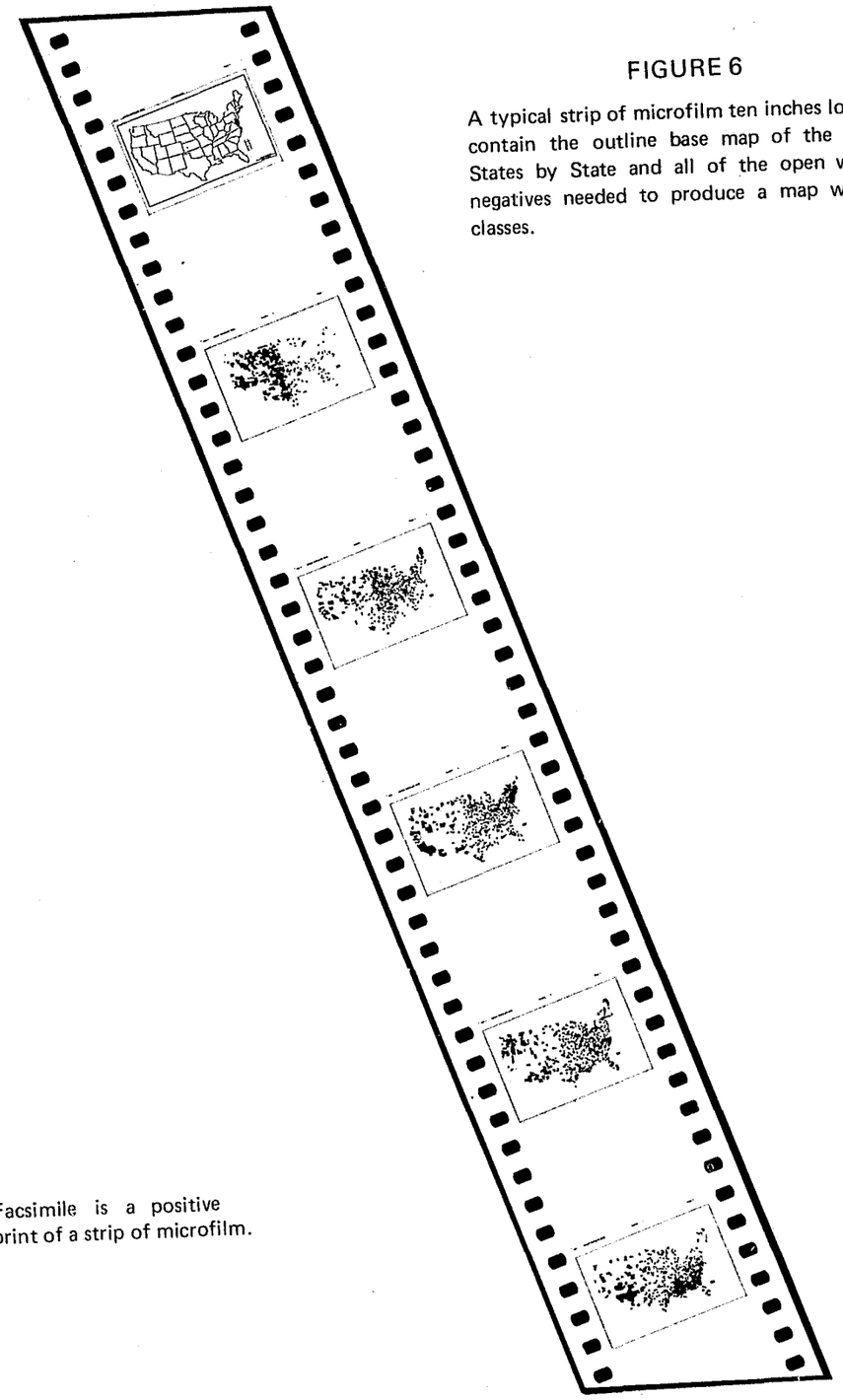
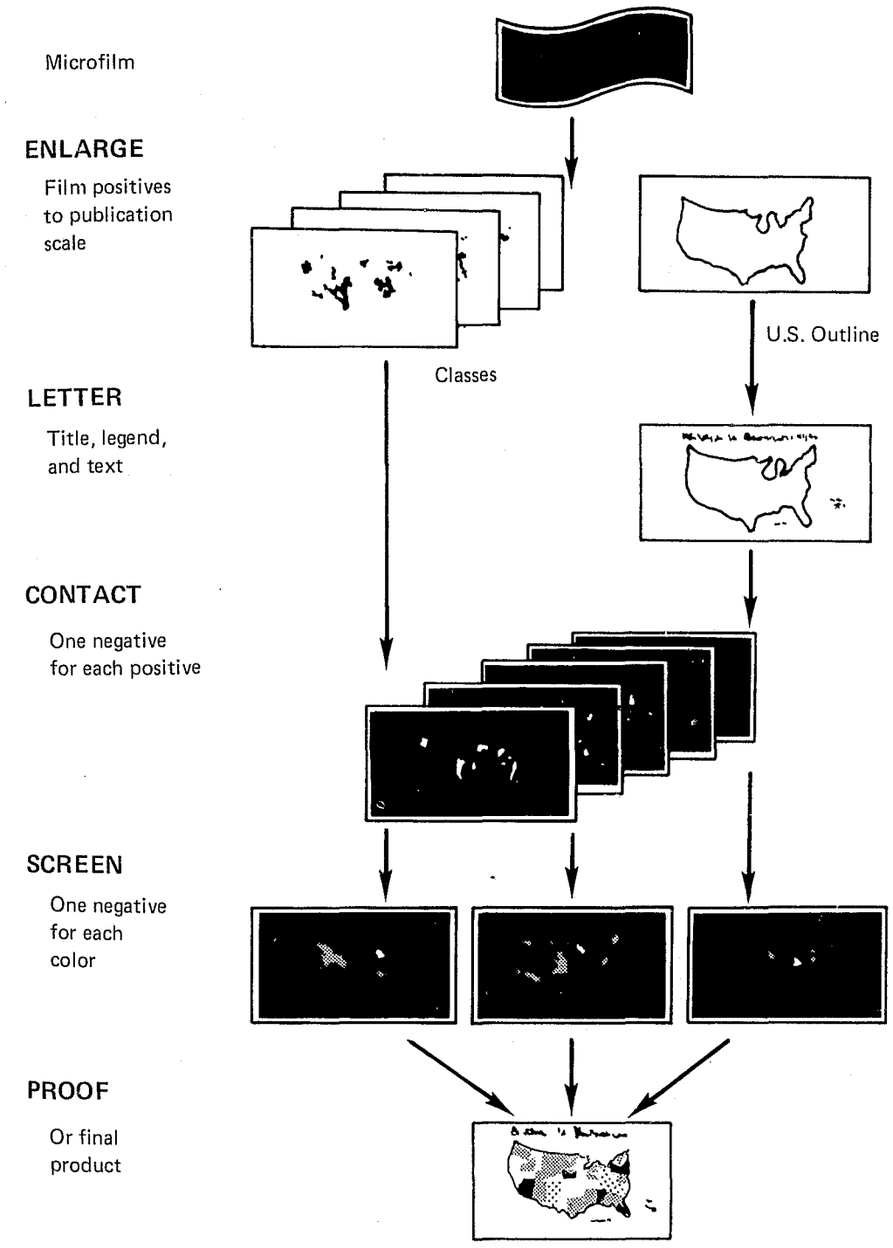


FIGURE 6

A typical strip of microfilm ten inches long will contain the outline base map of the United States by State and all of the open window negatives needed to produce a map with six classes.

Note: Facsimile is a positive print of a strip of microfilm.

FIGURE 7 Conventional Part of the System



Second, the resulting screen also would have been coarse and obtrusively visible after enlargement to publication size. Finally, having the screening permanently on the microfilm meant that any change in either publication size or screen values would require cycling all the way back for another computer run. Thus, screening was done most efficiently during the photographic operations by merely having the photographer insert the correct screen at time of composition—a procedure that takes about 30 seconds for each class interval.

The "open window" negative approach allows the cartographer complete flexibility to modify screens, change from color to black and white, alter scale, and mix the maps of two or more variables; all without going back to the computer for further processing.

SUMMARY

The micrographics mapping system developed by the Bureau of the Census for the **Agriculture Graphic Summary** publication is an example of a total system approach to real production problems. It is a blend of automated and conventional procedures structured to balance each technique, from data input to final negatives ready for the printer, and to balance them in such a way as to take maximum advantage of the benefits of each technique. The system permits complete flexibility in the form of the final product because modifications as required by the changing needs of a project may be entered without costly and time consuming reruns. And most important, it has reduced the cost of choropleth maps from one-fifth to one-twentieth of those produced entirely by conventional manual techniques.

Bibliography

Morrison, Joel L. "Changing Philosophical-Technical Aspects of Thematic Cartography." *The American Cartographer,* Vol. I, No. 1, 1974.

_____. "Automation's Effect on the Philosophy of Thematic Cartographers." *Proceedings of the American Congress on Surveying and Mapping—American Society of Photogrammetry, 1972 Fall Convention,* October 1972.

Peucker, Thomas. "Computer Cartography." *Commission on College Geography,* Resource Paper No. 17 (Association of American Geographers, Washington, D.C.) 1972.

Puhl, R. J. "An Approach to Micro-Publication of Geodetic Data Utilizing the Computer, Microfilm, and Standard Printing Techniques." *Proceeding of the American Congress on Surveying and Mapping—American Society of Photogrammetry, 1974 Spring Convention,* March 1974.

Robinson, Arthur J., and Sale, Randall D. *Elements of Cartography* (John Wiley & Son, Inc., New York) 3rd Edition, 1969.

Schweitzer, Richard H. "Mapping Urban America with Automated Cartography." *Proceedings of American Congress on Surveying and Mapping—American Society of Photogrammetry, 1973 Fall Convention,* October 1973.

U.S. Bureau of the Census. *Census of Agriculture,* 1964, Volume III, Special Reports, Part 5, "Maps" (U.S. Government Printing Office, Washington, D.C.) 1969.

_____. *Census of Agriculture,* 1969, Volume V, Special Reports, Part 15, "Graphic Summary" (U.S. Government Printing Office, Washington, D.C.) 1974.

Question Period

Mr. Alan Pisarski—There is always a problem in the way you cut off different strata of the variable you are plotting and where you draw the boundary lines for your different colors or shadings. You can distort, I guess you could say, even bias, information by selectively picking the strata, or simply just not portraying the information as appropriately as you might. One of the answers frequently has been identified as a kind of an interactive capability to display. Look at what kind of display you get given the stratification that you have established, and then modify slightly, and take a look again. Does your system have that capability, or how do you decide these boundary questions?

Mr. Fred Broome—I am going to partially side-step that by saying that the boundaries are provided by the sponsor of the map that we are making. If we are doing a population map, the Population Division tells us what the "cut" points are for the classes. However, I would like to elaborate a bit. No, our system is not interactive at this point in time. The micrographics part of it is batch-oriented. We input control cards which then give the class interval cuts, and the system then goes from there more or less in an automated fashion. I would agree with you that a bias can be introduced by the way you "cut" the data.

Mr. Alan Pisarski—Does that mean that given the flow chart of the system that you described, that you have to go all the way down to the bottom of the flow chart to get a map back before you know what it is going to look like, and then interact? Is there an opportunity somewhere in between to think about what you have accomplished and whether you want to change it in color tone or in the stratification of the variable?

Mr. Fred Broome—At this point in time we cannot interact with it. We do have to go through the processor procedures before we can see if it is acceptable.

Mr. Roger Haiflich—I was wondering whether this capability would be available to people who are using the CUE program so that when we do an extension, or an update, we could send in our files at that point and get a set of maps back. Then we could do a visual editing on our own.

Mr. Morton Meyer—You have posed two questions. As regards to the GBF/DIME-File updates, we are plotting the maps and reviewing the plots so that we may detect and correct erroneous coordinates. As part of the process, local agencies will receive a set of hard-copy plots covering all maps for which there is a GBF/DIME-File. As regards to the micrographic plotting programs that Mr. Broome has developed, these will be made available to local agencies, without cost, by the Bureau of the Census. Once the plotting programs are available, you can develop your own mapping capabilities and produce your own maps locally. COM units are generally available, and time can usually be obtained at very moderate costs.

Dr. Robert Aangeenbrug—There will be a conference held in the beginning of December that deals with automation in cartography. It is a rather important area of research and development, and I will try to distribute, it we can find the time, an invitation to some of you that are interested in the technical area that Mr. Broome is working on. I urge some of you to ask either Mr. Broome or myself about this conference. We will try to reproduce invitations for those of you that are interested in this.

Mr. Pisarski has also alluded to a number of problems that we have as users; one of these is the tyranny of automation, which raises some very serious questions with perception of this information. Some of you may be aware of the phenomenon of ecological correlation, and if there were some sociologists here we would probably hold a pretty good argument about some of these maps. I think they are somewhat peripheral to this particular conference, but we are currently negotiating, at the instigation of the Director of the Bureau of the Census, to hold a conference some time in the spring or summer. Currently we have a very wide interest of at least six organizations. We will distribute to the SMSA's any information that we can give you about this conference.

Statistical maps can present very vexing problems. I think Mr. Pisarski particularly, who has a lot of graphics experience, is aware of the problems of how to lie or tell the truth with maps. One of the problems with the four-way correlation you saw, for instance, is that it really tends to create more problems for certain analysts than you had before—the unequal size of counties, and so forth. I urge you to keep after Mr. Broome, or myself, or Mr. Meyer, or other people that have interest in the professional cartography aspects of this, so that we can get some of your feedback about this problem. There are some excellent opportunities on the horizon, and also some great problems. The costs to resolve these problems, however, are rather enormous at this time.

DIME Files and the Fort Wayne Experience

ROGER L. HAIFLICH

This paper deals with three subjects—our present and past projects in an overview, setting forth some of the myriad uses of these files; the failure of current technology to create interest enough (and hence to obtain enough money) to maintain the system, setting forth the problem areas; and the experimental system which proposes to solve these various problems.

Up to now, most of our requests have been for one-time runs, either in support of a project or to give insight into distribution of data for presentation and management review. Though not a complete listing by any means, the following examples give a good cross-section of these requests.

TRAFFIC ZONES

The first project (which induced study into the possiblities of the DIME file) was to assign 103,000 motor vehicle registrations to specific traffic zones. After approximately two months of study and work, about 90 percent of the task was accomplished—some 92,000 registrations had been assigned.

FOOD STAMPS AND WELFARE ASSISTANCE

These reports and displays served a dual purpose—they were used by the Director of Public Welfare to determine case load spreads by case worker, and they were used as input into the NIRCC over-all planning information file (which supports the HUD 701 program).

HOUSING AND BUILDING

Reports and displays on demolitions, houses brought to code, and new building permits were used in conjunction with the HUD and Regional Planning Agency's 701 program, and the new Housing and Community Development Plans and Programs. The data is also useful in showing growth patterns and is currently being used to update the transportation plan.

HOSPITAL EMERGENCY ROOM PATIENTS

These reports and displays are being used to help determine if there is a need for expansion of present hospital facilities or if, in fact, an entirely new facility should be built. If a new facility is decided upon, these reports will be of value in determining the best location.

YWCA

The question here was where to locate the new building. The YWCA provided a 20 percent sampling of their current membership, which was then displayed by tract. The results of that display dictated the abandonment of two suburban sites in favor of a downtown site.

SURVEY SAMPLE DISTRIBUTION

Numerous surveys are presently being conducted concerning public transportation and public feeling about urbanization trends. Here the purpose of each request is to verify the uniform spread of the sample and then to study the distribution of the returns, thereby yielding greater insight into the results. Input for selection has been either the vehicle registration file or the voter registration file. The output is mailing labels with area identifiers in the upper right corner.

FORT WAYNE COMMUNITY SCHOOLS

This has been the most encouraging and enthusiastic user so far. After seeing the potential, they decided to convert all student record files to Indexed Sequential Files utilizing a concatenation of tract, block, and street code as key. From this base they are now making forecasts as to student population in schools, as well as some boundary adjustments. Future uses include bus scheduling, racial balance, and student shift forecasting. We have been running a match for Community Schools on a quarterly basis, but they have asked for more timely reports.

This, then, brings us to our current status—and Part Two of this presentation.

One thing that should be stressed is that we labor under very difficult circumstances because one political party controls the city government and another controls the county government. I represent the County Data Processing Agency (a nonpolitical entity) and out work in the planning area is authorized by the Northeastern Indiana Regional Coordinating Council.

Major users are under the authority of city government and therein lies the crux of our problem—as well as the impetus for our current efforts. In order to persude the differing factions to sit down and draw up working agreements concerning the DIME file, we must show a system so valuable as to overcome the rivalry. The purpose of this presentation is to explain those efforts.

To quote Abraham Lincoln, "If we could first know where we are and whither we are tending, we could better judge what to do and how to do it." In an attempt to accomplish this, we have tried to discern the basic problems as they now stand and then, using this discernment as a starting point, to proceed toward solution. Following are the six basic technological problems presented by the current system.

1. **Cost**—the cost of setting up programs for varying types and sizes of records, as well as differing median for input. The staff of the Census Bureau deserves credit for attempting to make their programs universal, but there still remains a considerable job of setting up the report programs, the various sorts necessary, and the format cards used in the ADMATCH system. Also, as was previously discussed, the cost of maintaining the DIME file itself is significant.

2. **Nonconformance**—The lack of common areas for aggregation is a serious problem. To be effective, the change required in the methods currently being used or in the areas used for planning districts must not be too drastic. Most users are from three main sources—the school systems, the Board of Safety (fire, police, EMS), and the planning agencies. Of these, only the planning agencies are willing to stay with tract and block, and even they are interested in increased capabilities, such as traffic zones.

3. **Splits**—The school system splits boundaries down the center of streets, whereas the police want both sides of a street in one area. Planning areas for police and fire departments also split many census blocks due to differing land uses within the block. This is most prevalent in the suburban areas where the blocks are larger, but it is still a point of contention.

4. **Flexibility**—All users demand the flexibility to change their boundaries at any time, but they do not want massive cross-reference tables which must be updated and corrected with each boundary change or when additions of new streets create new blocks and segments.

5. **Timeliness**—The time lag in updating a DIME file is too long if current practices are continued. The most frequent comment received from the user departments is, "If we must maintain voluminous cross-reference tables of tract and block to our area, why not just give us an ACG which we can maintain outselves at much less cost. We don't care about coordinates, node numbers, enumeration districts or maintaining census block identifiers. We can simply add streets and address ranges to our files and forget the DIME file."

6. **Displays**—As for map displays, tracts are nice, but most agencies want to see data displayed to their areas. This presents a problem of developing a new map for each entry due to demand for different areas and to differing scales.

We now come to Part Three of this presentation—an outline of what we are currently doing to overcome these problems.

For this purpose, shortly after returning from the DIME workshop in Houston, I began developing the AIAD (Automated Information Aggregation and Display) system, which is now 70 percent complete.

The major innovation to this system is the use of segment centroids and segment side centroids for its aggregation parameters. Two attributes of a DIME file contribute to the system—one is the coordinates sets and the other is the fact that by examining the node

orientation of the segment, the quadrant can be determined as well as the direction of the segment. All identifiers are then split left and right, and the centroid can also be split, forming two separate parallel segments for aggregation purposes.

Returning to the six problem areas referred to, here is an explanation of how the AIAD system will affect each one.

No. 1 was the cost of customizing all programs and sorts for each job. The AIAD system eliminates this problem by setting up a front end program which allows for data to be entered on cards or tape (blocked or unblocked). The user indicates up to 40 characters of key information in one or two contiguous areas, along with the starting position of the address and its length up to 35 characters. This data is then used to reconfigure the input into a standard format for entry into the system.

This approach eliminates the need to modify any following programs as the location and size of the components is fixed. Also if the user desires, the final results can be appended to the source records by matching keys and transferring the geocoded data.

No. 2 is solved by requesting the user to bound his areas of interest using node numbers off of the metropolitan map series. Only those nodes at vertices or points of change along the boundaries need to be specified, thereby greatly reducing the task of bounding the areas.

For no. 3, the user merely indicates whether the boundaries are to aggregate to one area or split into separate areas, using only one option per submission. By utilizing segment centroids or segment side centroids, aggregation is controlled. Also, splitting of blocks does not affect aggregation, as the unit of aggregation is in effect segments, not blocks as in other systems.

No. 4 is alleviated by entering aggregation area boundaries using node numbers. All users are furnished copies of the latest update maps. As for new streets or additions, the user need not concern himself unless he wishes to move his boundary, as any occurrance will automatically aggregate to his current areas of interest with no input from him at all. If a change of boundary is desired along a new street, making the boundary change is as easy as moving a line and reentering only those new nodes which describe the new boundary.

No. 5 is affected only to the extent that this system can get sufficient support in the three areas of principal influence, which are the planning agencies, the school system, and the Board of Safety. Support from these agencies would enable us to exercise the update responsibilities. Purdue University has responded favorably to our using their digitizer for updating the coordinates, and we still have some of the people available who originally started the DIME file coding. However, to achieve this particular responsibility, we must receive more financial backing.

As for no. 6, the user must simply indicate whether a display is desired and if so, what scale is desired. The system then calculates from the nodes given the necessary map parameters and sets up the necessary punched output deck. A later addition to the system will be the ability to indicate display splits. The system now produces a map with ten shades of gray with equal ranges per shade.

There are many inherent advantages to be gained by utilizing this system, one of which is improved accuracy, since one agency maintains the updates. The Census Bureau gains by having an up-to-date accurate file available.

It will very likely be some time in December when we determine if support will be given to this system. We are quite confident, however, that the advantages are evident enough to warrant the necessary approval.

Question Period

Mr. Anthony Savoia—Is this software that you are developing proprietary software? Will it be in public domain once it is completed?

Mr. Roger Haiflich—Yes, it will be in public domain. When we are done, we hope to have it documented to the point that anyone who wishes can use it. There are a couple of things that I should point out. The one thing necessary to make the system work is the mapping program that is utilized, called AUTOMAP 2. It is available from Mr. Jack Dangerman at Stamford Research Institute. It costs $450.00. The reason I utilize it instead of SYMAP is the fact that Mr. Dangerman's program put out an intermediate file, which he calls a base map image file, and that is what I utilize for the aggregation portion of this program. Financially, you cannot go in on every individual data item and calculate the area it is going to aggregate into. What you do is set up a base map image file and then utilizing the coordinates on each data item, you calculate the row and column where this data item will aggregate and pick up from the base map image file the area that it belongs in. What you can do is calculate row and column in one program, sort them down and aggregate them together, and then only look them up once. It becomes a very cheap way of getting a multi-purpose aggregation program.

Mr. Anthony Savoia—What language are they written in? Is it FORTRAN?

Mr. Roger Haiflich—It is Assembler and FORTRAN both.

Mr. Anthony Savoia—Do you have an abstract of the system now?

Mr. Roger Haiflich—I have an abstract that I wrote on the theory of the system in May of this year. But the system has changed considerably since this abstract was written. I would really rather wait until December, at which time we should have a more current abstract written.

Mr. Anthony Savoia—How are you making that known to other individuals that might be interested?

Mr. Roger Haiflich—I was hoping that as a result of these workshops, the Census Bureau would form an information exchange. The Census Bureau could then take pertinent reports and abstracts, get copies made, and distribute them to those people who are on a mailing list or in a users group.

Mr. Alan Pisarski—I think this is really quite an important topic. At previous conferences, usually around the second day, we start to get into a fist fight about how one builds toward these kinds of aggregated areas. I guess there are a couple schools of thought. In my experience in the Council of Governments in Washington, we built the monster converter files which you talked about. Washington is a city of about 30,000 blocks, so it took a fair amount of doing to build these converters, that is to say, a block-to-zone, a block-to-school district, a block-to-whatever. We usually put the onus on whoever was proposing the use of that system for aggregation to do that work; it was their responsibility and that's one way we got out of it. Of course, physically and logistically that is a difficult job and tedious.

There is always a group of proponents at a meeting that talk about using a polygon search technique, the point-in-polygon kind of thing, and insist that is the rational way to do it. I always have difficulty with that. It seems to me perhaps more elegant than is required. If I understand what you are suggesting here, it sounds to me like you have found the kind of intermediate position that makes an awful lot of sense—if I understood it, I am not sure that I did. Having identified the boundary nodes around a zone, what processes are then used to aggregate the internal components of that zone?

Mr. Roger Haiflich—AUTOMAP, itself, has a point-in-polygon logic in it, which it then uses to produce a computer generated map like SYMAP does. What it generates is a matrix, built up by line, and it goes across each line and writes the area identifier into each print position. Utilizing that fact, all I did was use these same calculation formulas for each data item as it came in, to calculate row and column. Then I sort these by row and column and pass them against the base map image file, which then automatically picks off the aggregation area. By utilizing the segment centroid, you aggregate your data item into only one area because the last area identified along a boundary will control that whole boundary. Whereas, if you use what I call segment side centroids, where you have split a segment, you can actually split data items on that street into the two different areas along the boundary. As I have said, it becomes a very rapid method of the point-in-polygon approach, because you only actually calculate point-in-polygon once. We tried mathematically to do it the other way, and it was totally unfeasible from the standpoint of cost.

The other point I should mention is that this is written for a 370/145 under DOS/VS with FORTRAN IV, and basically the programs that are in assembler language are ones that would be rather simple to convert.

Mr. Alan Pisarski—The comment I would like to make here, and it really looks back to something that Mr. Broome said in his presentation, and that is, when you develop a technology, such as micrographics, one of the outputs that never seem to surface is the fact that you have learned how to do something cheaper. Computer technology just never seems to work that way. The reason for it is, and I think Mr. Broome's presentation really brought that home, that you learn to do something that you used to do in a new way and learn how to do it cheaper. What frequently happens is that that opens up horizons, and those horizons mean that now you can do things that you were never able to do before. You probably end up spending as much or perhaps more money. You really have greater capability but you spend as much money or more. It is probably a very important point to make when you are trying to justify budgets because you have a manual system that cost a $100,000 a year and you end up with an automated system that takes a quarter million dollars a year, and somebody says we really have not made progress. It is not too easy to quantify what it would have cost just to replicate the old technique.

It comes back, also to your presentation,—this is kind of a pet comment of mine that I like to throw out in discussions like this—we start out with computer technology frequently doing old things in a new way. We really do not change what we do, we just kind of replicate the old methodology and then have a new form of accounting. It takes a while after you have gotten accustomed to that—after you have accomplished that—to really begin to look at problems in a new way and begin to use the inherent logic of your new tool, as opposed simply to thinking in the same old way. I think this question of areas systems is a good example of it. What we are doing really, is having a lot of different agencies and organizations to work with fixed boundary areas systems—traffic zones, school districts or what have you—that they are accustomed to working with, and then we simply add up statistics for them. I submit that in many instances that kind of a backwards process is really left over from our inherent inability to do anything else years ago.

