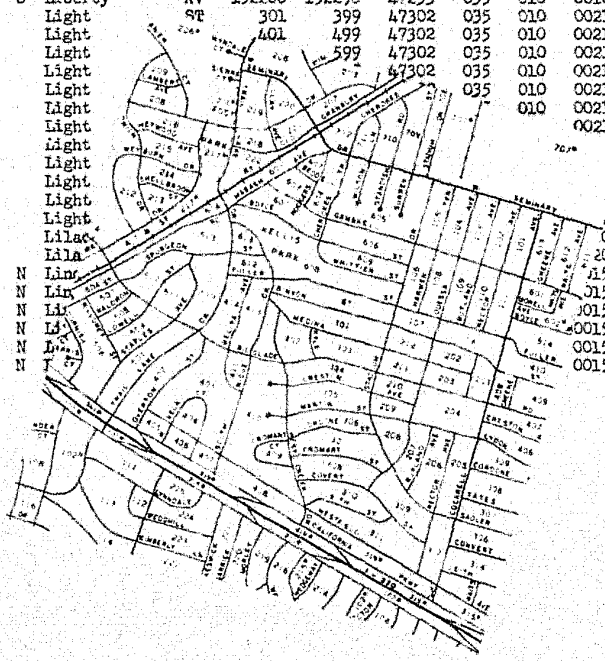
Geographic Base (DIME) File System: A Forward Look

Conference proceedings

April 16 and 17, 1974
Boston, Mass.

STREET DIR	STREET NAME	STREET TYPE	ADDRESS RANGE LOW	ADDRESS RANGE HIGH	ZIP CODE	COUNTY CODE	AREA CODE	TRACT NO.	BLOCK NO.
S	Liberty	AV	152101	152199	47233	035	010	001600	318
S	Liberty	AV	152201	152299	47233	035	010	001600	319
S	Liberty	AV	152000	152098	47233	035	010	001600	320
S	Liberty	AV	152100	152198	47233	035	010	001600	223
S	Liberty	AV	152200	152298	47233	035	010	001600	218
	Light	ST	301	399	47302	035	010	002101	124
	Light		401	499	47302	035	010	002101	113
	Light			599	47302	035	010	002101	112
	Light				47302	035	010	002101	111
	Light					035	010	002101	101
	Light						010	002101	201
	Light							002101	210
	Light								211
	Light							1	220
	Light							1	221
	Light							01	222
	Light							101	223
	Lilac							000	402
	Lila							2000	402
N	Lin							1500	304
N	Lin							01500	305
N	Li							001500	306
N	Li							001500	213
N	L							001500	212
N	L							001500	211



mputerized Geographic Coding
Series GE 60 No. 5

U.S. DEPARTMENT OF COMMERCE
Social and Economic Statistics Administration
BUREAU OF THE CENSUS

Geographic Base (DIME) File System— A Forward Look

Computerized Geographic Coding
Series GE 60 No. 5

Conference proceedings,

April 16 and 17, 1974,
Boston, Mass.

Issued October 1974

U. S. DEPARTMENT OF COMMERCE
Frederick B. Dent, Secretary

Social and Economic Statistics Administration
Edward D. Failor, Administrator

BUREAU OF THE CENSUS
Vincent P. Barabba, Director
Robert L. Hagan, Deputy 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 Gerald J. Post, then Assistant Division Chief for Operations, Geography Division. Ann Liddle and Jacob Silver were instrumental in organizing the conference and in compiling and editing these proceedings.

The Census Bureau is extremely grateful to the Bureau of Transportation Planning and Development of the Massachusetts Department of Public Works, the Transportation Systems Center of the U.S. Department of Transportation, and the Urban Systems Laboratory of the Massachusetts Institute of Technology for cosponsoring the conference and for their fine cooperation. We wish to give special thanks to Walter Kondo and Terry Anne Vigil of the Massachusetts Department of Public Works, to Henry Bruck of the Massachusetts Institute of Technology and to Robert Tap of the Transportation Systems Center 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 $2.00

Preface

This report presents the proceedings of the fifth in a series of conferences devoted to the Census Bureau's Geographic Base (DIME) Files. This conference was held in Boston, Mass., on April 16-17, 1974. The papers presented and the resulting discussions centered on the theme of the "Geographic Base (DIME) File Systems—A Forward Look."

The purposes 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. The question and answer sessions have been edited for conciseness of presentation.

Copies of the first four conferences 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. 1972 ($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)

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

James Andrade
Statewide Planning Program
Providence, Rhode Island

Thomas Atkins
Secretary
Executive Office of Communities and Development
Commonwealth of Massachusetts
Boston, Massachusetts

James Atkinson
Metropolitan Area Planning Council
Boston, Massachusetts

Howard Ball
Director
Office of Policy Planning
U.S. Department of Housing and Urban Development
Washington, D.C.

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

Constantine Ben
Office of Highway Planning
Federal Highway Administration
Washington, D.C.

Gerard Boetje
Charles Stark Draper Laboratory
Cambridge, Massachusetts

George Boyle
Planner
Lower Pioneer Valley Regional Planning Commission
West Springfield, Massachusetts

William G. Bozicas
Regional Field Specialist
Defense Civil Preparedness Agency
Maynard, Massachusetts

Susan A. Brayford
Research Support Specialist
International Association of Chiefs of Police
Gaithersburg, Maryland

Henry W. Bruck
Associate Director
Urban Systems Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

James Burwell
Central Midlands Regional Planning Council
Columbia, South Carolina

E. F. Carlberg
Geosystems Manager
Boeing Computer Services, Inc.
Seattle, Washington

Thomas Conboy
Statewide Planning Program
Providence, Rhode Island

James T. Controvich
Senior Planner
Springfield Planning Department
Springfield, Massachusetts

Charles Cook
Transportation Planner
Berkshire County Regional Planning Commission
Pittsfield, Massachusetts

Donald Cooke
Research Affiliate
Laboratory of Architecture and Planning
Massachusetts Institute of Technology
Cambridge, Massachusetts

LeRoy D. Dauterive
Senior Planner
Regional Planning Commission for Jefferson, Orleans,
St. Bernard and St. Tammany Parishes
New Orleans, Louisiana

Chris Davis
Transportation Information Group
Charles Stark Draper Laboratory
Cambridge, Massachusetts

Donald DeVorkin
Manager
Transportation Information Group
Charles Stark Draper Laboratory
Cambridge, Massachusetts

Jack Doolittle
Chairman
Massachusetts Bay Transit Authority
Boston, Massachusetts

Arthur Dukakis
Director
Boston Data Collection Center
Bureau of the Census
Boston, Massachusetts

Albert C. Etienne
Data Systems Section
Department of Police
New Orleans, Louisiana

Marilyn Fine
Program Officer
Office of Data Systems and Statistics
Department of Housing and Urban Development
Washington, D.C.

Roland J. Frappier
Principal Planner
Rhode Island Statewide Planning Program
Providence, Rhode Island

William Grippo
Planner
Montachusetts Regional Planning Council
Fitchburg, Massachusetts

Louis Hallisey
Bureau of Transportation Planning and Development
Massachusetts Department of Public Works
Boston, Massachusetts

Irwin P. Halpern
Deputy Assistant Secretary for Policy, Plans and
International Affairs
U.S. Department of Transportation
Washington, D.C.

John Halterman
Systems Programmer
Geography Division
Bureau of the Census
Washington, D.C.

William G. Harward
Bureau of Transportation Planning and Development
Massachusetts Department of Public Works
Boston, Massachusetts

David Hathaway
Principal Planner
Pittsfield City Planning Board
Pittsfield, Massachusetts

Joyce A. Healy
Assistant Secretary
Manufacturers Hanover Trust Company
New York, New York

Edward F. R. Hearle
Senior Vice President
Booz, Allen and Hamilton, Inc.
Washington, D.C.

Wayne Hill
Old Colony Planning Council
Brockton, Massachusetts

Robert Hinckley
Manager
Computer Operations
Transportation Systems Center
U.S. Department of Transportation
Cambridge, Massachusetts

Thomas Humphrey
Director
Bureau of Transportation Planning and Development
Massachusetts Department of Public Works
Boston, Massachusetts

Ruth A. Hunter
Senior Specialist/Group Leader
Kentron Hawaii LTD
Cambridge, Massachusetts

Michael J. Keidel
Chief
Planning Research Division
Southeastern Regional Planning Commission
Waukesha, Wisconsin

James A. F. Kelly
Assistant Director
International Association of Chiefs of Police
Gaithersburg, Maryland

S. Jeffrey Koch
Geographic Planning Specialist
Geography Division
Bureau of the Census
Washington, D.C.

Walter Kondo
Bureau of Transportation Planning and Development
Massachusetts Department of Public Works
Boston, Massachusetts

Arthur H. Kunz
Planning Coordinator
Nassau-Suffolk Regional Planning Board
Hauppauge, Long Island, New York

James M. Langley
Engineer Programmer
Department of Public Works and Highways
Concord, New Hampshire

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

Ellis Levin
Industry Development Manager
Headquarters West
Boeing Computer Services, Inc.
Seattle, Washington

Daniel B. Levine
Associate Director for Demographic Fields
Bureau of the Census
Washington, D.C.

Ann Liddle
Geographic Planning Specialist
Geography Division
Bureau of the Census
Washington, D.C.

Joseph Loria
City Planning Department
New Bedford, Massachusetts

Anthony Ma
Planning Director
Northeast Ohio Areawide Coordinating Agency
Cleveland, Ohio

Thomas E. Madigan
Transportation Planner
Central Massachusetts Regional Planning Commission
Worcester, Massachusetts

Alfred J. Marston
Economist
The Port Authority of New York and New Jersey
New York, New York

Robert Meader
Senior Planner
Fall River Planning Department
Fall River, Massachusetts

Walter Messcher
Computer Operations
Transportation Systems Center
U.S. Department of Transportation
Cambridge, Massachusetts

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

Raymond Molski
Principal Planner
Morris County Planning Board
Morristown, New Jersey

James Murdock
Traffic Records Specialist
New York State Department of Motor Vehicles
Albany, New York

John Musselman
National Civil Defense Computer Facility
Olney, Maryland

Thomas O'Brien
Director
Office of State Planning and Management
Boston, Massachusetts

William G. O'Hare
Assistant Planner
Northern Middlesex Area Commission
Lowell, Massachusetts

Tilghman O'Neill
Program Analyst
Defense Civil Preparedness Agency
Washington, D.C.

Maureen Osolnik
Special Assistant for Federal Regional Council
U.S. Department of Health, Education and Welfare
Boston, Massachusetts

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

Gerald J. Post
Assistant Chief for Operations
Geography Division
Bureau of the Census
Washington, D.C.

Jon Rickert
Urbanologist
Data User Services Office
Bureau of the Census
Washington, D.C.

Elias Samaan
Executive Director
Three Rivers Coordinating Council
Fort Wayne, Indiana

Alvin J. Sanders
Associate Planner
Office of State Planning and Management
Commonwealth of Massachusetts
Boston, Massachusetts

Jack Schofield
Coordinator of School Planning and Zoning
Clark County School District
Las Vegas, Nevada

Elizabeth Schumacker
Lecturer
Department of Civil Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts

Richard Schweitzer
Chief
Geographic Statistical Areas Branch
Geography Division
Bureau of the Census
Washington, D.C.

Barry Sherman
Senior Planner, Transportation
Southeastern Regional Planning and Economic Development District
Taunton, Massachusetts

Jacob Silver
Chief
Program Development Branch
Geography Division
Bureau of the Census
Washington, D.C.

Roger Smith
Delaware Valley Regional Planning Commission
Philadelphia, Pennsylvania

Raymond Soller
Senior Quantitative Analyst
City Planning Commission
New York, New York

Paul R. Squires
Associate Director for Data Collection and Processing
Bureau of the Census
Washington, D.C.

Maureen Strickler
Special Assistant for Federal Regional Council
Law Enforcement Assistance Administration
Boston, Massachusetts

Dennis Strohbusch
Planner I
Three Rivers Coordinating Council
Fort Wayne, Indiana

G. Paul Sylvestre
National Criminal Justice Information and Statistics Service
Law Enforcement Assistance Administration
Washington, D.C.

Robert N. Tap
Acting Chief
Information Division
Systems Research and Analysis Directorate
Transportation Systems Center
U.S. Department of Transportation
Cambridge, Massachusetts

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

S. Thyagarajan
Executive Director
Capital District Regional Planning Commission
Albany, New York

James W. Turbitt
Associate Director for Economic Fields
Bureau of the Census
Washington, D.C.

Terry Anne Vigil
Regional Planner
Bureau of Transportation Planning and Development
Massachusetts Department of Public Works
Boston, Massachusetts

Robert B. Voight
Chief
Data User Services Office
Bureau of the Census
Washington, D.C.

Regis Voye
Programmer/Analyst
Metropolitan Council of the Twin Cities Area
St. Paul, Minnesota

F. M. Weaver
Systems Specialist
Systems Development Corporation
Charlotte, North Carolina

Pamela Werner
Research Engineer
Urban Systems Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts

F. E. Westerfeld
Systems Analysis and Data Management
Regional Planning Council
Baltimore, Maryland

Paul T. Zeisset
Statistician
Data User Services Office
Bureau of the Census
Washington, D.C.

Agenda

April 16, 1974

MORNING SESSION

AFTERNOON SESSION

April 17, 1974

MORNING SESSION

April 17, 1974

AFTERNOON SESSION

Chairman's Introduction

EDWARD F. R. HEARLE

Data processing technology is almost trite to talk about, but there are at least three kinds that come together around the subject of our conference this morning: Computer technology, remote sensing technology, and cartography. Computing is the manipulation of growing volumes of data; remote sensing is the art of taking pictures of the world from various places above it, and the conversion of those images into data that can be manipulated by the computer. And cartography and computer graphics are what produce displays of data, which is one of the principal bases for the growth of geographic base files. So the technology is moving and giving us growing capability.

Additionally, the analytical interest of both managers and analysts is growing. In the last two or three decades, perhaps even before that, the principal dimension of analysis has been time-oriented. We compare this period of time with some prior period; we forecast this period compared with others. It is interesting to speculate where we may go on the other dimension of analysis, the spatial dimension, now that we have this growing capability to compare a to b to c, when a, b, and c are different chunks of geography defined and described in all sorts of ways through the kind of techniques the geographic base files make possible.

It is clear that there is greater interest in different arrangements of geographic data; for instance, the notion of the central city versus the suburbs which arose in the late 50's and through the 60's as compared to the change in the past half dozen years showing growing concern about parts of cities and neighborhoods and all sorts of combinations of urban and rural areas. This analytical interest involves more spatial comparision which is basically what the geographic systems facilitate. It is also interesting that programs with big dollars, like revenue sharing, are now explicitly connected to small geographic areas, and therefore a new technology has to be implemented in order to carry out the explicit will of the Congress in the allocation of vast amounts of public funds where the cost of error is both very high and very visible. So the importance of precision with respect to spatial definition is growing. Just in the last month two papers that illustrate the growing importance of geographic base systems crossed my desk; one a manual on "Point Referencing of Properties and Parcels of Land" from the Department of the Environment of Great Britain, the other, a new book called "Land Parcel Identifiers for Information Systems" coming out of a conference held under the auspices of the American Bar Foundation. So the field is in ferment and so this conference is at the forefront of the ferment.

Both sides of the supply-demand equation are moving fast. There is a greater demand for this kind of data and there is a greater technological capability through geographic base files to provide it. So the central question is "So what?" We have this wonderful technology. We have this ill-defined but intense interest. And that is really the issue of this conference, "A Forward Look." What are we going to do as a result of having this growing capability?

Opening Remarks

JACK DOOLITTLE and IRWIN P. HALPERN

JACK DOOLITTLE

This is the part of the program that, when I am in a convention, I usually take a look at the night before and say "Oh, the hell with it, I can miss that and get an extra half hour sleep." I get the opportunity this morning for those of you who did not take that opportunity to welcome you on behalf of Governor Sargent and apologize for the fact that Mr. Alan Altshuler is not here. Mr. Altshuler is in Washington today trying to collect on some of the promises of the 1973 Federal Highway Act which will give us here in Massachusetts a major opportunity to put meat on the skeleton of Transportation Policy which we have been building for the last 3 to 4 years. We have an application prepared now and our paper work is pretty much done on moving some $600 million worth of interstate highway funds in transit construction.

The importance of data and the importance of the census and the data-gathering processing to regional planning is something which you all are much more aware of than I. There are a couple of aspects of it which are of interest to me. When I was in Minnesota they had a slogan there for their regional planners. As you know, they have one of the better efforts of regional planning. They had a slogan there which was the title of their regional plan. It was, "Four Million by Two Thousand," expecting (based on the 1960'ish population change) that Minneapolis-St. Paul and the metropolitan area would reach 4 million by the year 2000. Their land-use planning efforts were geared to accommodate that kind of growth.

Just last Thursday, or so, I got the most recent monthly publication from the Metropolitan Council in which, based on the 1970 vintage data and massagings thereof, they have changed the slogan from "Four Million by Two Thousand" to "2.9 million by the year 2000." I think you can see here the potential impact of this data to land-use planning and to transportation is a very significant one indeed.

When I was called up from the bull pen to make this appearance last Friday, I wondered why they chose me to do this particular job. When I saw the title "DIME" on the materials that were sent to my home Saturday morning, I decided they chose me to come this morning because the MBTA has a thing called DIME TIME, and while you are here, I would recommend that you take advantage of it. There is a very nice station across the street, and between the hours of 10:00 AM and 2:00 PM you can get on our rapid system for 10 cents.

I found later in reading the material from the brochure that there wasn't a reference to DIME TIME. I was a little disappointed. I thought that our convention people had been doing a better job than I had anticipated that they would. One of the things that also arrived in my mail Saturday morning was Mr. E. J. Kahn's book, "The American People," which I read with some additional interest in view of this morning's meeting. I found that among other things, there are 9,053 regional urban planners in the United States according to Mr. Kahn's interpretation of the data. Something like a 7-to-1 ratio male to female among that group. Thereby, normatively, regional planning is somewhat ahead of some of the other professions.

One of the things that Mr. Khan says in looking at the data and interpreting the information from the 1970 census is that as he peers into the future he can assure with some self-confidence—he can assure himself at least—that there will be jobs for urban regional planners for many years to come, based on the turmoil not only that census taking is now in, but also the turmoil the basic population is in where one out of every five people moves every year—with the national leadership in Washington setting the pace.

Some of the interesting things which came out of that effort of reading Mr. Kahn the other day includes interestingly enough, the fact that there are as many people in this country who have never been to school as there are who live in California. This is a rather startling little piece of trivia that you can take back home and tell your bosses that you heard at the conference in Boston and, thereby, justify your expense accounts.

One of the other things that I learned with some lack of surprise was that three out of five Americans live in urbanized areas. Of course, that is one of the principal areas of concern for us here in Massachusetts. We are relatively an urbanized State. We have something like 12 Regional Planning Areas, with 10, 11, or 12 urbanized areas. They are a rather strange hodgepodge in some places. Our cities range in population from metropolitan areas like Boston's 2.5 million to very much smaller urbanized areas.

One of our cities here in Massachusetts was reported in the 1970 census to have a population of 2,100 and that happens to be Provincetown. In 1970 it had a population of 2,100; in 1972 it had a registered voter list of 3,400. Now there is something either wrong with the way we count our voters or the way the people were counted in Provincetown or perhaps the myth of Massachusetts

politics in graves registration extends out as far as that on the end of the Cape.

Tom Humphrey is going to speak to you later on and will be much more technical and proficient on the subject that you are gathered here to discuss than I. I would just like to give you about a 5-minute break from my allocated time. One of the things we in transit try to be is on schedule, so I will just close by thanking you all for coming to Boston. I hope that you all will enjoy your stay here. There are lots of things to do other than sit at these tables; although I don't want to incur the wrath of your Chairman, I would encourage you to get around the city a little bit, put some money in our fare boxes. We can use the money. Last year's deficit was $110 million and we might go to $130 million this year; but with good cooperation from you perhaps we can reduce that by a minor amount. Thank you very much and welcome.

IRWIN P. HALPERN

It is a pleasure indeed to be here and to represent my department which is one of the cosponsors of this important conference. A good set of opening remarks should have a beginning and an end and above all they should be very close together. So mine will be just that. I am more or less serving as chaplain here to provide some inspiring remarks. I do not have any substantive comments. These will be delivered by colleagues subsequently. I would like to emphasize the strong and long and continuing interest of the Department of Transportation in this project, the geocoding activity. I think that we have been in it from the beginning. The present Director of our Information Policy Program, my colleague Mr. Alan Pisarski, has been one of the guiding lights.

One of the problems that besets any bureaucrat such as myself is the large bill for collecting data. I keep emphasizing to Mr. Pisarski that if something is not worth doing, it is not worth doing well. How much data do we really want to buy? What kind of new approaches to collecting and using data do we want? I think that in this particular instance there is ample justification.

We are making very good use of the geocoding system already. The most recent use that I think is paying dividends is in our car pool locator program; specifically the computerized matching car pool programs that had been disseminated around the country, helping the local jurisdictions to get people to share rides to work. Thus, the geocoding project is paying dividends in energy conservation, which is also one of our responsibilities.

Some of you, if not many of you were probably at our Watergate meeting of 2 years ago. That lead to very specific decisions as to what should be done next in the geocoding area. I think that we can point to some accomplishments in the form of converter files that deal with the problem of lack of comparability or combatibility between the numerous geocoding systems being used in Washington and that provide a common reference base for pulling these together, using the county as the basic reference point. We also have developed a capability to do computer mapping of the county coordinates. These are accomplishments that will be made available to all of you. The products are sitting on the table in printed form in the back of the room and some of our colleagues will be talking about them at greater length, so I will not be saying anything more about it. We are now moving on to the next, more complex level of subcounty locator capability and again it will be enhancing our capability to integrate the disparate kinds of coding systems that are being used.

I am very much encouraged by the wide representation of users here at this conference and the fact that not only our department but other agencies and groups see considerable utility in this project, and, finally, our working together in the way the Federal Government, State and local governments should be working together to get the great imponderables dealt with properly. If this conference is as productive as the Watergate one of 2 years ago, it should be a very fruitful conference. So my best wishes to all of you for a highly productive and enjoyable meeting.

Present Status and Future Prospects of the Census Bureau's GBF/DIME CUE Program

MORTON A. MEYER AND JACOB SILVER

(Presented by Morton Meyer)

On behalf of the Census Bureau and the Geography Division, I would like to express our thanks to Mr. Thomas Humphrey of the Bureau of Transportation Planning and Development, State of Massachusetts; Mr. Robert Tap of the Transportation Systems Center; and Mr. Henry Bruck of the Urban Systems Laboratory; our cosponsors, without whose help and assistance this conference could not have taken place. I would also like to give special thanks to Mr. Walter Kondo of the Bureau of Transportation Planning and Development who truly was "our man in Boston" and Mr. Jacob Silver and Ms. Ann Liddle of the Geography Division. Together they took care of all the details required to inaugurate a successful conference. It is wonderful to be here in Boston.

Over the past several years members of the Geography Division have given numerous speeches, both formal and informal, and have participated in numerous meetings and conferences wherein they have described the Census Bureau's Metropolitan Map Series, the geographic base (DIME) files, and the CUE program developed for their correction, update and extension. It is pleasant to report that during this short span of time a very significant change in program emphasis has taken place.

At first, we spoke about needs to be met, work that had to be done, and offered promises of progress. Later came reports of some actual accomplishments coupled with somewhat optimistic forecasts of how rapidly development of the remaining steps in the CUE system would be completed. Today, we are able to state, with very few exceptions, that the entire system is operational, that we have begun preparing second- and even third-generation computer programs and procedures (that is, programs and procedures which incorporate improvements suggested by local agencies), and that these programs are being made available or will soon be made available on demand.

At one time our concern focused on whether local agencies of government would recognize the value of a geographic file as a unique and important operations, management and analytical tool. Today, local government is informing us of ongoing GBF/DIME activities which represent new and imaginative uses of the file, and our problem now is to make certain that we can meet the demands being placed upon us for updated files by the 104 SMSA's now in the program and the many more soon to follow.[1]

Although this particular situation has changed, and the activities of the Geography Division along with it, one salient fact remains constant. For the files to be most useful they must reflect, for each area they serve, the continuing modifications in statistical, geographic, and political boundaries which accompany the dynamics of urban growth and change. The Census Bureau's CUE program, Correction, Update, and Extension of the geographic base (DIME) files has been developed to capture and record this ever changing geographic panorama in a highly efficient and effective manner.

CUE PROGRAM

To this end, the CUE program provides for—

1. The correction of existing files as necessary to produce a complete and accurate GBF/DIME and Metropolitan Map Series (MMS).

2. The extension of the GBF/DIME and the MMS maps to cover the entire geographic area of the SMSA. (At present only the urbanized portions of SMSA's are covered.)

3. The establishment of a geographic base (DIME) file and Metropolitan Map Series in those SMSA's where the file and map series do not currently exist.

4. The establishment of a standardized methodology to systematically maintain a current and accurate file and map series on a continuing basis.

The program makes available to local agencies the clerical procedures, processing methodology, and the computer edit and correction programs needed to meet these goals. And through these programs, the technical expertise and resources of the Bureau not only become accessible to all, but most important, standardization of the GBF/DIME system nationwide can become a reality.

---

[1] A summary listing describing some of these varied activities is provided in appendix A. It makes for interesting reading.

Standardization is, in fact, the key ingredient of the system. Without it the Census Bureau would be unable to participate effectively with local agencies, and instead of a compatible nationwide series of geographic base (DIME) files there would exist hundreds of independent and largely noncompatible local systems; and the anticipated usefulness of the files, both in terms of information exchange within and between local governments, and the public and private sectors, and as input to the Bureau's current and future geographic and geocoding processes would be seriously impaired.