Today I think we can begin to use area as a dependent variable rather than as an independent variable. Instead of establishing what the area are before we start and then adding things up to it, we can actually begin to look at area as a dependent output variable and say, "Over what area does this information pertain? Over what area does this characteristic remain significant?"; and begin to generate areas, say for school district boundaries, or health districts. The example that you gave was the services area of public health people, but I think there is even more here when we get away from the inherent bias of aggregating statistics over fixed areas.

Mr. Jacob Silver—From what you have just said, the work that you have done is carried out as the people have come to you with a problem and you have helped them solve this problem. Several cities which have this same type of situation have been getting a number of agencies and organizations together in a consortium type of association. Is this the general direction that your agency is moving? Is it possible that you could go this way? If you have not, are there any barriers to this type of informal or formal association?

Mr. Roger Haiflich—I think when people originally approached us and we told them that they were going to have to develop large cross-reference tables, which were then going to become outdated, they tended to turn off. What we are hoping to do, once this system is proved out on all our data, is get everybody together in one meeting and start to set up guidelines. We have currently had meetings with the larger users of the system. I do not know, maybe I have not answered what you were really asking; but we are, in fact, planning on getting everybody together into one user group.

Mr. Jacob Silver—Has something been done already? Or is it planned?

Mr. Elias Samaan—I would like to answer that question. In 1971 an attempt was made between our Department and the Allen County Data Processing Department to get department heads within the City-County Building together and explain to them the GBF/DIME-File, and the advantages of using this file. The results were bad, and many department heads, at that time, thought it would cost their departments extra time and effort which they did not have. At that time, we knew that our approach was not proper. Later we started meeting separately with the differnet department heads that we thought could effectively use the DIME File and benefit from it. These dapartments, such as Welfare and Fort Wayne Community Schools, once having realized the advantages, agreed to work closely with us, and we now feel that shortly another general meeting or seminar would be beneficial. I think the key issue is the saving in time and money that will be incurred in addition to the planning statistics or facts that will help the policymakers in making decisions.

Mr. Alan Pisarski—I think there is a point in that that is worth referencing, and that is, there is a tendency toward standardizing on a fixed system. This is true in any coding structure. You would be astonished in Washington at the amount of emotion that gets involved in coding systems. People can absolutely get incandescent on the subject. I am talking about commodity coding and all kinds of strange coding processes. In the area of geographic coding, it seems to me to be a matter of little concern to get the world to standardize on a code, and say, "Everybody use the same."

Coming back to my previous point about generating areas of output, I think that areas are different for different functional purposes, and I think they ought to be. They should follow the boundaries of the problem. The crucial thing is to build the conversion capability—the matching capability—so that you can move between these different data sets when you need to for relating statistics. I think that is far more important. I know that in the Department of Transportation, at the national level, our program is much more keyed to developing conversion capability to move between different data sets, aggregate them, and relate them, than it is to fixing on a standard.

We held a survey conference on National Systems for Geocoding and there are 67, maybe 68, major national code systems in operation. Everybody came to the meeting to defend theirs, and to assure that we were not going to create another standard coding system for the country. It was obvious that the country did not need another standard coding system. So our whole approach has been to be able to relate between those sets and make each of them a little bit more comparable.

Law Enforcement Applications of Geographic Base Files

JAMES A. F. KELLY SUSAN A. BRAYFORD ALLEN L. PEARSON, II

Introduction

JAMES A. F. KELLY

We welcome the opportunity to visit with you and to tell you about what we are doing in this field. To start with, let me point out that we are taking a team effort approach to our project; this amounts to an integrated staff between the IACP and the Geography Division of the Bureau of the Census. We are going to do our presentation today in a team fashion as well. After I make my remarks, Ms. Susan Brayford will give us an overview of the staff work done to date on this project. Mr. Allen Pearson will discuss the applications of geobase files currently being employed by law enforcement agencies and their plans for the future. Now, it is my task to brief you on the nature and mission of the IACP in general.

The IACP, incidently, is the International Association of Chiefs of Police. Along with describing the organization, we will also try to describe to you our hopes and objectives for this specific project. The IACP is a not-for-profit membership association made up of approximately 10,000 law enforcement executives and security executives throughout the world; however, the bulk of the membership is from the United States. We have a full-time staff of 150 people who conduct research relevant to the law enforcement objectives and then transfer that knowledge through written reports, training sessions, and technical assistance to State and local law enforcement agencies. We serve as coordinators for professional and technical advancement by dealing internally with many law enforcement agencies. Our staff members go to various law enforcement agencies to help them on their projects and to present training programs. Therefore, we get feedback directly from the people who are involved at those locations. We are in a position then to share information gained through our own learning experiences.

Cost effectiveness is another important consideration when we are dealing with the law enforcement agency. We have heard some discussion about LEAA and the kind of money that is being pumped down into the State and local government, including funds for specific projects that are not operation oriented. The lessons from the various projects that we have worked with point out to us that the project efforts conducted at the local level, and experiments with systems at the local level, can be shared to find optimal needs and methods for performing these tasks at a later date by other agencies.

Even though there is considerable fragmentation of state and local law enforcement agencies (approximately 16,000 in the U.S. alone), there are many needs, functions, and procedures in common. Data processing equipment requirements and availability differ, however, which affects the outcome and even the planning of the particular project that we will be discussing further today. This is because some larger organizations have the capability to develop operational programs and the funding to experiment with other kinds of programs during development. There are a number of smaller agencies, however, that must talk about sharing and must talk about programs that will meet their particular need.

Overview of Work to Date

SUSAN A. BRAYFORD

The Research Division of the International Association of Chiefs of Police and the Geography Division of the United States Bureau of the Census are currently undertaking a project funded by the Law Enforcement Assistance Administration to examine existing law enforcement applications of Geographic Base Files.

The staff is being guided in its work by an Advisory Committee made up of 14 members representing criminal justice planners, geographic base file theorists, and law enforcement administrators in agencies that currently have or are planning to adopt such systems.

During recent years, a number of law enforcement agencies with access to advanced computer facilities have acquired a capability for handling the spatial dimension of crime or police operations on a computerized basis. Systems already developed include San Francisco's CABLE, New York's SPRINT, and Kansas City's ALERT. Dallas has developed a sophisticated method of computer assisted dispatch; St. Louis is experimenting with a vehicle locator system which they will tie in with their current criminal justice information system—the list of such system development continues to grow. A conservative estimate of cities planning to develop systems by 1976 will be over 200. However, most of these individual agencies are unable to undertake comprehensive system development by themselves due to lack of money or trained personnel.

In many cases, these law enforcement agencies will be seeking LEAA funds and will be paying for new requirements analysis and systems design and conceptualization studies. In effect, they will be duplicating efforts being carried out in other parts of the country. Even if an agency wanted to model its system on one already operational, a considerable amount of money would have to be spent on adaptation due to the lack of common criteria and a common operational environment.

In view of these extensive activities, there appears to be a need to consolidate and coordinate the efforts of various agencies in the field of geoprocessing in order to provide a more uniform basis for comprehensive systems development. Therefore, the major objectives of the GBF for law enforcement projects include inventorying all capabilities to be derived from GBF systems, synthesizing the best elements of each system into one or more optimum systems for large and medium-sized police agencies, preparing necessary manuals and programs, and presenting a series of workshops for law enforcement administrators on the system and its applications. The staff is also researching critical issues relating to law enforcement GBF applications such as file management, privacy-security of systems, lowest level of file detail, intergovernmental relationships, hardware sharing, and other issues to be identified during the project.

A survey has been conducted of approximately 250 law enforcement agencies to determine the status of GBF planning, implementation, and operation. Some results of this survey are presented later. Using our survey returns and with the assistance of the Advisory Committee, eight cities were identified for an indepth study. The purpose of these visits is to acquire detailed information on the operating capabilities of the systems, costs relating to development and maintenance, operational problems, and planned needs for additions or revisions of the GBF systems.

The sites selected were: Dallas, Tex.; Long Beach, Huntington Beach, San Francisco, and San Jose, Calif.; Kansas City, and St. Louis, Mo.; Jacksonville and Miami, Fla. In the upcoming year, a descriptive report will be issued that will include the survey results and site visit reports. Also, papers will be written on requirements analysis, systems inventory and conceptualization, and model system(s) design. Finally, there will be 10 regional workshops held to disseminate information on the planning and development of a GBF, relations with other elements of city government, and uses of GBF technology in law enforcement operations and management.

The project staff developed the Geographic Base File Survey that was sent to approximately 250 law enforcement agencies. Since a goal of the project is to develop a model GBF System for medium and large police agencies, the cities and counties surveyed all had populations of above 70,000. The survey had four major sections: section one gathered general information on the agency surveyed; section two referred to the status of GBF's in the geographic area of the agency; section three was aimed at obtaining information from police agencies that had a GBF system in operation; and section four requested information from agencies that were planning for the use of a GBF.

As of November 2, 76 percent of the agencies completed and returned their questionnaires. Data from these 190 questionnaires were analyzed and included in this report. Seventy percent of the responding agencies indicated that they are currently using or are planning to develop a GBF system. Of the 64 agencies currently using a GBF, 58 percent indicated that they are using DIME (Dual Independent Map Encoding) or DIME derived files. Those files developed in-house usually follow a format similar to either ACG (Address Coding Guide) or DIME but require additional information such as intersection data, exact addresses, or common place names. Nearly all geographic base files used by law enforcement agencies add geocodes such as police reporting areas and beats.

Several areas of potential use for a GBF system were identified: computer-assisted dispatch, crime trend analysis, planning, research/statistics, and resource allocation. The chart above indicates the current and planned uses of GBF systems in the departments which have an operational file.

Application	Operational	Proposed	No response	Total % operational or proposed
Computer-assisted dispatch	7 (11%)	42 (66%)	15	77%
Crime trend analysis	44 (69%)	17 (27%)	3	95%
Planning	48 (75%)	11 (17%)	5	92%
Research/statistics	48 (75%)	11 (17%)	5	92%
Resource allocation	36 (56%)	23 (36%)	5	92%

Seven of the 64 agencies currently have computer-assisted dispatch applications, while two-thirds of the agencies planned to add this application in order to reduce the time it takes to respond to citizens' calls for service. Forty-four of the police agencies have crime trend analysis capabilities while an additional 17 plan for implementation. As can be seen, most of the identified applications are geared toward information retrieval and report generating data that can be used for manpower allocation and distribution. In addition, other possible applications were indicated: address verification, address history, vehicle locator system, and crime mapping. Of the 64 agencies currently using a GBF, 73 percent operate on hardware shared with other local agencies. Seventy-three percent also indicated that the GBF served only one law enforcement agency. In addition, agencies were asked about the level of address detail that is currently contained in their file. Thirty-three agencies use files which have exact addresses, 36 systems have block face information, 33 have intersections, and 41 contain police districts or beats.

Similar questions were asked of the 73 anticipated users of a GBF. They foresaw the following applications:

Application	Planned
Computer-assisted dispatch	50 (68%)
Crime trend analysis	60 (82%)
Planning	59 (81%)
Research/statistics	60 (82%)
Resource Allocation	60 (82%)

Again, computer-assisted dispatch was anticipated for inclusion by a lower percentage of police agencies, while the other four applications were expected to be used by about 80 percent of the agencies. Concerning the level of address detail to be contained in the system, 43 agencies expected to include exact addresses, 43 departments planned to use block face, 45 agencies anticipated using intersection data, and 43 departments planned to include police districts and beats.

Of these 73 agencies, 65 percent expected to operate the GBF system on hardware shared with other local agencies, while 57 percent indicated the system would serve more than one law enforcement agency.

Applications and Operations

ALLEN L. PEARSON, II

There are several law enforcement applications of geographic base files as indicated earlier. Some of the more common uses are in the areas of computer-assisted dispatch, crime trend analysis, research/statistics, and resource allocation.

As major cities are remolded by expanding population and development, law enforcement agencies are becoming increasingly aware of the need to minimize response time for calls for assistance. Several major departments are using a GBF system to tie the address given by a citizen requesting police service to a particular beat and additionally to find the patrol units on that beat available to answer the call.

Currently this function is performed in the following manner: 1) a telephone clerk receives a call for assistance, calls up a complaint form on a Cathode Ray Tube (CRT), and enters the address data first; 2) the address data is transmitted to the central computer for verification and geocodes while the telephone clerk obtains additional information; 3) before the telephone complaint-taker has finished with the call, the computer either indicates there has been a match and gives appropriate geocodes or indicates a match could not be found, in which case the clerk asks for more information concerning the address; 4) the call for service with the added geocodes is relayed to a dispatcher who checks the unit availability records for that area and

assigns a car to respond to the call. Dispatching is being handled in this or a very similar manner in Huntington Beach, California and Jacksonville, Florida.

In discussions with agencies which are currently using computer-assisted dispatch or are preparing to go operational within the next year, police departments usually indicate the need for intersection data, a commonplace name file, and/or an exceptional address file. The need for an intersection file is apparent in that numerous calls for service, such as traffic accidents, are related to intersections rather than addresses. A commonplace name file is desired since many citizens refer to shopping malls and public buildings by name rather than by address; an example of this could be the ability to refer to the White House rather than 1600 Pennsylvania Avenue. The exceptional address file is a file which contains information of interest to police officers assigned to a call at a particular address or in a given area. For example, the file could contain addresses where police officers have been attacked, the number and type of calls received in the last 24 hours concerning the address, and warrants issued for residents. This information, it is felt, would greatly reduce the likelihood of an officer walking unprepared into an unusual or highly dangerous situation.

The use of geocoded files by police to analyze crime distribution and fluctuations in those patterns is becoming widespread. An example of a police department using GBF's for this function is the Washington, D.C. Metropolitan Police Department. After geocoding data from the detail offense reports, the statistical processing branch can produce detailed data relating to numbers and types of crime by various geographic areas. In addition to this, there is a report developed called the Crime Trend Alert. Whenever the offenses in a specific reporting area reach a maximum level over a given time period, a special printout is automatically produced containing a list of all offenses of that type that have occurred. All available information that has been fed into the computer which would be helpful in establishing patterns or modus operandi are retrieved on the printout. Examples of offenses and the maximum number of occurrences per month before a printout is produced for a reporting area are: Burglary—5, Robbery—3, Larceny—12, and Stolen Autos—3.

Another example of the investigative use of geobase files would be a recent murder in Washington, D.C. The informant indicated that he did not know the murderer, but knew that he had been picked up for a robbery of a Safeway store with a silver-handled pistol. With this information the detectives went to the research and statistics branch. The research and statistics branch took the address of all the Safeway stores in the area and went back 90 days pulling all of the armed robberies and their suspects. From this list of suspects, which was produced 3 hours after the initial request, the detectives had information which finally led to an arrest.

All police agencies have the problem of allocating patrol strength. The demand for services seems to increase at a rate faster than either money or manpower is made available; therefore, more effective use must be made of available manpower resources. Resource allocation involves balancing a variety of factors such as time of day, day of week, locations of crime-prone areas with more calls for service, number of units available for assignment, and other factors peculiar to individual police departments.

Probably the best known resource allocation system is LEMRAS (Law Enforcement Manpower Resource Allocation System), an IBM proprietary package that provides information to law enforcement managers on probable frequencies of police calls by time, day, and geographic area. This data is presented in various tables which allow supervisors to have a basis for predicting manpower requirements.

The San Francisco Police Department is currently using LEMRAS in connection with the Location System which is a part of their overall CABLE program—their reference file. This is a corrected and extended version of the Census Bureau's GBF which provides the geographical information necessary for LEMRAS. The data input for LEMRAS normally originates from the dispatcher and consists of elements about calls for service. These are: date, time a unit begins service, time a unit completes service, type of service performed, and location of the incident. This information is entered into a computerized data base where is is used in mathematical programs according to preset forecast specifications. Tables are then produced in various configurations by hour of day, day of week, and week of year, and may be done on a citywide basis or on any geographical level requested. The tables forecast both the average number of calls for service that can be expected from a given area during a particular time and the average amount of time required for an officer to complete the call.

The necessity for statistical reports has greatly increased in recent years. Data in this form includes not only the needs of cities as a whole, but also reports by individual wards for councilmen or areas around specific locations. Police agencies which have effectively geocoded their incident reports can better serve the growing and wideranging requests for data.

For example, the Hartford, Conn. Police Department currently produces crime reports requested by city council members or representatives of local civic groups. In order to process the requests, the individual indicates the crimes to be included, ranging from auto theft to homicide, and the boundaries of the area to be studied. Currently the Hartford Police Department is handling an average of two special requests per week. There can also be reports generated on incident versus cleared to give a check on the clearance rates.

The Hollywood, Fla. Police Department uses their geocoded files to produce reports for resident groups. In addition to routine Total Activity Reports, the Planning and Research Division develops studies on specific areas of the city approximately twice a month. They have just completed a report for a Homeowners Association on burglaries in an area comprising six reporting districts. In addition to the raw figures, the results are compared with other reporting areas of the city.

From the research conducted thus far, it seems apparent that geographic base files do have significance for the field of law enforcement. Law enforcement, more than most fields, requires a file with accurate, reliable, and up-to-date information, especially for dispatch operations. One idea that is being explored in some cities is a central dispatch center for all emergency vehicles. This would divide the responsibility for file update and maintenance among police, fire, and the agency that handles ambulance operations.

As was mentioned earlier, geographic base files in every department are slightly different due to agency needs and requirements. Intersection data, exact addresses, commonplace names, and other local data are often added to the DIME format. In other cases, officers have devised their own file within the department using various rationales for file development.

Each department then uses the geographic base file for different applications and interfaces the GBF to other reference or data files for additional information generation. As was seen in the discussion of the different applications, the specifics of file development and use varies from department to department depending on the priorities determined by agency administrators; however, the fundamental concern to effectively and efficiently perform the police function and respond to community needs is universal.

APPENDIX

Geocoding Standard 4.8

Where practical, police should establish a geographical coding system that allows addresses to be located on a coordinate system as a basis for collecting crime incidence statistics by beat, district, census tract, and by other "zoning" systems such as schools, planning zones, and ZIP codes.

Commentary

This standard calls for the development in medium and large cities (e.g., more than 100,000 population) of a computerized geographical coding system. The best and most readily available system in many cities is the Geographic Base File (GBF) developed by the Bureau of the Census. In this system, each record in the file describes a straight line segment with the following information:

1. Coordinates at each end;
2. Tract and block number on each side;
3. Other geographical codes (such as precinct) on each side;
4. Name of the line (such as street, railroad, water feature, or political boundary); and
5. Address ranges on each side (when applicable).

A GBF can provide data in fine geographic detail for planning and evaluation. More important, it becomes a useful tool in determining day-to-day and hour-by-hour allocation of patrol manpower. It can be used in the dispatch process or to retrieve geographic information for investigative purposes. It can be used to match or compare data from police operations against data from other departments.

Geographic Base Files, with the appropriate computer and ancillary equipment, can draw maps with crime rate information or spot maps recording specific incidents. Through crime maps or tables, the location-based file can highlight areas with concentrations of various incidents by seasons, as required by police administrators.

Under the most advanced application of such maps, the computer can suggest changes in beat assignments on the basis of expected workloads together with an administrator having other specialized information. Geocoding forms the basis for computer-aided dispatching systems, and is useful as an investigative tool for various geographical matching problems.

This detailed level of geocoding is essential in scientific experimentation or program evaluation where definition and measurement in experimental versus control areas is anticipated.

Geographic Base Files are used by other local agencies as well. Some of these may not be compatible with the needs of the police agency. However, the costs to the police agency of having an existing or a forthcoming system upgraded to meet its needs would be substantially less than the costs of completely independent projects. If the police are using the same GBF as other city agencies, it would be possible to compare police data with data from school systems, health and welfare departments, and engineering and building code departments. Also, results from the census, such as demographic and housing characteristics, will be available for areas defined by police administrators.

The census approach is a byproduct of the technique of taking the 1970 census by mail. The DIME (Dual Independent Map Encoding) file is developed by entering each street or road segment, other line segments such as railroads, streams, and political boundaries, the "nodes" where these line segments intersect, and the block number for the areas bounded by the segments. Further, the highest and lowest numbered street addresses in each street segment are entered, as well as tract number and local codes as selected by the local agencies.

This file has a unique self-editing feature. The computer can "chain" each block by its bounding line segments and nodes. If one of the segments has been improperly coded, the block will not be enclosed and the errors can be identified for correction. A second edit consisting of chaining every block around a "node" further identifies errors for correction.

The street address, which is usually on any incident report, is the only geographic code required for the DIME system. A capability is being developed to handle events reported only by intersections. In most large cities, an agency is already working with the Bureau of the Census to update and maintain this system.

Related Standards

The following standards may be applicable in implementing Standard 4.8:

4.1 Police Information Systems.
4.5 UCR Participation.

Source: National Advisory Commission on Criminal Justice Standards and Goals, **Criminal Justice System** (Washington, D.C.: U.S. Government Printing Office, 1974—528-395), pp 66-67.

Question Period

Mr. James A. F. Kelly—I would like to mention that some of the Census data can be applied to help in manpower distribution and allocation since we already know that there are a number of offenses not reported to the police. This has a great deal to do with the social-economic makeup of a particular community. Therefore, if we can break this information down to that point where we can correlate it with social data, we can pick up those areas and better distribute manpower in accordance with social predicators as well as past experience when responding to calls for service. Dr. James Stevens, who is the Project Leader of the Census Bureau's phase of this program is also represented here and will be able to answer questions.

Mr. Harvey Albond—First of all, Ms. Brayford and gentlemen, I am impressed, and I would like to make a comment. The comment is that the police department does not exist in a vacuum. We have very real needs, and LEAA has funds, fortunately, to allow us to get into these specific applications. However, the LEAA application and the police department application is going to be a far better application if the entire city data system is available— correct and up-to-date—for the police department itself to use. Consequently I would hope that rather than getting into areas of specialization—we went through this in transportation studies, we went through this in land use studies— let us not go through it in police enforcement activities, but rather deal with a series of building blocks where the total resources of our government are avialable for each other.

Dr. James Stevens—I would point out something very briefly here. For those of you who have asked whether there has been anything relating DIME file development to law enforcement agencies or police applications, there is a very nice one-page statement available in a publication called **Criminal Justice System**, which was recently published by the National Advisory Commission on Criminal Justice Standards and Goals. The one-page statement indicates the following: "This standard calls for the development in medium and large cities, more than 100,000 population, of a computerized geographical coding system. The best and most readily available system in many cities is the Geographic Base File (GBF) developed by the Bureau of the Census." It further goes on to delineate many of the applications that we have discussed today, as well as some additional items that that may be possible in coming years. (A copy of "Geocoding Standard 4.8" appears in the appendix preceding this discussion).

Mr. James Thomas—Are you compiling an inventory of software programs as part of this survey, and if you are, is there any indication of the person to contact, or how to get access to any working programs that are already in existence?

Dr. James Stevens—We will be during the second phase of the project. We are not during the first phase, primarily because we are looking at eight exemplary cities as opposed to looking at everything that is available in the United States.

Mr. James Thomas—What is the time frame then for the different phases that you referred to?

Dr. James Stevens—The first phase will be completed probably between February and April. We will have our next advisory committee meeting sometime during the spring, at which time, we will provide the advisory committee with a descriptive report on the eight cities that we have examined. In addition, we will have several "issue papers" which will address the major issues concerned with the development of systems operations, funding, management, security and privacy, and so on. The project in total is a 2-year effort; the research portion will be completed, we expect, next summer.

Mr. Don Brown—I was wondering if possibly those of you from Washington might care to briefly comment on the impact of security legislation, and what impact that legislation might have on general access to the GBF files by law enforcement systems, and vice versa, general access to law enforcement files by general government.

Dr. James Stevens—I think proposed legislation, at least that which is currently pending and being discussed, will have significant impact on police systems in general. Many of the systems that we have seen in the past two months have, or often have, components that would probably be in violation of some of the Federal legislation that is currently being discussed. Some of the State legislation is already having a similar kind of impact to that which we expect by some of the Federal legislation.

I think it is our general conclusion that the GBF/DIME, as a system, is not really a threat in the sense of providing anything that the police departments cannot do or do not already do. Right now some police agencies can inquire with an individual address, and they can pull records by address. The GBF/DIME system does not give them any increased capability

over this. As a matter of fact, I believe Mr. Pisarski mentioned earlier today, and it has been discussed quite extensively, that with the GBF/DIME type of application, police departments do not need the individual addresses. They do not need to have that specific an item of data in their files, or let us say, they do not need to use it quite like they are using it now or quite like they intend to use it, given the presence of a GBF/DIME as a tool.

Mr. Don Brown—If I could comment following up on that. The Regional Crime Information Center that is used in Cincinnati and Hamilton County, does not rely on a GBF file. It does, however, rely on the auditor's plat and parcel system for address location for the purpose of gleaning relevant information—as one of the speakers said—about known police assaults and warrants and all of that. However, the people in the Cincinnati-Hamilton County Center have stated that their reliance on that general-purpose file is one reason which prevents them from participating in national NCIC information with direct connect. If that is true, there is a definite implication, I think, about participation in general-purpose geo-files.

Dr. James Stevens—I am not sure I understand your full question. Now it is true that general-purpose governments, or computer systems that are not under control of law enforcement agencies cannot have direct access to NCIC files. That is a fact, I do not see how that ties in with a GBF/DIME system as a tool for analytical purposes. Were you saying it could not be used as a tool for analytical purposes by a system that is not under control of law enforcement agencies?

Mr. Don Brown—I guess I am trying to determine what is or might be national policy in terms of a law enforcement system's reliance on a general file that other units of government are also utilizing. In this case, Cincinnati and Hamilton County law enforcement are relying on the county auditor's plat and parcel system to obtain information. I think, likewise, the reliance of San Francisco's Police Department on a general-purpose geo-file system might have some bearing on whether they can participate directly in NCIC.

Dr. James Stevens—In the type of system that you are talking about, are they extracting individual items of data from the system to be used for some purpose or are they using it as a tool, say to locate where an address is in a general geographic area?

Mr. Don Brown—They are using if for computer-assisted dispatching, as in the example that the young lady just described.

Dr. James Stevens--And do they have every address in the city on it?

Mr. Don Brown—Yes, every address.

Dr. James Stevens—There may be some confusion in terms of the NCIC type of system as opposed to geographic base files. Is everyone knowledgeable about NCIC?

NCIC is strictly a set of data files on wanted persons, missing persons, stolen guns, stolen property of a certain value, and so on. It is specific items of data that have unique numbers that are wanted for some reason or need to be located by police agencies. Any terminal that has access to, or computer system that has a direct link with NCIC has to be under control of law enforcement agencies. They cannot be under control of the auditors office, or the city managers office, or any other type of department. As a result of this, most cities simply have terminals into State systems and the State systems, under the control of the Department of Public Safety, normally will have direct links with the NCIC.

Now that is for the retrieval of specific items of data on individuals or on items of property. Trying to tie back to this, there might well be some problem should national legislation prevent general-purpose government from analyzing or having access to the kind of data files you are talking about. For instance, if the national legislation allows arrest records to be maintained by police departments without disposition, it might also specify that those arrest records cannot be processed through systems other than those under the control of police agencies. If that is the case, what you are talking about probably would become a problem. If national legislation, however, does not specify that local data files cannot be processed by local systems, irrespective of who controls the computer system, then it would not become a problem.