It is equally important to point out, however, that the CUE procedures are flexible and do not demand a rigid, inflexible system, identical in format and use throughout the United States. Rather, the file is considered to be constructed in two parts: one part containing certain standard elements that will apply to all areas; the second part containing the local geographic elements which will vary from area to area, depending upon the local use of the file and local requirements.

Although the CUE program has been specifically designed to be carried out by local agencies of government, the Bureau, nevertheless, recognizes that some agencies, particularly during the initial phases of the CUE program, may not have available the technical personnel, computer facilities, or budgetary support to correct the tape files locally or even to run the computer edits. In this situation, the Census Bureau will assist the local agency in every way possible and, within the limits of the funding available, the Bureau, itself, will then carry out the various computer operations. Thus the local agency will then be responsible only for the clerical phase of the program and for reviewing the computer outputs to ensure that the Bureau has done its part of the job correctly.

Because of previous developments or commitments unique to the local situation, there may also be agencies which will undertake the correction, update, and extension activities using procedures and computer programs other than those that are being developed by the Census Bureau. The Census Bureau will, of course, also work with these agencies in a continuing cooperative effort. But in these cases special arrangements must be made in advance between the local cooperating agency and the Census Bureau to guarantee a product compatible with the needs of both organizations.

COMPUTER EDIT AND CORRECTION PROGRAMS

I would like to describe briefly the computer edit and correction techniques which are incorporated in the CUE program. They, perhaps better than any other phase of the operation, illustrate the extent and depth of the planning and development effort required for the establishment of a viable geographic base file update and maintenance technology.

All computer programs which are available for local agency use have been written in ANSI COBOL and are designed for small to medium computers having 60 to 100K or more of usable core. These programs include—

1. CREATE

   This program is designed to serve two separate functions: (1) In areas where a GBF/DIME does not currently exist, the CREATE program will establish a file using locally coded information transcribed on geographic coding worksheets. This is the "NEW" file option portion of the program. (2) In areas where a file already exists and is being extended, the program, using the "EXTEND" option, permits a large number of new records to be added to the file. The program—whether operating under the "NEW" or "EXTEND" option—requires a computer with 60K of usable core. (This program is currently available.)

2. FIXDIME II (2 and C Versions)[2]

   This program (1) edits the locally prepared correction inputs to the GBF/DIME for completeness and consistency; and (2) inserts the accepted corrections into the GBF/DIME. FIXDIME II is supplied in two versions: FIXDIME 2 for agencies which do not intend to correct or insert x-y coordinate information in the GBF/DIME during a correction pass of the file; and, FIXDIME C for agencies which can obtain from local sources known x-y coordinate information for node points and want to include this information with other corrections to the file. This program also checks the coordinate values being inserted to ensure that they are within the parameter of values that bound the area. FIXDIME 2 requires a computer with 47K of useable core; FIXDIME C requires a computer with 60K of useable core. (These programs are currently available.)

3. ADDEDIT-L[3]

   This program edits address ranges along a street feature, ZIP code consistency, and the orientation within and between segments on both street and nonstreet features. For example, street and nonstreet features are checked to determine whether or not all segments of the feature will chain together, approximating their relationship one to the other on the ground; whether the addresses at the "From" node end of the segments are lower than the addresses of the "To" node end of the segments, whether all odd address numbers are on one side of the street and even numbers are on the other; etc. Error outputs, appropriately flagged and sorted in map sheet number sequence, are listed. (Address corrections or other corrections, as needed, can then be inserted into the file using the FIXDIME II procedures.)

   ADDEDIT-L consists of two programs (a preprocessor and an edit program) separated by a system sort program. There are several built-in options available to the local user which permit varying the extent of the edit, process and/or printed output. Among these are (1) flagging (identifying) only selected types of errors and (2) printing either the complete computer-chained listing from node point to node point of all segments of a feature which contains a flagged record, or limiting the printing only to the flagged record and the segments on either side. The minimum core requirements of this program are approximately 100K. (This program will be available early this summer.)

4. TOPOEDIT

   As large numbers of records in the file are updated and/or added to the file, it becomes desirable to have the capability of editing the topologic consistency of the file. This program edits the network features of the file to determine their validity; for example, that a given block is bounded on all sides. It includes several options, including an option to edit only records in certain tracts, thus eliminating the necessity of editing the network within tracts in which no changes have occurred (Corrections required, if any, can be inserted into the GBF/DIME using the FIXDIME procedures). TOPOEDIT consists of two programs (PRESORT and TOPOEDIT) separated by a system sort; PRESORT requires approximately 18.5K of useable core; TOPOEDIT approximately 44K. (This program is currently available.)

5. FIXCORD

   This program provides the means to insert missing or to correct erroneous x-y coordinate values. The basic process involves inputting the distance (measured clerically) between specified base points of known value on each metropolitan map sheet to those nodes which need to have x-y coordinate values assigned. FIXCORD then uses these measurements of length (in 20ths of an inch) as input to calculate node values in each of the three coordinate systems used in the GBF/DIME (state plane, latitude-longitude, and map-set miles). The values for these three systems are then inserted into all appropriate segment records. (This program will be available later this summer.)

6. UPDIME

   This is a comprehensive program, designed for large computers (180K or larger of useable core). It is written in FORTRAN IV and includes many of the functions of FIXDIME II, FIXCORD, and TOPOEDIT. It does not include the functions of ADDEDIT which will need to be run separately. An early version of this program has been completed. However, changes in the CUE program and expansions in the scope of some of the editing programs require that the program be fairly extensively modified before it can be released. The final program is not expected to be available until late 1975.

[2] This is a second generation program. The original FIXDIME program was run with the 258-character GBF/DIME file and could not insert corrected coordinate values. A year ago the files were expanded to 300 characters and FIXDIME II can only be used with the expanded file.

[3] This is the version of the ADDEDIT program that is available for local use. The Census Bureau's "in-house" version does not contain the options to alter parts of the program which these paragraphs describe.

THE CUE METHODOLOGY

The CUE program involves three major operations—Correction, Update, and Extension.[4] I shall speak about each as if they were separated in time, with each operation being dependent upon the completion of the previous operation, although this is true only during the initial phase of the program. Once the full program is underway, all operations take place concurrently.

The first operation calls for the correction of the errors still remaining in the file as, based on our own experience and that of others who have updated and attempted to update files locally, updating a GBF/DIME becomes an extremely laborious, if not impossible, undertaking if the file errors are not first corrected. File inconsistencies are, of course, detected by the computer, but the necessary correction inputs must be supplied locally. Only the local agency has the knowledge to determine the actions needed to eliminate the errors which have been revealed by the various computer edit programs. The Census Bureau supplies all forms and procedures required to carry out this operation.

Once the file has been corrected, the first phase of the updating operation begins. This phase incorporates the changes in local geography which have taken place since the files were originally created (between 1969 and early 1971).

It is most important to note that since the geographic base (DIME) files are a computer image of the Metropolitan Maps, updating the computer files must be preceded by an updating of the maps. Each of the local coordinating agencies will, therefore, receive a reproducible set of the node-dotted and numbered map sheets for their area on which (following Census Bureau procedures) they add new street development, delete paper streets still appearing in the file, and correct street names and topologic features as required by changes in street patterns.

So that the Census Bureau will also be able to maintain an updated set of maps, at the appropriate points in the CUE update cycle (or as needed) the local agency will return a complete set of updated map reproducibles to the Bureau. The Bureau will add this set of map sheets to its inventory after having first produced and forwarded to the local agency a new set of reproducibles for use in future updating.

The second phase of the update operation is a repeat of the correction and update procedures described above, except that it is geared to be carried out as a continuing program. In other words, the local agency is expected to establish a mechanism through which any changes made to local geography, including changes to political boundaries, are routinely identified and added to the maps and files. Then, depending upon the needs of the local area, the actual file updating can be carried out monthly, quarterly, or annually. The Census Bureau's own interest in the updated GBF/DIME files and the Metropolitan Map Series is, in general, geared to an annual cycle.

[4] A complete description of CUE operations which illustrates in diagramatic form the "interface" and "interconnection" between the local coordinating agency and the Census Bureau operations is provided in appendix C.

The final phase of the CUE program consists of extending the Metropolitan Map Series and the geographic base (DIME) files out to the SMSA boundaries. At present the extension of the map series is well underway, and to date, drafts of extension maps have been submitted to approximately 72 areas for local review, correction, and updating. After completion of these operations and approval of the resulting maps by the Census Bureau, preparation of a geographic base (DIME) file covering the extension area can begin, including the assignment of street and road network identifications, block definition and numbering and, where possible, the inclusion of address ranges or equivalent rural area identifications.

LOCAL AGENCY PARTICIPATION

During the period 1969 to 1971, 196 standard metropolitan statistical areas joined in the development of a GBF/DIME file system. As reported earlier, as of April 1 of this year, 104 of these same areas are either participating actively or have agreed to participate with the Census Bureau in the CUE program. Of course, not all areas are proceeding through each phase at the same time. Some areas are just starting the correction operations. Some others, such as Columbus, Ohio, and Fort Lauderdale, Fla., are well on their way towards completing the updating and extension of the files. In total, approximately 20 percent of the participating areas have now completed the Correction phase of the program and have begun or will soon begin the update of the Metropolitan Map Series and the files.

To plan intelligently for CUE activities, local agencies, in addition to a knowledge of the technical aspects of the program, must also have information regarding the costs of the program. To this end, we have asked a number of the participating agencies to provide us with detailed progress reports, together with an indication of the number of man-hours required to complete each of the major clerical processes. As a result, we now can and do provide this information routinely to each area entering the program. Estimates are, of course, prepared for each area individually since workload and man-hour requirements vary greatly depending upon the quality of the original GBF/DIME file, the number of records in the file, and the availability of good reference information locally.

In general, however, the two correction phases of the CUE program (that is, correcting (1) inconsistent spellings of names and street types and illegal single-sided segments, and (2) erroneous address ranges and ZIP codes) will require approximately 50 man-hours per 1,000 file records. The Update phase, which requires that the Metropolitan Map Series be updated to the current date and, secondly, that file records be deleted, added, or corrected to reflect these changes is estimated to demand about 150 man-hours per map sheet involved. Assuming 10,000 file records and 7 map sheets at 35 square miles per map sheet (the average number of map sheets for an area containing 10,000 file records), a total of 1,550 man-hours would be expended locally in correcting and updating the file. Of this sum 500 man-hours would be required for file correction; 1,050 man-hours for the update process.

The man-hours involved in extending the Metropolitan Map Series and the GBF/DIME to the boundaries of the SMSA are much more difficult to estimate. However, assuming 4 extension map sheets (containing approximately 140 square miles per sheet), and about 3,000 GBF/DIME records per map sheet, extension of the maps and files would require approximately 400 man-hours per map sheet.[5]

It should be emphasized that these examples provide estimates of local clerical man-hour requirements only. They do not include the costs of the interrelated computer processes, for two reasons. First, those computer operations which are performed by the Census Bureau—since they are necessarily an integral part of the Bureau's GBF/DIME system—are carried out without charge to the local area. Second, the variety of computer systems and charging algorithms in use locally make it impossible to provide meaningful average estimates.

NEW GBF/DIME FILE AREAS

There are today, approximately 70 SMSA's for which a GBF/DIME file does not yet exist. Some areas never participated in the program originally; but most represent new SMSA's created by the Office of Management and Budget subsequent to the 1970 census. I am pleased to be able to report that 8 of these 70 SMSA's are now in the program; 2 are new SMSA areas; the remaining 6 represent previous nonparticipants. The unit cost of establishing a new GBF/DIME for these particular 8 areas (at 35 square miles per map sheet) is estimated to be about 250 man-hours per map sheet. As with the extension program, the estimates vary considerably area by area depending upon the size and geographic characteristics of the area concerned.

FUNDING THE CUE (GBF/DIME) PROGRAM

Participation in the CUE program is, of course, not an end in itself. The major purposes of CUE are solely (1) to ensure standardization of the product and (2) to make possible an efficient maintenance and updating vehicle for the GBF/DIME files and the Metropolitan Map Series. Nor, of course, are the GBF/DIME files, themselves, end products. They, in turn, serve only as inputs to a technology which permits for the first time (and I exaggerate only slightly) the development of effective procedures for carrying out research and for managing and conducting real time programs dealing with local problems whose solution requires ordering economic and social data spatially and geographically.

Since neither CUE nor the GBF/DIME files are end products, it is not surprising that funding for their development, which does not envision an end product, does not exist. Availability of funding for CUE, whether from Federal, State, or local sources, is predicated upon the use of the GBF/DIME file to carry out a specific Federal, State, or local program more efficiently and more effectively. This is readily illustrated. One Federal planning group, for example, has advised the local government agencies in its area that—

[5] Detailed estimates of the average clerical work-load and man-hours required for an area of 10,000 records (an area slightly smaller then the current Lowell, Mass. file) during the first correction and update cycle are illustrated in appendix B.

"...one of the eligible items that could be included in the overall work program...is the establishment of a geographic base file...(this)...would also permit...correction and maintenance..."

"The reason for inclusion of this as an eligible item is the recognition...of the importance of a geographic tool for identification of land use, social and economic data..."

"However...if a geographic base file is to be included in your planning process, a design program for such an effort be prepared first, before any substantial funding for the base file is committed."

Unfortunately, although the Census Bureau is the major actor in the CUE program area, it has not yet been able to follow the above example and provide funds to assist local areas in correcting, updating, and extending the GBF/DIME files and the Metropolitan Map Series. Hopefully, this situation will change as up-to-date files and maps are critical to the Bureau's geographic program. There is, in fact, now some reason for optimism as our fiscal 1975 budget, which is presently before the Congress, includes funds to be used to help defray the costs incurred locally in carrying out the CUE program objectives of the Census Bureau. Let me hasten to add, however, that even under the best of circumstances, such funds will be limited in amount and not available to each cooperating agency on an annual basis. Major funding for the GBF/DIME system will still be dependent on the usefulness of the files to meet Federal, State, and local program objectives. This is as it should be. In the long run, Census Bureau uses, vital as they are to us, represent but a small fraction of the full potential of a GBF/DIME system.

LOOKING FORWARD

The technical problems involved in establishing a nationwide, standardized set of geographic base (DIME) files have finally been overcome. It is fair to say that the difficulty of the undertaking had been underestimated originally. But at long last the delays are behind us. The clerical procedures, processing methodology, and computer programs being provided to local agencies are tested and proven packages and the Geography Division is proud of its products.

All that now remains is full implementation of the GBF/DIME at the local level and the organization of local users of the file into informal consortiums so that—in cooperation with the Census Bureau—the file and Metropolitan Map Series can be systematically maintained on a current basis. By so doing, the inefficiencies and high costs of parallel, noninterrelated, and perhaps competitive systems could be eliminated.

It is recognized that coordination of local activities in this area is not an easy task. Dr. Edgar M. Horwood, Codirector of the Urban Systems Research Center of the University of Washington may have expressed the difficulties of the undertaking most eloquently during his summary of our Seattle Conference. I would like to quote him. He said:

"... the Census Bureau and local agencies are both working on the same assembly lines in regard to development and improvement of the geographic base files and use systems. Added to the Federal and local arena, we have also the private sector, which may either be involved as another part of the production line or in fact manning its own production line in collaboration with or in competition with the public agencies.

"These three sets of actors—the Federal agencies, the local agencies, and the private sector—constitute an entirely new type of venture in the production, quality control, updating, standardization, and servicing of geographic base files for both local and national use. The Census Bureau is thus thrust into a joint venture with local actors without whose cooperation and interest its production will be compromised. It now remains to be seen how well all of these actors will be able to work together, or at least in constructive competition..."

The mechanisms through which a cooperative local effort can be established and maintained are just now beginning to emerge. Undoubtedly they will be different areas as each will be designed to meet the unique needs of the individual situation. This, too, is as it should be. The Census Bureau and the Geography Divison stand ready to assist local areas in furthering their development.

One final point remains to be made. Neither CUE nor GBF/DIME are intended to represent closed systems, complete and perfect from now to eternity. In the near future, we will be asking you for any suggestions which you may have which will lead to improvements in the program and that would be applicable on a nationwide basis. One such possibility, for example, which we are just beginning to explore, might be the addition of urban transportation zone codes to the file.

Some years ago we concluded that "the golden age of the GBF" was at hand. We know now that we were wrong. As you will hear over the next few days, the best years are yet to come.

APPENDIX A

A Sample of Letters and Other Documents Related to the Adoption and Use of Geographic Base (DIME) Files

The Census Bureau is often asked how the files are being used by local agencies and organizations. The following quotations were abstracted from a group of letters and documents received by the Geography Division. The names of agencies, persons, or areas have purposely been deleted. If anyone desires to follow up and contact an agency or person quoted below, please write for further information to:

Chief, Geography Division
Bureau of the Census
Washington, D.C. 20233

1. Letter June 8, 1971 (from New Mexico)

Our Urban/Regional Information System Committee which represents all of the local government entities and operating departments has adopted the GBF as the primary geographic base system along with land parcel records. The City of *** and *** County Data Processing Departments are beginning to use the GBF in the development and operation of functional subsystems which will service all of our local using agencies. As previously noted in our letter of May 11, 1971 to Mr. ***, the GBF will be serving as a base for such subsystems as street inventory, traffic census, water and sewer system inventories and for such management systems as routing and scheduling of solid waste collection and disposition and law enforcement activities.

2. Letter May 20, 1971 (from Texas)

We have and will have for the foreseeable future a continuing need for a rather regular use of geographic base files in such functional planning programs as public transportation, solid waste disposal, and airport-systems planning. Accordingly, we will shortly be commencing a map and G.B. file maintenance program to support these efforts on a continuing basis. Cooperating with you in this effort will certainly enhance the utility of the work which we would otherwise accomplish alone.

3. Letter March 27, 1973 (from Alabama)

As you know, Mr. *** of your staff has been working with the *** Police Department in setting up an application for an indexing system based on geographic location called the DIME File. This DIME File will be used by the *** Police Department for manpower allocation and crime-oriented planning.

Enclosed is a copy of the DIME File application. The City of *** anticipates receiving this grant within a short time. A contract will have to be drawn up for the services of the *** Regional Planning Commission to be provided as set out in the grant.

IMPACT AND RESULTS: This project will provide a geographic indexing system for all *** Police Department reports. Such an indexing system will allow the department to perform comprehensive crime analysis by location and develop crime profiles for all areas of the city. Crime analyses and profiles are necessary for effective planning. The analyses and profiles can be used to assign and allocate departmental resources, determine training needs, and justify budget requests. The department will be able to place task forces in areas of high crime or areas where specific types of crime are occurring most frequently. The number of men on duty each day of the week and the number working each shift will be determined by the analysis. Since this will put more men on the street when and where they are needed, the response time for the police department should drop. The present response time for mobile units is 5 minutes. Within 3 months of implementation of the DIME file, response time should drop to 3 minutes.

The DIME file will provide for future implementation of computer-aided dispatching of mobile units which should further reduce response time. Computer-aided dispatching is a long-range goal of the department.

The DIME file will enable the Department to have concrete, up-to-the-minute reports for all supervisory personnel. This will alow supervisors the opportunity to constantly evaluate programs in terms of their effect on crime. The reports will also be used by the Department to provide periodic feed back to the nonsupervisory personnel. This will give the officer on the street a true picture of the results of his efforts and allow him to see if and when new programs or policies are needed. This constant flow of information should improve communication

throughout the Department which will create an atmosphere that is conducive to change and experimentation.

It is anticipated that within 1 year of implementation of the DIME file, the Police Department will be able to effectuate a 2-percent decrease in the number of crimes committed and a 5-percent increase in crimes cleared. Within three years, the Department expects to bring about a 15-percent decrease in crime committed.

4. Letter December 3, 1973 (from South Carolina)

We have already used the file for numerous things. Some of these are:

(1) Neighborhood Analysis
(2) Enumeration District Street Directory
(3) Census Tract Street Directory
(4) Law Enforcement Needs Study
(5) Substance Abuse Study
(6) Users of *** County Bookmobiles
(7) The University of *** is using the GBF in some of their class course work.

5. Letter April 21, 1972 (from Minnesota)

Some of the programs which are, or will be developed to utilize the GBF's are listed below:

1. Metro Area Transportation System Analysis in cooperation with DOT and the *** State Highway Department.

2. Housing inventory and analysis of blight conditions. (Metro-Study)

3. Socioeconomic conditions by areas, as related to multi-agency social services, unemployment, etc. in a study being undertaken by the Human Resources Planning Coalition.

4. Juvenile crime survey, elderly population survey, etc.

5. Land use inventory, with initial emphasis on Tax Forfeit Land inventory.

6. Letter March 2, 1972 (from Minnesota)

As a side note, clearly such a file will be useful to our newspapers not only in helping us maintain identification of our subscribers/nonsubscribers through which we can help control our newspaper circulation, but also most helpful to us in better serving our advertisers with direct marketing capability to our specific nonsubscribers.

7. Letter April 17, 1973 (from Nebraska)

I might add that the University of *** has utilized our 1970 DIME file tape, without correction, to develop a NICKLE file and has run, with a great deal of success, the ADMATCH program (OS Version) to locate students residence by various geographic areas of the community. Our DIME file appears to be in very good shape. I might also add that we, in the city, intend to do the same with automobile registrations, as soon as the update is complete.

8. Letter March 7, 1972 (from Michigan)

After completion of this portion of the work program, the Council would begin to implement automatic geocoding on a large scale. Files currently being studied for possible goecoding on an areawide basis include an employment file from the *** Employment Security Commission, social services rendered by member agencies of United Community Services, and birth and death records for the *** Department of Health. Aside from the obvious benefits from additions to the regional data base, this experience will provide important inputs to the annual maintenance and update procedures described in Phase II. Also, at this point in the program the Council will publish, in report form, a tract level coding guide as a service to those agencies wishing to do clerical geocoding.

9. A copy of an "in-house" statement of local program development January 1974 (from Indiana)

Present and past applications of the DIME File include:

A. Transportation planning. Vehicle registrations were matched by address and allocated to traffic zones. The match rate was 90 percent of the 103,000 registrations in the urbanized area.

B. Housing information. Building permits since 1970 were matched (and mapped) to determine development patterns. The match rate was very low, but this was expected due to the time lag in updating the DIME File.

C. Welfare recipients. ADC families and food stamp recipients were matched and mapped from present files. It is hoped that a continuing system can be implemented in the future for the County Welfare Department.

D. Public Schools. The largest school district in the area used DIME/ADMATCH to locate all students. A match rate of 87 percent was obtained. The results are being used for both short-and long-range planning by the school district. In the future, the district plans to use the DIME file to locate students eligible for bus service, and to schedule the buses.

E. Carpooling. The Coordinating Council ran an experimental carpool system using DIME/ADMATCH. Approximately 86 percent of respondents who desired carpool information were matched. It is anticipated that the new Census Bureau program CARPOL will be utilized by this agency to provide carpooling services to the entire community.

F. Research. DIME/ADMATCH has been utilized by consultants and students to solve their unique data problems. Match rates have consistently been over 80 percent. In all cases, DIME technology was of enormous help in reducing costs as well as increasing the reliability of the data.

10. Letter December 11, 1973 (from Virginia)

After a thorough study of the material presented at the last meeting, this Commission is convinced that the GBF is an excellent, basic tool and standardized system for all future planning. The effort to update the GBF for the *** Urban Area and expand the coverage to include, not only the *** SMSA, but the whole Planning District is a tremendous task but justifiable.

It is visualized that the GBF, with an associated Transportation Data File Subsystem, would facilitate the surveillance process required in the annual updating of the *** Area Regional Transportation Study to comply with Federal and State requirements; it would provide the base for analyzing information necessary to continually appraise and update the *** Urban Mass Transit Program; it would assist in evaluating and implementing the myriad complexities of the rural and regional social services transportation programs, and it would enhance the implementation of a regional intermodal transportation staging plan. The GBF and related subsystems would be utilized by our other planning disciplines.

11. Letter November 9, 1973 (from Louisiana)

Mr. *** was also most generous with his time and answered my many questions regarding the Geographic Base File program. In my own case, the availability of this file enabled us to select an area probability sample of the *** area which will be used in a survey regarding the imported fire ant problem. Our report will be submitted to EPA by Dr. *** who is to testify in the hearings concerning the use of the insecticide Mirex.

The potential usefulness of this file for other workers in public health was readily apparent to me. Mr. *** I know will be contacted by two public health colleagues, one from the school and the other from the State who are most interested in exploring potential uses of the file relevant to their respective areas of endeavor. Mr. *** in turn may well find that they may be of help to him.

12. Letter August 10, 1971 (from Pennsylvania)

The potential utility of the GBF can hardly be overstated. It provides a computer-based tool with which address-specific data records relevant to highway and comprehensive planning can be geocoded. We will make first use of the base file during this coming year in the preparation of a regional employment location file using records from the two States' Bureaus of Employment Security, the *** Commercial Establishment File, and ***. Also we are tentatively committed to a project that would utilize the GBF to geocode motor vehicle registration records from *** and *** to produce an auto ownership data file for our urban area. This project, if feasible, would be undertaken during FY 1973.

13. Abstracted from an application for a Law Enforcement Assistance Grant (which was approved October 31, 1972) from Louisiana.

IMPACT AND RESULTS

This project will demonstrate feasibility of technology transfers, the cost relationship between original development and transferred technology arriving at the same basic results, and will also develop procedures for future transfers.

At the conclusion of this project a long overdue basic tool will be available, a geographic base file with x-y coordinates. The *** Police Department will benefit from the file more than any other agency. With this tool we will be able to correlate calls for service to reporting ares small enough to use LEMRAS, a manpower resource allocation system. The more effective utilization of men and vehicles should dramatically decrease response time on calls for service, thereby increasing the probability of on-scene apprehension. Beat structures will be aligned on an as needed basis, again affording better police protection. The file will eliminate the need to hand code reporting zones thus eliminating human error, and making statistical data available on a real time basis. Presently, the Police Department is spending thousands of dollars a year coding and keypunching reporting zones. This file will cut this cost and allow more efficient utilization of these funds.

Command and control functions can be developed with this file. These functions will provide the communications center with the historical data by location. Police officers better equipped with information are less likely to be injured. Dispatchers equipped with the same information can make better value judgements as to the number of personnel to assign to a call.

A future application, Computer Assigned Dispatch, depends upon the availability of a geographic base file with x-y coordinates. With this function police units can be picked for calls based on their relative location and priority of call. They can receive assignments in digital as opposed to voice form, insuring greater security.

Computer mapping is another function dependent on this type of file. At present all geographic statistics are done by hand. In all of the above areas this file will introduce a level of efficiency unattainable with present methods.

There are six other agencies interested in using this type of file: City Planning Commission, Regional Planning Commission, Sewerage and Water Board, Property Records, Alcohol Safety Action Program, and the Fire Department. The two planning agencies see it as a tool to aggregate data on flexible areas giving them accurate and timely information with which to plan in areas such as traffic studies, lighting studies, and street improvement. The Regional Planning Commission has agreed to assume the responsibility of maintaining the maps associated with the file.

The Fire Department is interested in converting its voluminous card files containing building contents, structural data, and company assignments to a computer based file. They are also interested in routing engines to fires via a coordinate system file.

The Sewerage and Water Board is interested in using this file to locate underground facilities that they maintain.

The Property Records group sees this file as a means to control titles and tax information when the file is brough to the parcel level.

The documentation of this project should make this system available for any city that wishes to develop a LOCUS system from the basic DIME file package.

APPENDIX B

Estimated Local Workloads and Clerical Man-hours Average Area, SMSA

CUE Operations

	Workload	Man-hours[1]
I. Total segment record in the file.	10,000	
II. Correction phase (This phase is subdivided into two parts.)		
A. Part 1 corrections (This includes correcting the segment name consistency and coding limit line/unmatched segment listings.)		
– The review and correction of these listings average about 18 man-hours for each 1000 records in the file.		180
B. Part 2 corrections (This includes correcting the Part 1 FIXDIME edit/transaction listing and the ADDEDIT address error listing.)		
1. The number of records that will require review is estimated at 38 percent of the total file.[2]	3,900	
2. A clerk reviewing the error listing can correct an estimated 12 records per hour.		325
Subtotal—Estimated clerical man-hours to complete the correction phase.		505
III. First update cycle (This includes updating the maps used during the original coding (1969-1971 period) to the current period (including redrafting the changed information on the reproducible maps); node dotting and numbering as the result of map changes; adding, deleting, or changing records as a result of map changes.)		
1. The average number of map sheets for a 10,000 record file.	7	
2. The update of both the maps and the file is estimated to take 150 clerical hours per map sheet.		1,050
Estimated clerical man-hours for both the correction and update phases.		1,555

Extension of GBF/DIME

The assumption made here for estimation purposes is that this is an area which includes 2,000 blocks and covers four map sheets (140 square miles per map sheet).

	Workload	Man-hours[1]
I. Map review and preparation (This phase is subdivided into two parts.)		
A. Correction and update of the Metropolitan Map Series. (This includes the review, correction, and update of a set of diazo prints.)[3]		
1. The number of map sheets.	4	
2. The average time for a clerk to review and correct a map sheet is 30 hours.		120
(The Census Bureau will prepare a new set of reproducibles to include the corrections made under part A.)		

[1] These figures do not include any time for clerical training or for supervisory time. The figures also assume that acceptable reference information (such as addresses) is available to the clerical staff. If any field listing or research time is required to obtain the necessary correction information, the estimated number of man-hours to carry out this search or field listing should be added to the total man-hour figure indicated.

[2] Although 39 percent of the file is listed in the edit/error listing for general review, not all of the segments shown have error flags. If one segment is indicated to contain an error, then all segments along that street or nonstreet feature will be listed.

[3] If this is an area in which block groups have not previously been defined and blocks numbered, the estimated number of man-hours to carry out this operation will have to be added to the total man-hour figure.

	Workload	Man-hours[1]
B. Map preparation prior to coding. (This includes node dotting and numbering of the reproducible map sheets and highlighting statistical and political boundaries on diazo prints.)		
– The average preparatory time for a clerk is 30 hours per map sheet.		120
II. Coding and quality checks		
A. The number of segments to be coded. (This is based on the estimated number of blocks and the average number of segments per block.)		
1. The number of blocks assumed is 2,000.		
2. The average number of segments per block is 6.		
3. The number of segments to be coded.	12,000	
B. Clerical man-hours of coding.		
1. A clerk can code an average of 16 segments per hour.		
2. The average number of hours of original coding.		750
C. Quality check of original coding. (This includes clerical editing, sample quality checks and the rework of rejected tracts.)		
– The man-hours involved in this stage is estimated to be 40 percent of the original coding man-hours.		300
III. Clerical correction of computer edits. (As the result of the file creation program (CREATE), the topologic edit program (TOPOEDIT) and the address edit program (ADDEDIT), a number of records will list out on a series of error listings and will require clerical review to determine if they are actually in error. If they are in error, a transcription record of the correction will need to be prepared.)		
A. Approximately 33 percent of the records in the total file will require review.	3,960	
B. A clerk can review the records and make corrections when necessary at an average rate of 12 segments per hour.		330
Estimated total man-hours		1,620

See footnotes on page 14.

APPENDIX C.

Introduction

The following paragraphs describe the Census Bureau's CUE program—the Correction, Update, and Extension of the geographic base (DIME) files. (Description of program written October 1972.) This program is designed to develop a complete and accurate geographic base file, and provide for its maintenance and update on a continuing basis. There are six parts to the program, one of which is optional. Each of these is outlined on the accompanying flow charts, which in turn are keyed to the text through the use of alphabetic codes. If any part of the chart is not immediately clear, the text should by referred to for a fuller explanation. The operations being carried out in the CUE program cover three major areas of the activity:

I. Correction Operations
   Part 1
   Part 2
   Part 3 (Optional)

II. Update Operations
   Phase 1—Update to Current Date
   Phase 2—Continuing Update

III. Extension Operation

Although some of the operations are illustrated in the diagrams as if they were separated in time, with each operation dependent upon the completion of the previous operation, this is true only during the initial phase of the program. Once the full program is underway, the update and extension operations may take place concurrently.

NOTE: There will be agencies which, because of previous developments or situations peculiar to the local operation, will out of necessity undertake correction, update, and extension activities using procedures and computer programs other than those outlined in the following document. The Census Bureau will also work with these agencies in a continuing cooperative effort. In each of these cases, however, special arrangements will be made between the local cooperating agency and the Census Bureau to ensure a product compatible with the needs of both organizations.

Additional copies of this appendix are available and can be obtained at no cost by writing to—

Chief, Geography Division
Bureau of the Census
Washington, D.C. 20233

Any questions concerning the CUE program, the related computer programs, or the GBF maintenance operations for any particular area should be directed to the same address.

LOCAL CUE OPERATIONS

I. CORRECTION OPERATIONS

PART 1

A. The Census Bureau provides a geographic base (DIME) file tape with x-y coordinate values to the local corrdinating agency. The Census Bureau at the same time provides two computer edit listings (the Segment Name Consistency Listing and the Coding Limit Line/Unmatched Segment Listing) to assist in locating certain types of errors.

B. The local agency reviews both listings for errors and enters appropriate corrections on transcription forms provided by the Census Bureau.

C. The local agency decides whether—

   1. It will insert the corrections into its file locally using the Bureau's FIXDIME program and, when completed, submit these corrections to the Census Bureau; or

   2. Submit the corrections to the Census Bureau without correcting the local file; or

   3. Insert the corrections into its file locally using the Bureau's FIXDIME program and, when completed, submit a copy of the corrected geographic base file to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

D. If the agency decides to just submit the corrections to the Census Bureau, it merely forwards the transcription sheets or keypunched cards containing the corrections to the Bureau.

E. If the agency decides to correct the file locally, it punches the corrections into data cards and uses FIXDIME to correct the GBF. The resulting output is a corrected GBF-File I and a FIXDIME EDIT/ Transactions Listing.

F The rejects from the FIXDIME Edit/Transaction Listing are reviewed. If the review shows that too many correction cards have been rejected, these cards are reviewed and corrected as necessary. Steps E and F are repeated.

G. The accepted name and segment correction punched cards (or computer tape) are forwarded to the Census Bureau or, alternatively, a copy of the corrected GBF-File I is forward to the Census Bureau.

H. Part 1 of the correction operation is completed after the Census Bureau receives either—

   1. The Segment Name and Unmatched Segment Correction worksheets or keypunched cards, or

   2. The accepted corrections on keypunched cards or tape after FIXDIME, or

   3. A tape copy of the corrected GBF-File I.

PART 2

A. The Census Bureau furnishes the local coordinating agency the Address Range Edit (ADDEDIT) Listing, and the FIXDIME Edit/Transaction Listing which includes any rejected records resulting from the Census Bureau's running of FIXDIME.

B. At this time the Bureau also furnishes the agency with a reformatted version of the corrected GBF.*

C. The local agency reviews the ADDEDIT Listing and the rejects of the Census Bureau's FIXDIME Edit/ Transaction Listing and enters corrections on transcription forms using procedures provided by the Census Bureau.

D. At this point the local agency must decide whether or not it wants to perform the optional task of adding or correcting the x-y coordinate values of the GBF nodes using local reference sources showing known "absolute" values. If it chooses to do so, it follows the series of steps indicated under section A of Correction Operations—Part 3 (Optional).

E. The local agency decides whether—

   1. It will insert the corrections (including the coordinate corrections) into its file locally using the FIXDIME II correction program and, when completed, send the corrections to the Bureau; or

   2. Submit the corrections to the Census Bureau without correcting the local file; or

   3. Insert the corrections into its file locally using the Bureau's FIXDIME II program and, when completed, submit a copy of the corrected GBF-File II to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

F. If the agency decides to just submit the corrections to the Census Bureau, it merely forwards the transcription sheets or keypunched cards containing the corrections to the Bureau. (These will also include absolute coordinates if the agency used the option to insert these values.)

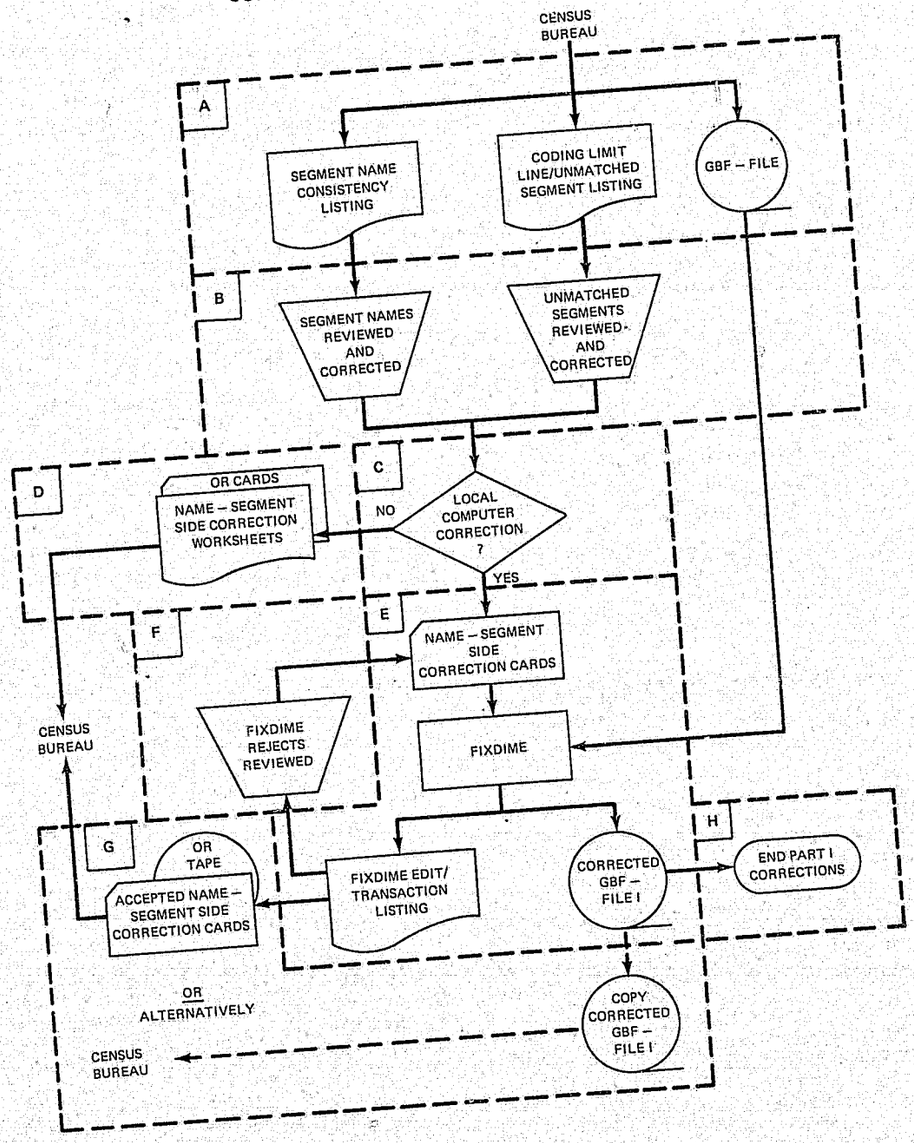
G. If the agency decides to correct the file locally, it punches the corrections into data cards and uses FIXDIME II to correct the reformatted GBF. The resulting output is a corrected GBF-File II and a FIXDIME II Edit/Transaction Listing.

---

*"Block number" and "map number" fields have been expanded to provide for suffixes in these fields: the "local ID" and "ward" fields have been deleted, and the overall record size has been increased to 300 characters.

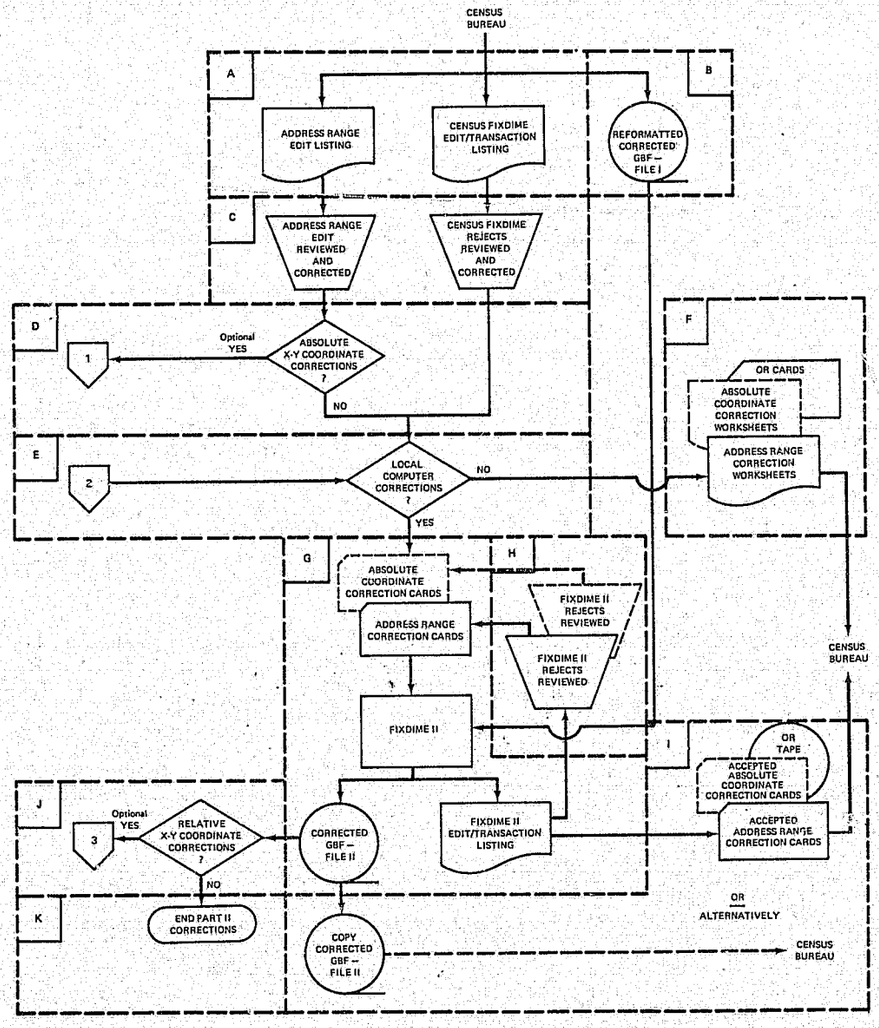
LOCAL CUE OPERATIONS

CORRECTION OPERATIONS - Part 1



LOCAL CUE OPERATIONS (CONTINUED)

CORRECTION OPERATIONS - Part 2



H. The rejects of the FIXDIME II Edit/Transaction Listing are reviewed. If the review shows that too many correction cards have been rejected, these cards are reviewed and corrected as necessary. Steps G and H are repeated.

I. The accepted ADDEDIT (and absolute coordinate) correction punched cards (or computer tape) are forwarded to the Census Bureau, or alternatively, a copy of the corrected GBF-File II is forwarded to the Census Bureau.

J. At this point the local agency must decide whether or not it wants to perform the optional task of adding or correcting the x-y coordinate values of the GBF nodes using the relative coordinate correction procedures and the FIXCORD program to insert coordinate value corrections into the corrected GBF-File II. If the agency chooses to do so, it follows the series of steps indicated under section B of Correction Operations—Part 3 (Optional). If the agency chooses not to insert relative coordinate value corrections, the correction operation is completed.

K. Part 2 of the correction operation is completed after the Census Bureau receives either—

1. The ADDEDIT (and Absolute Coordinate) Correction Worksheets or keypunched cards, or

2. The accepted corrections on keypunched cards or tape after FIXDIME II, or

3. A tape copy of the corrected GBF-File II.

NOTE: When an agency selects the option to send the corrected records to the Census Bureau, and these corrections are not entered locally into the file, it must be understood that no computer tape of the corrected geographic base file will be supplied to the agency. The Bureau will store the correction inputs until after the update operation is completed. At that time a corrected tape will be made available on a loan basis to those agencies that did not elect to insert the corrections and the updated information into their files. The loan tape is to be copied onto a local tape and returned to the Census Bureau. It should also be understood that the loan tapes can only be made available following a schedule which will not interfere with other operations of the Census Bureau.

PART 3 (OPTIONAL)

A. Absolute Coordinate Value Corrections

1. The agency reviews the ADDEDIT Listing to identify nodes with missing or erroneous coordinate values.

2. The correct absolute values (which are obtained from a source independent of the Census Bureau) are transcribed onto transcriptions sheets provided by the Census Bureau as part of the Correction Operations—Part 2.

3. The steps that follow are the same as, and form a part of, Correction Operations—Part 2, starting with step E.

B. Relative Coordinate Value Corrections

1. The ADDEDIT Listing is reviewed to identify nodes with missing or erroneous coordinate values.

2. The appropriate measurements to calculate x-y values are made and recorded on transcription forms provided by the Census Bureau.

3. The local agency decides whether—

a. It will insert the corrections into its file locally using the Bureau's FIXCORD program, and when completed, submit the corrections to the Census Bureau, or

b. Submit the corrections to the Census Bureau without correcting the local file, or

c. Insert the corrections into its file locally using the Bureau's FIXCORD program, and when completed, submit a copy of the corrected GBF-File III to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

4. If the agency decides to just submit the corrections to the Census Bureau, it merely forwards the transcription sheets or keypunched cards containing the corrections to the Bureau.

5. If the agency decides to correct the file locally, it punches the corrections into data cards and uses FIXCORD to correct the file. The resulting output is a corrected GBF-File III and a FIXCORD Edit/Transaction Listing.

6. The rejects of the FIXCORD Edit/Transaction Listing are reviewed. If the review shows that too many correction cards have been rejected, these cards are reviewed and corrected as necessary. Steps 5 and 6 are repeated.

7. The accepted relative coordinate corrections on keypunched cards (or computer tape) are forwarded to the Census Bureau, or alternatively, a copy of the corrected GBF-File III is forwarded to the Census Bureau.

8. Part 3 of the correction operation is completed after the Census Bureau received either—

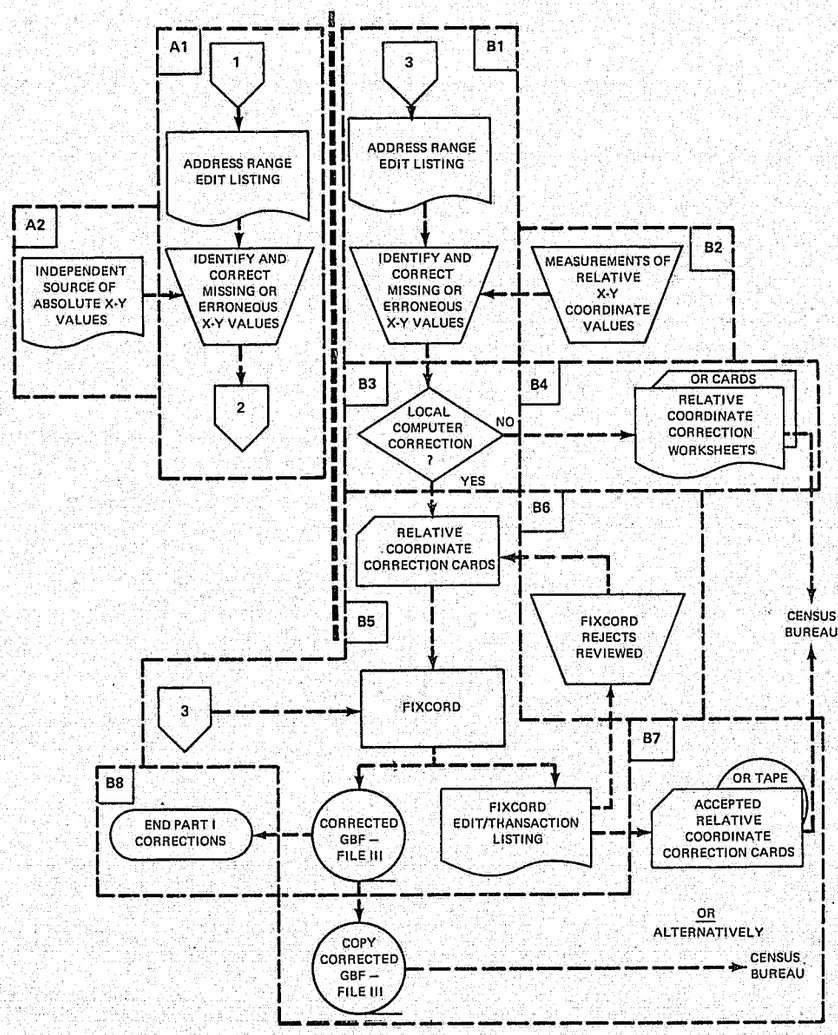
a. The FIXCORD Relative Coordinate Correction Worksheets or keypunched cards; or

b. The accepted corrections on keypunched cards or computer tape after FIXCORD, or

c. A copy of the corrected GBF-File III.

LOCAL CUE OPERATIONS(CONTINUED)

CORRECTION OPERATIONS - Part 3 (Optional)



II. UPDATE OPERATIONS

PHASE 1—UPDATE TO CURRENT DATE

A. In accordance with Census Bureau specifications, the local agency updates the Metropolitan Map Series sheets to the current date, and as necessary, enters new node numbers and new block numbers on the updated maps.

B. The Census Bureau provides the local coordinating agency the FIXDIME II (and FIXCORD) Edit/Transaction Listing, which includes any rejected records resulting from the Bureau's running of FIXDIME II (and FIXCORD).

C. The local agency reviews the rejects of the FIXDIME II (and FIXCORD) Edit/Transaction Listing and enters corrections on transcription forms using procedures provided by the Census Bureau. At this time, all new segments and corrections resulting from the map update and node and block numbering are likewise entered on the transcription forms, using procedures provided by the Census Bureau.

D. The local agency decides whether—

1. It will insert the corrected, updated segments into its file locally, and when completed, submit the updated and corrected information to the Census Bureau, or

2. Submit the corrected, updated information to the Census Bureau without inserting this information into the local file, or

3. Insert the corrected, updated information into its file locally, and when completed, submit a copy of the updated Geographic Base File to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

E. If the agency decides to first submit these data without updating its file locally, it merely forwards the transcription sheets or keypunched cards containing this information to the Bureau.

F. If the agency decides to update the file locally, it punches the corrected and updated information into the data cards.

G. At this point in processing the file, the agency must decide whether—

1. To use the Bureau's UPDIME and ADDEDIT computer programs for updating its file, or

2. To use a combination of the FIXDIME II, FIXCORD, TOPOEDIT, and ADDEDIT computer programs to update its file.

In either case the resulting output is an updated and edited GBF. The UPDIME program produces an Edit/Transaction Listing: The FIXDIME II, and FIXCORD programs will also produce separate Edit/Transaction Listings for each of the two respective programs at the end of each program run. The TOPOEDIT and ADDEDIT programs will produce Edit/Error Listings, with the different types of errors appropriately flagged.

H. The rejected records from each of the Edit/Transaction Listings are reviewed. If the review shows that too many data cards have been rejected, these cards are reviewed and corrected as necessary. At the same time, the flagged records in the Edit/Error Listings are also reviewed and corrected. In both instances, steps G and H are repeated.