Mr. Alan Pisarski—As a comment, this is very sophisticated area of concern which was alluded to in several of the other comments before; the gentleman from Fort Wayne, I think, brought it up, and Mr. Albond did. I think that the law enforcement work, in the future, is going to be really way ahead of some of the other applications. In many respects it makes much more sophisticated demands on the technology than do some of the other applications.

I think the users at the moment kind of break down into two categories, and this was what was referenced in the Fort Wayne presentation. The people who never heard about the Geographic Base (DIME) System, have not thought about spatial analysis ever before, typically health agencies maybe, educational agencies, and welfare agencies. To present them with this kind of technology is a whole new wonderful world.

There is another set of agencies, typically the best examples are the transportation agencies or law enforcement agencies, that have very long histories of experience and requirements in this area. They both deal continuously with the spatial dimension in statistics. As a result, they are frequently the least coordinated of the agencies, in the sense that when a new technology comes along, such as GBF/DIME, they have already got a heavy investment in some internal system. It is a little harder for them to join, and so, some of the most sophisticated transportation planning agencies have been the most reluctant, if you will, to participate in the GBF/DIME system.

It raises two questions. Number one, it is my feeling, and I am speaking now from the point of view of the DOT, that in the future LEAA will lead in the funding of the balance of technology here; and if LEAA's kinds of needs are met, the needs of DOT will also be met. I would draw attention to the fact that in addition to some of the things that were mentioned here, particularly with respect to Washington, D.C., like intersection files, major generator files, and special address files, there are classical transportation problems too.

In Washington, D.C., I guess in 1967, we built these kinds of geocoding files for our home-interview survey and had exactly the same kinds of problems. But it raises two areas of concern. One is the question of conversion costs, for instance, where an agency has a very good system, but it is not a generalized system available to other people, and they want to join. What are the kinds of conversion costs that are built into a system like that?

Finally, the last point, which I think concerns Mr. Albond, is how do we assure the interrelationship of these technologies. How do we assure that the public safety people who are going to be leading in this thing, not only lead, but also disseminate this information and these technologies to the other agencies? I think that is a real question to be considered tomorrow in the discussion, Agencies that receive funds, Federal, local or whatever, have a real responsibility in just letting the rest of the world know that they have done the work, that they created this kind of capability. Very frequently it just does not get around in the rest of the city that in fact it does exist. I think there is a whole area of public responsiblity here in the use of public funds. It is a very important question that we should address tomorrow.

Mr. John Evans—Our GBF/DIME file covers two counties. Around the Akron area there are quite a few suburbs where the police departments number from 1 to 10 men. In the district in which I live, there was a robbery in process to which the State Police, County Sheriff, Copley, Fairlawn and Akron Police Departments all responded at the same time. The problem arose as to who had the authority at the scene of the crime. Many of the suburban police departments make use of Akron's jail. Akron has no computer with GBF/DIME information and since the suburban police departments cannot afford such a computer, what is the feasibility of tying both into one unit—suburban into big city?

Mr. Alan Pisarski—Yes, it is an excellent question. There has been a significant effort in Washington to merge municipalities, but I certainly would defer to any of the gentlemen here to comment or respond to that question.

Mr. James A. F. Kelly—We do have a lot of problems in that area. Approximately 16,000 law enforcement agencies exist in this country, each having pride and independence. Consolidation has been suggested in many reports, but there have only been a few major efforts along these lines—Virginia Beach and Newport News in Virginia and Dade County and Jacksonville in Florida.

The sharing of resources offers the greatest potential for the future. Why should everybody in one county have a jail when the sheriff already has one? Problems of this nature have been overcome because nobody wants to run a jail. But in terms of computers and dispatching, this is a bit more difficult since it concerns operations directly under control of individual agencies. Despite the fact that consolidating operations can be cost-effective, progress along these lines has been slow.

Dr. James Stevens--Simply in terms of some of the systems we have looked at, I might refer you to a couple that have done exactly that, or are doing that. Santa Clara County right now is putting together a centralized county level dispatching system which will handle dispatching for the county, for the city of San Jose which is over 500,000, as well as for numerous small municipalities within the county. It will also handle fire dispatching and emergency services dispatching from the same system using mini-computers, but using the same Geographic Base (DIME) File. In St. Louis County, they are currently dispatching from the county level for the county law enforcement agency as well as something like 30 small municipalities of the size that you are talking about. They do not have a computerized system but are talking about it.

Another similar kind of example is the Huntington Beach type of experience where the police are using the Geographic Base File for dispatching for the city of Huntington Beach, and the fire department is tied in for using the Huntington Beach system as well as the Geographic Base (DIME) File for three other cities. They have one centralized dispatching system for fire dispatching for four cities under a joint-powers agreement. So these are three current examples of where actual systems are in operation or where the type of organizational configuration you are talking about is used.

Mr. Alan Pisarski—I would like to make a final comment on your question. The problem clearly goes well beyond geographic base files. In the Washington, D.C. area, I remember a very severe problem of this order that actually took legal agreements between each of the counties and between the States, because of the Washington metropolitan structure, to permit equipment to move; so that for instance, the fire department from one area could go into another area in the event of major fires. In fact, those agreements were signed about a month before the major riots in Washington in 1968, otherwise such movements would have been legally prohibited.

Mr. Morton Meyer—I would like to further reinforce one of the points touched upon by Mr. Kelly. The cost of modern technology is such that you now have to be very wealthy, indeed, to go it alone. The economics of cooperation, even if not the logic of it, comes to have a very powerful persuasive effect—particularly during inflationary times.

Real World Applications of the GBF/DIME System

JOHN E. ROWE

INTRODUCTION

The Unified Statistical Evaluation Study (USES) was established in late 1971 by agreement of the Director of the U.S. Census Bureau and the Mayor of the city of Indianapolis, Ind. It began functioning in June of 1972. Three professional employees are furnished by the Census Use Study and a clerical staff is furnished by the city of Indianapolis. Its office is located in the City-County Building, which is the seat of both city and county government in Indianapolis and Marion County, Ind.

The Project Manager interfaces with the Mayor, the Deputy Mayors, the members of the Mayor's cabinet, the Board of County Commissioners, the City-County Council, and other elected officials. The other professionals on the USES staff interface with division directors and middle-management personnel. An effort is made to respond to data requests regardless of whether the requests are received from units of government or from the private sector.

The purpose of the study is to become exposed to the data needs of the broadest possible cross-section of the community in order to document "real world" applications of census data and Census Bureau programs. To this end, a log is compiled detailing all of the data requests received. This log indicates the type of organization making the request, the intended purpose of the data, whether a satisfactory answer was supplied by the USES staff, and if so, from what source and by what methodology. If unable to handle a request, the reason for the failure is also stated.

Viewed with the knowledge that only 10 percent of the data collected during the 1970 Census of Population and Housing was published in eyeball readable form and assuming that a similar proportion probably will hold true for the 1980 census, this log of requests may prove to be a valuable input in deciding which data items to publish and which must remain on the summary tapes following completion of the 1980 census.

The USES operation consists of a great deal more than just supplying columns of figures. It frequently is involved in coordinating projects from their conception to their completion. Two criteria which must be met before a project is accepted is that the task involve small-area data anaylsis and that the project have transferable features. By "small area" we mean areas that are smaller than the city itself and are typically such units as wards, school districts, police beats, fire districts, traffic zones, or neighborhoods. The transferability feature is a must, as the Federal portion of the USES funding is earmarked for research and development that will lead to insight into solving problems that are typical of urban America. All successful programs are documented and made available to others who wish to take on a similar task in their own community.

APPLICATIONS

Let us take a look at a few specific examples in which data furnished by the USES office had an impact upon important decisions affecting local government in Indianapolis.

1. **Department of Labor Geocoding Project**—The Manpower Administration of the United States Department of Labor asked the Census Use Study to determine whether it is technically feasible to generate neighborhood level profiles of job seekers. They were hoping for development of an information system to help local manpower planners determine priorities should manpower revenue sharing become a reality.

   The best source of data on job seekers seemed to be the Employment Security Automated Reporting System (ESARS File). Consequently, it was used as a basic source. This file contains not only the names of persons who have asked the Employment Security Division to help them find work, but contains a great deal of information about the characteristics

of those job seekers. Consequently, by summarizing the data in the ESARS File, we were able to determine not only how many persons were seeking work in Marion County, Ind., but how many persons were within each age group, ethnic group, sex, education level, and within many other categories. In addition, we were able to determine how many were on welfare, how many were receiving food stamps, how many were eligible for the Work Incentive Program, and other information of this nature.

The Marion County portion of the ESARS File was extracted from the State file. The GBF/DIME System and a Census Use Study program called UNIMATCH were used to append census tract numbers to the records. An 86 percent match rate was achieved by the UNIMATCH program. The 14 percent that did not match were automatically punched on to interpreted punchcards. Clerks using current information obtained from the fire department and the local planning agency then appended the appropriate census tract numbers. Keypunchers applied the four necessary keystrokes to punch the tract numbers into the cards. The data was then sorted back into the master file. The combination of man-machine geocoding resulted in all but a handful of the 13,800 records being properly geocoded.

Output programs were written which displayed the resulting data for each of the census tracts of Marion County, as well as each of the 9 political townships, the 25 City-County Council districts, and 9 selected poverty neighborhoods.

Many other combinations would have been possible since using census tracts as the basic unit permits using those tracts as building blocks which, when grouped together, can encompass many different geographical units.

A similar process was followed to produce printouts detailing similar information on persons who were actually drawing unemployment compensation checks from the Employment Security Division. The levels of geography for which claimants' data were shown were identical to the levels employed in connection with the ESARS File, and many of the data items were similar.

The next step in the project was to learn whether there was any particular value to producing those profiles. So sample reports were distributed to many local groups, both public and private. Thus far, we have been informed that the Secretariat of the Manpower Area Planning Council (MAPC) used this data extensively in preparing its comprehensive plan for fiscal year 1975. The data helped them determine the funding levels needed for both placement services and training programs for various categories of persons such as youths, women, senior citizens, and minorities. It also helped them determine which neighborhoods of the city should receive special attention.

The Community Services Program (Model Cities) used the data to determine which Multiservice Centers should be staffed with employment counselors. We understand that the State Employment Security Division is cooperating by providing the necessary counselors. The Employment Security Division itself indicated that they might use the data to decide where to locate several outreach offices.

In addition, the planning staff of the Indianapolis Public Schools System used the data as an input to determine which inner-city schools should be open during the evening hours for the purpose of presenting adult education courses.

The Mayor of Indianapolis indicated that he appreciated having comparative data aggregated at the City-County Council district level so that he could impartially compare requests for facilities or services from various competing councilmen. Members of the Chamber of Commerce staff and Manpower planners have also offered various expressions of enthusiasm for the data produced by the project.

Consequently, a second round of printouts is being produced using 1974 data. Up to this point printouts of each of the geographical levels previously described have been produced using 1974 ESARS data and 1974 claimants data. We are presently in the process of working with the files of the Indianapolis Metropolitan Manpower Commission and plans include printouts based on welfare data. So in connection with 1974 data, four distinct types of printouts will be produced in an effort to determine whether this additional data adds anything of value to the manpower planning process.

2. **Carpooling**—During the height of the energy crisis, the Mayor asked the USES office to coordinate a local government carpooling effort. As luck would have it, the Los Angeles based staff of the Census Use Study had already developed a carpooling program in connection with county government in Los Angeles.

Indianapolis city employees who were willing to consider carpooling were asked to complete a questionnaire giving some pertinent information such as their name, home address, hours of work, department, and their work telephone. The data from the questionnaires were keypunched and the files were geocoded to the tract level using address matching program called ADMATCH in conjunction with the GBF/DIME System. The CARPOL program was then used to produce printouts.

Each individual who asked to participate received a list with his or her name at the top. The list showed the names, addresses, departments, hours, and work telephone numbers of persons who lived within the same tract and worked a similar shift in the same building. It was up to the employees themselves to make the final contact to arrange the carpools.

We also included a parameter which indicated that no person would be on any individual's candidate list if the distance between their respective houses exceeded 20 percent of the distance between home and work. That distance parameter could have been adjusted in a variety of ways to make it suitable for the circumstances involved.

No notable problems were encountered and the CARPOL program seemed to function well. The identifiable costs for keypunching, printing, computer time, and similar items were approximately 41 cents per participant.

Shortly after execution of the City-County Building pilot project, the Arab oil embargo was lifted and the crisis atmosphere seemed to pass. While we had made tentative plans for a community-wide carpooling effort and many had business firms had expressed interest, it was our feeling that a community-wide effort would not succeed at that particular time. But should a crisis recur, there are key people in the Chamber of Commerce, the Indianapolis Jaycees, numerous major employers, and members of the news media who are well aware of the successful pilot project and could be quickly marshalled into an all-out effort. So effective counter measures to cope with another shortage can probably be implemented in rather short order.

3. **Police Interactive GRIDS System**—The Indianapolis Police Department captures data pertaining to all runs that are initiated by the dispatching desk. They asked the USES office to help them develop a system for more efficient utilization of this data.

The GBF/DIME System was used to produce a listing of streets. An intersection code indicating the number of blocks in each direction from the central point of the city was added to each record. The GRIDS program of the Census Use Study was then modified slightly to fit the specific need. GRIDS is a computer mapping program.

The result was a system in which incidence of crime, either in total or by individual category, can be displayed on overlays. Rather than using a black/white shading system, a numeral appears for each five block square showing the number of runs into that square for the specified time period.

Several applications are contemplated. A decision has been made to display the data for an extended period of time, such as a full year, for the purpose of trying to equalize the workload between the various beats. Boundaries will be adjusted so if the pattern for the coming year is similar to that of the current year, approximately the same number of runs will be made by each police car.

In addition, plans call for plotting several types of felonies at hourly intervals. Each plot will be presented to the shift commander for his review. He can then determine whether any significant or unusual patterns are developing and thereby make judgements as to whether any of the policemen in the field or any of the mobile teams at his disposal should be shifted to a particular locality.

It is also contemplated that after functioning in this fashion for about a year, and eliminating any apparent flaws in the system, it can become an integral part of an automated dispatching system.

So, in this environment, the GRIDS system has evolved into the Police Interactive GRIDS System and shall be known by the acronym "PIGS."

Up to this point we have been discussing actual applications. However, I think many more governmental applications are feasible and I believe some of them will take place in Indianapolis during the next year or two. For instance, at the present time there are no precinct police stations in Indianapolis. All operations generate from headquarters located in the City-County Building in the heart of downtown Indianapolis. There seems to be a growing sentiment among some ranking police officials and council members for a series of precinct stations. If such is to take place, the PIGS system could be a very valuable tool in pinpointing the neighborhoods of highest activity, which I feel should be the prime sites for precinct stations.

Routing problems can be tackled with a well constructed GBF/DIME System. A systems analyst on the USES staff and a professor of Civil Engineering at the University of Illinois, Urbana Campus, worked out an algorithm for optimizing the routing of trash collection vehicles. Such algorithms are usually referred to as "least path programs." We tested the algorithm on two actual collection districts. If the routes our program suggested were followed, instead of the present routes, the saving in mileage would be 20 percent a day. When you consider that the average mileage for the city fleet of trash collection vehicles is 1.6 miles per gallon and that the cost per gallon of gasoline to the city has doubled in the last year, a 20 percent savings could be very important.

Snow removal is a very aggravating problem in Indianapolis. We are at the southern edge of the snow belt and only subject to three or four substantial snowfalls each winter. Consequently, the number of available snowplows is very limited. Which means that when we do have a major snow, the traffic tie-ups are considerable, and the economic loss caused by a majority of the city's work force being late for work is enormous. It seems that a similar algorithm could be applied to the task of routing snow removal vehicles in the most efficient possible manner, minimizing the dead segments they travel while not actually removing snow.

Plotting the incidence of fires on density maps or overlays so as to visually see their concentrations might be an aid in locating additional fire stations.

Plotting the location of children from current school records, from the annual school census, and/or from birth certificates might be a useful input in determining the location of future grade schools and high schools.

Plotting the location of all students within an existing school district might suggest how many children per district are really in need of school bus services. When coupled with a "least path program," perhaps the most efficient routes for picking up those children could be devised.

The Defense Civil Preparedness Agency has devised a scheme for allocating the working population to available fallout shelters. Doing this well would require a census of the daytime working population, a rather considerable task that has not been preformed as yet. However, two of the necessary ingredients for a successful shelter allocation program, a geographic base file and "least path programs" can provided to the agency through DIME technology.

The GBF/DIME System can be used to generate street directories. These are directories that print out the address range for each city block in eyeball-readable form. Usually the odd numbered and even numbered sides of the street are shown separately. By looking up an address in this directory, one can readily determine where that address is located with respect to such units as school districts, police beats, fire service districts, townships, city-council districts, wards, precincts, and whatever other important geographical or political subdivisions are coded into the GBF/DIME System by the agency that constructed it.

The ADMATCH program can be used to append traffic zone numbers to vehicle registrations. The number of vehicles is then totalled for each traffic zone. This can be a very valuable input for transportation planners in determining future thoroughfare routes.

Perhaps one of the more important situations in which a GBF/DIME System can be of value is in coping with the considerable problems caused by the fact that no two agencies seem to maintain their files in the same format. Some planners and statisticians have been screaming for some kind of law or regulation that will require all agencies to keep compatible files. This, of course, will never happen. Because of the very nature of the functions performed by different agencies, school district boundaries will never precisely correspond with police beats, which will continue to be different from areas serviced by welfare department case workers, and so on down the line. Standardization of statistics is just not likely to take place in the foreseeable future. But if all of the files you wish to review contain an address, then the problem becomes manageable. By using the Census Bureau's ZIPSTAN program, the spelling and manner in which street names are handled in various data files become standarized. This, in turn, enables either of the available address matching programs, ADMATCH or UNIMATCH, to operate at a relatively high degree of accuracy. In addition, if complete accuracy is necessary, it is possible to append a program modification of the type I described in connection with the Department of Labor Geocoding Project, which enables final corrections to be made by hand. So in spite of the level of geography for which the various agencies collect their data, by using the GBF/DIME System and an address matching program, we can tabulate or display data from many diverse agencies at exactly the same level of geography. Therefore, if your particular planning function or market research requires you to study information from half a dozen different agencies, but you want to be sure there is no overlap and no gaps in the geography represented by that data, you can do so.

This address matching capability also can be important to those of you who are becoming troubled by the obsolescence factor of census data. If, as the decade wears on, you no longer feel that census data is suitable for your function because of its obsolescence, you can use address matching techniques and the GBF/DIME System to enable you to reformat current data generated by State or local agencies to fit your need.

Up to this point, all of the examples I have discussed apply to the public sector. I realize, however, that the private sector played a major role in the development of the GBF/DIME System for Mid-Ohio. So perhaps it would now be appropriate to suggest some applications that are directly pertinent to the private sector.

Let's discuss carpooling again. I feel that computerized carpooling can be important to business firms that have a large number of employees. If the number of automobiles coming to and from a particular work location is reduced through carpooling there can be some improvement in traffic flow on the surrounding streets. There is a tendency for employees to be more prompt when carpooling than when driving individually. It is not nearly as easy to have one more cup of coffee if one knows that Herkimer will be sitting in his driveway honking at 7:30 sharp. So punctuality is sometimes improved through carpooling. Also, when large numbers of employees leave at the same time it is sometimes possible to reduce utility usage, such as lights and heat, in large sections of offices or plants earlier in the evening than would otherwise be the case. Sometimes, when a business is engaged in negotiations with EPA, promotion of carpooling can be pointed to as an example of the company's concern for air quality improvement.

Companies that provide parking for their employees may have the most important reason of all for promoting carpooling. While these costs fluctuate widely from city to city and area to area, it is not unusual for the acquisition and construction of each parking space to run as high as $3,000 to $6,000. Prices, of course, are especially high in congested urban areas. So if widespread carpooling makes it unnecessary to acquire the building next door so it can be torn down and converted into parking spaces, a rather dramatic savings can be effected.

It is my feeling that carpool matching schemes should be handled manually if the number of employees at a single location is fewer than 500. Some thought should be given to computerized carpooling in situations of 500 to 1,000 employees at a single location. In most instances involving more than 1,000, I do recommend a computerized approach.

While all insurance policies contain an address as well as a name, it is rare for an insurance company to try to get a graphic picture of the concentration of their policyholders. It seems that it might be interesting to ADMATCH or UNIMATCH the complete file of policyholders in a large metropolitan area and run a plot using GRIDS or some other mapping system. It then might be desirable to locate claims offices in areas of high concentrations of policyholders and establish new sales agencies in areas that currently have a low concentration.

With respect to the various insurance policies that provide fire and wind storm coverage, it might be desirable to plot the location of insured units to the individual city block level. This might suggest whether there is too large a concentration in any tightly compacted area; a concentration that would mean financial disaster in the event of a tornado or general conflagration.

Contractors doing business with various branches of the Federal government are now faced with affirmative action plans or contract compliance regulations which require that the racial balance of their employees correspond to the racial balance of the total population in the area from which they draw their work force. Large employers can have difficulty precisely pinpointing their work-force area. Consequently, plotting employees' home addresses might be desirable so employers will know which tracts, townships, counties, etc. comprise their work-force area. Once that has been determined, publications and summary tapes of the Bureau can be consulted to get a clear demographic picture of the work-force area, and the employer can determine whether the ethnic balance of his staff is in proper balance for the area.

Address matching functions can also serve as a valuable input in determining which portions of a community should be subjected to a direct mail advertising campaign, or deciding which weekly newspaper would serve a desired need, or even figuring out where a branch location should be established.

In connection with any of these three functions, I feel it is desirable to perform an address match on all known customers such as checking and savings account customers, charge account customers, etc. Once it is determined which census tracts have the largest concentration of customers, some demographic highlights of the tracts should be studied. I am thinking about items such as median home values, average monthly rent, the racial mix, a breakdown by age groups, the average number of children in each family, and median family income. Data items such as these give you a

fairly good composite picture of the types of people that are currently customers. The trade area should then be searched for other locations that have similar demographic characteristics but far fewer current customers. These locations then become the target neighborhoods for the next promotional effort.

Many mail order houses that handle direct mail on a commercial basis are beginning to adopt the census tract as an area for which their addresses are aggregated. Census tracts are nearly always much smaller than ZIP code areas. But because of zoning laws and tendencies toward natural selection, the enclaves of potential customers might all be nestled in one corner of a ZIP code area, in just one or two census tracts. Mailing to those tracts is invariably cheaper than mailing to an entire ZIP code area but could very well produce almost as many responses.

CONCLUSION

I know that these suggestions just scratch the surface, but perhaps one of them may have stimulated an idea that affects your particular operation. If so, I will be here during the balance of this function and will be happy to respond to requests for additional information.

I recently shared a platform with Gerald Fox who is the City Manager of Wichita Falls, Texas. After describing some rather sophisticated computer functions in which his city now engages, he suggested to the audience that they not go away with an exaggerated idea of the importance of what he had been explaining. But he offered the opinion that the line between running a *good* city administration and a *great* city administration could be crossed by city officials doing five percent of their jobs five percent better. I concur with his remarks and suggest that those of you who are engaged in government think about them from time to time.

I also suggest that in today's economic situation, those of you in the private sector who do five percent of your tasks five percent better may be the ones who are able to continue using black ink rather than red.

APPENDIX

A Dual Independent Map Encoding (DIME)

The DIME file is a system for representing map features numerically for processing by computer. It is essentially a file of segment records where a typical segment is a portion of a street defined by intersecting streets or civil boundaries. Other segments may be defined for natural features, railroad tracks, jurisdiction boundaries, and the like. The basic feature of a DIME file is that each node (intersection) and block (area bounded by segments) is uniquely identified. "Dual Independent" refers to the fact that each boundary segment in the network is described by specifying the nodes at the ends and the blocks to the right and left. With each node and block uniquely numbered, the computer can then construct two independent networks and match them to insure the existing network is completely represented and all land accounted for. Spatial information is added to the DIME file by assigning geographic coordinates to each intersection node.

Census Tracts

These Census-designed areal units are statistical subdivisions of SMSA's. Tracts are designed to be relatively homogeneous in population characteristics, economic status, and living conditions, though these conditions may change over time. The average tract has about 4,000 inhabitants. Census tract boundaries and boundary changes are recommended by local census tract committees and approved by the Census Bureau. As tracts increase in population, they may be split.

UNIMATCH (Universal Matcher)

An improved matching system that has many capabilities not available in ADMATCH such as the ability to handle building names, street intersections, and non-address matching. It is a generalized record linkage system which will compile, assemble, and execute a file matching system tailored to the specific needs of the user.

ADMATCH (Address Matcher)

A package of computer programs developed under contract to the New Haven Census Use Study that provides the capability of geographically coding computerized data records containing street addresses. The system compares the addresses on input data records (after standardization with a preprocessor) with the address ranges in a reference file. A "match" occurs when the street names are judged identical or equivalent and when the address falls within the defined range. Geographic codes from the reference file may be attached to the matched data records.

Grid-Related Information Display System (GRIDS)

A computer mapping system developed by CUS for producing character-printed maps from detailed data. GRIDS is written in ASA Basic FORTRAN IV and has several mapping options available.

Statistical Data—The Bureau of the Census as a Numerator and Denominator

VINCENT P. BARABBA

As an overview to this entire conference I would like to present a brief theoretical discussion of the current and future costs and benefits relating to certain types of information. I hope by this to provide you with some criteria for judging the data systems, techniques, and usages that will be presented to you during this course.

To begin with, it is self-evident that our country is entering a "margin of error" squeeze as to how both the governmental and private sectors conduct their affairs. The energy shortage indicates how narrow and ever decreasing this error margin really is. We know that viewed over time, business, manufacturing, and governmental decisions can have an irreversible ecological impact—information that is sufficiently accurate, timely, and easy to use must be available so that the best decisions can be made to maximize the efficient use of society's limited resources.

There are real costs in both time and money associated with the collection of reliable and useful data—but from a societal perspective the cost of statistical information is closely related to the use. The relative cost to society of any data varies, and is dependent on the number of times and the way that it is used. This assumes, of course, that we agree that society benefits when decisions relative to the utilization of its limited resources are based on accurate information.

The simple collection and publication of statistical data does not by itself fully meet the needs of today's decisionmaker, nor does the presentation of statistical data necessarily maximize its potential uses. Even though society's decisionmakers are furnished with accurate, timely, complete, and usable data, it is a disservice if they and society are surrounded by too much data. Too much data places the user in the same plight as Coleridge's Ancient Mariner, who cried, "Water, water, everywhere, nor any drop to drink." It is necessary to find ways of reducing the current oceans of data into meaningful information for decisionmaking.

To be truly useful, the data must be presented in a way which will achieve meaningful measures. Meaningful measures do not have to be complex. They can be as simple as basic mathematics. Some of the most meaningful statistical measures merely require the use of a numerator and denominator.

The numerator shows the actual number of events that have occurred—for instance the number of births, deaths, or divorces; the number of whites, blacks, or persons of Spanish origin; the number of families with incomes under $5,000 or over $15,000; or other similar measures.