I. The accepted, corrected and updated data (in the form of keypunched cards or computer tape) are forwarded to the Census Bureau, or alternatively, a copy of the updated GBF file is forwarded to the Census Bureau. Phase 1 of the Update Operation is completed after the Census Bureau receives either—

1. The corrected and updated data worksheets or keypunched cards, or

2. The corrected and updated data on keypunched cards after UPDIME (or, alternatively, the Quad-Program series), or

3. A tape copy of the updated GBF.

NOTE: If an agency has been cooperating with the Census Bureau and has been supplying corrections, updates and additions to the Bureau file, but has not inserted these corrections into its own file, the Census Bureau at the completion of the update operations will provide on a loan basis a copy of the corrected and updated file. This loan tape is to be copied onto a local tape and returned to the Census Bureau. It should be understood that the loan tapes can only be made available following a schedule which will not interfere with the Census Bureau's other operations.

PHASE 2—CONTINUING UPDATE

The coordinating agency establishes the local mechanism through which continuing changes in local geography that occur are systematically identified and reported. These changes and additions are incorporated into the file on a periodic basis.

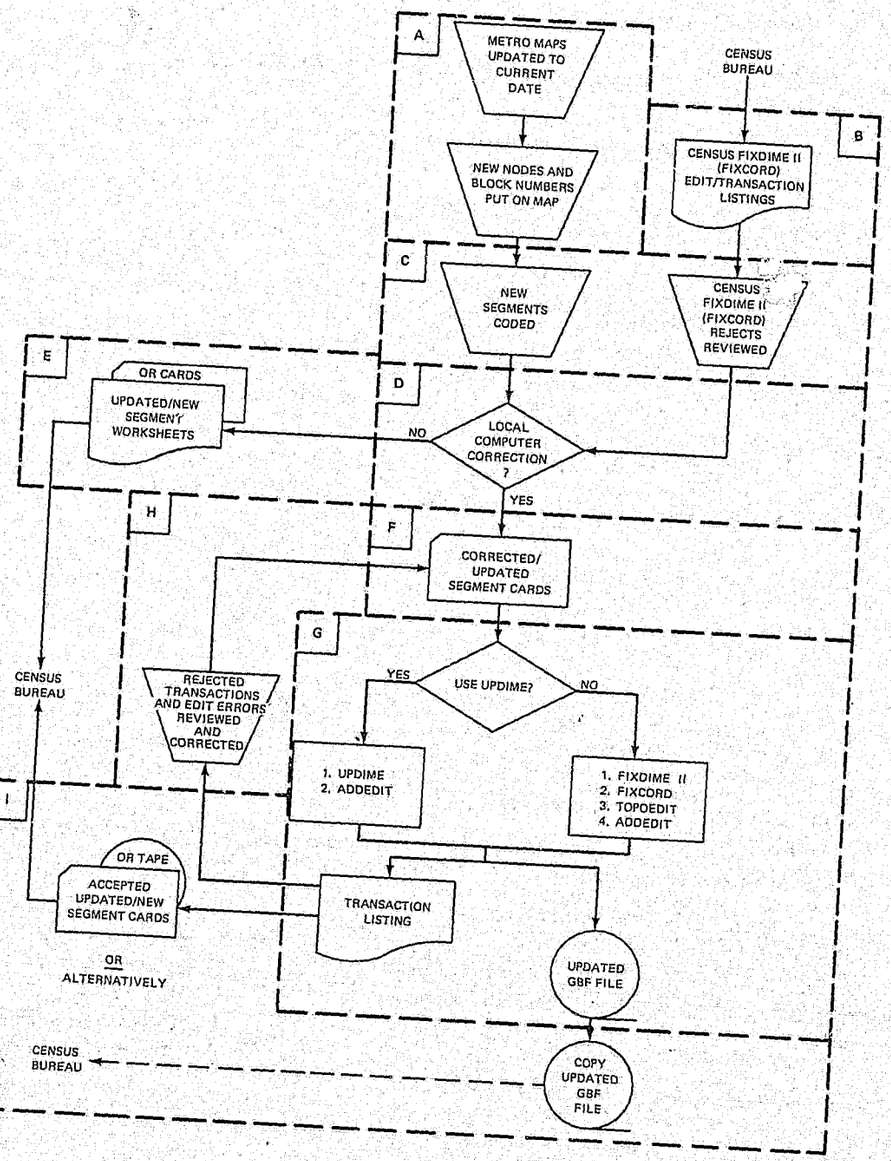
A. In accordance with Census Bureau specifications, the local agency updates and maintains the Metropolitan Map Series sheets to the current date on a regular and systematic basis, and as necessary, enters new node numbers and new block numbers on the updated maps.

B. The Census Bureau provides the local coordinating agency an UPDIME (or FIXDIME II and FIXCORD) Edit/Transaction Listing, and TOPOEDIT and ADDEDIT Edit/Error Listings which include any rejected records and other discernible errors resulting from the Census Bureau's program edit series.

C. The local agency reviews the rejects in the Edit/Transaction Listings and the errors in the Edit/Error

LOCAL CUE OPERATIONS(CONTINUED)

UPDATE OPERATIONS - Phase 1 - Update to Current Date



Listings and enters the still needed corrections on transcription forms using procedures provided by the Census Bureau. These are held by the local agency pending the next cycle of update operations. The local agency will also add the continuing updated information to the transcription forms following Census Bureau procedures.

D. The local agency decides whether—

1. It will insert the corrected and updated segments into its file locally and when completed, submit the updated and corrected data to the Census Bureau, or

2. Submit the corrected and updated information to the Census Bureau without updating the local file, or

3. Insert the corrected and updated information into its file locally, and when completed, submit a copy of the updated geographic base file to the Census Bureau. (Arrangements would have to be made on an individual basis between the local agency and the Census Bureau regarding this alternative.)

E. If the agency decides to submit these data to the Census Bureau without updating the file locally, it merely forwards the transcription sheets or keypunched cards containing this information to the Bureau.

F. If the agency decides to update the file locally, it punches the corrected and updated information into the data cards.

G. At this point in processing the file the agency must decide whether—

1. To use the Bureau's UPDIME and ADDEDIT computer programs for updating its file, or

2. To use a combination of the FIXDIME II, FIXCORD, TOPOEDIT and ADDEDIT computer programs to update its file.

In either case the resulting output is an updated and edited GBF. The UPDIME program produces an Edit/Transaction Listing: The FIXDIME II and FIXCORD programs will also produce separate Edit/Transaction Listings for each of the two respective programs at the end of each program run. The TOPOEDIT and ADDEDIT programs will produce Edit/Error Listings, with the different types of errors appropriately flagged.

H. The rejected records from each of the Edit/Transaction Listings are reviewed. If the review shows that too many data cards have been rejected, these cards are reviewed and corrected as necessary. At the same time, the flagged records in the Edit/Error Listings are also reviewed and corrected. In both instances, steps G and H are repeated.

I. The accepted, corrected and updated data (in the form of keypunched cards or computer tape) are forwarded to the Census Bureau or, alternatively, a copy of the updated GBF is forwarded to the Census Bureau. Phase 2 of the Update Operation is completed when, after each cycle, the Census Bureau receives one or more of the following:

1. The updated transcription worksheets or keypunched cards.

2. The updated data on keypunched cards or computer tape after UPDIME (or, alternatively, the Quad-Program Series).

3. A tape copy of the updated GRF.

III. EXTENSION OPERATION

A. The Census Bureau furnishes the local coordinating agency with map sheets for review that extend the Metropolitan Map Series to the SMSA boundary.

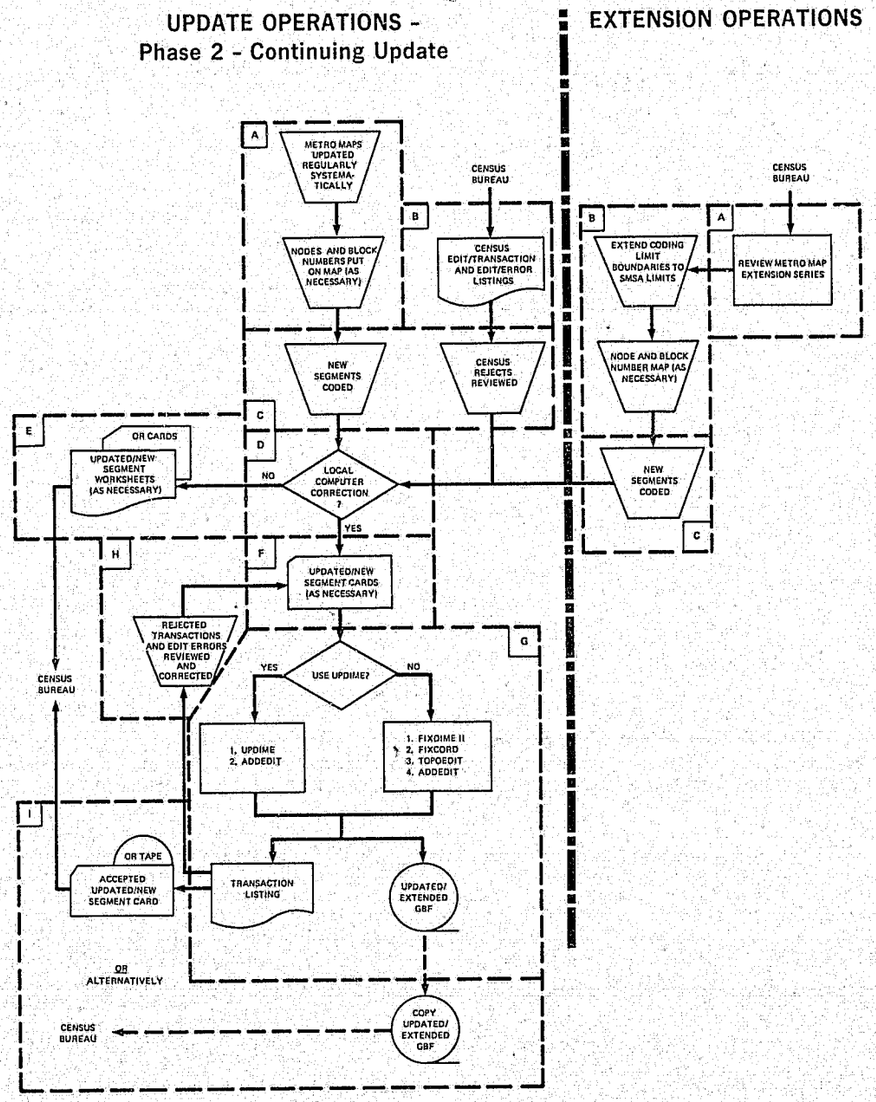
B. The local agency extends the coding limit line out to the SMSA boundary. In so doing, new node points and numbers are identified, and blocks beyond those previously numbered on the Metropolitan Map Series are defined and numbered using procedures provided by the Census Bureau.

NOTE: To promote standardization and to encourage universal use of these procedures, the Census Bureau will assist and review the locally defined blocks and block numbers. The Census Bureau cannot commit itself at this time to use the block definition or block numbers that will be established locally, particularly if the Census Bureau's procedures are not followed closely. However, if changes in block numbers later do take place, then the Census Bureau will provide the local coordinating agency with an equivalency table of the changes made.

C. New segments are entered on transcription forms using procedures provided by the Census Bureau.

D. The steps that follow in the extension operations are the same as, and form a part of, the Continuing Update Operations, starting with Step D. Once this stage is reached it becomes a matter of continuous recycling through the update and extension operations.

LOCAL CUE OPERATIONS (CONTINUED)



Question Period

Mr. Hearle—Thank you, for a magnificent overview of the status and that very appropriate closing. One of the thoughts that came most strongly to me was the fact that we are in a new kind of intergovernmental ball game now where the Federal Government and the State and local government are together trying to establish what could be called a statistical system; or at least a system that requires the active continuing substantial involvement of both to make it work for either. And that is an organizational, a management, and indeed a political problem as well as a technical problem.

Mr. Meyer observed that the technical problems that have inevitably risen to the fore in the last several years are largely solved. I suspect he means principally from the Census Bureau side; that is to say, the technology from the Census Bureau side is in good shape, but there may be a different variety of technology that has to arise from the local side in terms of the use, the application or, if you will, the meshing of the GBF products from the Census Bureau side with the local data and the local data systems.

I think one of the things which this conference might usefully explore, particularly since we have some living, breathing local officials here, is the extent to which you, ladies and gentlemen, who have to make it work in order to get the dollars from your political executives, how you see it proceeding and what you see being required both from the Federal side, other than more dollars (it's sterile to assert simply we need more dollars from Washington). What do you need from the Federal side to mesh what they are doing better with what you are doing and want to do to make it work for you so that your local political executives will see the value and, therefore, provide the support?

Mr. Carlberg—In your list of computer programs, I guess the implication was that all of these are currently available except UPDIME. Is ADDEDIT something that has been distributed or is the processing to be done only at the Census Bureau?

Mr. Meyer—A COBOL version of ADDEDIT designed for local use has been written and is now being tested. It will be available this summer.

Mr. Strohbusch—The question I want to ask is what level are you speaking about when you talk about financial assistance to local areas?

Mr. Meyer—That's a good question, and I really don't know the answer. It depends on the amount of funding that is finally appropriated and upon the size of the GBF/DIME file. If our funding request is fully granted we would expect to provide an average, over all areas, of approximately $10,000 per update cycle. It is not going to be enough in many areas to fund the entire program. However, it is going to enable you to get back to the people you work for and say "Look, the Census Bureau isn't just asking us to do something. They are trying to give us a little help so that we can get started."

Mr. Strohbusch—Can I ask one more question on that? The Three Rivers Coordinating Council is the area transportation planning agency and there seems to be a lot of tie-in or possible potential tie-in between DIME file and GBF with the concept of transportation planning as carried out by the Federal Highway Administration and UMTA. Has there been any cooperation between the Census Bureau and the Department of Transportation to mesh these two together, shall we say?

Mr. Meyer—The answer is definitely yes. The Department of Transportation was one of the agencies which provided major funding for the establishment of the GBF/DIME file originally. There has been and continues to be a great deal of cooperation between our agencies. DOT also considers GBF/DIME files to be an important tool in local transportation planning programs. However, with your permission, I would like to leave the specifics of the GBF/DIME-Transportation interaction to the people who are here from DOT and the transportation planning agencies can describe them much better than I can.

Mr. Kelly—Is direct technical assistance available to municipal agencies rather than just funding?

Mr. Meyer—Again the answer is yes. We do send out staff members from Geography Division to assist with local agencies with their GBF/DIME system operations. Other organizations within the Census Bureau provide programs and technical advice for using the GBF/DIME file. The Data User Services Division for example has on the shelf programs such as ADMATCH programs, designed to help local agencies utilize the file for management information and planning purposes.

Mr. Westerfeld—My question pertains to the Bureau's distribution of the Metropolitan Map Sheets, themselves, rather than updates of the geographic base file. I'm thinking of the availability of high quality updates of the maps in the nature of the 1970 census block maps, but before the 1980 series comes along. I want to distinguish between these that are more user oriented and maps that have the node numbers on them which become too "busy" and cluttered to make much sense out of them. If they were available, it wouldn't be for the people using and maintaining the files, but for those who need to use updated statistical maps in their own right. Do you have a distribution schedule?

Mr. Silver—Are you talking about having a separate set of maps from what we have now? A set with no node dots and numbers, etc?

Mr. Westerfeld—Yes, updated and revised from the ones that were out in 1970.

Mr. Silver—We could provide these. All we would have to do is take off that one layer of the cake, so to speak, the overlay that has the node dots and numbers on it. We could than produce the map sheets you described. But this would have to beat the cost of reproduction.

Mr. Meyer—Let me ask a question of the audience, the answer to which might lead to a modification of our present procedures. Would those of you who are here from local agencies find an updated set of maps without the node number overlay to be useful? Please raise your hands if you agree. Quite a few. Well, we will see what can be done about it. Now that we are aware of your interest we will look for the necessary funding.

Ms. Strickler—Law Enforcement Assistance Administration and also with Federal Regional Council in New England. My question is on the files. What is the smallest size city that you produce them for? Is there a population limit? You were talking about metropolitan areas such as SMSA's.

Mr. Meyer—Our plans now include development of files for the entire SMSA. This would of course cover all cities within the SMSA regardless of their size. The historical basis for the generation of address files were the mailout mailback requirements of the 1970 decennial census. In 1970, the Bureau enumerated about 60 percent of the population of the United States by mail. For the 1980 census, the Bureau hopes to enumerate about 90 percent of the population of the United States by mail. To reach this goal we will need to develop files for cities of 25,000 or more and their environs as well.

Ms. Strickler—So if a community is not within an SMSA at the present time, the file would not be applicable.

Mr. Meyer—That is correct.

Mr. Cooke—To clarify Ms. Strickler's question, couldn't a small city, let's say one that is of 20,000 population, use the Census Bureau programs on its own to produce a GBF?

Mr. Meyer—Oh, yes. I am sorry if I misunderstood you. The programs, themselves, are applicable to any size file regardless of how small the city. However, in terms of the resources that we could make available, our direct participation would have to be limited. But we would provide programs and procedures and help as much as we could.

Mr. Hearle—Could I follow up on that question? Do you have any experience of a city that has set about to establish a GBF outside the Census Bureau system? Has anybody done that, that can be reported on?

Mr. Meyer—The answer is yes, but I am not familiar enough with the details to report on it. Perhaps we could discuss this later in the session. I am sure that Mr. Cooke has knowledge of some of these areas, and perhaps there are others in the audience who could also contribute.

Mr. Carlberg—Could you give us a kind of thumbnail status on DIME file update; that is, how many DIME files there are all together and how many are in various stages of update, how many have been completely updated, how many of them are being extended or have been extended?

Mr. Meyer—When the files were established, at the time of the 1970 census, 233 SMSA's were eligible and could have established GBF/DIME files; 196 of them did. Of those 196, 105 are now currently active or have agreed to participate in the CUE Program. Eighty have completed part 1 or the correction process and 76 of the 80 are completing the address range edit operation of which 15 have run the files through the address range edit program a second time to locate and remove the residual errors from the file. In addition, eight "new" areas are now participating in the program and six of these represent areas which declined to participate in the original program. Incidentally, the 105 areas now in the program represent an increase of about 35 areas over this time about a year ago. The formal paper includes a detailed progress report.

Mr. Weaver—In your presentation, you quoted Dr. Horwood's remark about the competitive situation between private and public sector. Is the Bureau working to enhance the private sector's participation in the GBF area? Are there any trade organization contacts being made or anything that would make the local contact easier or more flexible for funding and participation?

Mr. Meyer—With fairly rare exceptions, the Census Bureau can work only with one agency per area. This is almost always the council of governments or some similar regional group. Otherwise, we couldn't exercise effective coordination on a nationwide basis. For example, there are now 270 SMSA's: Even as few as 10 different agencies, public and private, representing differing and perhaps overlapping geographic areas within the SMSA would mean attempting to integrate the activities of almost 3,000 organizations—a guarantee of complete collapse. So we refer inquiries to the appropriate local coordinator, remarking also that the coordinator will be delighted to hear of your interest and possible collaboration in the program.

Mr. Weaver—If I may proceed a little further. What I was really interested in is the promotion of the technology as a technology nationwide and to make other interests aware—the banking interests, the savings and loan interests—making them aware of the technology. Also promoting it so that the private sector can come to the Bureau to ask who is locally doing this, etc.

Mr. Meyer—Yes, the Bureau supports an active program in this area. This conference is one example. The Data Users Services Division, through its publications also promotes knowledge of the GBF/DIME files. Mr. Voight, the Chief of that organization, will be talking about his activities in that area. DUSD also conducts GBF/DIME workshops, at fairly regular intervals which include sessions directed to the private sector.

The Massachusetts Program

THOMAS HUMPHREY

INTRODUCTION

The Massachusetts Bureau of Transportation Planning and Development (BTP&D), an administrative unit of the Massachusetts Department of Public Works, is the State agency having the responsibility for undertaking comprehensive transportation planning for all modes of transportation throughout the Commonwealth. This responsibility includes both systems planning (the development of long-range plans and programs) as well as project planning (the development of short-range plans and programs) through the completion of environmental impact statements.

The BTP&D was established in 1964 by the State legislature, in part, as a result of the requirements of the 1962 Federal Aid Highway Act which mandated that all proposals for Federal-Aid highway programs must be based upon a comprehensive, cooperative, continuing (3C), transportation planning process. The early and mid-1960's saw the development of urban transportation studies that resulted in the collection of transportation, land use, social, economic and demographic data that filled numerous file cabinets and numerous reels of magnetic computer tape. Unfortunately in Massachusetts, as well as in most other States, the results of the "old" urban transportation studies were not used as effectively as they might have been. One of the reasons for this situation, in my view, was that it was very difficult to maintain the momentum generated by the initial urban studies. There were numerous sets of data and models available, but not nearly enough people or expertise available to use all that information effectively or on a continuing basis.

The utility of the products of the old urban studies was limited by questions of credibility. It took a long time to collect the data needed to do transportation planning, and sometimes longer to analyze the data. Thus, the question of timeliness was always present. Additionally, many questioned the results obtained from the urban studies in terms of adequate participation by citizens, local- and state-elected and appointed officials, and other interest groups. There was the question of the fiscal reality of some of the transportation plans developed; and finally, there were questions related to the social and environmental costs and impacts of transportation plans that had to be considered.

After the appointment of Alan Altshuler as the first Massachusetts Secretary of the Executive Office of Transportation and Construction (EOTC), and Bruce Campbell as Commissioner of the Department of Public Works (DPW), the BTP&D's transportation planning process was reviewed and evaluated and major institutional and technical changes were initiated to revitalize the 3C process statewide. Since the development of transportation plan and program alternatives depend not only upon future desired development patterns, it was decided that the transportation planning process would have to be expanded to include interdisciplinary inputs from other State, regional and local agencies; thereby, providing broader technical expertise upon which to base policy decisions. At the same time, it was decided that the process would have to be expanded to provide a mechanism for full and open participation by citizens and elected and appointed officials at the State, regional, and local levels. During the past year we have made great strides towards achieving the overall goals summarized above.

One of our goals is to pull together State programs and assist in the integration of meshing of Federal, State, regional, and local planning programs. Consequently, at the State level, the BTP&D is coordinating its planning and policy development with the policies and plans developed by the Governor's Resource Management Policy Council (RMPC). The RMPC, formed in 1973, is made up of those State Cabinet Secretaries and their representatives having responsibility for resource management planning. Ultimately, it is hoped that from the cooperative Resource Management Policy Council efforts, comprehensive statewide land use and related development policies and plans can be developed with full consideration of the interrelationships of all functional planning programs.

At the regional level, the BTP&D has endeavored to broaden transportation planning through the comprehensive, cooperative, continuing "3C" planning process guided by a joint Transportation Planning Advisory Group (TPAG) in each of the 13 planning regions that encompass the entire State. The TPAG which was established by a Momorandum of Understanding signed by the Secretary of Transportation and Construction, the Commissioner of the DPW, and the President of the RPA (plus the Chairman of the MBTA in the Boston Region) is a regional forum established to provide policy advice to the signatories on all matters related to the development of transportation goals, objectives, policies, issues, plans and programs.

Having taken what we feel is a lead role in providing expanded intergovernmental tools for comprehensive transportation planning, the EOTC/DPW has inititated a complementary effort in the field of data collection. A key element in this expanded continuing planning program is the maintenance and updating of planning data reflecting current and future conditions. The BTP&D has proposed that planning data files be maintained, updated and interrelated by using the Census Bureau's geographic base (DIME) file system. The BTP&D originally assisted in the development of the GBF for Massachusetts and we have a long standing commitment to using it as fully as possible.

We feel that such a geographically based system should be used to interrelate various agency files. It will provide for compatability between planning regions and facilitate the interchange of information between interested agencies at all levels—State, regional and local. It will reduce inefficiencies caused by duplication of efforts, stemming from use of incompatible data files; thus it should reduce data collection costs. Finally, we feel that it will act as a catalyst in developing a statewide planning process allowing ultimately for a federation of various agency files.

To test the potential for such a system, in 1973 the BTP&D continued its lead role in supporting GBF/DIME expansion by creating a pilot project in the communities of Agawam/Longmeadow, successfully linking road inventory data with census data through use of the GBF/DIME.

BUREAU OF TRANSPORTATION PLANNING AND DEVELOPMENT CENSUS NEEDS

We have introduced the concept of geographical collection and analysis of data to comprehensive statewide and regional planning through the RMPC. We hope that our commitment and investment will be a catalyst for interagency, multidisciplinary use of this technical supportive process. Full participation by other State agencies in the federation of data and information systems rationalized on a geographical basis was proposed to other State and regional agencies by the BTP&D in 1972. It is now being initiated as part of our 1974 work effort.

The BTP&D 5-year work program calls for a substantial commitment to updating and expanding the geographic base (DIME) file to cover the entire State. The work should be complete (or nearly so) by the end of calendar year 1975 (the FHWA funding year). Our plan to expand the geographic base (DIME) file to cover the entire State, not just urbanized areas, corresponds with our major policy thrust of the past 2 years—that is, the expansion of the 3C transportation planning process to cover each of the State's 13 planning regions (9 of which contain urbanized areas, and 4 do not).

While the Federal Government, through its "3C" process and Action Plan requirements, has called for an integrated coordinated, interdisciplinary transportation planning process in urbanized areas, the BTP&D has taken the Federal mandate one step further to better represent its State mandate to carry out comprehensive statewide transportation planning, by setting up a process statewide in all 13 regions. The statewide and regionwide approach recognizes the inadequacy for planning purposes of arbitrary urban-rural distinctions in a relatively small but highly urbanized State. As a result the BTP&D has broadened urbanized area studies to regional studies.

The expansion of the GBF file to cover the State is considered by us to be the first substantial step in building comprehensive and accessible information files for the newly expanded planning process. With the expanded GBF, regional and statewide studies can better analyze and adjust land use, socioeconomic, and environmental factors and transportation data and develop the predictive mechanisms necessary to forecast land use patterns and transportation needs.

EXPANDING THE GEOGRAPHIC BASE FILE SYSTEM

In both the BTP&D's annual and 5-year work programs, funds have been programmed to expand the GBF file to cover the entire State. For calendar year 1974, a modest $100,000 has been allocated and approved by the FHWA for this purpose. While coordinating the effort of editing and expanding the files, we are working closely with the 13 RPA's. Specific work elements with appropriate funding support are being developed in each RPA contract to undertake those tasks. Hopefully, total RPA contracts for comprehensive transportation planning will be expanded to $2 million per year, with the addition of UMTA to FHWA funds. A portion of the funds will be devoted to a systematic extension of the GBF over the next several years. In addition several municipalities, (Fall River, Springfield, and Pittsfield) have already begun some work on their own. We intend to cooperate fully with them and attempt to avoid duplication of effort whenever possible through the assistance of the Bureau of the Census.

THE EXPANDED GEOGRAPHIC BASE FILE SYSTEM IN OPERATION

For transportation planning and other functional planning needs, a basic geographic information collection unit must be established to facilitate interdisciplinary planning. For our purposes we are calling this unit a small area analysis zone (SAAZ).

The SAAZ will be the smallest unit within which geographically based comprehensive planning analysis can take place using all available forms of information. Each agency will be able to aggregate SAAZ's to whatever level necessary for its own analytical purposes. In the case of transportation planning, our lowest level of interest is the traffic zone, which incidentally will often be no larger than one SAAZ and, therefore, identical with it in most cases. It is intended to have general consistency and specific congruency where possible, using tracts, blocks, traffic zones and small area analysis zones so that the development of any system of regional and municipal subdivisions or data cells will provide most of the geographic delineation required for the others.

The SAAZ and geographic base (DIME) file provide a means of relating and defining different area units, while the address and street location attributes permit aggregation of data.

Through much of the State, the SAAZ will be the same size as, and thus identical with, a minor civil division (MCD) or a traffic zone. In highly urbanized areas, it will be, of necessity, smaller, often perhaps as small as and again often identical with a census block. The SAAZ's can be aggregated to MCD level or higher—its boundaries are also coterminus

with the standard units of the Census Bureau (blocks, block groups, enumeration districts, MCD's).

Future network and travel patterns and volumes are products of the multidisciplinary, comprehensive process now established in transportation planning. This process provides land use goals, objectives, policies, plans and programs from which transportation facilities will be planned, designed, and constructed. Of major importance to long-range transportation planning is the projected future functional classification of all transportation systems. Functional systems classification must relate to the State, regional, and local land use and socioeconomic forecasts and goals, and predicted patterns of development derived from the aggregated SAAZ, regional and State analyses.

CONCLUSION

To conclude, we believe that our comprehensive transportation planning program is taking a lead role in the development of geographically-oriented data base files. To determine future transportation needs, data and information must be located geographically and the various economic, environmental, demographic and land use factors for these geographic alternatives must be known and considered. The GBF provides a geographic framework in which the data can be tagged for identification.

Other agencies may have varying requirements for geographically based data. Since it is required for open space, economic development, and housing needs analysis, for example, it is obvious that others can and should use the same system—the geographic base (DIME) file. Meanwhile EOTC/DPW will continue to assist in the development of this file—considered of critical importance to our data needs—and to advocate interagency, multidisciplinary use of this basic resource for State and regional comprehensive planning.

Question Period

Mr. Hearle—How are you handling this decentralized transportation planning through the regions in communities where the Census Bureau had no geographic base file program?

Mr. Humphrey—That is a good question, and in order to give a good answer, I would like to call upon either Ms. Vigil or Mr. Harward, if you would like to respond to that. We do have plans for that covering areas outside the SMSA's.

Mr. Harward—Basically we intend to implement through the Regional Planning Agencies. The Census Bureau is going to furnish the training and overall supervision of the information. The Bureau of Transportation and Planning is going to fund this program through the Regional Planning Agencies. In some of the Planning Regions the core city seems to be taking the lead role in developing the GBF/DIME. We hope to get down to the level of the smallest towns in Massachusetts. Talking about 25,000 population, we have a town with 35 people, and we intend to get to this level. At some point we hope to update the existing file; extend it to the limits of the SMSA; and as a third effort, go outside the SMSA, where I feel we are going to have the most trouble.

Mr. Humphrey—We are talking about a 2- to 5-year time frame to try to complete this work.

Mr. Madigan—I might add a comment on what Mr. Harward said. I was involved in 1966 in the expansion of census tracts to cover all of Worcester County; at that time most of the 10 urban counties in the State were tracted, although there remained conspicuous gaps. Conforming to the Bureau of the Census rule that the average tract size in an area must be over 4,000 population, we were able to establish not only a separate census tract for each municipality of 1,000 or more people but also several special tracts of less than average population for institutional populations. We need to have a push between now and 1976 to finish tracting the rest of this State so that tract data—especially sample data—will be available from the next census for each Massachusetts municipality with a population over 1,000. As you know, the town is the significant governmental unit and, therefore, the significant data collection unit in New England, and we believe that this phenomenon should be reflected in the census tracting. We would urge that, where necessary, two or three small towns be placed in one tract which could later be split along municipal lines. In the case of Mount Washington with a 1970 population of 52, there are options of three or four towns with similar rural characteristics which would make one tract of adequate size (1,500 to 3,500).

Mr. Humphrey—I might point out that we are going to be depending upon the Regional Planning Agencies to provide us with a lot of assistance in the work that I have covered briefly this morning. They are best equipped to do a lot of the work. We hope to provide the guidelines, technical assistance and the money to fund this thing.

Mr. Hearle—Is it your plan to extend the Census Bureau model of the Metropolitan Map Series in terms of scale and other characteristics to the balance of the area that you cover with GBF's, or are you going to develop a different mapping system?

Mr. Harward—Another project that we have is updating our general highways series. We are converting it from a county to a regional series. The needs of the Census Bureau will be taken into full consideration when we develop these maps.

Mr. Thyagarejan—Is the State of Massachusetts thinking or has it thought about the funding of the files from sources besides transportation monies, such as LEAA funds and health planning funds, HEW, HUD and a whole variety of other sources? Or are you entirely dependent on transportation planning funds?

Mr. Humphrey—We are using the transportation funds that we have available as a catalyst to try and get this thing moving because we feel that this is a very important program for transportation planning purposes. We are working with the Resources Management Policy Council, which is a cabinet level ad hoc group of individuals, to see if we can't get additional funding, and hopefully Mr. Tom O'Brien will cover that tomorrow afternoon. Mr. O'Brien is the Director of Office of State Planning and Management, and he has a real interest in this very issue. I am not sure what other funding is available, but we are working very closely with Mr. Tom O'Brien, and I think we should get funds in order to complete this program.

Ms. Strickler—Are you coordinating with the seven Criminal Justice Planning Areas in Massachusetts?

Mr. Humphrey—I don't think so. We have all kinds of planning boundaries in Massachusetts. There is a book several inches thick, one page being a map of the State showing different kinds of boundaries of one purpose or another. There are health planning boundaries, regional planning boundaries (the regions that we are dealing with); there are all kinds of boundaries. Another effort that is under way within the State through the Resource Management Policy Council, is an attempt to try to unify these planning regions or these planning boundaries. Unless other members of my staff here know otherwise, I don't beleive we are.

Mr. Kondo—One of our problems is to find out exactly what sort of planning districts are being used by agencies in the State. One of the things that we have not yet considered is the Law Enforcement District because we really were not aware of its existence. I think we are interested in talking to people in other agencies who have different areal units, and to see how we can coordinate our work. As Mr. Hunphrey mentioned, the special area analysis zones, were developed with that thought in mind, that is, to try to develop some common basic unit of area by which we can interchange data.

At this point we have not considered the Law Enforcement Districts. However, we would be interested in learning more about them and how we can incorporate them into our zonal setup. If anybody else has other areal units that we are not cognizant of, please let us know.

Mr. Harward—I would like to add to what Mr. Humphrey has said. The Bureau of Transportation Planning and Development is running on half-throttle on this thing, deliberately so. If we go full speed we are nervous about just taking over the whole thing and turning small area analysis zones into traffic zones. That is one thing that we don't want to have happen. We wish other agencies would give us some input to this thing through the Resource Management Council. Otherwise, we are just going to control it, and that is not a thing that we want to do.

Mr. Humphrey—There are certain things we have to do as we attempt to develop our transportation planning program; we want to move carefully to make sure that what we do is compatible with other agencies, so we don't duplicate or overlap.

Mr. Cooke—If you worked with the GBF/DIME file you find out that it is a pretty simple technical matter to add special area codes, such as Crime Districts, to the GBF/DIME. The problem is one of coordination with the agency that is going to use the particular special area codes. There is, however, one very difficult technical problem, which we will face in Massachusetts; we don't even have to go to Mount Washington Town to face it. You face it within the Boston SMSA, in Bedford for instance, where there is rural route post office delivery rather than a house number/street name system.

To restate this, one of the biggest applications of the GBF/DIME is in address coding, that is, translating data that are address-related into census block or coordinates. Obviously, you cannot do address coding if you do not have addresses. I was wondering if you were considering attacking this problem somehow?

Mr. Humphrey—We anticipate a lot of problems, no question about that. Of course, we are hoping to learn something from this conference. We are also hoping to take advantage of the expertise that is available within the State government and within the various Regional Planning Regions to help us overcome some of these problems.

Mr. Strohbusch—We are having that same problem in Fort Wayne. We are surrounded by a very rural area, and we are working with the County Planning Commission. Before they will give a building permit they assign a street address, even though it may be 16000 Jones Road or something as such, but it is helping us. We feel it will take time to be accepted; however, I think if you work with the local planning agencies, they might be able to help with that problem.

Mr. Hearle—This is a very significant issue that Mr. Cooke raises. I wonder if we could advance the state of our collective knowledge by inquiring if the practice seems to be universal among the local agencies that you represent that the street address is a creation of the local agency, and which local agency. I understand in Indiana it is the Planning Agency. Is that correct? Do they officially number the street and everybody agrees? How about Massachusetts? It is a function of the planning agency or the city engineer, or who is the official who controls the designation of addresses?

Mr. Controvich—In Massachusetts it is normally the Department of Streets and Engineering or the Department of Public Works which assigns the address sequence.

Mr. Hearle—So in the case that you raised, Mr. Cooke, who do you ask to change the rural routes into street addresses? Is there a city engineer? Is that the appropriate official? Who is the guy that has to be persuaded to get out his drafting pen and do the job?

Mr. Kelly—They are outside of municipal limits. Most counties in the United States don't have that kind of action. It has got to come from the State since most of those roads are State maintained.

Mr. Hearle—That is helpful. In many sections, however, the county engineer performs the function for the unincorporated area outside of the municipality.

Mr. Kelly—Out of 3,000 counties in the United States, how many do you think have a county engineer?

Mr. Hearle—Several hundred?

Mr. Harward—In Massachusetts we don't have any unincorporated areas, which makes the Census Bureau and everybody else love us because it just confuses things.

Mr. Westerfeld—The practice in Maryland, which is a situation that has strong county government and no minor civil divisions of political importance, is that the responsibility is divided between either planning or public works agencies. Although in some instances neither moves, so the Post Office Department gets in and assigns on its own prerogative.

Mr. Kunz—The situation in the counties of Nassau and Suffolk, N.Y., where I am from, is that we don't get the numbers assigned unless we get an agreement from the Post Office to provide city delivery. I don't know whether there is a similar situation in other areas, but you first have to have that agreement here. Each of our 108 local municipalities has the power and the ability to assign numbers, but they will not do it unless they are pushed by the Post Office and get that agreement. At present 25 percent of our 115 postal areas do not have city delivery, and we cannot possibly code them until we get some kind of incentive.

Pragmatic Use of the DIME File

JACK SCHOFIELD

It gives me a great deal of pleasure and it is a privilege to have traveled from Las Vegas to be with you and to be one of you here at this important conference. I would like to begin my presentation with a welcome from the Honorable Mike O'Callaghan, Governor of the State of Nevada, whom I had the privilege of teaching school with back in the fifties, and the Honorable Mayor of Las Vegas, Oran Gragson, and welcome each of you to visit out beautiful city, and to please try and visit it often and leave some of your happiness there.

Those of you who have flown over that State know that it is a barren place except for the casinos. That reminds me of the attorney that drove to Las Vegas in his $10,000 Cadillac and returned home 3 days later in a $100,000 Greyhound bus.

Many of us in Las Vegas believe that it is a very progressive city and we are attempting to pragmatically use the geographic base (DIME) file, and we have used it. We also have had an exercise in cooperation with the State Highway Department of Nevada in creating the usefulness of the GBF/DIME. I produced a slide presentation and thought we would see this first, and then have questions and answers afterwards. (Slide presentation follows)

Las Vegas, Nev. is situated geographically in the southeast corner of the State of Nevada near Hoover Dam and Lake Mead. The county is named Clark, and geographically, one of the largest school districts in the United States. The State of Nevada is comprised of 17 counties, Clark County having the greatest population. The population of the county is approximately 330,000 plus, and the school enrollment as of March 22, 1974 was 77,550 students, in kindergarten through the 12th grade. There are 41,264 elementary students, kindergarten through the 6th grade; 18,842 students in the junior high schools, 7th through the 9th grade; and 14,679 students in senior high schools, 10th through the 12th grade; with 2,260 special education students.

In the city of Las Vegas there are 56 elementary schools, kindergarten through the 6th grade; 14 junior high schools, 7th through 9th grade; 9 senior high schools, 10th through 12th. In our rural areas, as seen on the map, in the north, south, and west quarter of the county there are 11 elementary schools and 5 junior/senior high schools. The superintendent of Clark County School District, Dr. Kenny Guinn, a young capable educational leader, has had this position at this time for 4-plus years.

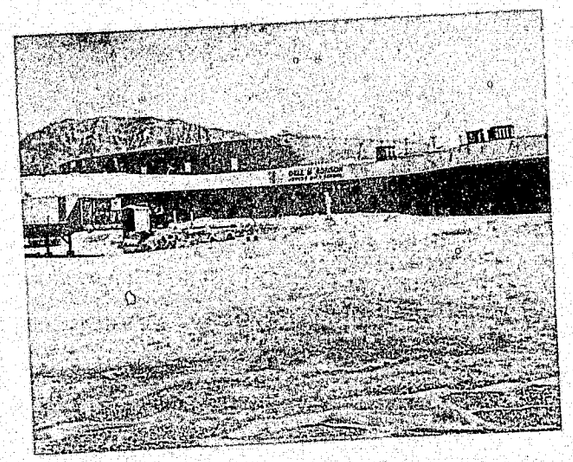
Las Vegas has earned the title of the entertainment capital of the world and is better known for this type of thing. It is difficult for those who live elsewhere to realize the normality of this city, where the people live and work in the typical type of location and industry, or that we have one of the largest school districts in the United States. A number of us, who have lived in the county for many years, take pride in the fantastic growth of this unusual city because of its unique and beautiful buildings such as hotels, motels, churches, schools, lakes, parks, ski resorts, and recreational areas, such as hunting, fishing, hiking, and an unlimited supply of sunshine.

The majority of the population of Clark County resides in metropolitan Las Vegas, as shown here. The green dots denote property owned by the Clark County School District. The blue dots denote existing school buildings.

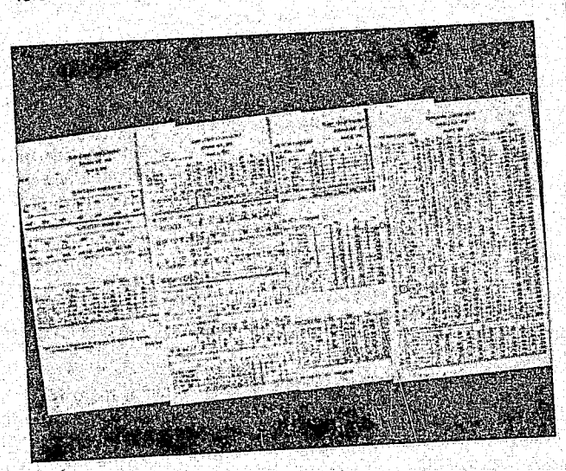
The foundation for the budget that determines the needed money to run a school district of this size must be determined by the number of students that will attend schools for the next 2 years. These complicated projections must be calculated in order to create this budget. The projected number of students that will be in school the following year and how these students will be zoned, so that each school will enroll the number of students within its enrollment capacity, is the responsibility of the Planning and Zoning Department.

A great deal of planning must take place prior to building new schools. First of all, a determination must be made of the needs in any given geographical area. Dr. Charles Neely, on the left, and Mr. Stanley Bokelmann on the right, are examining plans of one of the schools prior to the final approval by the school board so as to minimize mistakes in drafting, concepts, engineering, playgrounds, etc. Prior to this, Dr. Neely has had to write educational specifications for each type of school, and its current concept and philosophy prior to awarding the architectural contract to draw the plans.

As indicated here in several slides following, there is a great deal of construction continuously taking place in Las Vegas and Clark County. Scenes following show the beautiful clear air, blue skies, and scenery that prevails in Las Vegas. This is a typical junior high school, currently under construction. The concept of a school all under one roof. The name of this junior high school is Dell Robison, and it will be completed in September 1974. Robison Junior High School is shown here with a portion of Sunrise Mountain on the right.

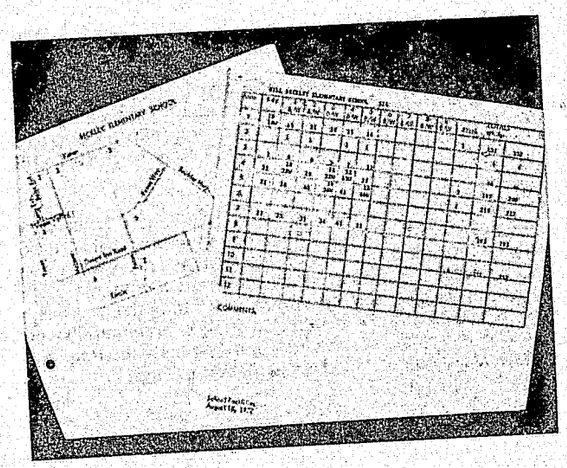


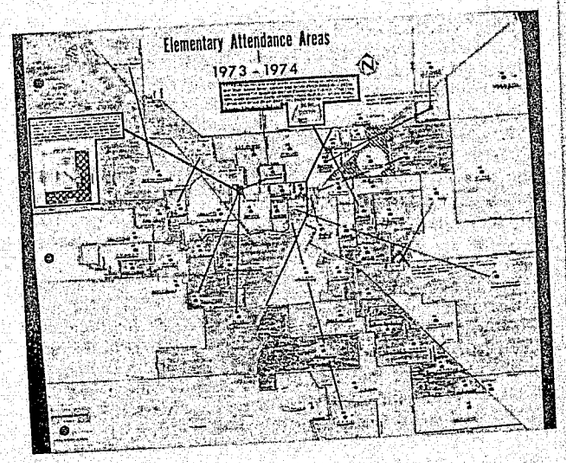
These four pages show the ultimate result of projecting students for the following year. These projections must be made on a grade-by-grade level of advancing students by using what is called a grade-progression ratio. This is calculated by historically reviewing the last 5 years of growth by grade level from kindergarten to 12th grade and by grade average using the percentage of growth. Also considered are the births in the past 5 years. This percentage ratio has been applied to each grade level in each school by determining the following year's projected enrollment. It is interesting to note that last years's projections came within 49 students of actual enrollment.



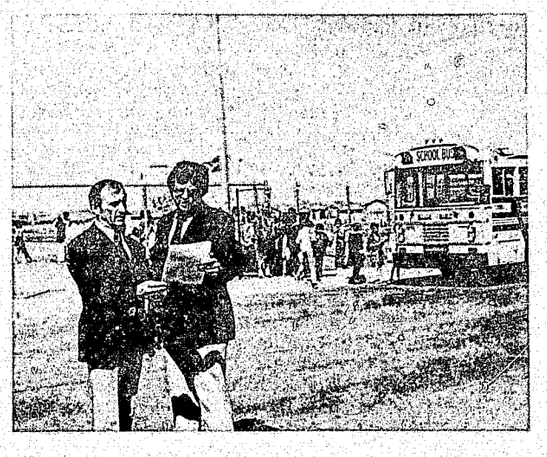
This is our elementary projection on the right. The one on the left shows our sixth grade centers, junior highs and senior highs in the Las Vegas area. The one on the left is the rural attendance area schools, kindergarten to 12, and the one on the right is a summarized total of all schools in the Clark County School District.

The method used in the past was at best inaccurate and inefficient, and was accomplished with resentment by secretaries, principals and teachers in each of the schools. Prior to our introduction to the DIME file, created by the Census Bureau, this system was as efficient as any could be. As indicated by the page on the left, the geographic boundaries of each school zone in Las Vegas were drawn by secretaries in the School Facilities Division's zoning office. The secretaries in each of the schools would in turn count the number of students in each geographic zone and record it on the form shown on the right. Not only would they sometimes be inaccurate in the count, but if a smaller zone was needed many decisions had to be left to a calculated guess which was traumatic at times. This slide shows the elementary attendance areas of each school, color coded so as to make it easier to read. A number of these areas must be changed annually. Some, however, remain fairly stable. This is one of our junior high school zones and this, our senior high school zones.

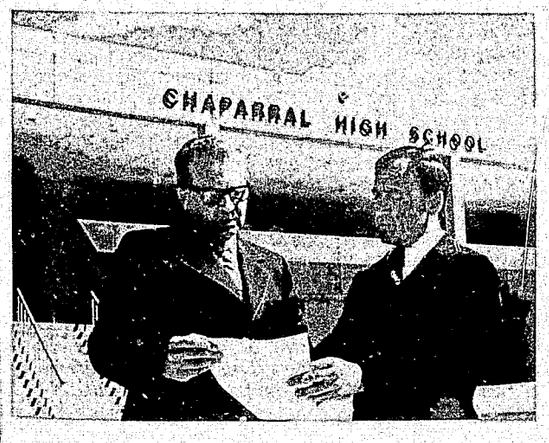




The slide here indicates a court-ordered integration plan in order to satisfy a Federal court's mandate that all class rooms in every school in Clark County School District be not more than 50-percent black. The gray shaded area indicates the seven sixth grade centers that are located in the geographical area of Las Vegas. The residents living in this area are approximately 95-percent black. In order to achieve a Federal court-ordered mandate of 50-percent black-white ratio, a plan was devised to use the seven existing school buildings in the predominately black residential area as sixth grade centers. This was implemented by busing all white sixth grade children from the so-called white schools to the seven sixth grade centers and busing the black students who reside within the sixth grade centers' school zones to the matching white schools as indicated by the lines drawn from each school. The most important criteria was that feeder-school alignment would be followed as much as possible. This we have been able to accomplish with very few exceptions.



Here is a typical elementary school dismissal at the end of the day with the black and white students both boarding the bus to their homes. The principal of William E. Ferron Elementary School, Mr. Roger Gehring, is assisting in gathering data for projections and zoning. The C.W. Woodbury Junior High School principal, Mr. Francis Cortney, is assisting.



This is a typical model of the junior high schools that we are constructing in Las Vegas. Mr. Carroll Russell, Assistant Principal at Chaparral, is assisting. Chaparral High School is the newest high school in Las Vegas, and is the two-story concept. It will house approximately 2,500 students.

Dr. Thurman White, Associate Superintendent of Facilities Division, Clark County School District, was a key party to the DIME file innovation. Under his direction and that of Mr. Ed Greer, Associate Superintendent of the Business and Finance Services Division where the computer department is housed, and his Director, Mr. Ron Jones, the concept of establishing a method of matching students to given geographical areas was born. Dr. Neely, second on the left, Director of Planning and Zoning, School Facilities Division, was also one of the key contributors in adopting the DIME concept along with Ms. Barbara Van Norsdall, Coordinator of Programming in the Data Processing Department. Our department was directed by Dr. White and Mr. Greer to research and discover a way to address match students in order to determine the exact number of students by grade level living on any given block within the Las Vegas attendance area.

After visting the school district in Costa Mesa, Calif., and the Pasadena School District, and inquiring into the Sacramento School District and several others in northern California, a decision was reached that developing geocoding by x-y coordinates, using these systems, would not be feasible or economically feasible for Las Vegas. An improvement over our existing program would take a great deal of money and man-hours to digitize and create a useful efficient mapping system.

Our agency did not know that the DIME file existed in the manner that it could be used as a tool for ADMATCHING students until the Clark County School District, through a trial and error approach, made contact with the Nevada State Highway Department. When the connection was made by the two agencies, their representatives, Mr. George Westenhoefer, Chief Planning and Surveying Engineer, and Mr. William Flaxa, Highway Engineer, could see the potential benefits of a coordinated effort to update the outdated GBF/DIME file for the Las Vegas area. The Las Vegas GBF/DIME file was current to June 1968 and needed to be updated to a more current status, which at that time was June 1973. Through a cooperative effort from both the Clark County School District and the Nevada State Highway Department a united effort was enacted to expedite this monumental task.

The project cost approximately $20,000. This included five hand-picked college students, four Nevada State Highway Department employees, and four Clark County School District employees who work together updating the file to a useful condition so that the Clark County School District computer department could begin ADMATCHING students with the DIME file. The project took approximately 6 months, from June to November 1973.

The first ADMATCHING involved 77,000 records and reflected a 12-percent fallout error with a goal of 98-percent match rate. A subsequent match achieved this 98-percent goal. This initial ADMATCH was for students grades 7 through 12. More corrections and refinements were necessary to meet the criteria agreed upon. This achievement allowed the Zoning Department of the Clark County School District to begin and complete the projected zoning for the 1974-1975 school year. Also, the projected number of students for the same period of time.