The denominator, on the other hand, will show the population within which the events can possibly occur. It shows the population at risk to the event. The denominator may be the total population, the number of married women of childbearing age, the total number of families, or some similar population group among which the events in the numerator occur. A ratio can show actual to potential and serve as a utilization rate.

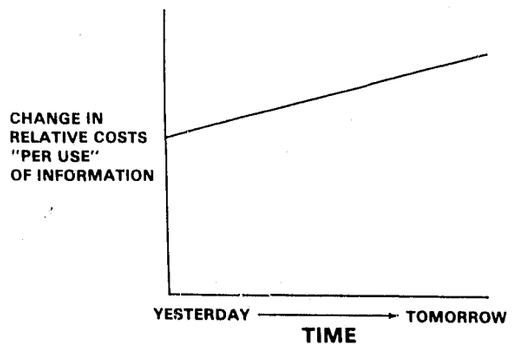
For example, in health statistics it is important to know the number of people who die from a specified cause, such as cancer. The total number alone, the numerator, can provide useful information on the magnitude of occurrence of this disease. Additional understanding can be obtained by looking at the denominator. The denominator could be the total population, the total *dying* from *all* causes, or the total in one of the age, race, or sex groups. Knowing the denominator permits the development of analytical data showing the percent of the total population dying from this cause, the percent *these* deaths are of all deaths, the rates of occurrence for various population components, and the changes in these rates.

In order to understand the potential contribution of statistical information and to compare the changing relative costs of collecting differing types of data, I'd like to examine our current information systems by

looking at three different ways we now generate statistical information. The first is through specific-purpose surveys that are designed especially for the problem to be explored. The second is through the use of a more comprehensive general-purpose survey or census; which, rather than having a design intended to produce information on one particular problem, is constructed in such a way as to provide insight into several areas of concern. The third method is through the synthesis and analysis of data already collected. The costs of these information system components vary greatly. The costs will change over time, and again, the cost *per use* will vary.

Figure 1 illustrates the change in relative costs per each use of information acquired through the collection of statistical data in a specific-purpose survey. The specific-purpose survey may include in its universe both the numerator and the denominator that are of primary interest to the decisionmaker. The costs of collecting this type of data have steadily increased over the past years and are likely to continue to rise in the foreseeable future. The measurement of specific actions, attitudes, or individual characteristics, in relation to a particular problem or objective, requires sample survey design questionnaire development, the development and implementation of a collection methodology, and the utilization of analytical techniques.

FIGURE 1. RELATIVE COST OF SPECIFIC PURPOSE DATA

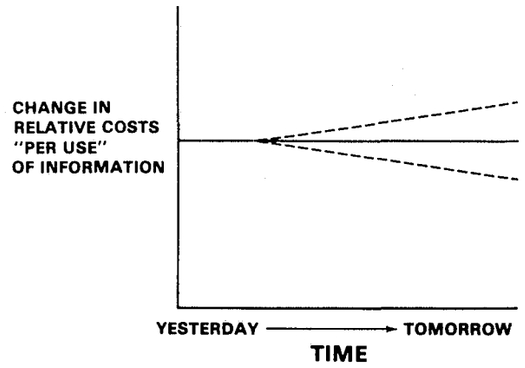


Specific-purpose surveys may also be directed at an already identified population such as patients in hospitals, persons who have previously been identified as recipients of pensions, or similar unique universes. In some cases the population data collected in a general-purpose census can then be used as the denominator.

Although extremely useful information can be obtained through specific-purpose surveys, they are often expensive undertakings, (even though the cost is held down by sampling). Survey costs will continue to increase due to the rising cost of labor, transportation, the use of more sophisticated and expensive methodologies, and the additional constraints that have been—and will be—placed on data collection activities by the public's concern for individual privacy. One of the major reasons for the relatively high cost of specially collected data is that the information obtained cannot be used for other analysis and decisionmaking needs. A specific-purpose survey has, by definition, a narrow range of uses and the cost of generating the data is borne by a limited set of users. Unless a data set can be used for a variety of purposes, the cost to society remains the same as the cost and value to the original user.

Figure 2 provides an illustration of the relative "per use" cost of information collected in a general-purpose type statistical system. Census data are a prime example of general-purpose statistics collected for a multitude of uses. General-purpose data can provide insight into a variety of specific problems through the use of proper comparison and extrapolation. General-purpose data provide benchmarks describing the general social or economic situation, they establish and define parameters, and provide a vital time reference.

FIGURE 2. RELATIVE COST OF GENERAL PURPOSE DATA



Costs related to the collection and analysis of general-purpose data face the same pressures to increase that we find with specific-type data. The collection of information in a major decennial census, which requires contact with every household in the nation, is, of course, an expensive undertaking. However, the greater the number of users of general-purpose data, the less the cost per use. Census information is widely used by Congress, Federal, State, and local governments, businessmen, educators, and researchers. Many others are unaware that they use it, since census statistics are presented in a pervasive abundance in textbooks, almanacs, magazines, and newspapers.

These myriad uses make census statistics comparatively less expensive on a "per use" basis. In addition, the use of sampling within the census framework further reduces census costs, since every household does not have to be contacted for the same information. The use of self-enumeration, whereby respondents receive and return the census questionnaire by mail, provides for lower costs and more accurate information.

The Census Bureau's Current Population Survey is sometimes perceived as a specific-purpose survey. Each month some 47,000 households are interviewed for the very specific purpose of obtaining information on the magnitude of unemployment in the Nation. If this was the only information collected and its use was limited to the statistics released by the Bureau of Labor Statistics, it would obviously be a limited-purpose survey. However, the CPS is used as a general-purpose survey to update, on a national basis, information collected in the decennial census on such matters as family income, educational attainment, school enrollment, marital status, households and families, migration, and similar subjects. In addition, this ongoing survey permits the addition of unique questions on such subjects as birth expectations and voter participation. Further analysis from annual average data as well as special tabulations from CPS data can be made.

The use of administrative records, available from government and business, in the production of general-purpose, anonymous statistics is one of the factors which helps reduce the cost of data collection. The value of information from administrative records is considerable because the expense of collecting the information has already been "costed" to other functions. These records—whether used by public officials in their planning efforts or by the business community—can be very valuable, particularly on the local level, through the use of the Census Bureau's geocoding system—the GBF/DIME System. It is also the case that the hardware and software expenses of the GBF/DIME System have already been costed out to research and development. Again, the greater the usage the lower the cost of each use.

One of the major assets of the GBF/DIME System is the fact that it permits locally generated data to be related to census data, which opens unlimited opportunities for the user. The numerator, in this case, is the administrative records and the denominator the census data.

I have to point out, however, that making use of the abundant information which is available creates a problem. It is the concern over confidentiality, which has been the subject of considerable recent discussion in the media and Congress. The Census Bureau, as you undoubtedly know, has an obligation to protect the privacy of individuals—both under Federal law and as a long-standing matter of principle. There also is a built-in protection at the local level. This protection is the GBF/DIME System itself, because it allows the assignment of characteristics to specified geographic areas without releasing information which could be used to identify a particular person. Yet the system has the merit of permitting any number of applications; from Federal to local and private areas there are uses which can be of immeasurable benefit to society.

An example of this system's application is that it allows the establishment of a health information system that will enable communities to locate and target segments of their population who require health services.

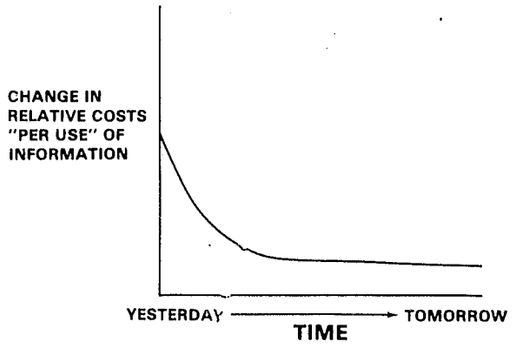
The GBF/DIME System permits a community to take its local files such as those relating to births, deaths, morbidity, and health services—the numerators—and geocode them to census tracts or smaller areas if necessary—where there is a population-at-risk. This can produce a wealth of information about every small area in the community, since the file contains the number of deaths and their causes, births, the number of health services, and the people who receive the services.

Once these local data are structured through an information system they are useful in identifying populations which are most vulnerable to health problems. The health information system can also identify the specific needs of these people, and who is receiving what services.

A very important factor in this system is that if these identifications are noted periodically, it is possible to monitor trends to determine if needs are being met, services delivered, or other indications of quality health care are present. This ability to monitor health trends is particularly valuable in predicting the probable health conditions of those segments of the population which are the most prone to poor health.

Figure 3 illustrates the likelihood that the relative cost of analyzing data already collected may decrease in the future. One means of making greater analytical use of data already collected is through the increased application of computer capabilities. Computer capability has kept ahead of the increased costs of purchasing or renting computer systems. In addition, more qualified personnel are being trained to use this advanced equipment, while the generalized software packages have increased efficiency, and access through remote terminals permits more people to use expensive main frames.

FIGURE 3. RELATIVE COST OF ANALYSIS OF AVAILABLE DATA

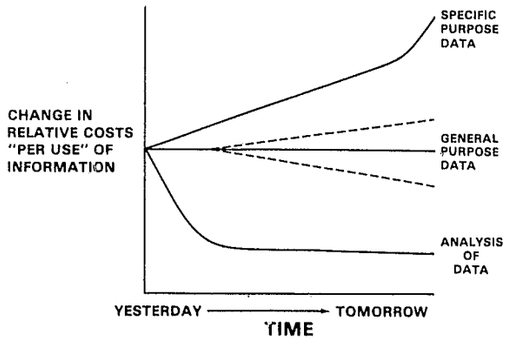


The further analysis and synthesis of information already collected also helps provide some cost reduction per use of information in that it permits the continued building on prior knowledge concerning the phenomenon being studied. Building on what is already known helps reduce the costs of new information by eliminating false starts and needless repetitions.

A tremendous amount of statistical information has been collected and published. The further analysis of these data can provide valuable information to serve the social and economic needs of our Nation. New mathematical techniques, such as model building, simulation, synthetic estimates, and path analysis, can help provide more meaningful measures in the further analysis of previously collected material. Moreover, the further analysis and refinement of data, even when new mathematical techniques are not used, can still provide additional understanding. The presentation of newly computed rates and ratios, and the selection of the most appropriate numerator and denominator can also help provide meaningful measures.

The message of figure 4 is quite straightforward. It summarizes the three previous charts and clearly indicates that relative costs must be carefully considered. At some point in time, when the costs of general-purpose data and the further analysis of previously collected data become sufficiently less than the cost of specific-purpose data, the decisionmaker may very well decide to use the less costly approach. He will, however, still need to give serious consideration to the quality, availability, cost, timeliness, and respondent burden, in determining the type of data to select. For example, administrative records may be available and valuable as a resource, yet privacy and confidentiality rules or considerations may prohibit access.

FIGURE 4. RELATIVE COST OF DATA



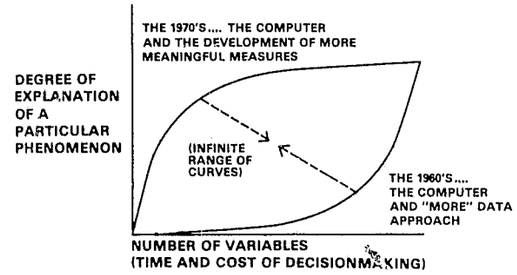
Even though this chart may indicate some cost-benefit gains through the use of other types of data, specific-purpose surveys will continue to be conducted for there will always be the need for unique information. This chart demonstrates that there will be several alternatives in the selection of information needed for decisionmaking processes.

Implicit in this presentation has been the idea that there is a relationship between the degree of data-certainty a decisionmaker will accept and the cost of that certainty—both in time and money. And, with most problems the degree of explanation is contingent on the number of variables to be used.

During the 1960's most decisionmakers found their solutions in the enormous amount of data generated by the computer. These individuals used the "give me everything you can get on the subject" approach. This method required a large number of variables to adequately explain a particular phenomenon. And, one of the problems with this approach was the difficulty in using the reams of data that poured out of these very efficient data generators. This difficulty led to increased efforts to develop methodologies which would allow decisionmakers to deal with only the more relevant and meaningful variables.

This latter approach means that a limited number of variables, properly calculated, can contribute to a greater degree of explanation. In a sense this approach is an effort to synthesize and selectively edit the masses of available data into meaningful information for decisionmaking. The development of economic indicators is an example of turning data into meaningful information for decisionmaking (figure 5).

FIGURE 5. DATA FOR THE DECISIONMAKER



In summary, there will continue to be specific-purpose surveys, but greater cost efficiency and public concern over privacy invasions will cause the further growth and development of large-scale general-purpose surveys and the continued refinement and analysis of data already collected. The growing statistical needs of the Nation—such as the social and economic status of the population, and measures of our economic activity and natural resources—clearly indicate the vital need for the continued development of statistical information systems to meet the varied needs of both national and local decisionmakers. A prime example of better information utilization is the GBF/DIME System which is the subject of this two day workshop.

In closing I'd like to share with you a poem by Edna St. Vincent Millay—which very well captures the problem we are attempting to confront.

Upon this gifted age, in its dark hour. . .
Falls from the sky a meteoric shower
Of Facts. . .they lie unquestioned, uncombined
Wisdom enough to leech us of our ill
Is daily spun. . .but there exists no loom
To weave it into fabric. . .[1]

[1] **Huntsman, What Quarry?** p. 92, 1939.

Great Expectations

HARVEY N. ALBOND, AIP

In 1969 the city of Niagara Falls began studies under the Community Renewal Program financed in part by the U.S. Department of Housing and Urban Development to inventory and analyze its problems and its needs. As a basic part of that inventory the city of Niagara Falls determined that eventually this information would be part of an information system that could form the basis for planning as a foundation for management decisions. This, then, is a two-part report: (1) the basis of its system; and (2) the concept employed to develop such a system.

At the outset we must advise you of the appropriateness of the title of this paper, "Great Exectations" in that we do not yet have an effective Geographic Code Base/DIME-File although we expect through the cooperation of the Census Bureau to have the final polished product in hand shortly.

Fundamentally we employed city records, both existing and specifically developed by the Planning Department, Census records, and the R.L. Polk City Directory tapes obtained from that company under our "701" Comprehensive Planning and Management Program. May I point out to you our specific indebtedness to the Department of Housing and Urban Development, initially under the Community Renewal Program and the successive years of assistance under the "701" Comprehensive Planning and Management Program.

We are pleased to report that our data, even in the early and incomplete stage, has consistently formed the basis for management decisions, otherwise referred to as "policy" decisions or, within city government context, as "political" decisions. In any event, they form the basis for decisionmaking, an appropriate and proper role for City Planning.

The Planning and Development Department of the city of Niagara Falls also acts as Federal Aid Coordinator, as well as transportation planner, civil defense shelter planning, etc. This gives us the opportunity of participating in and being responsible for research and the production of nearly all data and information required of the city.

We are thus able to combine Federal grants obtained under LEAA (Law Enforcement Assistance Administration) programs with systems needed and evolved for planning processes to produce the information and techniques explained in this paper, as well as form the basis for future expectations. We also will point out to you certain applied management functions as a result of these capabilities, not specifically related to, but derived from these early foundations.

MANAGEMENT INFORMATION SYSTEM

BACKGROUND

During 1969 an inventory of each parcel of property within the city of Niagara Falls was done as part of the Niagara Falls Community Renewal Program. The inventory was intended to serve two primary functions: first, as a file of basic data to serve as a source of information upon which to base subsequent analyses; and second, to serve as a permanent record of all land and property contained in the city for the purposes of day-to-day planning administration.

The data gathered in this inventory was compiled and recorded on computer tape. This computer tape is known as the Property Inventory Tape, and the data which it contains is the Property Inventory File. The basic pieces of information contained in the Property Inventory for each parcel of land in the city are the following:

1. Ward number
2. Property number
3. Census tract
4. Census block
5. Zoning category

6. Address
7. Street
8. General land use
9. Number of dwelling units
10. Number of parking spaces

11. Structure type
12. Building condition
13. Occupancy status
14. Area of parcel
15. Total building area (nonresidential)

16. Ground floor building area (nonresidential)
17. Number of floors
18. Use of ground floor
19. Use of second floor
20. Primary use above second floor

21. Year built
22. Assessed land value
23. Total assessed value of property
24. Tax exempt status
25. Year of last assessment

Basic sources of information employed in the compilation of the property inventory included a field survey, Niagara Falls tax maps, census tract maps, Sanborn maps, building record cards, Niagara Falls general tax and assessment roll, the Niagara Falls Industrial Survey, the Property Inventory Coding Manual, and the Niagara Falls City Directory.

PROBLEMS

In the CRP data processing activity, minimum information was collected to determine physical needs for renewal action. Because of limited availability of financial resources, two significant data processing activities were not undertaken: (a) development of a system for channeling information to a central point, including a system for keeping existing data files up-to-date; and (b) provisions of additional information relevant to solving renewal problems to assess alternative solutions, particularly in the area of services and social needs as opposed to strictly physical improvements.

GOALS

"Policies, programs and techniques in neighborhood improvement and housing services depend for their formulation, execution and evaluation upon a solid foundation of *up-to-date, reliable and relevant facts about neighborhood and housing conditions.*" (**New Approaches to Housing Code Administration,** National Commission on Urban Problems, Research No. 17).

Thus, it was the goal of this portion of the proposed Niagara Falls CPM Program to establish an ongoing system of data collection and management that will permit the city government to anticipate and respond to the specific needs for housing and community services of various components of the community. The basic files are stored and processed by computer.

CENTRAL DATA COLLECTION AND MANAGEMENT SYSTEM

OBJECTIVE

Information relevant to CRP activities is being or has already been collected by city departments as routine administration activity (building permits, water billing, etc.), by special agencies (Urban Renewal, Housing Authority, Building Housing and Rehabilitation, Planning and Development) and by departments and agencies for whom data collection and processing is an integral part of their functional activity (city data processing, city planning, regional transportation planning, etc.). However, there is currently lacking a system of channeling information to a central source and a systematic collection and storage of information in a form that is conducive to efficient and expeditious analysis, retrieval, and output.

It is the objective of this expanded CRP activity to help rectify the current inefficiency by establishing a centralized multipurpose information system to be made available in a form that can be used by all agencies, which will provide efficient and expeditious data collection, retrieval, analysis, and output.

PROGRESS TOWARD OBJECTIVE

Stage I

(a) System Design—In Stage I a study was performed of the immediate and long-range needs and requirements of the Department of Planning and Development as relating to a computerized Data Management System. The end-product of this study was a documented report of the findings, including recommendations, which was the basis for the beginning implementation of that system in State II.

(b) Additional Identifiers—Two more identification systems were added to the basic Property Inventory file. These were the 1970 Census block numbers and Planning Unit designations. Work on a geographic grid system, such as the DIME file, was to be incorporated into the system.

(c) Report Feature—A computer program was written that allows the Property Inventory file to be sorted

according to any specified identifier. It then aggregates specified data fields by either numerical totals, mode or median, and prints a report in the form of a list or cross-tabulation.

(d) Staff Training—The member of the staff who will be in charge of the Planning Department's portion of the data management system began to receive training in Computer Science at the State University of New York at Buffalo. This will better enable that staff member to communicate with the computer consultants, make minor programing changes when needed, and to oversee the department's data system when it becomes operational.

Stage II

During the course of the system-design study performed in Stage I, it was found that past efforts by other local governments to first develop a complete and detailed design of a Centralized Data Management System and then to implement that system have not succeeded. It was determined that this is basically because a complete system design is too rigid to be used as a long-term implementation plan for a new concept. In Stage II the design and implementation of the Centralized Data Management System in sections was begun.

The primary thrust in this stage and Stage III is in the area of computer mapping since the display of various kinds of data on maps is not only an integral and indispensable part of the planning process but the function of other city departments as well. The quick and efficient production of maps by computer will directly contribute toward realizing the objective of this portion of the CPM program, which is to provide efficient and expeditious retrieval, analysis, and output of information.

There were specific work elements completed in Stage II toward the above objective:

(a) Geographic Grid System—The DIME file, a type of Geographic Grid System, was completed and forwarded to the Census Bureau for clean-up.

(b) Base Maps—Computer base maps of the city were created from the DIME file for use in computer mapping of the City data files. These are inaccurate to some degree inasmuch as Census Bureau clean-up is incomplete.

(c) Mapping Capability—A general-purpose computer mapping program was written which can extract information from the Property Inventory Tape and plot it on a map of the city.

(d) Tax Maps—Work on designing a method for automatic computer updating of the city's tax maps was begun. This is in abeyance since a complete new tax mapping of the county is currently underway. It is assumed that this will produce the digitized nodes necessary to permit updating of tax maps to the parcel level.

Work Elements for Stage III

(a) Tax Maps—A method for automatic computer updating and drawing of the city's tax maps will be completed by the end of Stage III. This will permit access to the individual parcel.

(b) Zoning—Design of a method for handling a part of the rezoning process by computer will begin in Stage III. This will include machine-drawn maps of the area proposed to be rezoned and the surrounding properties derived from the automated tax maps, and computer printed form letters for notification of public hearings on a proposed rezoning. Presently, matching property numbers with Assessor's information is producing form letters for rezoning on a routine basis at significant savings in time, efficiency, and accuracy.

(c) Census Analysis—Programs needed for analysis and retrieval of data from the 1970 Census tapes will be written or otherwise acquired in Stage III.

These work elements are in direct conformance with the goals of this portion of the CPM Program since it will allow more accurate data to be available and it will allow for more efficient data retrieval and analysis.

DEVELOPMENT OF ADDITIONAL RELEVANT DATA FILES AND UPDATING OF PRESENT DATA FILES

OBJECTIVE

This activity has two objectives: first, to provide an up-to-date data base and, second, to add the necessary data to the existing data base needed to provide appropriate cross-tabulations to determine the interrelationships among various human, physical, and economic factors. One of the basic sources of this data will be the 1970 Census which was not available during the preparation of the CRP. Data to be added to the existing system are of two types:

(a) Dynamic activities occurring to structure, i.e., changes in housing use and conditions (from regular surveys, special surveys, area and cycle inspections, license and certificate inspections, permit inspections, complaint inspections, violation data, enforcement data, demolition data, rehabilitation data, etc.), and

(b) Demographic, economic, and social statistics, including use of R.L. Polk City Directory files.

PROGRESS TOWARD OBJECTIVE

(a) Update of Property Inventory Tape—The 1969 CRP Property Inventory computer tape was updated. This is a continuous updating process for that data file. The basic source of current information is the City Assessor's surveys and records. Therefore, future updating of the Property Inventory File will be coordinated with the updating of the Assessor's files.

(b) Update of CRP Population Report—The Population Report, written in 1969 as part of the city's CRP Program, was updated to include the 1970 Census data.

(c) Data from the 1970 United States Census was added to the Data Management System.

Work Elements for Stage III

(a) Tax Maps—An automated tax map file will be added to the Data Management System.

(b) File Maintenance--City Data files which are presently included in the Data Management System will continue to be updated.

Both of the above work elements are in direct accordance with the objective of this section of the CPM Program.

HOUSING AND POPULATION ANALYSIS

OBJECTIVE

The objective of this portion of the CPM Program is to identify and analyze the needs and problems of housing in the city of Niagara Falls in relation to present and future population characteristics.

PROGRESS TOWARD OBJECTIVE

In Stage II, R.L. Polk & Company used their data files to conduct an analysis of the population and housing characteristics of Niagara Falls. This provided the city with a substantial amount of current, detailed information on these two areas in the form of a computer data tape and a bound report.

An example of the type of information we will be mapping shortly is the relationship of retired heads of household to substandard housing, as well as migrational factors. We will match and map place of residence last year and this year for female heads of household with children who moved over the past 2 years to analyze the implied social and econonic problems.

It is anticipated that this capability will form the primary means of obtaining supporting data for the comprehensive planning and housing programs required by the Community Development Act of 1974. The extraordinarily tight time requirements of the Community Development Act, in our opinion, could not be met by our community to the level that we deem necessary and the law sets forth without the availability of such information and sytems.

COMPUTER TECHNOLOGY METHODS ADOPTED BY THE CITY OF NIAGARA FALLS, N.Y.

This section is basically a technological section for the purposes of briefly explaining concepts employed by the city of Niagara Falls to fulfill its goal of a fully integrated Management Information System. Being a long-term plan, all the concepts are not presently implemented, but at the present rate of progress by the city of Niagara Falls, the ultimate goals are in sight.

The city does not own a computer. Computer time is leased on a yearly contract with a local service bureau. The majority of the data processing for the city is computed through consultants for the city of Niagara Falls. Because in the future the city may either have its own computer or may transfer to an alternate site, all programing is designed to be flexible to enable such a move to be done quickly and efficiently.

The programing was done in ANS COBOL[1] and FORTRAN IV to help facilitate such a move. The serv-

[1] ANS COBOL - American and National Standard COBOL.

ice bureau has the following equipment and software: IBM 370-135 with 196 K[2] real storage, and 1/2 megabytes[3] of Virtual Memory. They are presently under DOS[4], VS[5], release 29. CICS[6] entry level is employed as well as 3270 type terminals. The city systems employ four 2314 type discs fully on a daily basis. These are primarily used for CICS files. POWER[7] is used for partition handling. Four partitions are presently employed, thus limiting the amount of storage in each partition.

Storage limitations are mainly the basis for the intermix of both languages. Also, the file handling of blocked, lengthy character records, such as the DIME[8] file is handled in ANS COBOL more easily than in FORTRAN, whereas the FORTRAN more efficiently handles number grinding as needed in such applications as plot generations. The greatest advantage of the languages used is interchangeability between third generation equipment.

The Department of Planning and Development, as well as other departments of the city, have present and future applications which rely upon the DIME information as part of the integration of Data Base files for more efficient management information.

The basic logic used in a majority of the programing is described in several steps. Foremost is the Plot Variables General-Purpose System (a Custom Package) presently adapted for City Planning use. The inputs for this system are a planning property inventory file, which fully defines all parcels in the city limits, and a DIME file. The system selectes variable numbers of census tracts or census blocks as requested by the user. The next step extracts various variables from the Property Inventory File, such as age of dwellings, conditions of dwellings, residential, nonresidential, and/or zoning requirements. These variables are summarized and comparisons are generated and applied by address ranges to DIME information.

Interim audit steps help support final results. This output is then passed to plotter generation programs that create plotter control files, which are readable by a CALCOMP 936 Digital Incremental Plotter. The coordinate data used in the above generations is map set miles.[9]

The same basic concepts will be used in the near future with Polk directory files and/or U.S. Census Count tapes in conjunction with the property inventory tapes.

In other city departments similar logic was used to produce fire maps showing locations of fires, as well as supporting edits to verify results. This has proven extremely valuable in pointing out graphically high-potential fire areas and has resulted in reevaluation of fire districts and manpower. Initial savings are in the order of $280,000 annually. Other concepts as mentioned above will be used to quick-reference property inventory, particularly in analysis of structures upon fire calls in order to reduce fireman injuries due to unsafe dwellings.

Other major areas covered, such as matching of high accident areas (figure 1), lie in the law enforcement area. This has resulted in justification and installation of traffic control devices under an LEAA grant. Future police systems are being developed to utilize the above concepts for optimizing posts, assigning manpower, and analyzing high crime areas and potential crime areas that could arise in the near future.

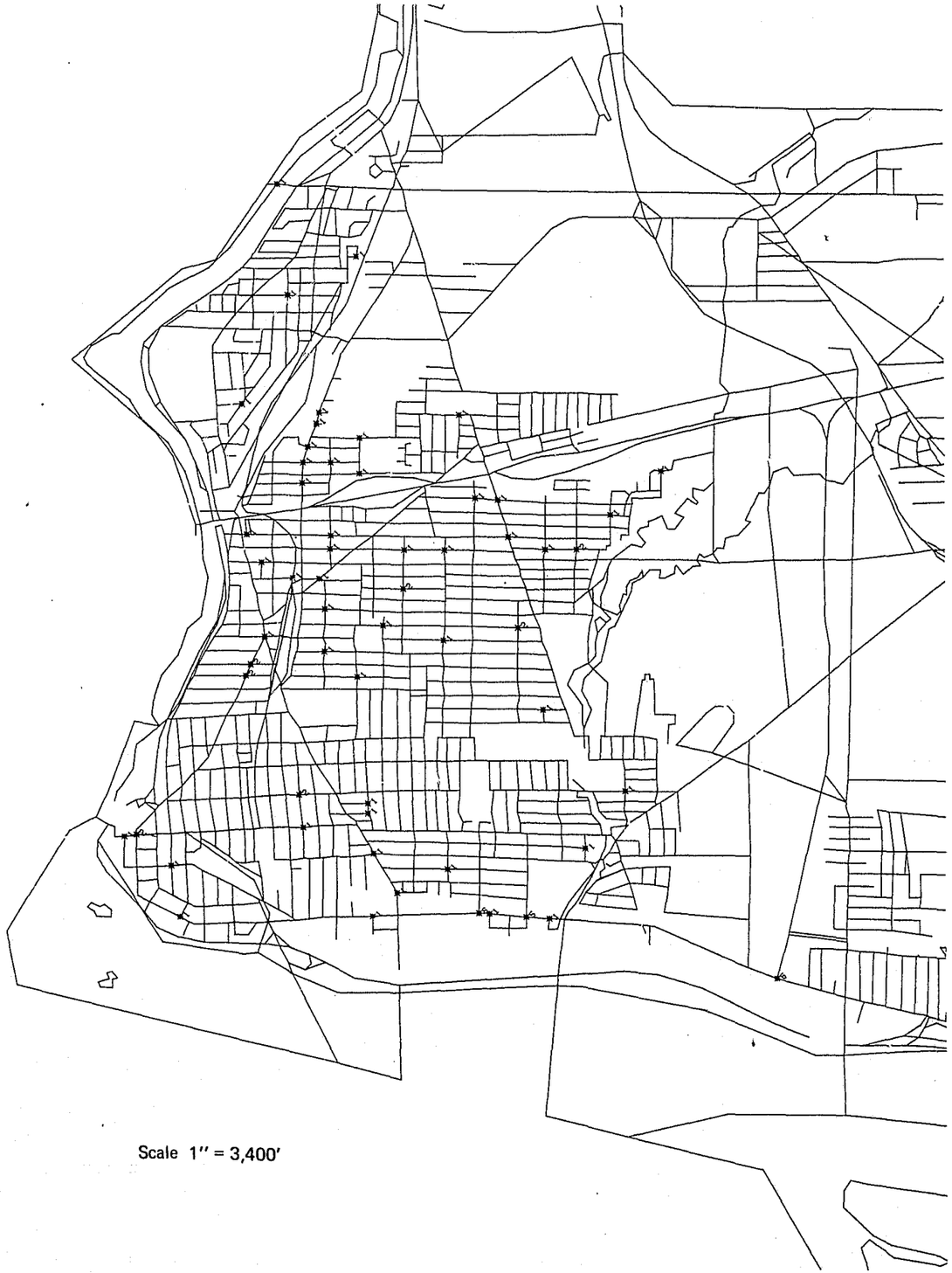
In conclusion, may I point out to you that the approach in Niagara Falls has been to develop specialized programs to meet specific applications with an overall appreciation that they will be integrated into a total system. Development of any one application thus became a building block for future applications. Thus, duplication is avoided and new opportunities are created. This calls for professional cooperation and management direction seldom found at the municipal level of government.

Specific recognition for the development of these systems should be accorded to:

Herbert J. Hoelter, Director of Office Administration, city of Niagara Falls
Ms. Donna J. Peuquet, Senior Planner, city of Niagara Falls
Ralph Giambra, Consultant, Dataware

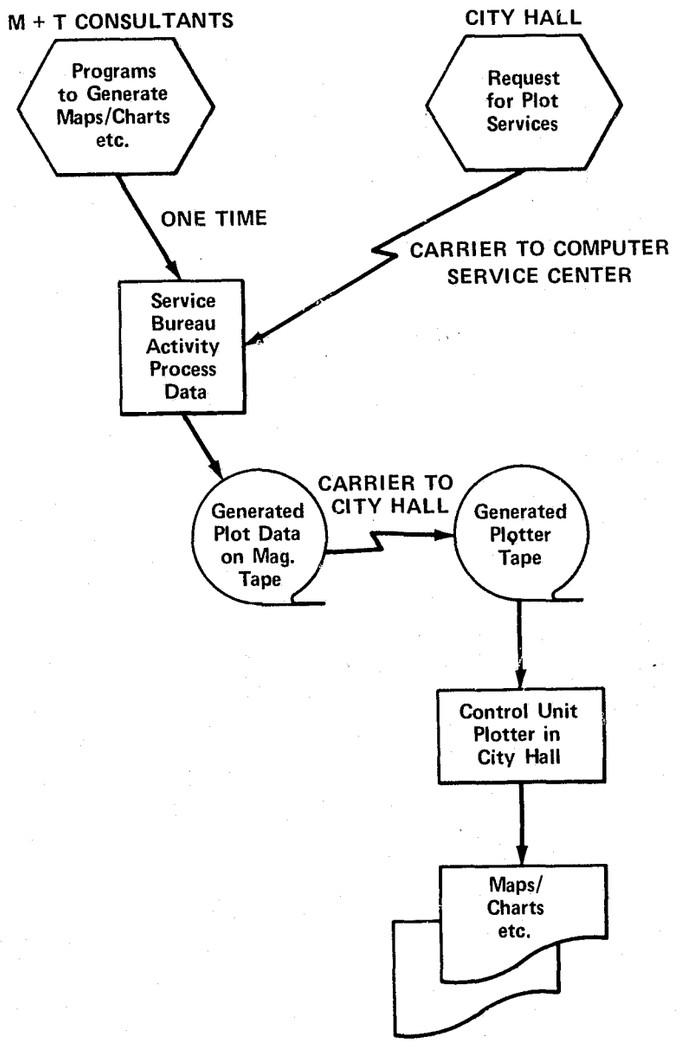
[2] K - 1,000 bytes core storage. (Byte - a sequence of eight adjacent binary digits operated as a unit.)
[3] Megabytes - millions of bytes.
[4] DOS - Disc Operating System.
[5] VS - Virtual Storage.
[6] CICS Entry Level - Customer Information Communication System (an IBM package for teleprocessing).
[7] POWER - Priority Output Writer (an IBM package for multiprograming).
[8] DIME - Dual Independent Map Encoding System.
[9] Map set miles - relationship of miles to inches within DIME file record structure.

FIGURE 1. **1973 Building Fires—City of Niagara Falls**

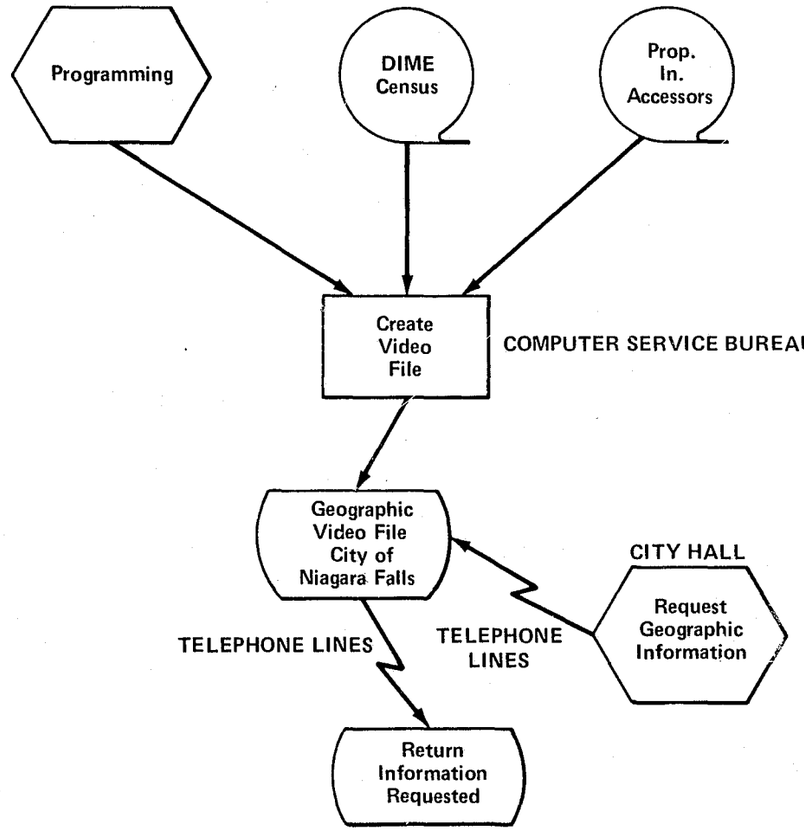


Scale 1" = 3,400'

APPENDIX A Plotting Activity Layout



APPENDIX B Video Display System
(Near Future)



Question Period

Mr. Jacob Silver—You mentioned quite a number of uses by the public sector. Have you been able to show the same types of examples or provide your services or ideas to the private sector? How has the private sector gotten involved with the use of this "tool" in your city?

Mr. Harvey Albond—No, with the exception possibly of "hooking" the Hooker Company to CALCOMP planning—they came over and played with ours and liked it so well that they now have one of their own, and also they took away the free digitizer (which they bought) that I could use—we have not basically involved the private sector yet. For one thing, we are concerned with the confidentiality of Polk directory data. The Polk people, after all, are in business and their tapes are available to private enterprises. Consequently, we have some concerns in this area. We haven't worked out a policy—at this point we are playing shy.

In terms of the private sector and other aspects, I suppose it is more indirect than direct. For example, we do meet regularly with the banks to discuss housing and mortgage needs, real estate, etc., but this is more or less the planner meeting with them, armed, if you will, with the basis of research. I meet with the Real Estate Board. As a matter of fact, based on some of this research, I lecture and serve on the faculty of the Graduate Realtors Institute in New York State. But, again our involvement with the private sector is a rather limited relationship at this time.

I am intimately aware, as a former marketing man of many years ago, how the data we could have, could turn out a basis for a variety of marketing decisions. I have thought, for example, of staging a publicity game for our newspaper, where we would ask them to supply a list of their subscribers, and we would immediately give them a map showing them a picture of all their nonsubscribers, and things of this nature. Obviously, this quite a tool for private enterprise, but I am not sure that this is a role of city planning.

Mr. Jacob Silver—The GBF/DIME-File is a tool that does not involve privacy. As "tools" in the management of data systems, wouldn't the Geographic Base (DIME) Files, the ADMATCH type of operation, and the mapping operation be of importance indefinitely to the local, private sector?

Mr. Harvey Albond—Certainly, if you are looking for a missionary, I could and would be one, and will be one very shortly. It is just a question, I suppose, of time. This is probably my greatest limitation when I run such a simple operation; we just never have enough time to do all the things that we want to do. I can see fantastic opportunities in the entire area of private statistical analysis just from what we, not relatively large, but simple business and government have been able to do for ourselves in a very short time, at a very modest expense, frankly.

Mr. Michael Hogan—In response to Mr. Silver's question or comments, I would like to make several additional comments. My first comment relates to the fact that I think it is interesting to attend a conference such as this one, where persons such as myself from the private sector have an opportunity to meet and work with government planners. I am here for two purposes: the first is to meet some of the government planners in Ohio, and the second is to learn some of the techniques that can be used in marketing decisions in the private sector.

I am, however, concerned about several factors. The first area of concern is that there is insufficient representation from the private sector. A second area of concern is that there needs to be a better mechanism to make the geocoding software more accessible to representatives from the private sector. As a result of the availability of public funds, governmental agencies are the only ones that can afford the luxury of some of the programing necessary to make these systems work. Once again, I would like to emphasize the point that I think there needs to be a mechanism, whether it is through the Census Bureau or whatever, to make these systems readily available to the private sector. My third concern is directly related to and is an extension of the availability-accessibility problem. Everything that comes out of Washington, for the most part, is an IBM type application and many of us do not have IBM equipment and cannot afford the luxury of converting existing software so that is will be compatible with our particular computer systems.

Mr. Harvey Albond—I will ask Mr Giambra to comment on the last part in just a moment. May I just say that I have just turned the corner. I now have more government time than I have private enterprise time. I, however, appreciate the position you make, I might point out that is kind of unique to hear that we have all the prototype for research money. We have always had very little compared to what has been used in the private sector. Particularly, when I think of some of the research capabilities that have come out of aerospace industries in the defense budget.

Frankly, we could be a service bureau, but on the other hand, I think we are competitive with private enterprise. There might be some uproars there. I do not think it necessarily costs an arm and a leg, if you know what you are doing. Our programs and our data by and large are public information. It has perhaps been manipulated and massaged, but it is public information. Theoretically, you could have gone through the same process, and therefore, I probably have no problem making my data available to you. The Polk directory does not cost an arm and a leg. In our city it would cost you about $5,000 to get the tape. A consortium of several banks, if you wanted, the newspaper, or somebody else could very easily purchase that tape from them. The only reason I do not sell it is because they are in the business of selling it. Perhaps I might have an ethical relationship there, but again it is public information. You could hire somebody to go through the directory and take out every name of this year and last year, but obviously that is the way they print the directory so that tape is available.

Mr. Michael Hogan—I was directing my question more towards the software.

Mr. Harvey Albond—Any software I have belongs to the world. It was developed with Federal funds and it belongs to the world. If it is worth it. If if is useful to you, we will give you the documentation on it. If you want the tapes, you can have the tapes. You ask the question about conversion to other types of equipment, we have had those problems in the past in our city government. I will let Mr. Giambra answer that question.

Mr. Ralph Giambra—All the programs that we have designed for the city of Niagara Falls are flexible enough so that they require a minimal amount of effort to convert to any other system, hopefully, that has an American Standard COBOL, such as UNIVAC, Honeywell and various other vendors. In fact that is the reason why we used ANSI COBOL and FORTRAN IV. Our company has done large conversions, and from the knowledge that we have gained through these conversions, we have designed programs for the city of Niagara Falls that are as flexible as possible. I feel that the programs that the city now has you could put up on your system, whether it be IBM or not, probably with 2 or 3 days effort and conversion. There are SYNTAX converters available at reasonable rates that would do the conversion automatically for you also.

Mr. Alan Pisarski—I suggest that during the discussion phase we come back to this "private" versus "public" question. It is an important question. There is a lot of room to pick on everybody. I think government planning people, particularly at the local level, really do not meet their responsibilities, having received public funds to develop something, to take the initiative to disseminate and let the world know, in fact, that the public product exists. I have tried to stimulate interest over the last couple of years in a requirement that any Federal grant would carry with it the responsibility for broad public dissemination of the fact that something has been done, praticularly when several agencies in Washington are funding exactly the same thing in the same city. We need some process that lets the police department know that the planning agency is being funded to do a GBF/DIME program, etc.

On the other hand, regarding the private sector, I am just astonished at how slow they are, and really how unsophisticated they are. I really do not think that the lack of funds is the answer. As Mr. Albond said, the tolls are not that expensive. They are there. There is a great deal of difficulty in providing this kind of information to the private sector. If you are in a local planning agency, it really gets to be a proprietary question of whether you are in the consulting business yourself, or whether your whole planning agency program gets bent out of shape because two or three private sector firms have discovered that you are a gold mine and the rest have not.

The Department of Transportation is going through a process right now deciding what its whole public information responsibility is; because, while it is a very important function, it can really modify the whole program of the Department.

Public and Private Uses of a Geographically Oriented Date Base

MITCHELL TUCKER, JR.

It's a pleasure for me to be here and tell you something of the GBF related work going on in Tulsa. My presentation will touch on three areas:

1. The Integrated Municipal Information System (IMIS) program in Tulsa (which is non-USAC);
2. The development of a locational subsystem derived primarily from our GBF; and
3. A brief summary of some of the systems which use or will use the locational subsystem and the impact on the public and private sectors.

Tulsa's IMIS development began in 1966-67 with a functional systems study of all city departments and related agencies. The computer hardware was expanded and the existing critical systems were redesigned to fit into a telecommunications and data base approach.

In 1970 the development program got under way. Three major subsystems were devised.

1. The Municipal Administration and Control Subsystem, which provides information to managers and operating personnel about the internal workings of the government itself. Much of this information is expressed in accounting terms, but a wide variety of performance resource management and planning data is also supported.
2. The Public Safety Subsystem, which provides information on police, fire, courts, corrections, investigation and enforcement.
3. The Community Development Subsystem, which provides information about the community itself, its condition, its reaction to various programs, how it is changing, and its general vitality. The primary outputs are in the form of feedback to urban planners and decisionmakers. It supports comprehensive community planning at all levels.

The GBF data was originally planned to reside exclusively in the Community Development Data Base as much of the data was by necessity geographic in nature. However, as the design and development process went on, it became more evident than ever that geocoding and geoprocessing is the common thread that holds a municipal information system together.

Thus a Locational Subsystem was designed. Its main purpose was to centralize and standardize all locational information, update, and retrieval functions, as well as to eliminate the redundancy and related inaccuracies of separate geographic systems.

Figure 1 is a data structure diagram of this subsystem data base. Data structure diagraming is a design tool we use in conjunction with our data base management software.

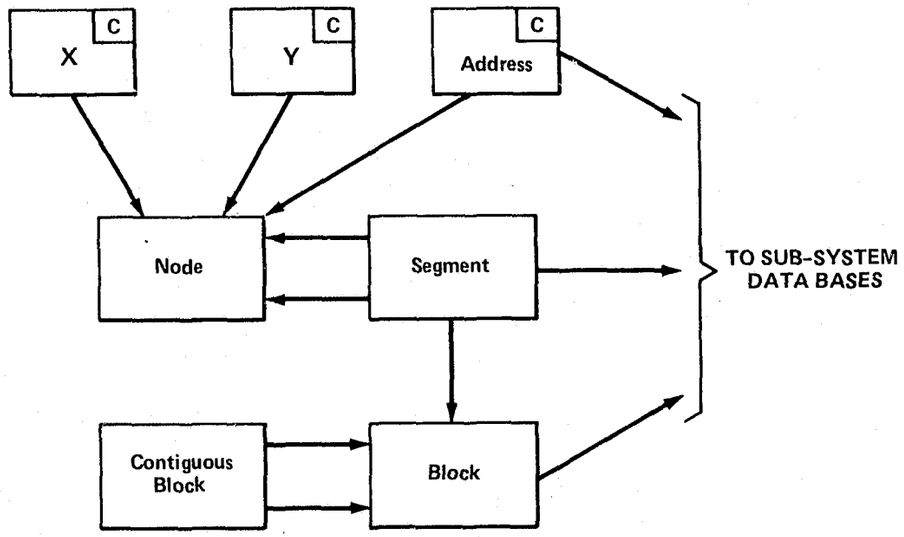
Each box represents a series of records with the same type of information in it. The arrows indicate the linkages between the various records.

There are three primary points of access. These are: the X coordinate; the Y coordinate; and address. The small "C's" in the upper right hand corner of the record indicate that they are directly accessable through a randomizing and search technique.

The node record is the conjunction of a particular X and Y coordinate pair. The address record is also linked to the node record even though any given node may or may not be associated with an address, intersection, or address range.

The segment record then can link any two nodes together to define a segment in geographic space. This gives the flexibility to define special segments for particular purposes which may or may not exist through the life of the file. This is useful for special studies or monitoring efforts. The base segments, such as streets, are protected from such manipulation by special software locks.

FIGURE 1



Associated with the segments are one or more blocks represented by the block record linked to the segment record. Any set of contiguous blocks may be linked together through that record providing the solution to many of the undefined boundary aggregation problems. We have found that running blocks instead of running or matching streets is more efficient in an interactive environment even though this subsystem allows both techniques.

Through the various linkages, address to node to X, Y coordinate conversions are handled quickly, as well as X, Y to node to address. A standard point in polygon algorithm is utilized for that problem, but we see the undefined areal aggregation problem as the primary power of such a structure.

A set of linkages to the three other subsystems provide the ability to extract data of various categories (i.e., crimes, financial resources for a particular program, and social/economic factors) for analysis and input into the next level of the planning process.

Figure 2 shows a conceptual view of the entire IMIS including a retrieval and analysis subsystem. This allows the noncomputer person to compose questions, yet undefined, to the computer and receive replies, free of computer jargon, the results of a variety of analysis routines, or in graphic format.

This subsystem is simply another way to represent and use the basic GBF data and to harness the power of its geocoding principles.

One of the biggest problems in our minds was the communication of various kinds of geographically oriented data to nontechnical people in a form that would have a high probability of affecting the decision-making and planning process.

Research, practical experience, and consultation with behavioral scientists convinced us that, indeed, a picture was worth a thousand words and was certainly of more value than 14 pounds of computer print out. We have designed and installed a multimedia information display system. A schematic diagram is depicted in figure 3.

A variety of media including 35mm slides, microfilm from a computer operated retrieval unit, and computer generated graphic information can be mixed together and displayed on a large-screen (5 feet by 7 feet) color television projection system. This system can be used for both internal planning sessions or for communications to private citizens in public meetings. Further, because it is TV based, we have tied it into the cable TV network for either live or taped broadcasts.

FIGURE 2

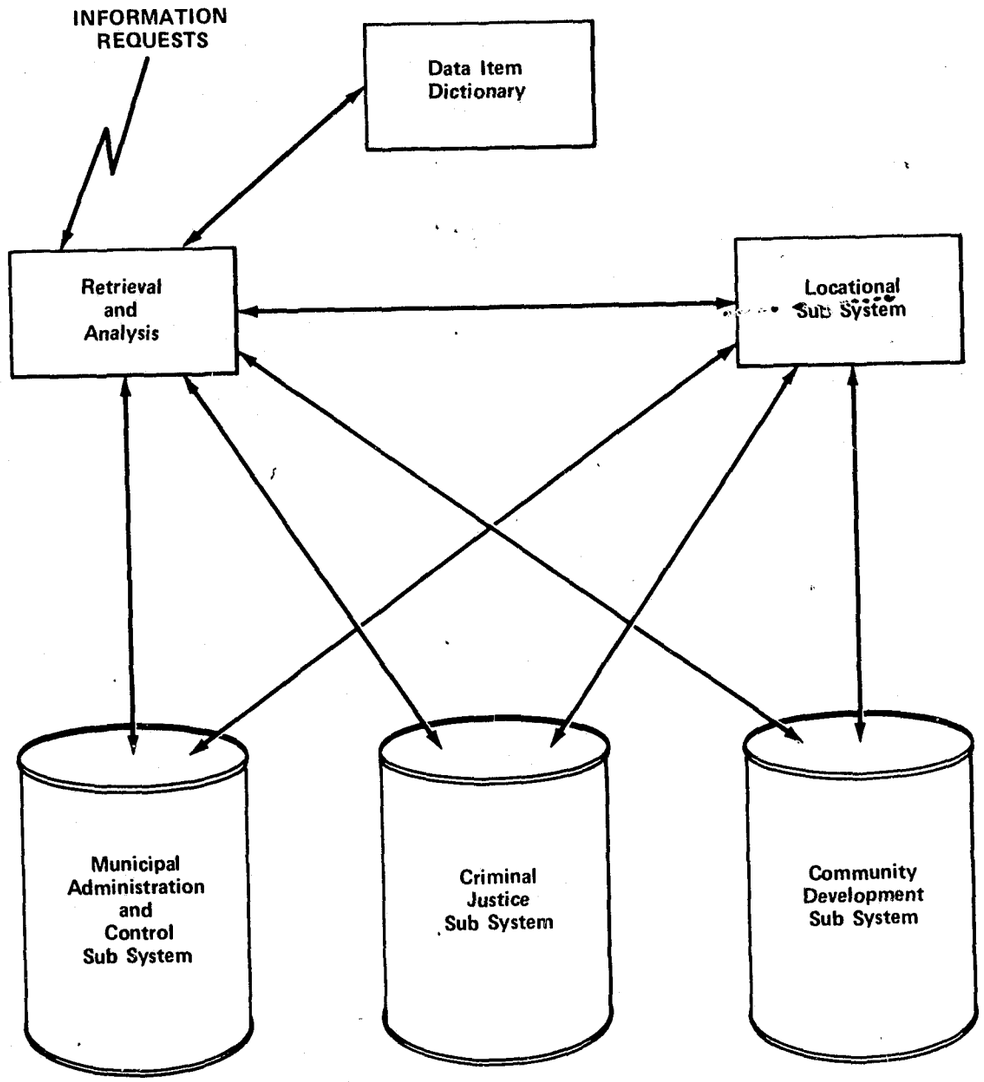
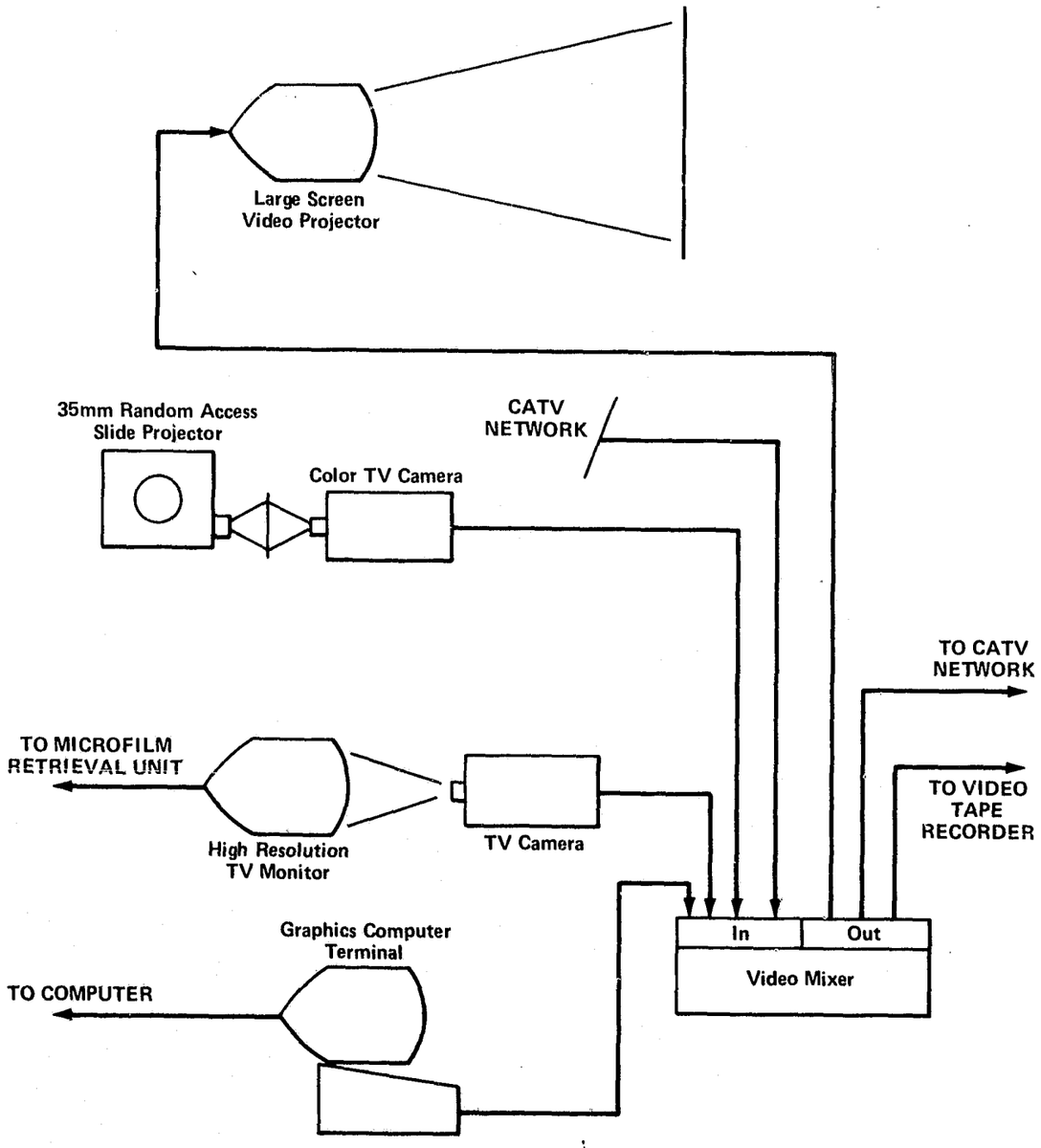


FIGURE 3



At this time, we have too limited knowledge concerning all the things this kind of system can do. But again, an extreme amount of flexibility has been built into it and we are getting new ideas every day. We do know that it is an invaluable tool to close the communications gap between the public and private sectors in geographically oriented information.