The Planning and Zoning Department of the Clark County School District is responsible for continuing the update of the GBF/DIME file, for ADMATCHING, and for serving as liaison between the school district, Data Processing Department and the Nevada State Highway Department.

The Clark County School District Data Processing Department is under the direction of Mr. Ron Jones; the Highway Department of Planning is under the direction of Mr. George Westenhoefer, in Carson City, statewide; and Mr. Bill Flaxa in Las Vegas. Ms. Barbara Van Norsdall developed a student zoning procedure manual to assist the system. Subsequent to the secondary student ADMATCH the next major undertaking was to ADMATCH the kindergarten through sixth grade students to the GBF/DIME file. Thanks to Dr. Thurman White and Mr. Ed Greer, approval to continue zoning students on the secondary level, junior high and high school was given. Zoning students to a degree of accuracy never before achieved in the zoning department of the school district was accomplished because of our ADMATCH accuracy. Immediately thereafter, the zoning was accomplished on the elementary grades with the same success. The greatest benefits achieved from this program was an accuracy level that was considered possible, but still not known until we actually implemented the zoning procedures using the GBF/DIME file in ADMATCH. By zoning students, using this method, we will be able to determine specific zones which will provide us with accurate, current data on students from grades kindergarten through 12th for immediate planning purposes. In addition we will be able to utilize the CRAM program for the purpose of rezoning for more efficient use of the taxpayer's dollar and more efficient educational needs for students. The changes the school district had to make to use the GBF/DIME file moved this agency from an archaic, obsolete method of zoning and projecting students to a much more sophisticated, efficient, accurate method, which will hopefully save the taxpayers in Las Vegas and the State of Nevada, a great deal of money in the future and provide for a more streamlined efficient timesaving program.

Prior to our involvement in this program in 1973, our experience in computer mapping was nil. At the present time, we now have a sophisticated system of mapping in our Zoning and Planning office with 4-foot by 8-foot wings. These wings are tackboards held in vertical position by a metal pole. The method of installation allows all eight wings to rotate in either direction so as to view both sides of the tackboards. The GBF/DIME file maps are able to be viewed most efficiently with this system. We can now use the DIME file with elementary school zones color coded, thus indicating the exact number, block number and nodes so that at any given time a member of the staff can record and request given geographical point-in-polygon numbers. Then with computer time available, we can provide exact up-to-the-minute data on students in that geographical zone that meet the criteria requested. The same information for two more wings is available to all junior high school zones and two more wings for the high school zones. The last two wings are a most helpful device. It is a one-half reduction of the GBF/DIME file which we call the nickel map, pieced together into one map which under glass indicates by color code, one for elementary, a second for junior high and a third one for high school zones. This gives an overview of the three zone boundaries in the three grade levels. Plans are in progress for coordinating data with other agencies, such as the County Data Processing Department, the Regional Planning Department of the Clark County area, the county police and fire departments, and others to provide a special data bank for the use of fire departments, police departments, sewage, sanitation, etc. This way, we effectively coordinate data to meet the many complex needs of local governments.

In conclusion the people in the Las Vegas area, Clark County, in our opinion, are being served better, more economically, more efficiently and accurately by the acceptance of the Census Bureau concept of the GBF/DIME file in a more pragmatic way than ever before and here is an outstanding example of two or more public agencies cooperating, one with another, to serve the taxpayers in a political geographical area. It can be accomplished and is in Las Vegas, Nev.

Question Period

Mr. Hearle—A couple of questions on the mechanics of making it happen. You have slides of a number of people and you mentioned the number of dollars and a little bit of the amount of time it took. Would you review that again, so that people can have in mind for a city the size of Las Vegas, how much manpower, how many dollars and how much calendar time is required?

Mr. Schofield—Last spring, about a year ago, our School Board directed our Department, the School Facilities Department, to get with this thing because at best we knew that we were spinning our wheels as far as numbers were concerned. From the practical standpoint, they wanted us to upgrade the estimating and assignment procedures and do something better than what we had been doing. They directed us to look into the possibility of zoning students through the use of computers.

I was happy to be involved in this because I am a member of the Nevada State legislature. I wear two hats. I am a member of a public agency, the school district, and also, I am a member of the State legislature. When we are in session (our legislature meets on a biennial basis—we must for 3 months out of every 24), I take a leave of absence from the school district and go into the legislature. In the Assembly, that is our House of Representatives, we have to be re-elected every 2 years. Our legislature is made up of 20 Senators and 40 Assemblymen. My philosophy of government is that we need a balance of all types of people in our legislature; we even need public employees. This is my strong belief.

Through this position that I hold in the 1971 session of the legislature, I had my first taste of the practical use of computer data. As a decision maker in the legislature I appreciate data that I can trust in and rely on. It seems that if you give somebody data off of the computer, it is like saying God said it. We take the data and we look at it, and we make our decision based on it.

We did this when we reapportioned the legislature. We knew that there was a disparity because of population in the State of Nevada, and since 1864, when Abraham Lincoln allowed Nevada to become a State, the cow counties (we call all the rural counties the cow counties; they don't like it, but we call them that anyway) have controlled the legislature because they had one Senator from each county—17 counties. There might be only 20 people in one of the counties, but they still had a Senator. They were very strong, and the practice was very unfair. In the 1971 session we were to reapportion the legislature, so we budgeted $30,000 for the "computer people" to come in and give us data, so that we could more fairly reapportion our legislature. We used this data; it was efficient; and it was fast. I was convinced (right or wrong) that they gave us good firm figures. As a result of this we did a good job in reapportioning. Now each of our 40 Assemblymen represent 13,500 people in a given geographic area.

As I previously said, the School Board advised us to get underway and improve our procedures, which we did. This was about a year ago. Through trial and error we finally discovered that the State Highway Department was updating the GBF/DIME file. We could see after checking with the Costa Mesa School District and with Pasadena, etc., that they were doing their own x-y coordinate digitizing program and also developing their own node systems. We couldn't handle what they were doing because it was too expensive and it would take too long. That is when we discovered the Highway Department was working with the Census Bureau and the GBF/DIME files.

We had several meetings with them and made a decision to cooperate with the Highway Department and participate in the CUE program. I believe that this is an exercise in cooperation between public agencies that will benefit the State of Nevada, the people, the school children and everybody involved. At the present time we are working with the other agencies in the county and the State to form an information system, a central information system, so that the police department and all of the public agencies can make use of the data.

It has been a very helpful, useful, tool for us. We are convinced that it is the way to go, and we are going to continue in that direction. I think that in the long run we are going to be far better off, and the decision makers will have more accurate facts for them to make their decisions.

New Federalism and You

VINCENT P. BARABBA

I realize that many of you have a general idea of the Bureau's overall mission, but as a backdrop I would like to give you a one-sentence description of what I think the Bureau's mission is. I would say that our job is to collect and disseminate accurate, complete, and timely information for statistical purposes—from and to individuals, business and industry, and State and local governments. That's it.

We have about 7,000 people to do the job. We work in five general areas of activity. Two of these are subject matter areas, I refer to demography, which involves people and households, and the economic institutions. The other three are the service-oriented areas—data collection, electronic data processing, and statistical standards and methodology. These responsibilities are handled through 27 divisions and offices. The end result is thousands of printed publications and a library of more than 200,000 computer tapes.

When I came to the Bureau about a year ago to be its Director, I did not imagine that I would or could absorb so much so quickly. In some ways I already feel that I have been there for years. The Bureau is an exciting place to work, with an excellent, hard-working staff. The quality of our professional staff is excellent.

There are many things that are going on at the Census Bureau that I would like to talk to you about, but time won't permit. Therefore, I will just touch upon three areas which I believe will be of interest to you. First, I want to talk a little about the transition that the Bureau has been undergoing, particularly in respect to our mushrooming workload and our reorganization.

Next I want to discuss the geographic base files and the cooperative efforts that will be required so that they will continue to be an effective tool. And third, I would like to present some innovative graphic displays which we are developing at the Census Bureau for presenting statistical data.

Right now, the Bureau is in a period of transition which, by itself, generates a unique set of problems. People usually think of Census as a very old and stable institution. Even though the first census was taken in 1790, the Bureau is only 70 years old as a permanent organization. And don't let that 70 years fool you. A lot of our activity is of very recent origin. I am going to give you two examples, and they are not adjusted for inflation.

First, the Census Bureau's work for other agencies has become more than half of the work we do each year and the demands on our expertise and services continue to grow rapidly. This reimbursable work has grown from a level of about $20 million 4 or 5 years ago to more than $50 million this fiscal year. What is less well known is that the annual program which is funded directly by the Congress has also grown dramatically in the past few years. More than half of the work supported by appropriations has come on board in the last 5 years.

My point is, at the Bureau it is no longer business as usual; new work arrives every month and there is no end in sight. We are proud of our capacity to absorb this growth, but we also worry about keeping up with the pace.

A good example is our role in revenue sharing. Our annual surveys, which produce the data for revenue sharing, reach more than 38,000 State and local governmental units. We are servicing many requests for help, and we review the returned forms and compile and edit the data.

Our new role in revenue sharing demonstrates the expansion of the Bureau's role from a statistical-gathering agency to a government accounting agency. Since the equity and accuracy of the revenue allocations must be assured in all cases, we have to view the data not as statistics but as ledger entries in each government's account.

Our computer workload has become enormous. Our daily movement of tapes from the library to the computers and back to the library amounts to about 10 tons. We print more than a million lines a day. The annual rate of growth in our tape library is equivalent to the total size of the library of the Internal Revenue Service. Put another way, the annual growth of our footage of magnetic tape would stretch about half way around the world.

Compounding the situation is the fact that we are a labor intensive computer system. We need more people than most centers because of the volume of tapes we run past our computers.

We are coping with our workload in a number of ways. We have a major program to update our hardware. We are moving toward mass memory storage and are looking forward to the installation of the latest tape drives for greater speed and more data per tape. We are looking at new publications routes to get away from printing everything on computer line printers. Greater data storage will help us retrieve less expensively and more quickly.

These transitions in the area of computerization are designed to help us be more responsive to our data users in light of the kinds of requests that we are getting.

Another major transition in the Bureau is in terms of organization and staffing. There have been two reorganizations in little more than 2 years—the major one being the consolidation of Census and the Bureau of Economic Analysis into the Social and Economic Statistics Administration. These changes have not been fully assimilated. We are still polishing our working relationships and adjusting to the problems that arise from organizational change. Our supreme objective here is to maintain the levels of professional competence and integrity for which the Bureau is noted. We are working hard at this.

Let us turn now to the development and maintenance of the geographic base files and the associated Metropolitan Map Series. These two programs were developed initially to furnish geographic tools to meet the needs of the 1970 Census of Population and Housing. They were not designed to encourage or implement the goals of the New Federalism. Nevertheless, the files have, in their own way, spawned major activities which are bringing the idea of the New Federalism home to local communities. We are seeing an increase in cooperation among Federal, State and local agencies and organizations in both the maintenance and the use of the files and maps.

For those of you who are not familiar with the ingredients of the New Federalism, let me give you a brief description.

There are two basic principles involved. One is cooperative effort to improve management. And the other is increased involvement of the private sector. More specifically, three goals should be met to make the New Federalism a success:

First, Federal, State and local governments should cooperate in developing skills in management, planning, and evaluation—to enable all levels of government to function more effectively and efficiently.

Second, the Federal Government should promote research to improve government operations at all levels and to improve State and local management capability.

And third, the Federal Government should create the proper climate so that the private sector can generate ideas and use resources to create more effective government.

The Census Bureau—with interagency support—has worked with local and State agencies to establish and maintain the management and operational tools of the geographic base files and the Metropolitan Map Series. These agencies represent a wide spectrum. They include HUD, DOT, HEW, and Law Enforcement Assistance Administration, and the Defense Civil Preparedness Agency. These geographic tools, in turn, are being used to organize the voluminous amounts of data contained in local administrative records. Organizing the data into a geographic framework is producing answers to a number of local questions—for instance, the distribution of low income housing and new housing starts; which parts of the city and suburbs have the welfare recipients; where the school-age children live in relation to the school they attend; changes in the geographic patterns of crime; where people live and where people work. I am sure that most town fathers in the State of Massachusetts have sought or will be seeking such answers.

So the Bureau has taken a major step in helping to make local data more understandable. Besides providing the GBF/DIME program—and the system for maintaining files and maps—we are offering additional user-oriented programs such as ADMATCH (or UNIMATCH) and GRIDS. The end result is that local data become more usable to those in decisionmaking positions—people like the mayor, the councilman, the public health director, the president of a bank, or even the chairman of a neighborhood interest group.

Mort Meyer ended his presentation this morning on the thought that users of the file would benefit by establishing a local program to maintain the GBF/DIME and the Metropolitan Map Series on a systematic and continuing basis. I would like to expand on this and describe the cooperative effort that will be required. Cooperation between local agencies and organizations is essential. Cooperation among Federal, State, and local governments is essential. This cooperation is necessary to make the GBF/DIME an even more useful, more workable tool.

Let me be more specific: Take the seemingly simple subject of street names and address systems. Local communities and counties should take a good look at their street-naming and address-numbering systems. In order to make optional use of the "tools"—in other words, in order for the GBF/DIME to carry out the service being asked of it—the street name and address numbering systems must be more logical and systematic. For instance, streets which have the same name but are located in different parts of the same community make it extremely difficult to work with the GBF/DIME. So does an address-numbering system which varies from block to block. These irregularities diminish the capability of finding the right geographic "pigeon hole" to which one can identify and aggregate data.

Along this same line, address systems used in the city need to be extended so that the files can serve a greater number of regional and county programs. As an example, it would be pretty difficult to locate geographical information concerning babies born last year if the address information were limited to, say, "Box 42 along Rural Route 6." Many of you who are working with the urban transportation study programs here in Massachusetts have been aware of this problem in trying to determine the distribution patterns of home-to-work trips near the fingers of your study area. Dealing with this situation requires the assistance of the Postal Service to help establish and implement city types of address systems.

A second area where cooperation is needed is in obtaining information concerning geographic change. This in my opinion is one of the most critical areas in any continuing maintenance program. There is a need to get the various local agencies and organizations to funnel this information on a systematic and periodic basis to the agency or group which is maintaining the GBF/DIME file and the Metropolitan Map Series. I'm referring to agencies which know about changes in street addresses and names, or changes in political, operational and statistical boundaries—agencies such as the assessor's office, city or township engineers, or the utility companies. On the other side of the coin, the agency or group maintaining the file must be capable of taking the information and updating the files on a systematic and periodic basis. It also must be capable of responding with a corrected and updated file upon request.

A third area of essential cooperation concerns the Federal and State levels. They must recognize the local uses being made of the GBF/DIME to meet the requirements and necessary inputs into the Federal and State programs. There are excellent examples of such recognition within several Federal agencies. These agencies use a geographic reference file to determine the distribution of social, economic, physical, or action-oriented characteristics essential to their programs. Examples are the Federal Highway Administration in its comprehensive urban transportation studies and continuing surveillance programs, the LEAA in its law enforcement management programs, HUD in its housing survey programs, the EPA in its environmental impact programs.

As part of this recognition I recommend that collective support to local CUE maintenance programs be instituted. Such support will permit a more central approach in handling the maintenance, update, and extension of files—and a more standard approach, locally. This will help avoid duplication of effort and cost which occurs when each organization operates independently—but uses basically the same types of geographic data. With collective support, the cost to each agency is less, while the number and value of the benefits increase.

Collective support permits each Federal, State, and local agency or private organization to operate from the same basic set—the same reservoir, if you wish—of geographic data. They all use the same street name, address number range, block, Census Bureau and local community geographic identifiers. The cooperative approach also will provide the maximum benefit to the local user. And if he benefits, each of us benefit as well.

Let me make a very important point here. I have been talking about working with people's addresses. This could raise the question of an invasion of privacy. One of the values of the GBF/DIME is that it can maintain private information in confidence while at the same time provide meaningful aggregate information. The process takes the actual address and assigns it to area units, such as blocks, tracts, transportation zones, school districts, or health districts, and eliminates the address itself. Likewise with information that could personally identify the respondent. In other words, instead of being the culprit, the GBF/DIME in this case literally supplies anonymity by converting the "personal" information into "area" information.

This would hold true whether we collected the information or somebody else collects it. A case in point would be information from local physicians. Doctors in an area could furnish actual addresses where a specific health problem occurs—this could be cases of V.D. or cases of severe emphysema. They could run the information through their local medical society computer—most of the societies have access to one—and turn the aggregate information over—minus the addresses—for use in a health information system.

The final subject on my agenda today is a description of some new color mapping techniques being developed at the Bureau for displaying statistical data. Mr. Schweitzer will be discussing the technical aspects of these procedures during his presentation tomorrow. What I would like to point out today is the potential of these techniques and their visual impact for displaying socioeconomic characteristics.

A slide presentation followed. Examples of the illustrations may be obtained by writing to:

Chief, Geography Division
Bureau of the Census
Washington, D.C. 20233

Columbus Announces

JAMES R. THOMAS

Columbus, Ohio (SMSA 1840) has a corrected and updated geographic base file (GBF). We have progressed from an address coding guide (ACG) to an uncorrected digitized GBF/DIME file through two correction phases and an update.

Our current GBF file contains 37,088 records that define the city of Columbus and immediate adjacent jurisdictions which are census block numbered with geographic locations per block face and in the standard Census Bureau format.

The creation of the ACG and then the initial DIME file was accomplished through the combined efforts of the Bureau of Census and Mid-Ohio Regional Planning Commission with the Census Bureau performing map maintenance and data processing and Mid-Ohio coding source documents for input.

Two years ago, there seemed to be no way that the Columbus GBF could be corrected and updated. There was no money available from the Census Bureau to continue the program. Among the local organizations with money, there was little interest, while the organizations with a keen interest seemed woefully short of funds for this purpose. One of the most enthusiastic supporters of the GBF, Chief of Data Processing Mid-Ohio Regional Planning Commission (MORPC), began probing the community to identify available resources. He found a ready ally at one of the savings and loans, and together they approached the Chamber of Commerce seeking assistance. These enthusiasts for the GBF found an understanding audience with the Chamber President. The Chamber President listened attentively and was favorably impressed with what he heard.

The Columbus Area Chamber of Commerce became involved with the GBF for its value as—(1) a planning tool, (2) a marketing device for branch locations, (3) a mapping facilitator, (4) an aid to school districting, and (5) an asset to the community.

The Chamber then launched a program to measure the community interest and to obtain the necessary financial support to correct and update the Columbus GBF. The first step in this process was to convene a meeting of data processing users in the immediate area to inform them of potential utility of this new census tool, determine their interest and willingness to support the development of a program for maintenance, and obtain support for the people promoting the local GBF.

Invitations were sent to approximately 250 data processing users in the community. Mr. Silver of the Geography Division of the Census Bureau, as well as several other knowledgable people in the community, were invited to make presentations to the assembled group. As a result of these presentations enthusiastic support was received from the 50 people in attendance while many others, unable to attend, indicated their interest in the program. With that vote of confidence, it seemed appropriate to go about the task of raising the necessary money from the Columbus community. A finance committee was named and the Chamber President, GBF Finance Committee Chairman, and I visited community leaders.

The city's chief finance officer was approached and briefed, and saw the value of this tool to the community. An ordinance was prepared, submitted to Council, and approved and signed by the Mayor, providing the keystone financial support needed in the community.

Concurrently with city government developments, funds were being solicited from local businesses and city schools. The Board of Education approved funding, contingent upon the city's approving their appropriation. The presidents of major companies were contacted, one by one, until all of the requested money ($12,000) had been raised, thus enabling the program to get underway.

From the group of contributors, a Chamber committee was established with the responsibility of developing the geographic base file for Central Ohio. The committee was composed of representatives of Mid-Ohio Regional Planning Commission, city of Columbus, Columbus Public Schools, two financial institutions, and the staff of the Columbus Area Chamber of Commerce. The committee consisted of planners, programers, and marketing specialists.

The purpose of the committee was to provide for the development of plans and policies regulating the effective use and application of the Columbus geographic base file. Specific objectives were developed:

To oversee the control and maintenance of the file

To determine broad community needs

To ascertain applications for the file

To set priorities for meeting data needs

To maintain a user software library

To establish a cost base for users

To report the activities and responsibility of this committee to the appropriate Federal agencies

To promote the use of the GBF by appropriate public and private organizations

To investigate alternative means of funding to insure development and adequate maintenance of the GBF

The GBF Committee began to function in mid-1972. As news of the committee activities were circulated others became involved from a radio station, banks, Battelle Memorial Institute, the Ohio State University, the utilities companies, and many others. The Chamber committee formed the hub for all activities and, as problems arose, it was the committee and its members that reduced each problem to its basic elements and then proceeded to resolve each element in turn.

Most of our original effort was consumed by the correction process. Our initial digitized GBF contained 33,463 records (258 characters per record). The first edit lists (Correction Phase I) provided by the Bureau were concerned with segment name consistency and coding limit line. The program FIXDIME developed by the Bureau of the Census was used. Corrections were coded during a period of approximately 3½ months. The correction input of 2,109 records, resulted in a new GBF containing 33,341 records.

Because of the timing of the Columbus correction phase, we were hampered by many problems, which today probably would not interfere with any new correction phase. For example, after completing our correction we submitted the material to the Census Bureau for their FIXDIME work. This came at the very time that the Census Bureau was caught in the squeeze of revenue sharing allocations. GBF activity, of necessity, was delayed while that work was completed. Shortly after receiving our corrected material from the Census Bureau we learned of the revisions in the Metropolitan Map Series. We were advised to wait for the new maps to begin our update.

At about the same time, the record size was expanded to 300 characters. At last the address edit listing, which flagged address data inconsistencies, was prepared by the Bureau of the Census. We found 33 percent of our total file was flagged for review but not all records flagged were necessarily in error. FIXDIME II was the program developed by the Bureau of the Census for this phase and after reviewing all 15,107 flagged records, our new GBF contained 33,282 records.

It was early in 1973 when the pieces of our program seemed to be coming together. New maps for Columbus were received from Jeffersonville. File corrections were made. Our tapes were returned by Washington. It seemed as though we were about ready to proceed with updating. Because of the long delay experienced with sending our work out of the area, we decided to explore our own community to see what services were available, so as to provide us with an independence we had not felt previously.

The GBF Committee members renewed their contacts seeking help from among their associates within the area. After extensive checking we located an organization willing to provide digitizing. We learned that the Department of Preventive Medicine at Ohio State University was using an ACG to plot incidence of disease in the city. Discussions with the manager revealed that he had developed a mapping program displayed on a CRT which could be used to make corrections with an electronic pen on the face of the CRT in the coordinates on our file. He agreed to use his program to correct the file. When we discovered a high error rate in our ZIP code file, we contacted W. A. Storing, a direct mail marketer who examined the file and provided ZIP code corrections for street features. As a result of these efforts, we felt we could complete our correction and update efforts relatively independently, and proceeded to do so, using our local resources.

As we embarked upon our program of updating the file, it became very clear that the information being provided by the city of Columbus would need to be expanded. It became necessary to locate the sources for this information within the city operations, create a mechanism for obtaining the information, and begin incorporating this data in our GBF file. We also felt the keystone for obtaining similar information from suburban communities and county government would hinge upon the success we achieved in working with our central city, the largest population entity in our SMSA.

The city Data Processing Department located the necessary information, devised a flow system for use within the city of Columbus, and we then attempted its implementation. It is interesting note that when we began this implementation, a process which had been working for years was questioned, and even the data flow which had been occurring, suddenly stopped. It became necessary to gain approval, not only for the new system, but the previous system as well. The City Development Department arranged this approval, and our information flow resumed with complete information being made available for our update process.

As you can tell, for our initial update phase there were many organizations involved: The Columbus Area Chamber of Commerce, City of Columbus Data Processing, Columbus Board of Education, Mid-Ohio Regional Planning Commission under the Urban Transportation Study Planning Process, and numerous potential users from private industry working together in a great deal of probing, begging and borrowing to establish a fairly satisfactory data input flow.

It was during this stage of our development that our original funds were exhausted, and it became necessary to seek alternative sources of work or additional funding to continue. We have numerous computer installations within our area from which we were able to obtain support for varying amounts of effort. Much of the work which had originally been contracted for was now contributed by interested parties within the area.

We had been purchasing computer time from a time-share facility, but when our funds were exhausted we discussed our problem with the city, and were given the necessary machine time to enable our work to continue. To meet certain time commitments we decided to accelerate the update program. This called for additional coding and keypunch personnel. Each committee member checked at his own facilities and the Columbus schools were able to provide additional help to speed our update process.

When the updating, coding and keypunching were done, we tried to add the new data to our GBF. The version of FIXDIME II/C we had was not working well on our Burroughs equipment. We attempted to copy FIXDIME II/C under program control to get by the tape marks between files, but the program had to be revised for different tapes and this tape had different blocking factors. We found the most expedient way to get a copy was to use IBM equipment, converting the tape to one continuous file by replacing tape marks with dummy records containing special characters to indicate separation of files. Because of local operating schedules among committee members there was no computer time available. We checked the Chamber of Commerce Data Processing Directory and found a Chamber member, Highlights for Children, with the appropriate IBM equipment, and we were able to arrange the necessary machine time. The original file was dropped off one evening on the way home from work and picked up the next morning on the way to work, file copied and program operable on our Burroughs machine.

The corrected FIXDIME II was ready in time for us to use for our final process run. Update records, plus some additional correction records (resulting from updates) provided 21,830 records; the new GBF output contained 37,088 records. For this final run computer facilities and time were furnished by the city of Columbus.