There are several systems which are or will be supported by these geographically oriented tools. One is a Community Activity Indicator Project. In cooperation with HUD and LBJ School of Public Affairs in Austin, Tex., the city of Tulsa has designed and implemented a set of community indicators that reside in the Community Development Data Base. These indicators are tools which make inferences about various physical, economic, and social conditions in our community. Problem areas can be recognized and resources can be allocated to remedy their condition.

The indicators are not cure-alls, but as the word implies, are "an indication" of a condition in a particular geographic area.

In order to monitor the various broad-base areas of the community, the Indicator System has been divided into 11 categories. These categories represent a cross-sectional view of the total community. These categories are:

Sense of community
Economic
Education
Environment
Health and well-being
Housing
Land use and recreation
Public safety
Transportation
City service delivery
Demography

Within each category, there are several indicators that comprise each particular category. For example, under the category of housing, some of these indicators are:

Vacancy
Single/multifamily building starts
Number of demolitions
Number of abandoned units
Number of subsidized housing units

These indicators, collectively, would generate a variety of information to infer conditions related to housing. Each of the 11 categories is described in the same manner.

At the present time, the indicators are aggregated at a planning district level. These are homogeneous neighborhoods and do not necessarily coincide with standard census geography. Hence, the use of our Locational Subsystem comes into play by relating data elements by common geographic areas to be used to construct an indicator. As two simple examples, figure 4 shows how the indicator "physicians per 1,000 population" is calculated using outside medical data and standard census demographic population data with projected updating. Figure 5 shows an indicator which reflects intracity mobility. In both examples, data is aggregated at levels and areas different than those reported. The desired results were obtained through a geographic conversion scheme based on the GBF-based Location Subsystem.

FIGURE 4. Number of Physicians per 1,000 Population

DESCRIPTIVE VALUE

Indicates the availability of doctors to the population. This indicator will show where the city stands in relation to other cities and States, but the indicator does not show whether all those who need the attention of a doctor are getting it.

METHOD OF CALCULATION

$$\frac{\text{Total number of physicians}}{\text{Population}} \times 1{,}000$$

DATA SOURCE, FREQUENCY OF AVAILABILITY, AND LEVELS OF DISSAGGREGATION

Oklahoma State Medical Association, Oklahoma Osteopathic Association, Yearly, Planning District/City.

REFERENCE

LBJ "Universe of Community Indicators," p. 48.

FIGURE 5. **Percent of Households in City Experiencing Intracity Mobility**

DESCRIPTIVE VALUE

The indicator is proffered on the assumption that a community having a greater rate of mobility (intracity) will concomitantly have a lesser "sense of community."

METHOD OF CALCULATION

$$\frac{\text{Total number "off" utility connections}}{\text{Total number of households}}$$

DATA SOURCE, FREQUENCY OF AVAILABILITY, AND LEVELS OF DISAGGREGATION

City of Tulsa, Utilities Services Department, Monthly, Planning District.

REFERENCE

LBJ "Universe of Community Indicators," p. 117.

Figure 6 shows a graphic representation for the indicator "communicable disease index". You can see that the new planning district boundaries do not coincide with census tracts or square miles.

Another system which shows great promise for the decisionmaking process is called The Proposed Development System. This is a projection system which is used to evaluate the impact of proposed developments on various municipal service delivery systems. Currently it indicates the impact on liquid and solid waste, air pollution, school children, water supply, parking, and other ancillary factors. It has been designed to be modular so that other service delivery functions can be added as we develop the linkages.

Again, this system, which has a large potential in the private sector, is geographically based. It must draw data from a variety of sources aggregated at different levels. The more complex the calculations, the need for a common geographic base grows. By maintaining a centralized, standard location system, we can alleviate many of the problems.

Going full circle, once the indicator and projection systems have shown a need or problem in a particular area of the city, the planning process will develop programs to correct the situation. These, of course, are translated into budget terms in the next program budget. The locational subsystem is also linked to the Municipal Administration Data Base so we can compare the actual fiscal resources being applied to a particular area for a particular program for monitoring and evaluation.

I have tried to give some indication of how the basic GBF data and processes can be converted to an integrated data base orientation, and give you some ideas on possible applications.

This is certainly not a static system and must be flexible enough to aid the ever-changing decision problems of local municipal management. With such new responsibilities as general and special revenue sharing bring, the local government must have better and more timely geographic data to develop community policies and evaluate their effectiveness. We believe that our data bases revolving around our centralized locational system give us a sound base to support out municipal management.

FIGURE 5. Percent of Households in City Experiencing Intracity Mobility

DESCRIPTIVE VALUE

The indicator is proffered on the assumption that a community having a greater rate of mobility (intracity) will concomitantly have a lesser "sense of community."

METHOD OF CALCULATION

$$\frac{\text{Total number "off" utility connections}}{\text{Total number of households}}$$

DATA SOURCE, FREQUENCY OF AVAILABILITY, AND LEVELS OF DISAGGREGATION

City of Tulsa, Utilities Services Department, Monthly, Planning District.

REFERENCE

LBJ "Universe of Community Indicators," p. 117.

Figure 6 shows a graphic representation for the indicator "communicable disease index". You can see that the new planning district boundaries do not coincide with census tracts or square miles.

Another system which shows great promise for the decisionmaking process is called The Proposed Development System. This is a projection system which is used to evaluate the impact of proposed developments on various municipal service delivery systems. Currently it indicates the impact on liquid and solid waste, air pollution, school children, water supply, parking, and other ancillary factors. It has been designed to be modular so that other service delivery functions can be added as we develop the linkages.

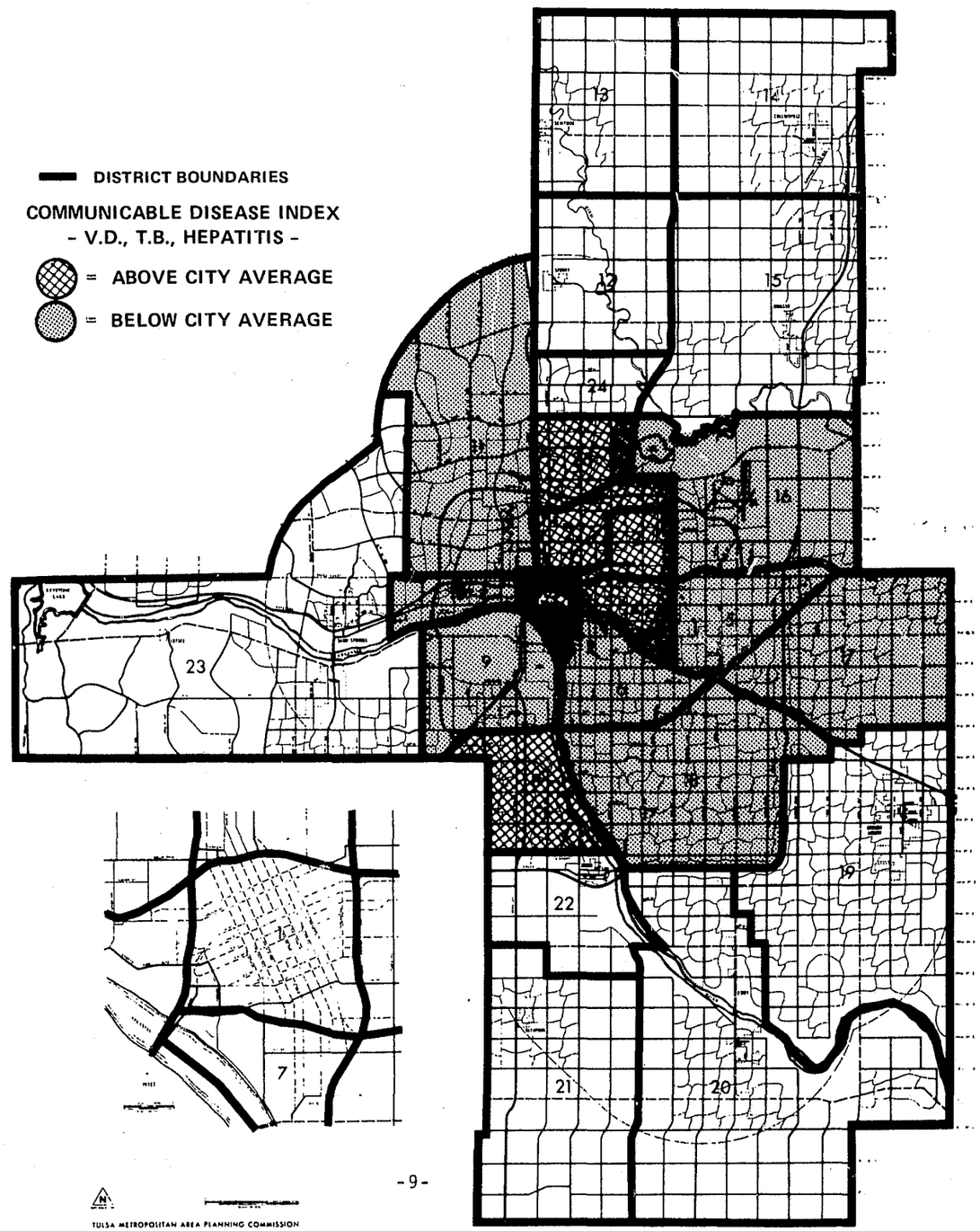
Again, this system, which has a large potential in the private sector, is geographically based. It must draw data from a variety of sources aggregated at different levels. The more complex the calculations, the need for a common geographic base grows. By maintaining a centralized, standard location system, we can alleviate many of the problems.

Going full circle, once the indicator and projection systems have shown a need or problem in a particular area of the city, the planning process will develop programs to correct the situation. These, of course, are translated into budget terms in the next program budget. The locational subsystem is also linked to the Municipal Administration Data Base so we can compare the actual fiscal resources being applied to a particular area for a particular program for monitoring and evaluation.

I have tried to give some indication of how the basic GBF data and processes can be converted to an integrated data base orientation, and give you some ideas on possible applications.

This is certainly not a static system and must be flexible enough to aid the ever-changing decision problems of local municipal management. With such new responsibilities as general and special revenue sharing bring, the local government must have better and more timely geographic data to develop community policies and evaluate their effectiveness. We believe that our data bases revolving around our centralized locational system give us a sound base to support out municipal management.

FIGURE 6 Tulsa Metropolitan Area Planning Districts



Question Period

Mr. John Meyers—You mention the use of a cable as one of the possible subsystems to be used with the GBF. I am wondering what percentage of the population in Tulsa is wired for cable, and what kind of success you have had in obtaining citizen feedback via this cable on any particular problem put forth for mass response.

Mr. Mitchell Tucker—The cable installation is still going on in Tulsa. The percentage of people signing up for it is very high, and as I said, that particular system is too new to get any kind of feedback at all yet. What we are finding is that the presentation methods using large screen TV in the room are better for communication.

Dr. Robert Aangeenbrug—One of the difficulties we had in Wichita Falls 4 years ago leads us back into an issue which we may discuss later. In Wichita Falls it was pointed out that, theoretically, two-way cable TV's would be acquired. This resulted in a project discussion that really affected the staff. Do you know whether the superdesign in Tulsa also allows for two-way TV? We discovered that technically we could reach a point where a city could turn one's TV off because of an overdue water bill. Or, we could broadcast an individual message on the screen for delinquent utility bills. You can see how we moved, theoretically at least, into some very difficult policy questions. It brings back not too pleasant memories. Did any of that discussion take place in Tulsa?

Mr. Mitchell Tucker—We do have two-way capability, but in terms of turning them off if they didn't pay their bill, that has not been discussed yet. They keep raising their rates and that turns enough off.

Mr. Alan Pisarski—I guess I found the whole discussion of your system very attractive. The last element, the proposed development system, I think is an excellent approach. An unnamed county in the Washington Metropolitan area got together with an unnamed consulting firm and tried to develop a cost benefit system for any kind of an initial development in the city, to find out what the total cost to the city would be, or the county, and what the total benefits in tax base would be.

I will never forget the output of about a half million dollar-research program which came up with a sophisticated series of equations in a model structure, and the summary said that the design was conceptually correct, but there was not a data base anywhere in the world to feed it. At that stage they just dropped the whole subject.

It seems from the way that you were describing it that you have taken a very straightforward and kind of clean approach to it. Is your approach based on existing rates of cost and output per unit input? Is that how you do this kind of work? In other words, the air pollution per unit of travel, or sewer cost per square foot?

Mr. Mitchell Tucker—Each segment, each delivery system has a set of tables that are changeable, and that is really how it works—the capacity, for example, of our water system, which does change of course; the capacity of our sewage handling facilities, which can remain stable but does change as we add to the treatment facility capabilities, and so on.

Mr. Alan Pisarski—How do you calculate the new demand that some development would create? Do you have simple rates to work with?

Mr. Mitchell Tucker—There are a set of tables that we use in our calculating. We have two ways of doing it. One at the medium standard for this kind of development or the maximum of this kind of development. Again, it is not magic or anything like that. It is a pretty straightforward way of doing it, but it gives a pretty good idea of what is going to happen.

Mr. James Thomas—What kind of utilization do you have for the community development aspect of this? What kind of demand do you have? How available is it? Who uses it?

Mr. Mitchell Tucker—Are you talking about the proposed development or the community indicators?

Mr. James Thomas—The community indicators.

Mr. Mitchell Tucker—To give you some idea of utilization, the library has been deluged to the point that they want to put a computer terminal in the library for public access. We have had requests from everybody, from private companies to the Salvation Army. We are now battling, by the way, a policy problem of publication and distribution; it has been running our presses into the ground. We so far have been able to ward off many special computer runs. We have a policy that we will not be a service bureau. Any standard reports that can be published, extra copies for example, of course, are public record, unless it is confidential data.

Mr. Alan Pisarski—Would you comment, in rough scale, on the cost picture over the years of your program.

Mr. Mitchell Tucker—The Mayor says it is too costly, and I say it is a bargain. These things are not, of course, a major development program, which is looking at all of the data bases and all of the conversion problems; massive amounts of data cost a lot of money, no question about it. We have utilized grants, but once the data bases are in place then these other systems come along rather inexpensively. Even the display system, itself, was rather inexpensive. This TV equipment is getting to the point that anybody can afford it now.

My department, which does all the systems development for all the city, the Planning Commission, and the COG has a budget this year of $970,000 and that includes all of the computers, everything. The city of Tulsa had a population of about 335,000 in 1970. The SMSA was 502,000, that includes the three counties that are in the COG area. That $970,000 is a high. In fact, the systems I have been discussing are a very small part and the indicator systems did not consume much cost at all. I cannot tell you exact figures. The Proposed Development System has taken a long time and a lot of coordination, and really boils down to the time and effort of the people both in the public and private sector. They chipped in their own time, and the computer programing was really nothing.

Mr. James Thomas—When you mention the population that you are servicing, is that the city of Tulsa, the 335,000, or is that the three-county area of 500,000?

Mr. Mitchell Tucker—I am serving not only the 335,000 within the corporate limits, but also the three-county region, the 502,000.

Mr. Alan Pisarski—I think we have had two really very interesting discussions. One of the things that we might play with during the discussion later on is the question of unit cost per capita for cities of various sizes and try to figure out if there is some magical number. I know in the early days of urban transportation planning, we used to use a very convenient number of $1 per capita total cost for a large scale transportation study; that was in the very large cities. I do not know how well that worked in the smaller cities. I will be coming back to both you gentlemen during the discussion stage. I guess I can warn you now to comment on the nature of GBF/DIME building problems in your city and particularly with respect to whether the prime focus of the problem is institutional, financial, or technological, and perhaps look at it in that light.

Market Area and Branch Location Analysis for a Large Commercial Bank Using GBF Systems

ROY K. CAMPBELL

Geoprocessing (data gathering and manipulation within a geographic framework) is rapidly gaining use and users. For many years geoprocessing applications were found principally in public and quasi-public organizations, but in recent years private business has become increasingly aware of applications of geoprocessing technology. Commercial banks are among the largest, if not the largest, single group of private business users of geoprocessing. The structure of retail banking (individual and family banking needs) provides numerous opportunities for using geocoded data: market share and penetration analysis, product and service cross-sell pattern development, service area delineation, branch location analysis, and strategy planning for branch network evolution. Thus geoprocessing of existing data sources can provide timely and spatially accurate input to actionable marketing and branching programs.

The experiences of Manufacturers Hanover Trust Company (MHTCo.) are illustrative of problems which have occurred, or can be anticipated in, many business-oriented uses of geoprocessing as a marketing tool. MHTCo. is a New York City headquartered commercial bank with deposits of well over $26 billion, it ranks third by deposit size in New York City and fourth in the country. Ninety-five percent of its 1.8 million accounts are retail, and deposits in these retail accounts provide a key portion of the money necessary for the lending function by which a bank makes its income.

These retail accounts are distributed among nearly 200 branch offices serving the New York metropolitan area and some upstate New York communities. The large volume of geographically-oriented data spread over an extensive area provides the prerequisites for geoprocessing, but also generates the first set of problems—availability, accessibility, and processing of internal data.

Virtually all banks provide a variety of account types—regular and special checking, savings, loans, credit cards, etc. A banking unit (an individual, family, business, etc.) may have several types of accounts and, often, several accounts of the same type (for example, a special checking account, two savings accounts, a personal loan, and a bank credit card). At MHTCo., each account type constitutes a separate data base with between 50,000 and 1,000,000 account records; to further confound the matter, most account bases consist of two files—a name/address file and a ledger file, which contains balance and account activity information. Almost all marketing applications of these data bases require both address and balance information, thereby necessitating a merger of the two separate files into a single file containing name, address, and balance data.

The uses to which marketing research wishes to put bank customer files suggests compression of the now merged account files into household, or more accurately, banking-unit records. This can be done for each banking unit by combining all accounts with the same last name using the same mailing address into a single record containing name, address, and account types with their respective balances. At MHTCo., compression reduces the original number of accounts by a third with each new "master record" representing a unique banking unit. This compression produces a high degree of analytical accuracy in cross-sell pattern research and cost efficiency in direct-mail efforts by minimizing duplicate mailings. Unfortunately, our production-oriented data processing center is not equipped to handle our file merging and compression needs, a fact which forces us to find an outside supplier, who must become familiar with the file structure and idiosyncrasies of each of the several account bases. This familiarization process is both expensive and time consuming, thereby placing a premium on finding a supplier with the capability and, more importantly, the expertise to process data beyond the simple file merging stages. Urban Data Processing, Inc., supplies the necessary file familiarity, software availability, and experience for our needs at

reasonable time and dollar costs. Our use of UDP illustrates a problem solving approach we've found most effective: purchase the solution to the problem whenever possible, rather than risk the time and budget overruns inherent in most developmental efforts.

Although the merged and compressed master banking-unit file, called the Customer Data Base, serves many purposes in and of itself, its implicit research capabilities require additional, external data. Using an outside supplier to manipulate our large volumes of internal data allows our research and planning staff more time to become involved in the tasks of finding, evaluating, and using these external data sources. The various geographic base files for the New York metropolitan area present a case in point. Virtually all spatially related research and planning for branch banking activity require locational detail below the ZIP code level. The required detail is available through the use of address coding guides, which assign census and other area tie-in codes to a record based on the address in the record. While not a conceptually difficult task, care is necessary because inaccurate geographic coding can limit, or worse, distort conclusions drawn from research based on the data. The ACG published by the Bureau of the Census for New York City would be the most obvious source of census-related data; however, it has been found wholly inadequate for large-scale geocoding at the tract level because of an unacceptable number of errors and omissions. A reject rate of 20 percent would leave well over 200,000 records to be manually coded, which is not a trivial problem. Furthermore, while sampling might be appropriate for some applications, any research conclusions used as input to spatial management decisions would be questionable, because the extent of errors due to improper coding would be unknown.

Rather than attempt to correct the Census Bureau address coding guide, we sought out an alternate source. The Reuben H. Donnelley Corporation maintains a very accurate ACG for the New York area, which we were able to obtain as a part of a larger research project. After validating their guide and evaluating its limitations, we used it to geocode the Customer Data Base with excellent results. The cost of obtaining this file was minimal relative to the time and manpower saved over attempting to edit and improve the Census Bureau guide or to manually geocode numerous rejects. We also added several other descriptive data elements to our ACG, the most important for our purposes being the branch of MHTCo. into whose territory the particular street-address-range record falls.

Another research application employing geographic base files is our ABLES (Automated Branch Location and Evaluation System). ABLES allows us to take an overview of an area, determine banking potential for each simulated market we postulate in the area, and rank them in order of most to least promising.

The ABLES computer programs combine four basic elements: census data on population and housing units, lists of commercial business, survey data on bank usage, and a file of all competing banking institutions. To a great degree, all of this data is integrated through the use of the bank's GBF—the Donnelly New York City ACG.

Each household is assigned a banking potential, based on income. A Bank Usage Survey had been taken in an earlier year to gather detailed information on such banking habits as family income, number and types of accounts maintained, average balances, and how far people traveled to their bank. The survey data was broken down by area type (suburban, city residential, poverty, etc.) and translated into a series of probabilities. For example, "the average" suburban family with an income of $12,000 has a .997 chance of using a bank for any service and a .946 probability of maintaining at least one savings account at a commercial bank. The average number of accounts is 1.66.

Multiplying all the factors together with expected balance figures results in a "commercial bank savings potential" of $1,039.67 for a household. From the survey results, a table of potentials for income ranges was developed; these were translated into linear equations for a smoother progression. (That is, a man's balance would not suddenly jump several hundred dollars as his income changed from $14,999 to $15,000, or wherever the break occurred). The resulting deposit potential of the blockgroup's "average household" was multiplied by the number of households to determine the total residential potential of the blockgroup.

Similiar logic was used to determine business potential. A commercial list of business in the area, classified by size and type of each business, was obtained; previous studies had given an indication of what percentage of net worth was available to banks from different types of businesses. The individual business values were summed by type and multiplied by the appropriate percentage to determine total business potential.

The final data element of the system was a computerized file of all competing banking institutions. Since an area might already be "overbanked" by

competitors, total potential would not give an accurate representation of an area's ability to support a new branch. Potential had to be adjusted for this competition factor. From various reference sources, a record was established and verified for each institution; each bank was coded for status, type, geography, etc. All approved applications and existing banks were included in the file; different weights were assigned to savings institutions and commercial banks.

The four data factors discussed above were aggregated into simulated market areas by means of a geographic coordinate system. Data was "collected" at a series of points; these points were "collected" into market grid totals according to the distances between them.

The Medlist coordinate file was used for the population data. (This file had been created by the Census Bureau for the Office of Civil Defense. From street maps, a population center for each enumeration district was estimated, and a point plotted. This was translated into computer-readable form by using a digitizer; coordinates were attached to census data and released as a special file. Thus, the population data came "prepackaged"). The commercial business list had to be coded to the appropriate geographic unit by first computer-matching it against our address coding guide, and then hand-coding as necessary, to aggregate business potential to the same units as residential potential. Bank locations were precisely plotted on a map and then digitized; coordinates were combined with identifying information.

The large area to be evaluated (a county, for example) was divided into grid squares; the size of these squares was determined by population densities and travel-to-bank distances that had been indicated in the surveys. Potential for a new branch at the center of each grid was measured by "collecting" all block group points within the prescribed distance and dividing by competition in the immediate area. (Diagram on page 00 illustrates this concept). The grid pattern was shifted several times to obtain a number of combinations of blockgroups. Finally, all of the individual grid totals were ranked to determine the parts of the county that showed the greatest potential and that should be more closely examined for new branch sites.