We noted that we increased our file by 4,257 segments over a 2½-year time period; this averaged 1,703 segments per year. Estimating an average of three segments per feature, this gives us an average of 568 features per year. The impact of each new segment averaged two old segments that required recoding (8,514 segments in 2½ years).

One of our problems has been the diversity of equipment that we find ourselves involved with within our user group. Our largest user have equipment such as Honeywell, Burroughs, Control Data, and to a very limited degree IBM. Consequently, programs received from the Census Bureau have not always been operable when received.

Apparently we had also been receiving developmental programs prior to the completion of documentation or modified programs which did not include the modification documentation. At any rate, when we were ready to run our ADMATCH program, we were unable to operate either the OS or DOS program on our Burroughs, Honeywell or CDC equipment. We did locate a CDC ADMATCH program and, after verifying the file input, were able to make a match. However, for our purposes several modifications seemed necessary. We rewrote the SPLIT program consisting of less than 20 COBOL instructions. The ADPREP program was modified as follows:

(1) File descriptions were modified to reflect known record lengths rather than the more generalized original version which accepted run-time parameters defining record lengths. The nature of these changes could be somewhat machine dependent. The changes were made to achieve a more efficient file organization on the test computer.

(2) All working storage variables which could be designated "COMPUTATIONAL-1" were so changed. This resulted in better execution times for the test runs.

(3) Portions of the program required solely to process options not being utilized for the test cases were removed. This included the State and place code processing options and the change/expand option.

(4) The program was divided into overlayable segments to reduce core requirements.

(5) Several minor corrections were made to input-output instructions.

(6) Several apparent program logic errors were corrected.

CONTROL-CARD SORT MODIFICATIONS

The number of key fields was increased to include the following:

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.

The reason for expanding the number of key fields is explained below.

ADMATCH PROGRAM MODIFICATIONS

(1) File descriptions were modified to reflect known record lengths rather than the more generalized original version which accepted run-time parameters defining record lengths.

(2) All working storage variables which could be designated "COMPUTATIONAL-1" were so changed.

(3) The program was divided into overlayable segments to reduce core requirements.

(4) Several minor corrections were made to input-output instructions.

(5) Several apparent program logic errors were corrected.

(6) The most significant change was made in the logic of the record-matching process. The original version of this process is described in detail in Appendix B of the ADMATCH Users Manual published by the Bureau of the Census. In the original version of this process, all reference file records for a given street are retrieved and stored in memory at one time. (A tape file is used for overflow.) Each data file record for that street is then compared to each entry in the table to locate the best match. This process proved to be too time consuming in the test cases. Therefore, both the reference file (GBF) and the data files were sorted on the keys listed above under Control-Card Sort Modifications. This then permits a record-by-record comparison of the data file and reference file for matches and eliminates the requirement for building and searching street tables.

As a result of this modification, processing time for address matching was significantly reduced.

Other problems that we have encountered have occurred when multiple file tapes are received. We had difficulty, at some installations without IBM equipment, in copying these files. Many times we did not know we had received a multiple file tape until after being set up on the machine. Whenever this problem was encountered, we located appropriate IBM equipment, copied the files needed, and proceeded to run the programs. Prior to the Census Bureau's publication on CARPOL programs, we ran our ADMATCH GBF employee listings to create a carpool system, which provided us with one of the first computerized carpools in the area. When news of our success with the ADMATCH program became known, even to a limited degree, within the Ohio area, requests for copies of the program were received. Two years of concentrated efforts to establish input procedures, implementation procedures and actual implementation have brought us to this point, and was made possible by the cooperation of many.

We look forward to constantly improving and refining the procedure we have already established, adding to them, and making them as automatic as possible for the ongoing updating.

Throughout the development of our continuing correction and update program, we felt it was important to derive a process which would not only be financially, but operationally independent. Through the cooperative effort within the community we have received many in-kind services. Other organizations have contributed both time and money to the support of our program. We are very proud of what has been done and how it has been accomplished.

There has been little full-time effort expended except by the paid staff at Mid-Ohio Regional Planning Commission, but there has been extensive weekend and evening effort made by many people in our community. It is the efforts of these dedicated people that has made our GBF what it is today. No one person or organization can take the credit for this development. It has truly been a community effort. And having been a community effort here is what we have gained:

A GBF.

A communitywide interest.

The possibility of a comprehensive planning tool.

An information file that will help business and community relate to one another.

We have begun an education process of certain business leaders showing what computerized information can do for them.

We have opened the door for countywide integration of business, economic, and social information.

A willingness on the part of many individuals and organizations to discuss common problems.

All of this has been accomplished as a result of correcting and updating our geographic base file and we expect to continue in operation on an ongoing basis.

Question Period

Mr. Hearle—One question that occurred to me is with respect to your plans for the ongoing update-maintenance. Do you have any idea of the cost?

Mr. Thomas—We are estimating a cost of approximately $16,000 for our 1974 GBF update.

Mr. Hearle—How does that work out? Is it one person full time or is it different people?

Mr. Thomas—The expenditures will be spent by many people performing brief part-time tasks, primarily staff at Mid-Ohio Regional Planning Commission.

Mr. Barabba—Who has control over the file? Who controls its use and availability?

Mr. Thomas—The GBF Committee of the Chamber of Commerce.

Mr. Barabba—Is the Chamber the operating agency?

Mr. Thomas—Yes, sir.

Mr. Molski—What are your plans for immediate application?

Mr. Thomas—The Ohio State Geography Department is interested in experimenting with social and economic trends. The banking community is interested in using it for marketing and branch site location. The direct mailers are interested for its application in preparation of street directories and direct mailing. The MORPC wants to use it for transportation planning and land use. The city of Columbus is interested in resource allocation for fire and police and housing uses. The Board of Education wants to use it for assigning children to school districts.

Mr. Molski—Are you going to charge any kind of fee for the use of the file?

Mr. Thomas—We have what we call a subscription fee which covers the cost of the annual update of our files, and provides an independent source of support from our users.

Mr. Thyagarajan—Two questions. One, have you computed an average cost of update based on growth and population so that we have an estimate of what to expect in the future in terms of new development? And the other is concerning the rules and regulations regarding usage and so forth that you mentioned. Is there a published document that is available to people from which you can get ideas on

guidelines as to how such projects can be administered in the long term?

Mr. Thomas—Our cost is based on our experience in updating the file from 1970 to 1973. We feel $16,000 will cover our 1974 costs. Regarding rules and regulations on usage, we have devoted time thus far in getting our file up and running, and do not have everything documented. Copies of our user agreement form can be made available. However, the intent of our user form is to localize distribution of our GBF and to assist us in raising funds for our continuing update.

Mr. Barabba—You mentioned that other areas were also interested. Have other communities contacted you about using the private sector to assist in developing a GBF/DIME system?

Mr. Thomas—Yes, so far we have talked to the Dayton Chamber of Commerce and Youngstown Chamber of Commerce. We feel that it is our community effort and cooperation in utilizing available resources that made our success possible.

Mr. Barabba—You indicated that many people collaborated in the development of the file. Is it your impression that they are going to use the file more because it is theirs?

Mr. Thomas—Yes, I feel that our user community, both within the private as well as the public sector, will be more anxious to make use of the file because they have been part of the file development, both financially as well as technically and operationally. I think we have at least 15 different organizations waiting for us to make the GBF available, so they can begin their own applications. One of the things we have been trying to do is promote transferability between the facilities of our users, Honeywell, Burroughs, CDC and IBM. We have been relatively successful so far.

Mr. Weaver—Two questions come to mind. The first has to do with the cooperation, which I think is outstanding. Do you think that the eagerness of the users now, the 10 or 15 users to get their hands on the file and start to use it, is due to their early cooperation or perhaps just to their current awareness that the GBF is available? The cooperation, certainly from a financial standpoint, was welcomed, but whether or not they participate financially they can still relate an end product or use of GBF.

Mr. Thomas—One of the things that I did not mention, which I think has tended to maintain a high level of interest within the community, for the past 2 years our GBF committee has met approximately every 2 weeks and there has been involvement by many people in our area. We have had 10 to 20 people turn out for each of these meetings. We mail a post card, to anyone who has shown any interest, notifying them that there is going to be a meeting, telling where it will be and inviting their participation. If they have user interest we talk about their needs and what they are interested in and how they can make use of the GBF. It is this involvement which has maintained the interest within the community.

Mr. Weaver—One last question. I assume the operational sources that you mentioned are from the city and county in terms of determining new addresses, and new streets and so forth. Am I correct in this assumption? Did you find that the city of Columbus and the county have immediately then grasped hold of the GBF and are using it internally, to perhaps supplement or take the place of some of their internal processes, or are you getting first an extra copy of each transaction that takes place internally? For a new street or an address?

Mr. Thomas—A copy of everything that goes through the city also goes to Regional Planning and is entered into their files. This information is used to update the MMS map series and GBF records. The city itself has not taken as much advantage of the data as they might, but they are providing information to us.

National Geographic Base Files: Future Developments

PAMELA WERNER AND HENRY BRUCK*

(Presented by Henry Bruck)

INTRODUCTION

In a paper prepared for the Census Geographic Base File Conference last year we discussed some aspects of a national geocoding converter and our experience in the development of such a structure. We pointed out in that presentation that the converter may well serve as the central element in the development of a national geographic base file system. We assume everyone here knows that the development of national geographic base files lags far behind the state of development already attained in urban geographic base file systems. We are, however, pleased to report that the first steps toward a national GBF, the county-level converter and associated mapping capability, are now operational.

It is, from our point of view, particularly appropriate that the theme of this year's conference is Geographic Base File Systems—A Forward Look. Just as there are new directions and developments in regard to urban geocoding, the gradual movement toward national geographic base files is entering a new stage. The emphasis in this next phase of development will be on subcounty entities and the extension of the converter to include geographic units at that level.

PROBLEMS OF SUBCOUNTY GEOCODING

The problems of developing a conversion capability at the subcounty level are easily an order of magnitude more complex and less tractable than those difficulties encountered at the county level. The principal problem areas are the following:

1. **Geodefinition.** In last year's paper, we noted the difference between geographic entities that are geodefined, that is, have explicit and defined boundaries, and entities that are geoidentified, that is, do not have firmly established boundaries. While there are many subcounty entities such as minor civil divisions, census tracts, wards, and others that have clearly delineated boundaries, there is a vast number of subcounty locational entities that are simply named places without specific boundaries. Place name entities are widely used and given sanction by the American National Standards Institute which is preparing a comprehensive listing of such place names in the United States. These amorphous entities cannot be described geographically with any degree of specificity and are generally represented by an arbitrarily assigned centroid. Thus, in developing a national geographic base file, one of the critical choices one is confronted with is, what is to be done with entities that lack geodefinition in any real sense? Should they simply be thrown out? If so, one is confronted with the danger of a severe loss of information. Or should arbitrary boundaries be assigned to these? Or, should some other way of dealing with them be invented?

2. **Instability of Boundaries.** Geographic entities at the county or greater level tend to have relatively stable boundaries. While there are occasional changes in the configuration of standard metropolitan statistical areas or independent cities and there are even occasional county boundary changes, the aggregate number of such changes in any single year is rather insignificant and relatively easily handled with an ongoing and operational update and maintenance system.

   Quite the contrary is true of geographic units at the subcounty level. According to the 1972 Boundary and Annexation Survey conducted by the Bureau of the Census, 29 percent of the places responding to the survey reported boundary changes occurring during the period January 2, 1970 through January 1, 1971 and 30 percent reported boundary changes during the period January 2, 1971 to January 1, 1972. It may be that, after a time, and if revenue sharing remains a part of the income redistribution policy of the United States, greater stability of boundaries will be achieved than has been the case traditionally. For the time being, however, the sheer volume of changes implies a heroic level of effort in maintaing relative currency for a file of this kind.

3. **Disparities in Size.** While there are very substantial disparities in size between the units that were incorporated into the county-level converter, some of the size disparities that must be confronted at the subcounty level are absolutely enormous. The variation between urban and rural land use activities, settlement patterns, and population densities create these

*The writers are, respectively, Research Engineer and Associate Director, Urban Systems Laboratory, M.I.T. We would like to acknowledge with thanks the major role played by Alan E. Pisarski, Department of Transportation, in the work described.

size disparities both within a single national geographic system and among the various nationwide systems. For example, the largest State in the United States, Alaska, is a single congressional district. Other congressional districts are made up of handfuls of blocks in Manhattan. There are counties, such as Los Angeles, that contain many congressional districts and there are many congressional districts that contain many counties. Ways must be found of dealing with these geographic disparities.

If these problems were not enough, there is no single comprehensive computer mapping capability which comprehends a significant part, much less all, of the geodefined subcounty entities in the United States. While working on the county-level converter, we did find an available digitized file of county boundaries. It had been developed initially by the highway administration and subsequently made operational by the Bureau of the Census. Nothing comparable exists at the subcounty level. To be sure, computer mapping capabilities have been developed for some SMSAs, even some States, and perhaps for rather large portions of the United States. However, there is no reason to expect any consistency among these individually created and oriented systems. Were one to assemble all of them, it is not at all certain that the cost of making them operational in the same computer environment would not exceed the very large projected cost of digitizing the boundaries of subcounty entities.

A STAGED APPROACH TO THE SUBCOUNTY CONVERTER

Our current thinking on an approach to the subcounty problem visualizes three separate, progressive phases.

Phase I. This will be an inventory of all subcounty entities contained in a selected universe of national geocoding systems. The file structure is a simple one, consisting of a fixed set of elements which identify a county within a State, and a number of periodic sets which list, in alphabetical order, the names and codes of the subcounty entities located within the given county. Each national geocoding system included in the file comprises one periodic set of names and codes for every county and each set is completely independent with no implied nominal or geographic relationship between sets.

Phase II. In the second phase, the periodic sets generated previously are to be compared, matched by name, and consolidated. File structure remains essentially the same with county components listed alphabetically by State and county. The entities will have been only nominally matched and there will be no attempt to define any existing geographic conformality among the subcounty units at this stage.

Phase III. The final phase of this file will contain a list of all the various county component entities spatially related at the subcounty level in as exact a geographic frame of reference as the nature of the units permit. The actual configuration of this file is undetermined. Each county component might be related to a single existing set of subcounty units such as minor civil divisions. Or, several types of computer programs for polygon analyses might be utilized to extrapolate and define areas of spatial conformality.

The development of the proposed county component converter file progresses from a very simple, nominal inventory of subcounty entities towards the delineation and definition of any existing geographic conformality among the complex of overlapping spatial units. As the file evolves through the three phases, the cost and effort involved increase dramatically. While the first phase could be constructed in a short time with relatively minimal efforts, the third phase may well be too costly and too time consuming to complete. The critical factors involved will be the current and expected levels of resources available for development of this file and the degree of geographic explicitness desired in the final stage. At this time there is no doubt that achieving the final phase of this file will require large-scale commitment. It is, however, important to note that even if all three phases cannot be implemented, each phase is separable and independent, with value as a geocoding product in and of itself.

THE COMING TOGETHER OF THE MICRO AND MACRO LEVELS

We want to conclude this forward look at the development of national geographic base files with some comments on the coming together of the macro and micro level of GBF development. Even though we have not addressed the issue specifically, we have not avoided thinking about the manner in which the macro level, from the top down, geocoding system on which we have been working, would interface ultimately with the micro, from the bottom up, approach the characterized DIME files.

The block faces so lovingly coded for DIME file purposes do, of course, aggregate up into minor civil divisions, census tracts, wards and other geodefined units. Were one to be able to give geodefinition to places, one could also aggregate to this level, as long as the context remains urban. The crucial point is that DIME coding remains relatively straightforward as long as the context is urban. But as the boundaries of urbanization are reached, the level of uncertainty requiring relatively arbitrary choices increases. Where one block face ends and another begins, where the intersection is properly coded and so forth, are questions that sometimes need to be resolved rather arbitrarily.

There is talk now of extending DIME coding beyond the context of urban areas, even to DIME coding entire States. We would suggest that, as soon as DIME coding is extended beyond zones of urbanization to areas which are traversed by roadways only at great intervals, where there are few intersections that can be easily identified as points to be encoded, much the same set of problems that we are encountering will tend to arise. Aggregation upwards must be to meaningful entities, and these will tend to be the same ill-assorted lost we have identified in our subcounty work. DIME users wanting to aggregate up to the congressional district level will, for example, have to think very hard about how properly to encode a section of a congressional district boundary that runs through the middle of a cornfield.

We leave you then, in looking forward, with the thought that both those working outward from urban areas using the DIME system and those of us working downward from larger geographic aggregates are about to meet in a grey area inhabited in part by curious entities having neither the constancy nor the certainty we would prefer to attribute to well-defined and well-mannered geographic entities.

Question Period

Mr. Hearle—That set of comments is the kind of insightful observation that those of us who have known Mr. Bruck for a decade have always found from his remarks. I could observe sitting here watching the audience that the elation of the Columbus story was settling down and everybody was moving to the reality of the tough problems ahead. Now, one of the pioneers who got us to where we are is telling us that the next decade may not be as easy as the last one.

Mr. Voye—I have a comment rather than a question. We are in the development stage in GBF and looking at potential uses for it. One of them being land use. We thought we would use the automated county real estate files. Unfortunately, we can't match range, tier, and section numbers to a node number. Whenever you come up with the answer to that, we would like to hear from you.

Mr. Bruck—I know exactly what you are talking about. That's the sort of thing that, of course, is bound to happen consistently. Among other things if you are trying to, even if you are trying to get a half way decent clean land use file outside the city, to go on the assumption that there are accurate boundary records for property holdings is by and large untenable.

My dentist is a great student of Massachusetts history, and he has been going into great detail into various claims of transfers of Indian land here during the very early settlement days. He has looked at the old property records, and what they say is that this property goes from yon large rock to yon large oak tree. Neither of which is there anymore. But that is the property record. There very often has been no survey since then.

Mr. Kelley—Could you give us any clues as to which directions you may go? What kinds of alternatives have you looked into in the rural areas?

Mr. Bruck—Well, we have come to no definite decision. As our program is laid out, the first part of it consists basically of organizing entities within counties, and then going on to attempt some geographic matching of these. But the problem of coding, in a sense of expanding a DIME file, is something that is going to need a whole other approach. A friend of ours, some you might know, Mr. Bob Baraclough, set about to make a DIME file for the State of Israel. Of course, he had little problem in the cities, but when it came to going out to the Sinai, he was looking for intersections of camel trails.

The problem is one of setting down a consistent group of rules in which there are no obvious features that cannot be coded, cannot be dealt with. What this set of rules should be, I think, is really opened to question at this point. There are efforts in Massachusetts, as Mr. Humphrey said this morning, to extend the DIME file statewide. There is talk about this in California. It has been done for Rhode Island, but Rhode Island is essentially one urbanized area. As soon as the statewide effort really gets going, I think, the fat will be in the fire. Decisions will have to be made. I think that the best thing that we can hope for is that these decisions will not come to be made uniquely so that you get one set of decision rules for Massachusetts and another set of decision rules for California.

Mr. Dauterive—Are you using the U.S. Coast and Geodetic maps to establish the state plane coordinates on a statewide basis so that you can identify locations outside of the metropolitan area, or landmarks outside the metropolitan areas?

Mr. Bruck—We haven't made a decision on any specific set of maps or any specific way of doing the coordinates. Very likely, one of the steps in doing this sort of work that I mentioned very briefly would be to obtain a very comprehensive set of maps including the USGS maps.

Mr. Dauterive—Have you all had any contact with the NASA, National Aeronautics and Space Administration, on the maps that they are doing for the satellite projects? It seems like everything that you are talking about pertains to establishing a standard set of maps which could be used for all geocoding. If the maps are not correct, at least everybody is using the same identification number. Are you proceeding that way, or are there some recommendations along that line?

Mr. Bruck—No. At this point, it is still in the "looking at the problem" stage. Which set of maps will be used really depends on a number of variables that are still going to have to be resolved.

Mr. Thyagarajan—It seems to me that the rigorousness with which you pursue this problem has got to do with how much use is expected to be gotten out of it. In the State of New York for instance, they have developed a data file called the Land Use Natural Resources File (LUNR) which is on the Universal Transverse Mercator projection system. It is an abstract grid, but we can pretty much pinpoint the location. It is possible to think about that kind of an abstract concept for areas that are essentially untracted and use it at a different level of information rather than trying to match up a DIME system to a national system or vice versa.

Mr. Bruck—It is possible. I think that it is really too early to resolve some of those specific claims. Projection can always be converted into another kind of projection, which is I think, a major problem in itself. The world has inherited from the CIA some marvelous things, including a computer program which if it ever becomes operational will permit changing from one projection to another quite readily as well as mapping the world in three dimensions and other good things, so that is not really the issue. I think that there is a real issue in trying to achieve at least some matches of boundaries. It is almost a certainty that, at the subcounty level, it will not be possible to achieve the level of matching that we did achieve through the county level converter. But at least there will be some matching, and particularly matching between the areas which can be aggregated out of a DIME-coded block-face type of file and the ones that are coded for the subcounty converter.

27007

Applications of Geographic Base Files to Law Enforcement - The LEAA Interest

G. PAUL SYLVESTRE (Law Enforcement Assistance Administration)

LEAA is not an operating agency, which is one way of saying that it doesn't do any work, so what I have to say here this afternoon won't take quite as much time. You are going to hear, after I finish, from a working agency, the New Orleans Police Department, and they will be describing actual hands-on applications of GBF. LEAA though, and specifically the National Criminal Justice Statistics and Information Service, has been aware for some time about the type of program that New Orleans is working on. As far back as 1966, the St. Louis Police Department was using its own GBF, or as they called it, the New Location Code for analyzing calls for services. This made possible, among other things, a manpower resource allocation project for that agency which received quite a lot of notice. Dallas is one of the several cities which is using a GBF in what is called Computer System Dispatch, or sometimes, Command and Control System.

It is no longer a novelty to see a room full of complaint clerks receiving telephone complaints and dispatchers sending cards out to the scene of the incident, all of this going on without anybody looking at a map. The status map with the flashing lights is no longer there. It is all in the computer.

In April of 1971, Sergeant Eddie Bradford of the Office of the Sheriff, consolidated city of Jacksonville, described to a conference similar to this one the use of a GBF in facilitating dispatch in the newly combined Jacksonville Police Agency for their combined city-county.

Many of you, I am sure, are familar with other specific projects. Some of which I do not know about even though they may have been funded by LEAA. That's the nature of the block grant beast. We are sure that there is going to be considerably more new activity in this area involving a lot of LEAA money. Out of this we have several concerns.

These are: One, duplication of effort. That is, a city may pay a consultant to do such work as requirements analysis and programing, which should already be on the shelf from work done in another city, perhaps by a different consultant. Two, unnecessary effort. That is extensive geocoding work not taking advantage of the DIME or Census GBF, and requiring additional debugging and editing and perhaps not even coming up with the level of quality of the GBF. Three, a less than optimum system. Some of the systems that we have looked at do not seem to take full advantage of the full capability of a GBF. For instance, the agency may have designed the system to assist dispatchers and that is largely what it does. All of the information is there, that is, fine geographical detail, time of the event, and so forth, to provide sophisticated spatial and temporal analysis of crime or to provide workload analysis on individual patrolmen. But nobody is using it. Or sometimes, if data is being produced which would seem to serve these pruposes, because it was designed by "the systems types" and the consultant, the operational people don't look upon it as their system; "It is something for headquarters, you know, and we don't use that." Four, compatibility with other systems. There may also occur a loss of usefulness if the data from other agencies cannot be matched to crime data.

I want to be careful what I say here; I don't mean that information of a specific address from the health department, or the school department should be in a police file. But it would be useful to match data on truancy or housing conditions against patterns of specific crimes on a small-area basis. But if the development of a police GBF was independent of that used by the general government, this will be made more dificult. Also, if a police GBF is developed without the attention to what is going on elsewhere, certain useful programs of one department cannot be readily adapted to another. So LEAA has begun to address these concerns.

We have made a small grant to the police department in New Orleans, so it could work with Long Beach in the transfer of its LOCUS system. This is the report that you will be hearing more about this afternoon. While this was under consideration, the Birmingham Police Department asked our Atlanta Regional Office for funds for an improved method of coding and tabulating crime data. In this instance, the regional office asked for our advice. We made a hurried one-day trip to Birmingham, and after calling Mr. Silver to get the name of the right person, we actually introduced the police people to the regional planning people. I won't say they did not know each other existed, but they did not know that they could help each other out. And this resulted in the decision that the police would subgrant part of the funds to the planning commission to speed up and upgrade its CUE effort to meet the needs of the police.

At the same time we informed Birmingham of the New Orleans project, and Birmingham decided to schedule its project a few months behind New Orleans to obtain benefit of their experience and obtain some nearly free programs. Unfortunately from the NCJISS point of view, many of the grants for such projects come directly from the State-controlled block funds. It is only accidently that we find out there is something going on in the area. So, if a project

comes to your attention, where you think that an agency should be working more closely with a regional planning commission or COG, let us know, or let our regional office know. We have phones to the Census Bureau and Mr. Silver is very good in telling us whom to contact and sometimes this works and sometimes it does not. After they get the money, they are very hard to control.

We are now giving serious consideration to a grant to a nonprofit agency or organization, which will permit an advisory committee consisting largely of police administrators, but including some criminal justice planners and geoprocessing specialists, to take a good look at existing projects. They will consider other practical applications as perceived by police administrators. Out of this, we hope that we can synthesize and document an optimum system. By disseminating such a concept widely we hope to affect not only new systems to be developed in the future, but to enhance such existing systems that presently fail to include some of the more desirable but less obvious features of GBF's applications. This is presently planned as a 2-year effort, but there will be interim reports to be published at the end of 33 weeks and approximately at the end of 44 weeks. The Geography Division of the Census Bureau will participate in all phases of this project under a separate interagency agreement. This should allow us next year to have a lot more to tell you about what LEAA has been able to do in GBF.