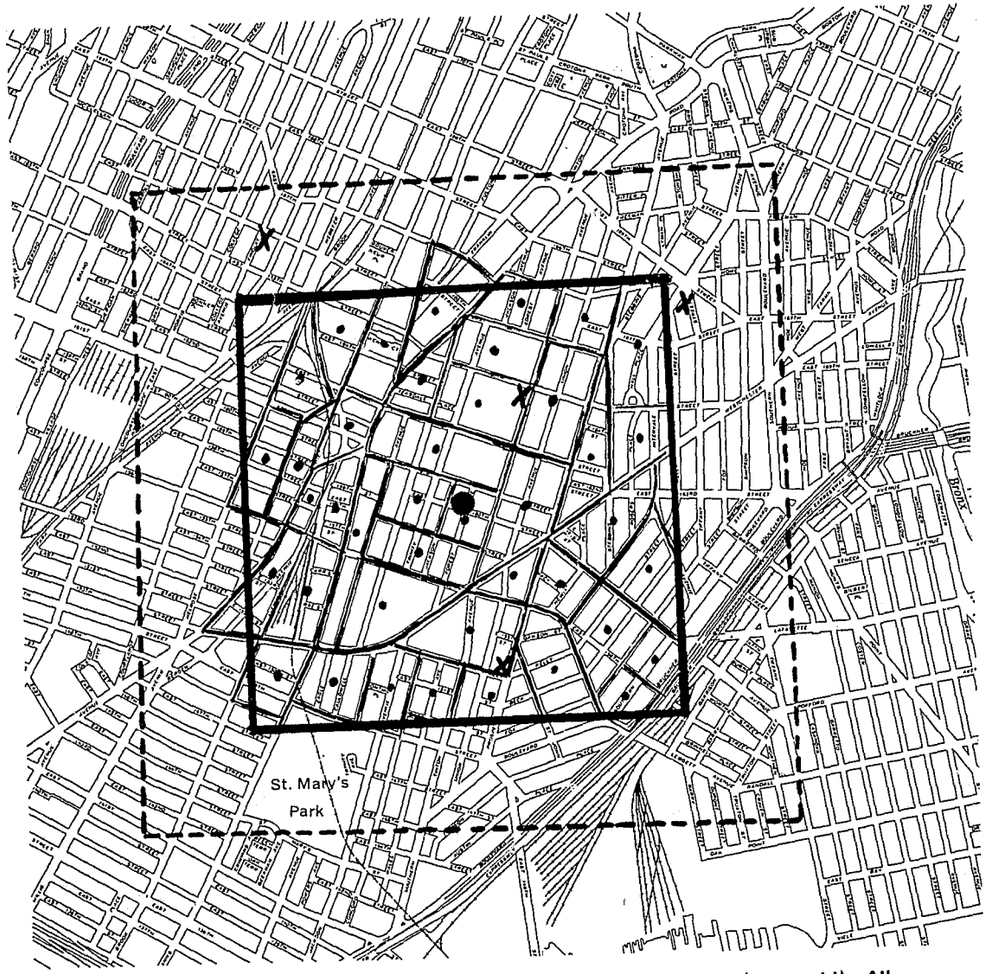
An ABLES run for Queens County can be used to illustrate the points made above. Because it is a densely populated urban area, a grid size of one mile was agreed upon. (This meant that all "potential points" within one-half mile of the center of the grid, at which potential is measured, would be included in the grid totals). The competition area was defined as an "inner bank grid" of one mile on a side (same as the basic potential grid) within which competition would be given full weight, and an "outer bank grid" of one-quarter mile beyond. In this fringe area, competition would be considered at half its usual value. Grid centers were established at one-quarter mile intervals to obtain as extensive a coverage as possible. Results were plotted on a map, (color-coded according to amount of potential) and showed clear patterns of which areas could probably support a new branch and deserved further investigation.

ABLES is essentially a simple system. We have not tried to make it say "locate a branch on the North East corner of Front and Main streets", rather it says "look for a site in or around the sourtheastern part of Bayshore." Human judgement is heavily built in, permitting off-the-model evaluation of spatial anomalies involving rivers, cemetaries, parks, and the like. We've been able to update the Census population and income data through commercially available sources while keeping the bank file up-to-date ourselves. Since the Donnelly-derived ACG is only a little over a year old, we've not had to face updating it yet.

I joined Manufacturers Hanover in 1968. One of my first projects was to find a method of relating the then upcoming 1970 Census data to the bank's marketing problems. I first heard of the address coding guide in 1969 and over the years, the concept of geographic base file, as realized in the ACG, has been a foundation about which our research people have erected a sophisticated approach to integrating customer records, external marketing data, and geographically packaged population information into analytical tools that have both saved the bank money and generated profits in their own right.

We do not have a DIME file for New York City yet. The complexities of our ever changing metropolis are such that we may never have one. We'd certainly like to have one—the DIME approach is the "top of the line" in geographic base files and could be easily used in our pragmatic approach. The graphic capabilities alone stagger the imagination. We'll probably have to wait until a broader segment of private industry joins its weight (and dollars) with the academic and civil groups who are already knowledgable geographic practitioners, then the aggregate demand will cover the dollars which have not been supplied to date.

ILLUSTRATION: How Potential and Competition is Aggregated



Map shows urban area with simulated market area of one square mile (inner grid). All block group coordinates within the inner grid are aggregated into the grid potential total; banks in the ¼ mile outer grid, as well as the inner grid are considered as competition.

- ● Hypothetical site for new branch bank
- X Location of competing bank
- • Block group centroid
- - - - Boundary of outer grid (for competition)
- ▬▬ Boundary of inner grid

Question Period

Mr. Alan Pisarski—I think I would like to comment. There was very little in that presentation that I would say was unique to banks. If you are talking about service-area analysis, defining service areas around some activity, a supermarket, or a fire station for that matter, whether it's a public or private unit that provides a service, it seems to me that this kind of spatial distribution approach is still functional and still quite adequate. I think that the notion of a vast difference in the approach of the private and public sector is really not necessarily true. A great deal of site analysis that is done for public facilities where there is an optimization function that may be slightly different, but not really terribly different, you are still trying to locate the best set of facilities in the best locations.

Mr. John Meyers—Did this particular procedure hold true in your expansion to other areas in New York State, that is, outside the New York City metropolitan area?

Mr. Roy Campbell—We have just taken a simpler approach. We have taken an area-analysis approach first. We are in your area now in Buffalo with a very small bank. The regulatory agencies would only allow banks like ours to take over fairly small banks, and most of them are located outside the urban areas. We are now in Buffalo, Rochester, Albany, down in Orange County. We expect to be in the Syracuse and Binghamton areas. The approach we take in these locations is to go in and in effect do an overview, kind of melting together the whole geographic economic approach to look at it. Then we work in small territories like that to the ZIP code level because data at that level is easier to work with. At that level we can certainly get the business data, especially down to a ZIP code territory, for anywhere in the United States using the ZIP code record.

We had one of our fellows go out to the University of Michigan and get his doctor's degree, and he worked on a series of models checking bank usage, relating the 1970 Census with a big survey that was done by the Survey Research Center at the University of Michigan. He derived another way of approaching bank potential different than the ABLES file that we have found works very nicely in territories outside of New York City. Given the size of banks we have, we do not have to know all of the data in painstaking detail. In most cases they will only allow us to put up two or three branches a year until 1976 when it becomes statewide. We do not need to know the fine detail in areas like Buffalo. The business potential becomes considerably more important, so we run our business list very heavily in those areas.

Conceptually it is pretty much the same approach; it is the idea of going from the top down. The particular details of application upstate are a little bit different. Again, the pragmatic has to come in, and we are looking at various cost constraints. It has worked very nicely, though; it has turned around our whole relationship with what we call the holding company clients in the division. That is our sales guide, to convince them that this kind of information is useful to them, and that it will give them a better concept of the ways they spend their dollars. It has worked very nicely for upstate.

General Discussion

Mr. Alan Pisarski—Some of the points that were raised yesterday deserve consideration. The institutional question came up again and again of how do you organize to make these things happen. It came up regarding two levels of coordinating mechanisms. Within a city, within an urban region, how do you put together the set of players in some formative fashion, hopefully point them in the correct direction, and keep them pointed in that direction? But also, I think, the question arose of coordination between cities, among cities. How do cities become exposed to what is happening in other places? How do the developments in individual cities become exchanged? We talked about clearinghouse functions perhaps to be carried out by some of the existing professional associations, which I think probably popped into the minds of some people. I was absolutely astonished that we did not get into the discussion, at any time, of what do you do outside the area of existing addresses, or nonstandard areas.

This morning we heard an additional series of papers on differing applications and experiences. One of the common threads that I noted in them was what I would call a very sensible sense of scale, modest efforts, not in their accomplishments, but in their pace. I did not see grandiose designs where the program design component takes all the money and 3 years. People started out working in a very pragmatic fashion.

I have asked Mr. Morton Meyer from the Geography Division to join me on the podium, and Mr. Aangeenbrug, who will be wrapping up this conference at noon, to help lead the discussion. I will be calling on people representing cities here to perhaps get some sense of feel of their experience in their particular area. We would like to relate as much experience of different cities as we can so that we have some sense of levels of progress and the problems that are happening around the country.

Mr. Charles Wenner—I want to express my gratitude to this group for coming to the city of Columbus. I have heard an awful lot of things, some I understood and some I did not. Of course, that is why we have people employed, that are here today from the city of Columbus, that can answer questions prior to decision-making problems. When I first started listening to this conference, I questioned why I should be here. It was not too long until it was very clear to me that if I have a problem or a question regarding computers I can pick up the phone and call Mr. Savoia, and he can come up with the answer.

I think the benefit has been that it has made the people in the city of Columbus that work with this on a daily basis, cognizant of the fact that there are other people all over the country who have gone through what they are going through; and maybe, as indicated here yesterday and today, that Columbus is far advanced in this field. Maybe we can share some of the expertise that we have with other communities throughout the country.

I would hope that as a result of this conference, the papers that have been produced might be made available to any of the next conferences that you might have because I believe a lot of information has been given out to the people. I am sure these same people will not be available at future conferences. If you could compile some of the history of conferences and where we are in Geographic Base (DIME) Files, and so forth, I think it would be a benefit for all concerned.

I might also mention that there were a couple of words used yesterday and today that might be a barrier to what some people are trying to perform. One was "consolidation." Here in the city of Columbus, we have managed in the field of emergency medical services to share and cooperate with the surrounding communities. If you are in a remote point in the city of Columbus and the next community has emergency medical services, and a call goes out, they will dispatch the nearest emergency squad to serve you. This, of course, is a step towards regionalization, but it is not consolidation.

Everbody wants to keep their positions, keep their own community, but it is working down that road; so the word "consolidation" is not too good.

The other word I heard was a "welfare bank," and I hope all those depositors that deposit there don't ever hear about that.

Again, I do want to express my appreciation to the people who put this conference together and to all the people who participated. I think they all did a very excellent job, and we would be most happy to entertain any future conferences.

Mr. Alan Pisarski—Perhaps I might ask some of the other people from Columbus—Mr. Henderson, Mr. Thomas—to talk a little bit about the issue that came up about how you consolidate and how you coordinate these activities. Perhaps more importantly, how do you get the word out to people that this tool exists, and let them know what you have got and what the capabilities are?

Mr. Ronald Henderson—We have had a problem here in Columbus in getting the word out. We have been concentrating on maintaining the file. A few of us went to Indianapolis to talk to Mr. John Rowe. We were really impressed with how they concentrated on using their GBF/DIME, an uncorrected file. They consequently got their city-county government to fund the CUE program, because they realized the value. We have taken the other approach—maintaining the file. I think we have a better file to work with, but it is a little tough getting people turned on to using the file. We are concentrating on that right now. That is one of the reasons why we were really excited about this conference—to bring some of our local public officials and private firms to this conference to show them what the GBF file really is and what can be done with it.

We have plans for conducting seminars internally in our own agencies and also for the State of Ohio, to help bring all of Ohio's GBF/DIME's up to date, to share applications with everyone in Ohio and, of course, everyone nationally. We have had calls from as far away as Dade County in Miami from criminal justice people interested in our on-line address matching program. We have taken advantage of application programs from Mr. Rowe and his people. One of the things we do not believe in is reinventing the wheel. I do not mind stealing someone else's program, and I certainly do not mind someone else stealing my program.

We are really gearing up, since we have the file pretty well on the way maintenance-wise, to concentrate on applications; so we are going to try "selling" the file in Columbus now.

Mr. Alan Pisarski—I meant to comment on a previous remark. I am sure there will be published proceedings of this conference. So the papers presented here and the discussions will be available. I do not know what the distribution on it or the previous proceedings happens to be. Mr. Silver, could you tell us a little bit about that? As one mechanism for disseminating some of this information, I think the proceedings themselves are a significant document.

Mr. Jacob Silver—All of the people that are present at this conference will be getting a copy of the proceedings. In addition, we provide copies to a number of advisory committees (such as the Small Area Data Advisory Comittee), to all of our coordinating agencies, to our census tract key person, plus various professional organizations (such as URISA), to the Washington, D.C. regional and State offices of FHWA, HUD, LEAA, and all of the Department of Commerce offices. That really covers a good number of people and agencies. Just to give you some idea how important and how useful people think these proceedings are, we have had the proceedings of each of our five other conferences published and have sold out each time. We now have gone through several reprintings of the earlier conferences.

Mr. Alan Pisarski—Do the conference proceedings get announced in both the **Data User News** and other data user publications?

Mr. Jacob Silver—Yes, **Data User News,** through the AIP newsletter, the URISA newsletter, and through the newsletters of several other professional organizations.

Mr. Alan Pisarski—Coming back to the question of the internal coordination within a city and how the different elements of the city get together, I had mentioned previously that I was going to ask Mr. Albond and Mr. Tucker to comment on that.

Mr. Harvey Albond—Contrary to public expectations, the "grant" is not dead. "Grantsmanship" is very much with us, whether it is in the discretionary funds of the Community Development Act or whether it is in all of the ongoing programs. Even OEO is not dead, the Community Actions Programs are alive and well. Transportation—I chair our regional transportation study—is in the throes of a required and continuous update. What I am trying to say, is that there are various nonprofit social organizations in the public and private sector, for example, the United Way in our particular case, that are looking for economic development assistance grants. The source of basic data that could be available within the system is the thing. Consequently, if you get the word out at all, you become an invaluable resource, sine qua non, you become something that they cannot do without.

In reconciling this public-private thing, which I was wrestling with earlier, I think in listening to our friend from the bank, the fact of the matter is that we are dealing with a consumer product. Government is a series of consumer products and services, and therefore, the various justifications and measurements we do, whether in the public or private sector, are almost identical—the techniques are almost identical. With a great deal of pride, I can now say that I do think government is ahead of private enterprise, but there is no reason why, again, with appropriate safeguards, and hopefully support, especially at budget time, we can't be of service in that regard also.

Mr. Alan Pisarski—I was struck by the thought, after hearing the Director's remarks—you remember that set of curves he put up on the screen demonstrating that costs of a general-purpose survey decline with increased usage—that the Geographic Base (DIME) File, which is obviously not a survey but a general-purpose tool, parallels his point regarding shared costs; and so I think that the same cost relationships hold. We have the same problem with Geographic Base Files; which is, that frequently coordinating institutions do not exist within a city to bring together a set of people to fund and to direct a tool that is of common value to all of them. This is particularly true when some of the players do not even know that the tool exists, or if they did know, it would not impress them unless they could see that tool in action—see its application.

Each city has to take its own approach on whether it is going to have a very heavy public dissemination program, or not. But, one of the functions of Federal Government is to assure that the local agency that it is supporting is doing the job of letting the rest of the city know that it exists.

Mr. Mitchell Tucker—In spite of that magnificent piece of software that we have developed, some of the data is just downright lousy. This comes from the problem of updating and that is now centered in a city-county agency. That is an institutional problem. They, of course, are having other problems, such as workload problems, which causes them in turn not to contribute all that much.

The impact of these newer systems is turning this around, and the Chamber of Commerce in Tulsa has been a big help in emphasizing, or causing a renewed emphasis, in this updating and maintaining of accurate geographic data. The Chamber, of course, understands now that this can be an immensely favorable impact in the private sector, the community.

Mr. Alan Pisarski—Would anyone else like to comment on the question of dissemination within a city and getting the word out?

Mr. Richard Luckay—The W.A. Storing Company is in the private sector. I believe we have the unique distinction of not only contributing financially to the upgrading of the Columbus GBF, but we also have had the opportunity, and the pleasure, to contribute hard data to the updating itself.

We provided Mid-Ohio Regional Planning with address and ZIP code data that we use regularly in our direct-mail operations. In effect, we were giving away part of our "family store" when we contributed. But we are interested in this GBF, and in making sure that Columbus has the best GBF in the country. What it amounts to is that data is data and without a viable application it is virtually useless. Once information regarding the GBF—and its potential— is disseminated by means of this convention to the local area and beyond, basic recognition of GBF capabilities will lead to requirements for specific applications. We have many specialized applications that are, today, among the best in the country. But unless the GBF remains in the public domain and is made available for everyone's use as a base data file, our applications will not have a market. Consequently, we are more than happy to be here today and to be able to contribute to the GBF in any way that we can, either financially or data-wise. You might say we are proof-positive that the private sector has the capability to help not only the GBF/DIME update (and the entire program), but conversely to help itself in the process.

Mr. Alan Pisarski—I think that is a very pertinent comment. We talked in the past about the notion of a kind of geographic utility, if you will, a public service function of the individual urban area that will provide this general service to the world with perhaps some group like the Post Office providing kind of in-and-out service.

Mr. Donald Bufkin—In the Tucson area, the Urban Transportation Study agency got into the creation of a GBF/DIME file for the region very early.

As I recall, we assisted by providing local mapping assistance and editing on early versions of the metropolitan map series and, of course, completed the GBF development and coding prior to the 1970 Census. The prime reason for the early involvement of the transportation study in the Tucson area was that we were the only feasible agency with regional planning responsibilities and capabilities. Our region had and still has sometimes sharp and sensitive divisions between city, county, and State planning agencies.

The transportation agency initially, and now since late 1973, the Pima Association of Governments (the designated Metropolitan Planning Agency) is the logical agency to function as local coordinator for GBF/DIME and as repository for the information.

Since the initial work of the transportation agency in GBF/DIME, the evolving and ever more structured "3C" urban transportation planning process has gotten considerably more confining. May I say for the benefit of the US-DOT people who are here today that you have created the Intermodal Planning Groups and you have created a certification procedure that has forced us to address ourselves more directly to transportations. Our role now in GBF/DIME is more that of coordinating, and attempting to encourage commitment and cooperation between city, county and State agencies. That function can be a frustrating and discouraging thing.

I wish that I could report that we have a key or that there is a solution waiting somewhere down the way, but the transportation studies that I am familiar with have little staff time to direct their efforts toward GBF/DIME development and maintenance. I have found that the attempts to encourage cooperation among city and county agencies is not bearing the fruit we would like to see it bear in regard to voluntary effort in correcting, updating, and expanding the GBF/DIME.

Mr. Alan Pisarski—Those are very appropriate comments. Perhaps not all of that is necessarily accidental. I think there was some conscious feeling in the Department of Transportation that we went through a period, in the mid-60's particularly, where in many cities the only data game in town was the transportation agency. We reached the stage of financial, if not mental, exhaustion. There were some really very bad things done. By that I mean, transportation data sets—which were quite comprehensive, quite sophisticated, and in many cities, as I say, the only set in town—were used for the most awful purposes, with a result that the Department was in a very uncomfortable position. It comes back again to the Director's point of how do you share the common cost of the independent data set. The way it was being shared was DOT did it and everybody else used it. There was a need, and there was pressure, to create a more balanced, let us say, data collection process.

I might mention two programs of the Department that I think would be appropriate to some of the concerns here. The one is the journey-to-work program and that has several facets to it. One, is the tabulation of the journey-to-work data from the 1970 census. That is somewhat aged, but I think it is a good example of what can be done in a kind of local-Federal-Census cooperation. It is to my knowledge, the first time that the Census has been in a position to produce data sets where the requester defined the area system for the data set.

What we did was design a set of basic tabulations of the journey-to-work, send those out to the cities and the States for comment, and put that back in for a couple of literations, and develope a standard tab set. The Federal government paid for the programing so that these data were available to run, and then a city could define, let us say, traffic zones, school districts, or whatever area system they wanted to use, and it could request the data by that area system. It had to provide the Census Bureau with a block-to-zone or a block-to-area converter. The Census Bureau would then run the summary tabulations up to that zone level, and provide that to the city. The cost to the city was the actual direct cost of the data processing as opposed to any of the start-up costs. I think out of the 200 and some SMSA's these days, about 180 or so have taken part in that program in one fashion or another, with varying degrees of success as a function of the geographic coding.

One of the very valuable elements too—because they actually did do origin and destination work between those zones—is that is was the first time that the Census Bureau, in effect, inverted its population file and did tabulations of the work address as opposed to the home address. Thus, a lot of nontransportation people have become interested in the data because it does provide some look at employment statistics.

Following that, we are now getting into the melding of journey-to-work data into the Census Housing Survey, which is an annual survey. Twenty cities in the country are selected for fairly intensive sampling each year and that is rotated on a 3-year basis so 60 major cities get covered—6 cities get 15,000 interviews each year and 15 or 14 get 5,000 interviews as an update of housing and journey-to-work statistics. I think there is another program which will be of some use.

A final program I might mention, because it fits very well with what Mr. Campbell was talking about, is something that we call "Special Area Analysis," which is a battery of computer programs that provide the capacity to do a kind of contouring as a function of travel times. It calculates people's accessibility to different services, whether it is accessibility to jobs or accessibility to hospitals. We are talking about the percent of population that can get to a facility in what proportion of time. It leans very heavily on Census data and on network data for the city. Mr. Meyer, I have made several references to the things that I thought were unique. Do you want to comment on them?

Mr. Morton Meyer—Mr. Pisarski has touched upon a use of the GBF which is limited only by the imagination of the user. Once the computer has available to it the elements of a mapped network and a way of relating data to the network, much that was formerly impossible becomes practical; and with additions to the GBF file, can even become routine. We plan to add building names to the file for "address matching" purposes and because buildings are also major trip generators. Add building coordinates locally, plus one-way street indicators, plus travel times by street segment during rush and nonrush hours, and the basis for truly sophisticated origin and destination analyses exist. The same techniques apply, of course, to accessibility to hospitals, location of doctors in relation to the population served, etc. The converse is also available, that is, where should facilities be sited to best serve the population of risk. I should mention here that in order to better serve needs such as these, the Census Bureau plans, within the limits imposed by confidentiality of individual information, to make available as much summary statistical information as possible for small geographic areas.

Mr. Michael Plett—I have a comment on something that has been stressed, but I want to restress it. I am from the greater Cincinnati area, from the Ohio-Kentucky-Indiana Regional Council of Governments, and if our agency had been asked to give a report here, instead of calling it the "Columbus Success," we would have called it the "Cincinnati Failure." The difference between the Columbus project and the Cincinnati project is basically that the Cincinnati project lacked two ingredients that were mentioned yesterday.

The first, of course, is cooperation. In Cincinnati, one agency was stuck with the role and nobody was around to really coordinate an extensive geocoding operation. Although we started out as a multiagency, multicounty operation, monies dried up in various agencies, other agencies had no reason to support the continuing effort, and finally only one group was left to carry out the activity—and that was a transportation planning group.

The other problem that occurred was that the whole approach was not applications-oriented. People kept asking, "Well, why are we doing this?" and there was no good answer. There was nothing to show the decisionmakers to justify our continued involvement with DIME technology. As happens in all Federally-funded transportation planning activities, as has been mentioned by the gentleman from Tucson, there is a strict set of certification requirement—a, b, and c. At every monthly staff meeting someone would say "Let us spend some time doing some updating on the Geographic Base (DIME) File," and the response would be "What does that have to do with a, b, and c?" Gradually, the commitment of time for the project was reduced. No time was spent on training people to continue this effort, knowledgeable people drifted away, and the project went fallow.

We have just spent the past month finding out that there is an application. We are currently required to extensively reappraise our existing (1965 Wilbur Smith) plan, and to adequately achieve this, we must utilize local data. For example, we have 35,000 employer addresses and over a million automobile registrations to geocode. If this geocoding had to be done manually, it could not be done; so, we must have an automated geocoding capability. This implies a need for the most up-to-date DIME files we can get. The trouble is that we have just expended the last month in trying to figure out where we stopped a year and a half ago.

The two ingredients that I think cannot be overemphasized are that you must have a cooperative sustained effort—you need a consortium by a number of agencies, a number of funding sources, and a number of people who can provide constructive input to let you know exactly how to best develop the file—and secondly, you must show initially that you are going to do something with the file. You have got to have applications to begin with; so that you can say "We either develop DIME technology or we are in big trouble."

Mr. Alan Pisarski—I think that is really an excellent comment. It focuses on something I said the first day. There are several kinds of applications, several kinds of activities, that really do justify the existence of coding guides and its support, as opposed to having a program that says, "Yes, we will fund the maintenance of coding guides on an annual basis." These specific applications should be the appropriate focus. I wanted to ask you or anyone else to comment on the question of whether there is a logical agency that we can point at in any town that ought to be doing this, or whether, each city should be flexible and try to go wherever the money is or wherever the capability and interest is.

Dr. Robert Aangeenbrug—One of the recommendations that I will have to make is do not lean too heavily on higher education. I think in the Cincinnati example, for instance, part of the life that was breathed into the organization was in the hands of one single academi-

cian, a very capable one, I am sure, and a very successful one. The first priority in higher education when tenure, promotion, and salary are considered, does not necessarily stress your local involvement in the community. I would very strongly suggest that you beware. This does not mean that these institutions are not an invaluable resource. I think Ohio State, University of Washington, and a number of other places have indicated this. But I really feel that unless you can convince the local higher education administration to go into some kind of contractual agreement, you are in a dangerous position. In the case of many institutions of higher education, and I am speaking now as a university planner, currently on leave, as one's priorities change in higher education, and they will, local community operations support may be quite vulnerable. It is one of several tragedies, which you have reported on here that have occurred, but it is a difficult thing. On the other hand, I do not think there is a best individual agency as far as I can tell, and I have meddled or have been involved in at least five separate city activities.

In New York, I do not think anyone wants to be in charge. Even the Port of New York Authority with all of its money was not interested, it politically could not cope, I suppose. This, I think, may be a case of diseconomy of zeal in a large organization.

I like the idea of the Chamber of Commerce personally, because I think that they are one of the few private agencies that have a community interest at a very high level. I was somewhat intrigued to hear the United Way mentioned. I think we might cultivate that particular cooperative semiprivate-public corporation a little more. I wonder if anybody could react. I really do not think there is a good model.

Sometimes a strong individual public administrator will create the same problem that happened many years ago when the GEO-PLANS and related projects were developed with high hopes marketed by a strong individual, with a great personality, fantastic fund raising capacities, who then got job offers elsewhere, and the project went some other place. I simply want to make you aware of that because we are playing with the hopes and aspirations of citizens, and there are a lot of problems out there. They are not getting solved in some of these places, after the prophets have left, so I want you to have a rather juandiced view.

Mr. Harvey Albond—There is no doubt that we are in the individual or personality cult. These things are evolving at various levels due almost to the happy coincidence of some leadership, some personalities, some capability coming together and sitting down to say what is the ideal solution. Looking at our own case study, if you will, we have struggled with the New York State Department of Transportation to decentralize the transportation planning process down to the regional level, to have our own study staffs, in cooperation with the State, in order to not only participate in local decisionmaking, but perhaps as importantly, to have some capability of spinning off this data that is evolved for transportation purposes to do planning purposes as "differentiated" from it, and vice versa. It is a feedback system that makes the transportation better because it works with the planning, as such. Now, I would be delighted, frankly, with a limited staff and budget, to dump my process, or shall we say aggregate it upwards, into a regional transportation study process. I happen to be the chairman of the transportation study.

On the other hand, I have a problem. My problem is that the money is going to come from UMPTA and from Federal Highway, and so on and so forth, and I am going to have problems getting additional work of a local character back out of them. We are going to have to have a certain degree of duplication. Perhaps we can merge programs and formats, we can allocate our information upward. I think the answer in our case is that we encourage regional models, we encourage State models, we encourage national models; but on the other hand, we are still going to have to have local responsibilities if we want to have local convenience. I guess that we have to pay a price for that.

Mr. Alan Pisarski—I would like to shift the focus a little bit and talk about coordination between cities and how the word gets out. I guess I will start off with a general question: Do you feel that you are being served, that you are being sufficiently informed about what other cities are doing? Rather than having a group of people stand up and say "No," I would like you to get up and comment specifically on how that might be resolved, even more specifically, on what might the Federal government do. I am afraid that we might be at that stage in the discussions where we ought to be talking about what should the Federal government be doing? Is the word getting out? Do you feel that you are getting enough information about what is happening in the country? What a) should be done, and b) what should the Feds be doing about it?

Mr. Jack Pascoli—You alluded several times to the 15 percent sample of work trip information of the 1970 Census. I am wondering if with 10 years of improvements in hardware, 10 years of improvements in software, and supposedly complete GBF/DIME Files for the 1980 Census, if we can get the work trip information out closer to the Census year instead of 3 years later the next time around.

Mr. Alan Pisarski—It is really a good case study. We started designing that process, probably in 1967, I would say. I think the first conference at the National Academy of Science was in 1968, and the purposes were really twofold. The experience from the 1960 Census had been that the Urban Transportation Studies which actively wanted to use the data had ended up in very long queues at the Census Bureau waiting for their turn to do some specialized processing, and at tremendous cost. The idea was that by developing a generalized base set of tabulations, we would reduce the cost, we would reduce the queue time. This was not to say that these tabs would be everything that an individual city might choose to have, and the avenue of special requests at the Census was still open, but simply to speed up this first-round process. Assuming that we were successful, and that 3 years out is successful, I guess if we had gone the old route of the queue and the individual requests, I do not know when we would have gotten through the process.

We had, as I say, about 180 or 190 special tabulations run. The total cost now, I guess, is close to a $1 million—around $800,000 or $900,000 spent on the processing outside of the programing—an incredibly expensive operation. If we had gone the individual route it would have probably taken 5 or 6 years and I do not know how much money. It is not to say that it was well done. I think it was poorly done. I think we have learned a great deal and hopefully we will do better in 1980. I guess the Council on the 1980 Census is just being formed now. The focus will be on a better set of journey-to-work statistics, on better geographic coding, particularly the employment end.

Mr. Ronald Henderson—This question of how to get cities to communicate, cooperate, how to spread the GBF work, is a very interesting question. It seems to me that there are only two alternatives for the GBF/DIME. One is for the Federal government to require that the GBF/DIME be installed and provide subsequent funding. In a sense they are doing that—the new LEAA directive, HUD requirements in some cases, are a round-about way of requiring GBF/DIME development. There are a lot of things that are happening, but either the government must require GBF/DIME development, or we will have to go the slow route with trying to talk to local public and private organizations about the advantages and benefits of the file. It is going to take a lot longer that way. Every city is going to have different approaches. There will be a Chamber of Commerce in one city; it will be the city offices in another; the State may be interested in another location.

Being of data processing origin and having read Orwell, I often have conflicting opinions on the situation. In a way, to me it would be so much easier if everything were required, and if everyone would have a number, and you had a board with everyone wired so that administrators could watch the board and know who was going to the bathroom; but, on the other hand, we lose quite a bit if we adopt that approach.

I do not see any other solution, either it is required or you try to convince people. Obviously, you get more cooperation if you try to convince people of the benefits. I think in a sense, that the government is requiring it in a round-about way. Hopefully the criminal justice people, planners, city officials, and private industry will see the advantages without having it forced down their throats. I think that the approach that is being taken is probably the best one that there is. I think that it should be the responsibility, the goal of everyone, at least here at this conference, to spread the work, to work in their own local area to try to make people realize, so that with what the government is doing and with what private industry is doing, I think we are moving along.

Mr. Robert Gebhart—Possibly what I have been thinking of is somewhat simplistic, but what is the possibility of the government agency most interested in promoting the GBF/DIME putting together a monthly or quarterly newsletter which would attempt to collect and then disseminate uses and utilization of the GBF/DIME, actual cases of various people around the country using it. Over a period of time this would give a prospective user who wanted to sell the concept a good documentation of possible uses and examples of people who are actually doing something. It could also promote an interchange of information between users which could help alleviate the problem of reinventing.

Mr. Edward Kutevac—As a director of an agency who has been attempting to keep alive local involvement in the Geographic Base (DIME) File in our area, it certainly was welcome news that possible there would be some funding available to agencies such as ours. I say this not so much because of the amount of money available but to point out to some of the decisionmakers that somebody knows that you are alive and appreciate the benefits of local involvement. We, in our area, Trumbull-Mahoning County, do not want to reinvent the wheel. It has been the two county planning commissions that have been working within their existing local budgets to supply the various coders and technical personnel to update the base file. I want to commend the city of Columbus and thank them for the aids that they have promised to help sell our Chamber of Commerce and Community Improvements Corporation to the possibility of assisting the local county planning commissions in completing and taking advantage of others experiences in establishing a useable geographic base file.

Mr. Alan Pisarski—I think the last two comments with respect to the newsletter and the funding are very appropriate introductions to Mr. Meyer to comment in this discussion on those and on anything else that he would like to toss into the discussion.

Mr. Morton Meyer—We know that in the GBF system we have the greatest thing going since the invention of the wheel, and maybe since the discovery of fire. It is true that its adoption to date has been less rapid than we may have wished, but that has also been true of many other innovations. Electronic computers, for example, were strange and forbidding beasts when they first arrived on the scene and only the most innovative agencies or corporations recognized their potential and underestimated the problems of bringing them on line. But today, no matter how much you may bless your computer and its perverseness, it is impossible to get along without one.

What I am leading up to is that the demands for spatially-oriented data now being placed upon you—not only by Federal and State governments but, and of equal importance, by the need for local governments to operate more efficiently in an ever more costly environment—lead automatically to the development and maintenance of geographic systems as really the only tool through which the need can be effectively met. By way of further illustration, let me again quote Standard 4.8 of the National Advisory Commission on Criminal Justice Standards and Goals which Dr. Stevens referred to in his comments. It states that "where practical, police should establish a geographical coding system that allows addresses to be located on a coordinate system as a basis for collecting crime incidence statistics by beat, district, census tract, and by other 'zoning' systems such as schools, planning zones and ZIP codes." It further recommends that police should use the same GBF as other city agencies and describes the benefits of using the Census Bureau GBF/DIME (a copy of Geocoding Standard 4.8 appears in appendix 1 following the Question Period of the paper by Ms. Brayford and Messrs. Kelly and Pearson). As regards "spreading the word," we in the Federal system can and do use newsletters, we can and will set up a clearing house, we can and do promote "benefits and uses," but we are in a sense located on the wrong side of the tracks. That is, we are not part of the local user community. Communication directly between cities is, by far, the most effective way of disseminating GBF/DIME information. For examble, if Mayor Moody, the next time he speaks publicly, can talk about how Columbus is improving its operations through the use of the Geographic Base (DIME) File System, and the importance of the file to a more effective city management, he will do more to promulgate the use of the system locally than anything that we in the Federal Government can say because he is an elected official, and equally important, because he has the credibility that comes from working with and trying to solve problems that are common to all cities.

I would like to conclude my remarks by thanking our host city, Columbus, for its hospitality and again expressing our thanks to Mid-Ohio and Columbus Area Chamber of Commerce without whose yeoman work this conference could not have taken place. I also want to express our sincere appreciation to our distinquished and gracious chairman for his superb direction of the conference during the past 2 days.

Summary of Proceedings

ROBERT T. AANGEENBRUG

INTRODUCTION

There have been times at which my behavior at these meetings surprises some people because I usually have a hard time sleeping at these conferences, since they provide one of the rare opportunities to see real live users. Now I am a pusher, as Mr. Barabba knows, of this concept. One of the alleged great-great-grandfathers of DIME was a part-time professor of mine, Mr. William Garrison. I also waddled around, trying to solve New York's problem, and got out, although I got paid as a consultant. It is interesting, I went back last night, being unable to sleep, and since I am now a bureaucrat, could not behave like I did at previous conferences. I actually went to my room and studied. I went back through the proceedings of these various workshops, and found out some interesting things. I think they kind of reflect what is happening, and what has happened, and what is going to happen. First of all, the first one of these conferences was held and planned and schemed by Mr. Jacob Silver, who together with Mr. Pisarski, has a perfect attendance record throughout these conferences. That is, they have attended all of them, as best I can tell from the program. The interesting thing is that these are actually rather active conferences that deal with dissemination of a product that to some extent we still do not quite understand. GBF's are a generic concept, and have been around for a long time. They have never been so exciting in terms of their potential, and we have not had any hits scored until just a few years ago. This particular session is an excellent example of a "consciousness-raising" session. The subject is local use of the GBF/DIME System.

Geographic base files have been and will continue to be in existence for years. A street map with a index is a geographic base file, so are various commercial street directories. DIME is a conceptual model of such a file. More specifically, it is a geographic file structure based on topologic referencing of map features. By itself DIME is an encoding convention of enormous value in point, line, and area referencing. Its feature of boundary encoding is especially important for area analysis. The term GBF/DIME System implies an operative file system suitable for geographic referencing based on dual independent map encoding. It is our purpose here to see to it that this "system" is of *use* to local decisionmakers and hence to our citizens. Thus far implementation of such "systems" is limited. Local uses on a regular basis are rare. In other words, we are just getting started. These conferences have documented some of these starts.

A quick review from previous conferences indicates that the number and the composition of the attendees reflect not only the increasing awareness of local users to this opportunity, but also finally documented case studies. It is true that we have, except for the Cincinnati case, little documentation of failures.

The mix of attendees at this conference shows a greater proportion of local users with the "Feds" less prominent numerically. The dissemination process of this technology still indicates a limited use of these GBF/DIME Systems. The real theme of this conference is about local programs. It is interesting, we only had three new local programs reported, the others had already been reported at previous conferences although there were new update and extension activities. I will give you a brief review of previous conferences as I think there are three things I want to talk about in terms of needs, and these are needs for local and other users. I do not think we can really separate them. First, output—technological needs; second, policy needs; and third, communication needs, which really grow out of both of these.

SOME REFLECTIONS ON THE GBF/DIME CONFERENCES

Since I am currently at the Bureau I felt compelled to take my own census. I compiled a list of attendees at each conference and counted a total of 376. Surrounded as I am in Suitland by tabulators I felt a cross tabulation was in order (see table 1). Not only

TABLE 1. A Classification of Conference Participants

Place	Date	Participants					
		Total	Local	Federal	State	Research	Private
Wichita, Kans.	November 1970	35	11	15	3	5	1
Jacksonville, Fla.	April 1971	41	15	16	5	2	3
Arlington, Tex.	November 1971	52	19	17	7	5	4
Seattle, Wash.	January 1973	62	20	21	4	6	3
Boston, Mass.	April 1974	88	30	30	13	10	5
Columbus, Ohio	November 1974	98	48	18	10	14	8
	Totals	376	151	117	42	42	24

does this reflect the changing mix of attendees, but also the expansion of the number of registrants. Mr. Meyer informed me that many requests for attendance at the current session had to be denied. From a modest size conference of 35 at the beginning, at the current session we have nearly 100 attendees. Using the address list from each of the conference proceedings, I arbitrarily categorized the participants in five classes: 1) local users—those employed by a metropolitan, municipal, or other local government; 2) Federal users—those employed by Federal agencies; 3) State users—those employed by State agencies; 4) research users—those employed in higher education or by nonprofit research and development institutes; and 5) private sector users—those employed by private consulting companies.

Although the largest absolute and relative number of participants are local users, a look at the presentations (table 2) indicates thus far experiences from 26 metropolitan areas have been documented. A further inspection of the table of contents of these proceedings also shows presentations made about several statewide systems and a number of methodological papers by the Geography Division including some critical looks at current problems. I recommend you read these proceedings.

I think we had fewer than five women at the first four conferences. Representatives of minorities have been rather scarce, although at this meeting 11 women and not so token number of minority attendees are present here today. To me it is a reflection of where we are. One of the big problems we face in this area is the dissemination of technical skills to the nonmajority, and I think that is a reflection, somewhat personal on my part, that I find rather interesting. Maybe you will be able to do something about it because, let me tell you, if you complete an academic program that combines geography, data processing, and do not ever forget, at least a course in calculus and statistics, you could probably find a job some place no matter how hard times get economically. I think there's some interesting opportunity areas for affirmative action here.

TECHNOLOGICAL NEED IN LOCAL AREAS

The specific technical requirements of a GBF/DIME System are not widely understood by most data processing managers or city planners. Perhaps they should not be, but one may ask—Is there in each SMSA one key person who could conduct a 2-hour discussion of this system for, say, the planning and data processing staffs in various metropolitan agencies? Could this person identify the specific hardware, software, personnel, and monetary requirements? And, could he or she make a case before the budget officers that would convince them of the utility of this system? I would venture to say in only one-quarter (about 60) of the metropolitan areas could such a person be found to give such a speech, and in only one-eighth of the SMSA's might he or she convince someone to provide additional personnel for control of files, cartographic programs, and statistical analysis.

URBAN BASE MAPS

The base map problem is quite often overstated. People vent their spleen against the USGS (United States Geological Survey), Mr. Meyer, Mr. Fay, and whoever else they can find. It is often an irrelevant argument. If you understand the local scene, you can in some crude form develop a local map. If you cannot, you do not belong in your current job. You do not even understand your own environment, and you have no capacity to fail. (Or, to put it in other words, you don't want to assume responsibility to do it yourself.) If you have no capacity to fail you do not belong in the public sector, or for that matter, in the private sector either. I think the base map question is a serious one, with some serious academic and other problems related to it, but I think it is sometimes overstated. It is on the other hand, also understated. If you have a file that has technical requirements for say, land record-keeping, the accuracy of the file becomes important. A case, that I can remember in southern California required large scale accuracy for say, water systems, fire hydrants, supporting the public safety system. You will have to kind of define the accuracy needs for each problem, some of these scale requirements are very detailed. In this southern California city it was discovered that its map had not been corrected after the earthquake. They had some fairly serious problems with this. I would hope that we can address ourselves at some technical conference not too long from now, to this question of accurate base maps. It is also a national problem, but I do not think it's fair, nor proper, to expect the Bureau of the Census to solve this problem by itself. Perhaps at a later time, we will be allowed to report on some of the Federal cooperative efforts that the Bureau has been very much involved in that address this area also from a policy point of view.

There are some other problems. However, many of these technical questions can be solved. I think as Mr. Pierce has shown in Albuquerque some time ago, if you

TABLE 2. Metropolitan Areas Documented

Conference site	Date	Metropolitan areas documented
Wichita, Kans.	1970	Wichita, Kans. Washington, D.C. Orange County, Calif. Atlanta, Ga. Kansas City, Mo. (see also Arlington, 1971)
Jacksonville, Fla.	1971	Jacksonville, Fla. Charlotte, N. C. Dallas, Tex. (see also Arlington, 1971) Minneapolis-St. Paul, Minn. (see also Boston, 1974) Tulsa, Okla. (see also Columbus, 1974)
Arlington, Tex.	1971	Cincinnati, Ohio Pittsburgh, Pa. Wichita Falls, Tex. New York, N. Y. Santa Clara County (see also Wichita, 1970) Kansas City, Mo. Dallas, Tex. (see also Jacksonville, 1971)
Seattle, Wash.	1973	Seattle, Wash. St. Louis, Mo. Albuquerque, N.Mex. Miami, Fla.
Boston, Mass.	1974	Las Vegas, Nev. Columbus, Ohio New Orleans, La. Cleveland, Ohio Minneapolis-St. Paul, Minn. (see also Jacksonville, 1971)
Columbus, Ohio	1974	Columbus, Ohio (see also Boston, 1974) Fort Wayne, Ind. Indianapolis, Ind. Niagara Falls, N. Y. Tulsa, Okla. (see also Jacksonville, 1971)

want it bad enough, you can get rather good and accurate maps by the use of current technology, and you do not need remote sensing necessarily. Orthophotoquads and current technology will allow you to do it. All you need is the will and the money which can be raised by creating broadbased State and local support.

For each metropolitan area one agency should be able to provide a single base map series which can support the local and Bureau needs for the GBF/DIME System. Whether this is the case or not is a political rather than a technical problem. The problem of keeping a map current is also a management rather than a technical problem.

SOFTWARE

One of the other problem areas is that of software, and it is a serious one. Again, I find it rather surprising that at this particular conference, not too much public discussion about the software took place. I think part of this is also based on your own professionalism. With this kind of a mixed group you are somewhat hesitant to publicly address possibly the wrong audience. I did find a whole lot of informal communication going on about software problems. I would certainly urge that special sessions on specific software problems be held at future meetings.

Just looking back through some of the documentation, we are still talking about some software being developed that needs 256K bytes to run. We are talking about large programs. Some of these are hard to maintain in the current environment. Yes, they are technologically possible, but they are not always locally practical. I think we must try at some future time, and I hope very soon, to address this question. Because if the Bureau makes a decision to develop a BPR 'type package (developed by the Highway Administration for transportation planning and analysis) which is an excellent example of a highly transferable complex system; the Geography Division, which would certainly be involved, would very much need your input. The Census Bureau needs very much to structure such a program based on its own as well as local, State, and Federal agency requirements. Aside from the question of making such a package IBM compatible, there are some other questions. For example, what type of an operational environment can be assumed or, should be supported.

By 1980 we do have some serious problems in terms of what kind of languages will be used. I know from a practical point of view, we will probably stick with COBOL and FORTRAN, but I am not really sure it is wise. After all, we are now talking about manipulating potentially large files with extremely complex geometry and geography. I am not sure it is wise for us to think about general-purpose COBOL programs. On the other hand, I do not know what the solution is. As Mr. Meyer says, and he is quite proper and I do agree with him, the basic requirements of Census geography are not necessarily those of your local environment. There is no way in which the census can serve all with endless capacity. I do not think that is efficient. I do not think that it is proper.

At the risk of preaching I would like to list some needs of local users which should be addressed at the next conference and hopefully documented in reports:

1. Document uses of each software module in the GBF/DIME System & CUE, for each SMSA; identify each agency, personnel, and hardware environment; include estimates of costs.

2. Develop software for use of GBF/DIME in nonmail (nonstreet address) areas.

3. Provide a test module for interactive graphic editing (small, documented, and transferable).

4. Provide additional special-purpose packages (i.e. CRAM, CARPOL) for intersection analysis, point-in-polygon search, and selected polygon equivalency boundary files.

5. Urge the Census Bureau to hold a technical conference for advanced users on, for example, address matching problems, computer graphics, small-area statistical mapping, and utilization of imperfect and GBF/DIME-Files.

QUALITY CONTROL QUESTIONS

I have not yet seen sufficient evaluation of quality control reported about some of these files. I think some of us are beginning to get nervous about 1980, or should I say 1979, or 1978, when the final decisions are made. My current information shows we will have the same things we had in 1970—"You like it? You used it?" There are some rather serious problems as far as I am concerned. It would be useful to know if and when quality control evaluation will be done and what the results of the Bureau's research and development efforts and plans are. I think that the people in the front line, particularly the Geography Division, will have an extremely difficult time answering all those questions unless they stand in the larger body politic of Federal programs and the Bureau's overall programs. Several specific questions highlight my concern:

1. How do we define metro-map accuracy?

2. What are the standards of quality for street-address files?

3. What is an acceptable GBF/DIME-File? What are the objective criteria?

4. What is a match (as in match rate in GBF/DIME System or UNIMATCH)?

5. What is an acceptable match rate per block, per tract, per street?

URBAN CARTOGRAPHY AND SMALL-AREA ANALYSIS

There is another rather vexing problem that has been glossed over thus far, and that's the question of cartographic process; the danger of a new whiz-bang approach. That reminds me of GEO-PLANS and the slides we had of the three-dimensional map of the population of Wichita. In color yet, rotating population densities through the computer. It was very exciting. We did not know what it meant, necessarily, but it looked pretty and people thought it was great. It was the only map that we could show to the governor and the mayor, and God knows who else, but its meaning was not very well understood. With the advent of improved communication we can tell messages that may not be accurate and cannot always be proper. I have some concerns about, for instance, what I referred to yesterday—the problem of ecological correlation. It is not new, but it has not recently been addressed. I have not seen any recent contract research that addresses this issue. We could indeed portray in map form some enormous misunderstandings among our policymakers. I think if you go back to Firey's work in Boston a long time ago, one becomes aware of the danger of making the wrong decisions based on that kind of communication.

We know very little about small-area data analysis. At the Bureau, sensitivity to confidentiality and a concern for statistical standards are well documented, but little or no documentation about small-area data problems is available. For example, the variance of a sample increases with a decreasing sample size. This implies that at the tract and especially at the block level our statistical estimates are subject to a large relative variance. Although there is a great demand for transferring current population estimation methodology to small-area forecasts, one can easily appreciate the difficulties of such estimates. In addition, the degrees of freedom available for using regression techniques further limits forecasts for very small areas. But with recent legislation the demand for detailed area data, including forecasts, grows.

I have asked recently in the Bureau who uses small-area data and what do they need, and to be perfectly honest with you, I could not get a complete answer. And answers are needed to plan for the 1980 Census. These are some real tough questions, only some of which are statistical, and I think the Bureau is probably one of the places I know that can possibly answer such questions. It is, by the way, an important issue particularly with rapid communication, including, say, cable TV.

In even a small city like Lawrence, Kans., where I am very much involved, we have local communications through the cable TV, and at such meetings we can show results of some of our statistical analysis of small-area data. To be perfectly frank, we have not always examined such presentations carefully. I would urge the Bureau and other Federal agencies to examine these questions. Now, do not think the Bureau is not sensitive to the problems of small-area data. I am simply saying I would want you to be involved in part of this dialogue also. As a matter of fact, I think the Bureau is probably the one Federal agency that is most sensitive about data use, with the exception of the journey-to-work information where I think DOT (Dept. of Transportation) has historically led.

POLICY CONSIDERATIONS

We are all here supposedly to improve the citizens' lot, and in so doing, keep the cost of local government down, not necessarily decrease it—I am not naive, that may not be possible—but keep it from growing faster than it is. We need proofs of utility, we need more mayors, elected officials, and senior level managers at these conferences. We need to prove that we have *solved* some problems. As one of my favorite city managers said, "I will never take this to the council unless I can tell them that I have saved them some money that they would not have saved otherwise." Armed with an improved taxation and evaluation in the assessment file created for his city, he convinced the council that they would be collecting more taxes, more equitably—it sold. I think these are the kinds of proof we need. Case studies could document some of these experiences. I think we need to encourage public administrators outside of the Bureau to be prepared to present results of problems solved through GBF/DIME System technology to elected officials and to the voters.

The ownership of the files also raises some very important questions. One which we discussed and discovered in the USAC project cities is the right of public inquiry. Please do not whip up too much emotion about the greatest municipal information systems. I have been there. The citizens may expect you to deliver. The other problem is, are you prepared as a manager to give them the information they desire? Not only are there legal questions but other political questions. For example, would you want to slow down a city government's computer by allowing time-sharing inquiry of the status of all citizens' municipal utility bills, or all the land use records, or whatever else a citizen would like to know? These are rather serious problems. I hope that in any systems design work you do, you involve where you can, people who have policy experience and responsibility. I would urge you to particularly read up on this if you can. I think **Public Administration Review** and a few other journals, for instance, may be able to provide you with this. Another aspect of file ownership is, of course, the question of regional versus local and the question of consolidation of public information systems.

An important dimension of new public technology deals with the "credits" of its results. Usually the chief excutive of any corporation owns anything he or she likes; I think that is only proper. When you are using this new technology you get into the serious question of having an extremely powerful tool which could lead to an interesting tug of war as to who is going to take the credit for it. Now, I am not telling you very much new, but I do think and I want, in this conciousness-raising session, to leave you with the impression that you should try to push the credit as far down and up as you possibly can. I find that the most impressive managers I have seen are those who have their staff take a lot of the credit, very far down in the organization and inform their superiors. I think if you talk about regional cooperative models, having somebody else take the credit is an excellent strategy.

Another real serious policy problem is the lack of adequate theory and models. We may be creating some kind of a monster. We do not really understand how local economies work. Nor do we understand some of our medical and other types of problems, including welfare and related phenomena. We do need some more theory. It is not really your business, per se, to deliver that. Most of you are from operating environments where you do not have time to deal with these, but decisionmakers need to have these kinds of models. They have them or invent them, you know. They may be such models as "All those folks over there always try to give me trouble, but they do not have much clout, so I'm able to ignore them." This is a model. I think we must as professionals see to it that the academes deliver us understandable and better models. It will be much better than getting them involved in taking over GBF/DIME-File maintenance relationships within urban areas. It is not really their responsibility, as far as I am concerned.

COMMUNICATION OF GBF/DIME TECHNOLOGY

Obviously this kind of conference provides excellent opportunities for face-to-face and formal communication. But what happens tomorrow? I would like to briefly touch on some opportunities available to you. In terms of local-to-local communication, URISA (The Urban and Regional Information Systems Association) is an organization that has a SIG-GEO organization, that means a special interest group devoted to geographic base files. For $2 per year you can join and receive a newsletter. Mr. Richard Schweitzer of the Geography Division is chairman of this group. You can write to him at the Bureau.

I was afraid I was going to have an argument with some of the Federal bureaucrats who were going to develop some stupendous proposals for more Federal dollars to provide communication among the locals. I really like what I heard from Mr. Meyer. Such communication must probably be informal and you the user should control it. It's not that expensive to do. I think Mr. Barb and Mr. Cooke, who preceded Mr. Schweitzer in this SIG, kind of invented the use of NTIS (National Technical Information Service) for disseminating proceedings of the annual URISA Conference.

I would urge you to attend some of the annual meetings or special workshops of URISA. Additional geographic interest groups are found in ASPO (American Society of Planning Officials) and AIP (American Institute of Planners).

Another way of expanding the utility of the GBF/DIME System technology is to attend the meetings and/or workshops sponsored by the Census Bureau. You might also consider pushing the technology by preparing your supervisors to read papers at their professional societies. So what if they steal your thunder. If you provide the script they are bound to be interested in supporting the GBF/DIME System technology. At the international level there is a great deal of interest in this technology. The URISA newsletter and the International Geographical Union can provide you information about your foreign colleagues' experiences.

Communication is a problem, and local communication is no exception. I will not belabor it because most of you probably understand it better than I do, but it certainly involves problems of cooperation, coordination, and again the cult of personality. Learn how to eat humble pie, don't get in the way and take too much credit. Get your subordinates or your superiors to take the credit for what you have done. You will enjoy it. I think it is one of the opportunities we have missed in some of the experiences I have watched.

In parting I have one more suggestion. The geographic base file technology is ultimately a local tool. It is a new tool, somewhat complex, but an ideal index to analyzing and monitoring local and regional phenomena. Once the standards for geographic information systems are set, the dependence on local update and maintenance will produce a "live map" for each major metropolitan area. At future conferences we, the "Feds," will be learning more from you.