Question Period

Mr. Hearle—Do any of you know of any law enforcement-related GBF-type activity going on in your areas apart from the presentation?

Mr. Voye—The St. Paul Police Department is utilizing the GBF, and the Police Department of the City of Minneapolis is looking into the applications at present.

Mr. Hearle—Are they working closely with you in terms of your central role?

Mr. Voye—Yes, they are.

Mr. Weaver—The Police Department with the city of Charlotte is part of the IMIS USAC project and is working directly with the GBF/DIME file. We are breaking the file into two different subfiles. One is a street segment that has been area coded automatically to response areas. We are also taking the same GBF/DIME file and building an intersection file. We are area coding the files so that when the address matches a block face they immediately know what unit to assign and, likewise, an intersection pair.

Mr. Sylvestre—I am glad to hear of those that are taking place with the cooperation of the law enforcement agency and the others working in the GBF area.

Mr. Voye—Some of the work in St. Paul was done with LEAA funds.

Mr. Sylvestre—Well, I am sure that most of it is, and that's not bad. I have no objection to LEAA funds going to it. I would just like to get as much GBF out of it as I can without duplicating or doing over work that has already been done elsewhere in the city.

Mr. Controvich—Springfield Police Department is also using the GBF/DIME, and they are currently thinking about using the nodes for accident location purposes.

27008

Securing Multi-Agency Participation in the Development of the New Orleans Geographic Base File

LEROY D. DAUTERIVE AND ALBERT C. ETIENNE

LEROY D. DAUTERIVE

The topic of discussion is "Securing Multi-Agency Participation in the Development of Geographic Base File in the New Orleans, Louisiana Region." This is primarily a joint effort being undertaken through the Regional Planning Commission, local parish planning agencies, and the New Orleans Police Department in an effort to reduce the cost to any one agency. Also, this joint effort was undertaken to take advantage of the capabilities and resources of each agency. Specifically, the New Orleans Police Department has a very good computer programing department, with very highly skilled people. The Regional Planning Commission, on the other hand, does not have a computer programing section, but it does have the map review and editing resources which the Police Department does not have. In this type of situation we have been able to work together very effectively.

When Mr. Etienne and I were asked to share with you our experiences in developing the geographic base file for the New Orleans area, we were somewhat surprised. The way we are developing the geographic base file is anything but an exemplary model for other areas to follow. The process that we are following is in a word "tacky." It is held together with bits of "time and materials" taken from a number of agencies to create the products needed to form the geographic base file.

At one time or another, we have borrowed staff and material from local city planning agencies, private utility companies, the U.S. Post Office, the Alcholic Driver Control Program, the New Orleans Police Department, the Jefferson Parish Sheriff's Department, the Regional Planning Commission summer staff, local public libraries, health planning agencies, local universities, the National Aeronautics and Space Administration, city and parish building regulations department, and the finance and data processing facilities of Jefferson and New Orleans Parishes.

To date we have financed the building of the geographic base file with funds from the Department of Housing and Urban Development, the Law Enforcement Assistance Administration, the Department of Transportation, local city and parish governments, and the State of Louisiana.

Before going any further into what we are doing to develop the geographic base file, let me give you a brief outline of the New Orleans area planning process. The New Orleans SMSA is composed of four county-type areas which are called parishes. These four parishes are named Jefferson, Orleans, St. Bernard, and St. Tammany. Since the boundaries of the Parish of Orleans are the same as those of the city of New Orleans, the two terms are used interchangeably to describe the city of New Orleans. The total population of the SMSA is slightly over one million people, with approximately 55 percent living in Orleans Parish, 35 percent living in Jefferson Parish, and the remaining 10 percent split equally between St. Bernard and St. Tammany Parishes.

Until the 1950's over 70 percent of the region's population lived in Orleans Parish. As cheaper land became available for home building, in the parishes of Jefferson and St. Bernard, large numbers of families moved annually to Jefferson and St. Bernard. Today almost one-half of the region's popualtion lives outside of Orleans Parish.

In 1962, the Regional Planning Commission for the New Orleans area was founded. Although it was founded in 1962, the Regional Planning Program did not get underway until 1967. One of the first components of the regional plan identified for special attention was the development of an easily updatable data base.

In 1968, the address coding guide for the metropolitan New Orleans area was developed. Also in 1968, the Regional Planning Commission began to explore the establishment of a regional 7.3-acre grid system for identifying and displaying regional data. After much study, this grid system has been put aside for two reasons; number one, the local planning agencies collect and need data by block and parcel identification, not on a 7.3-acre grid basis, and number two, converting field data to the grid identification system involves a costly, duplicative process. We would need to have the geographic base file in operation plus a conversion file to get data identified by the proper 7.3-acre grid.

We are focusing our effort on developing and maintaining the geographic base file for two reasons. First, planning data needed by the Regional Plainning Commission and local planning agencies is compiled by municpal address which is a good common denominator. Second, no elaborate changes in the present data-collection process are necessary if the geographic base file is used. If any of you have gone out in the field, you know what happens when you talk to the building inspectors and ask, "If you could just change the way you collect your data just a little bit." And he says, "Not me, not me, I've got union rules to contend with." There are so many obstacles to changing the methods of data collection that the fewer changes you have to make in the primary data collection the easier the process is.

Because the geographic base file is geared to use address information, we have a natural basis for agreement with data-collecting agencies. We don't have to ask them to change anything, or relatively little.

At this point, I would like to tell you briefly what has been accomplished to date in the New Orleans area. First, an address coding guide reflecting the 1968 metropolitan street pattern to which the DIME features have been added has been completed utilizing Department of Housing and Urban Development funds. Second, the New Orleans Police Department under a Law Enforcement Assistance Administration grant has edited and made corrections to be used successfully to sort employment data to traffic zones by the University of New Orleans as part of the Unified Transportation Planning Work Program for the New Orleans area. Fourth, over 25 informational presentations have been made to private and public agencies and groups who might be assisted by the geographic base file. Agencies and groups contacted included local school boards, Chamber of Commerce research personnel, housing finance groups, telephone company forecasters, local banking representatives, health planning staffs, and law enforcement department personnel, to name a few. Fifth, backing for the geographic base file development has been secured from each parish's planning agency. Sixth, funds have been secured from to update the geographic base file during fiscal year 1975 from the Department of Housing and Urban Development. We have gone out and done the things that were necessary to acquaint the people with what the geographic base file is all about.

To obtain multiagency participation in the regional development of the geographic base file, we have, based upon our experience, formulated an approach to secure each agency's assistance. This approach is based upon the premise that the priorities of the day-to-day operations of government and private agencies will override any serious attempt to set up a better way of completing a job, if the better way hampers the day-to-day operations. With this thought in mind, it was decided to sandwich in, as best could be done, the completion of the various pieces of information needed to update and maintain a functional geographic base file for the New Orleans area as part of the various agencies ongoing programs.

The present "piecemeal" effort focuses on first getting Orleans Parish operational; then going into Jefferson Parish, which is the second largest parish; and then tackling the two suburban-type smaller parishes. This approach, admittedly, is more costly and more time consuming, but faced with the impossible task of securing sufficient funds in any one year it is felt to be the most sensible approach for our area.

To carry out this approach, we have employed the following rules when soliciting an agency's support. One, when you approach the head of an agency and try to sell him on participating in the development of the GBF, you might have about 5 or 10 minutes, so you want to get in there and keep the approach as **simple** as possible. Number two, you want to keep it **practical**. You are not trying to come up with some blue-sky ideas about what can be done with this system. You want to keep it right down to where the agency head can relate to you. The third thing is, you want to keep it **pertinent** in terms of what it will do for his agency.

In keeping it simple, we sell one goal. We are not trying to develop a data bank; we are trying to develop a standardized tool for data manipulation. In keeping it simple, we only ask for what we need, not what we might like to have. In keeping it simple, we don't go into elaborate detail. And most important, in keeping it simple, we ask for only that product which an agency can produce as part of its day-to-day operations.

In keeping our approach practical, we try to make the work sequential to avoid going over the same ground twice. We sell the geographic base file on the basis that it is the key to developing a data-flow system at the least cost per agency.

In making our solicitations pertinent, we approach public and private agencies to determine if the geographic base file can (1) save them money, (2) save them time, (3) provide them more accuracy and (4) make them better able to respond to the problems in their respective agencies. If any of these four points can be determined as applicable, particularly if the geographic base file can save money, we are then able to more effectively solicit the agency's support. When these points were not applicable or have not been applicable, but the agency needs information, we proceed to pay for the product needed. If we don't have anything to trade, we have to find the dollars.

At this point, the question you may be asking is, "How are you going to set up the geographic base file for the New Orleans area?" The geographic base file will be made fully operational using the following approach. The updating of the metropolitan map series will be accomplished by the Regional Planning Commission and the local planning agencies in concert with the Census Bureau—street changes, new address ranges, node numbers, and new census block numbers. The coding of the updated map information in order to insert these changes into the file will be accomplished by the Regional Planning Commission in concert with the local parish Finance and Data Processing Departments. The actual insertion of the changes in the geographic base file will be accomplished by the New Orleans Police Department with support from the Regional Planning Commission.

Once the geographic base file has been updated it will be employed by the Regional Planning Commission.

The New Orleans Police Department has satisfied the respective agencies, needs during 1974 and transmitted the GBF in its present stage to other participating agencies.

In closing, I would like to emphasize that the approach we are using in the New Orleans area to convince many agencies to jointly participate in the development of the geographic base file is sound. While this process may try the patience of many, we feel that it is the best way to realize geographic base file development when Federal, State and community resources are limited. The longest journey can be accomplished one step at a time.

ALBERT ETIENNE

We in New Orleans have transferred the LOCUS system from Long Beach, Calif. It is basically a geographic base file, developed from DIME, with the extension on records of intersection information. I have done a file create on a test file, and we have that portion of it completed. One gentleman over here was talking about an intersection record which does create intersection records for you. I have also developed a couple of programs which might be helpful in the correction of DIME. I took the logic of the ADDEDIT program and did a segment link system which can list a file by feature name and node number sequence to make map review a lot easier. I have also a small system that will add things like traffic zone and ZIP code and mass corrections to the file.

Question Period

Mr. Ma—I am really amazed how you can coordinate all these agencies to be involved in the geographic base file. Can you elaborate a little further? How can you pool them all together? That was one of our biggest frustrations.

Mr. Dauterive—Who's coordinating? What we are trying to do is coordinate. When you go to an agency, you have got to know what you want. The only way to really coordinate the whole process is to do two things: Number one, set up a sequence of products that you need, and number two, divide the products needed among appropriate agencies.

We divided the products into two areas. The first area is the mapping work that has to be done, and the second area is all the computer work that has to be done.

In setting up coordination, I guess it would be better described as stealing "time and material," from different agencies. You look at what the agency is producing first, then go back to your office and determine where the products fit in terms of what you need. If it is not exactly what you need, you don't go back; but if it is exactly what you need or if it can be made into exactly what you need, then you go back and approach the agency.

That's when you find out if you can realize those four points that I mentioned previously: Can you save them money with the geographic base file, can you save them time, make their work more accurate, etc? If you can realize any of those four points then you go in and you say to the agency head, "I can give you a trade. If you produce just one little piece of data in this form for me, then I will give you this complete geographic base file. I'll just hand it to you, and it will be free, because you are producing this data anyway."

On the other hand, if the agency does not get anything out of the geographic base file, then you sit down and say, "I need this piece of information from you. How much will it cost me?" Once you have that price determined, you can look around for HUD funds or LEAA funds, or go to the city government and see if you can get the dollars necessary to buy that piece of information.

That is essentially what we try to do. To coordinate successfully you have to break the program down into the subparts that you need.

The map work, for example, has been broken down. If you look at the four basic mapping products needed, you need to update the Metropolitan Map Series; you then need that map enlarged (that is what we are going to do, enlarge it to 1"=400' scale); have someone put address ranges on the enlarged map and then prepare another overlay and put node numbers on it, which is the third part; the fourth part is to add the census block numbers. If you can get four agencies to do these four pieces in sequence then you have got it.

Mr. Hearle—How about these applications that you are going to hang on the GBF?

Mr. Dauterive—Let met tell you what we try not to do. We're kind of strange people, people who fool around with data—it's like the somewhat corny, but applicable joke. One man comes up to another and says, "Hey, everybody thinks that I am kind of strange because I like pancakes." The other man says, "Oh, you do? Why don't you come over with me? I have a whole drawer full." That is the way we get with data. You can like pancakes, but you don't go overboard. It's the same way we are doing with GBF program. We are not trying to sell elaborate data tools and elaborate information systems. We are trying to sell one thing—the GBF system and program, and we are trying to keep it simple. We are trying to say if you have this one thing, this one tool, then you could do many, many more things. We are not trying to tell you what you could do. If you think about it, however, there is one thing that you do need; you need this common denominator; you need the geographic base file.

That is all we are trying to sell, so we don't get directly into the use of data. All we are trying to do right now is to get that one piece, that common denominator, up and running. Once we do that, once we can show some products, then we can go into some graphic displays. We will then be able to get into what I call the more sophisticated data analysis.

Mr. Kondo—As I understand it, most of the agencies you work with are in the city of New Orleans. Have you had any contact with agencies with operate in a different jurisdiction than a city such as a State agency?

Mr. Dauterive—We have utilized primarily the city of New Orleans personnel, but we also have gotten support from the University of New Orleans, which is multiparish. We have gotten assistance from the computer services of the National Aeronautics and Space Administration, which is in the Slidell area of St. Tammany Parish. We have utilized address information collected by the Department of Roads and Bridges in Jefferson Parish where they record the addresses of all the streets they have under maintenance. While we are using primarily New Orleans agencies, we are not limited. The point is that we are trying to get to the agencies that have data.

In going to other agencies, the hardest job is finding out what there is they have that can be used. For example, we never realized that the Roads and Bridges Department had a set of maps with address ranges on it, until we went out there and happened to see them, and the only way we found that out is by going out and telling them the "GBF story." And once you tell your story and see what they have, you go back to the "drawing board" and say, "Is that agency going to be able to contribute to the program?"

Mr. Molski—I have a problem with the economies of scale. We are developing a GBF/DIME for the entire county, but the largest city in the county has a population of 55,000. I am wondering if you have had any experience with trying to approach a smaller city, such as that of 20,000 people; for instance approaching the police department, or the fire department in trying to streamline their operations with the use of a GBF? In other words is there a cut-off point to which the DIME file applies?

Mr. Dauterive—The economy of scale lies in the regional basis. Being the project manager, responsible to set up the address coding guide in 1968, I had 25 people working for me. They were sort of moving me around in about four different places: one in Jefferson Parish on the West Bank, one in Jefferson Parish on the East Bank, one in the city of New Orleans and one in St. Bernard Parish.

The economy of scale came in using the same people who became very skilled at what they were doing to develop the information, the address coding guide and the DIME information. I think that setting up for a small town is more of a problem because 50 percent of the problem in training the people to be accurate or getting an accurate crew to do the job. So I think if you are going to approach the problem of setting up the GBF you should approach it on a regional basis.

Go into the local municipality of about 20,000. Solicit their support and solicit their participation, but do it on the basis that "Here is something I am going to do for you as part of your participation in the regional process." Don't approach them on the point that they should do this for their city only. I do not think it is going to work nor could it be done that way.

Mr. Weaver—I have more of a comment that a question. What I have heard in this presentation is exellent. It really boils down to selling the user on the benefits that the GBF has to offer, and I think you have summarized beautifully how we could go about selling the benefits of the GBF; you have to do that for this ongoing cooperation, funding, etc., so congratulations in that area.

Mr. Etienne—In the very beginning we would go to a place and give them the pitch about DIME, ask them for clerks. The people would tell us, "It is a beautiful idea; when you get it running, come back and we will use it." As we struggled along by ourselves, just Mr. Dauterive and I mostly doing all of the work, borrowing a clerk here, doing something there, the closer it got to being complete, the more and more people were willing to give us two clerks for 10 days so that they could use it.

In the beginning it was a very hard row to hoe; everybody thought it was a good idea, but nobody wanted to invest anything in it. But then when it got close to being usable, the people were more willing to invest time or personnel as opposed to dollars. We usually did not ask for money, we asked for bodies.

UNIMATCH Experience With Truck Survey Data

EDWARD F. CARLBERG

In 1972 the Federal Highway Administration (FHWA) of the U.S. Department of Transportation initiated a National Truck Commodity Flow Study (NTCFS) to obtain information about the flows of commodities on the Nation's highways. The study was built around a questionnaire mailed out to a sampling of truck owners, asking them for details about their travels on specific days. Items of information collected on each questionnaire included types of commodities carried, weights carried, places from which the commodities were carried, distances traveled, and details about the vehicle involved. For every stop to pick up and deliver goods (up to eight stops per day), the questionnaire asked:

Where the stop occurred (by address, city or town, county, and State)

What sort of place the stop was (pier, factory, warehouse, construction site, store, etc.)

How many miles from the last stop

The time of arrival

The purpose of the stop (deliver, pick up, render service, etc.)

What commodities were picked up or delivered, by weight or quantity.

SAMPLE SELECTION

About 150,000 questionnaires have been mailed out over the past year. The truck sample selected was stratified by several factors, including a weighting of three to one in favor of the larger trucks. (Otherwise the study would have collected a mass of data on where sportsmen drive their campers, since 75 percent of the registered trucks are small pickups or panels.) The truck licenses of each State were sampled, and questionnaires mailed to registered owners in all 50 States and the Commonwealth of Puerto Rico. Each owner was asked to fill out the form on a specific day in the near future.

About 113,000 valid questionnaires have been returned—those where the truck was in service and the data was legible. Because details about the truck were also asked for in the questionnaire, the results became a valuable truck inventory. The owners were asked about gross weights, number of axles, body type, trailer details, and operating class (for hire, private, etc.). The questionnaire also allowed the compilation of statistics as to type of haul (local or long distance), whether commodities were carried, special kinds of commodities carried (including milk, oil, cattle, garbage, autos, repair equipment, beverages, people, etc.), annual miles traveled, age of truck, fuel used, and repair practices, among other things.

Even though the valid returns are only 113,000 out of 20-million registered trucks, the information has a high statistical probability of being representative of the universe of trucks.

Who will use all these data? The FHWA and other organizations within DOT, as well as trucking associations, economists, highway engineers, market analysts, traffic departments, truck builders, and others. The data might be used in determining highway needs, truck facility needs, regulations, investments, marketing plans, and other such decisions.

THE GEOCODING STEP

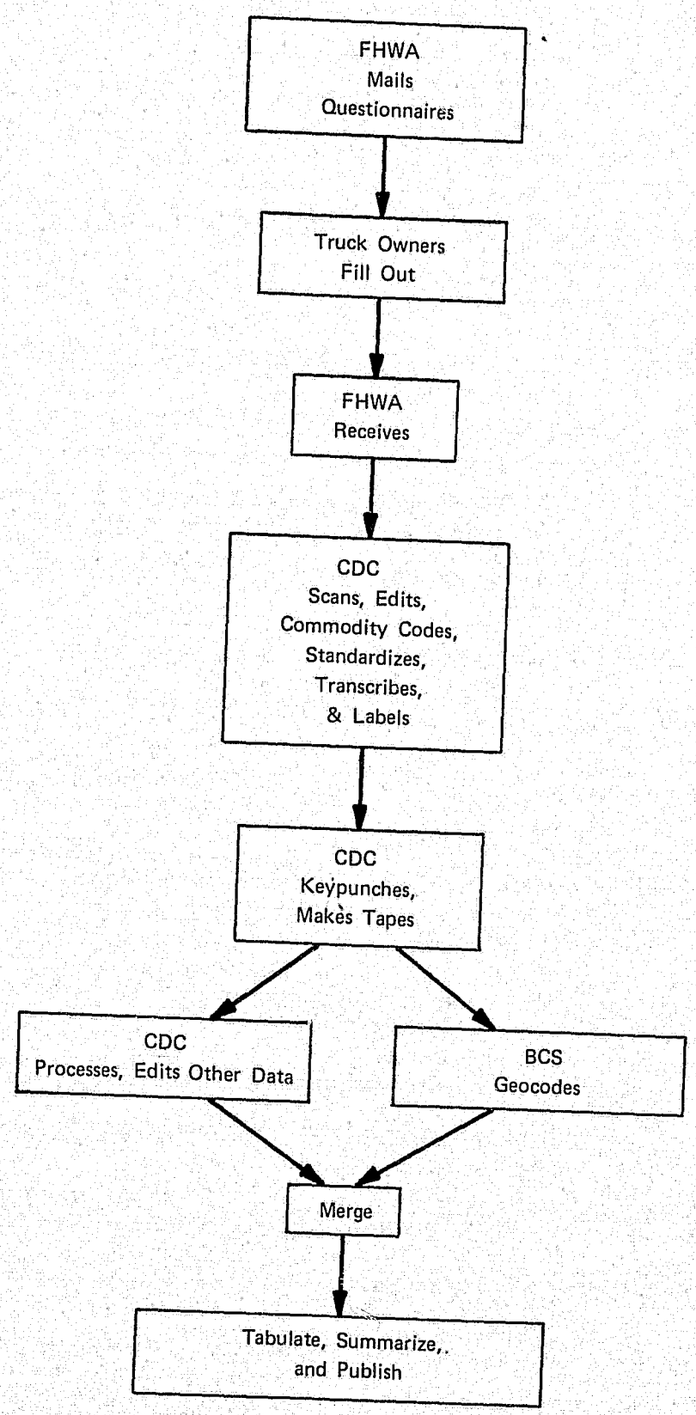
As may be seen from the following flowchart, Boeing Computer Services has only a small role in the process. It is an interesting role, however, because it represents probably the first major production use of UNIMATCH to do geocoding.

The geocoding problem itself is quite straightforward. The individual addresses of the stops are not used. The only geocoding required is to find a set of codes, from the town, county, and State given on the questionnaire, that will fix these locations geographically. To do this, DOT and the Census Bureau created a special combination file of the American Trucking Association's Standard Point Location Codes (SPLC) and the Census Geographic Area Code Index (GACI).

The SPLC is a grid-oriented system of six digits per code. The first digit is a region of the United States; the second digit is a State or portion of a State; the third and fourth digits represent a county or portion thereof; and the fifth and sixth digits identify either

(a) An area within a radius of 2-1/2 miles from a center point (which may be a post office, city center, crossroads, etc.) or
(b) Some specific location of transportation significance, such as a railroad yard.

The GACI (compatible with Federal Information Processing Standards) identifies States, counties, SMSA's, and places, plus codes for place size and place description (i.e., central city of an SMSA, incorporated town, unincorporated hamlet, etc.).



These two basic codes were merged into a single reference file of 103,000 separate locations. It was our job at BCS to take this reference file, match it to the data from the questionnaires, and append the appropriate matching codes to each location stated in the returned questionnaires.

Of the thousands of valid questionnaires returned, about a third indicated that the truck was not operated that day. Another third showed no commodities carried. Many of the last third, however, showed multiple stops, each of which required geocoding. Thus almost a million separate geocoding matches were anticipated. To date BCS has processed over 125,000 data records in 12 batches on our IBM 370-168 located in McLean, Va.

THE PACE SUBSET

The matching job uses only a portion of UNIMATCH. The Census Use Study people have lifted some modules out of UNIMATCH and tailored them to do this specific job very efficiently. This portion is called PACE, which stands for "place and county encoder." Its sole purpose is to assign FIPS State, county, SMSA, place, and SPLC codes to data records containing State abbreviations and place or county names, or both.

The Census Bureau people made some ingenious trade-offs between efficiency, accuracy, and completeness. The problem they faced was this: All data records must ultimately be identified to a location (matched), but not all data records are accurate or complete. The majority of the records are perfect in their spelling of the State abbreviation and the place name, and can thus be easily matched through efficient programing. In a few cases there may be more than one identically named place within a State; in these cases the county listed on the questionnaire is the tiebreaker. So on the first pass through the file, only exact matches are attempted, and all others are rejected. The reject pile is typically less than one quarter of the size of the original data file (approximately 75 percent exact match rate).

On the second pass the UNIMATCH uncertainty comparator is invoked to code as many places with minor spelling, phonetic, and transposition errors as possible. The uncertainty comparator compares each character in the reference file place names with the characters in the data file place names. Each time a character in the reference file matches one in the data file, a weighting factor is applied. Every mismatch causes a penalty to be applied. Care must be taken, however; for instance, WASHINGTON must not be matched to ASHINGTOWN. Thus the uncertainty comparator looks down only half of the place name to find matches. It will match WASHINGTON with WASHINGQTON or with WASHINGGON or with WASHINTON because such added, transposed, or missing letters still result in a net weight-penalty score that is high enough to exceed the match probability threshold. At the end of this second computer pass, another 10 to 15 percent of the file is accepted.

The third pass uses another reference file, the Permanent Supplementary Reference File, which contains an ever-growing list of common misspellings and abbreviations. PHILLY, for instance, is probably equated to PHILADELPHIA, and NYC to NEW YORK CITY. I do know that "TORKTOWN" and "WLKHART" are in the supplementary file; these probably are due to poor handwriting on the questionnaires. Since this pass also invokes the uncertainty comparator, the match rate on the previous reject file is now commonly over 90 percent.

At this point the remaining rejects are printed out and turned over to BCS clerical people. Using atlases and directories, they attempt to find suitable codes for each one. Some are obvious: a city listed in the county column or vice versa. Some are keypunch errors in the State codes, which must always be an exact match. Some are obvious to the human, such as the truck that went to Mount Rainier (a very large "noplace").

The fourth and final pass is made to the file (each file representing 1 month's data) to clean up the remaining rejects by matching the manually made list with the last of the machine rejects and thus coding every record to at least a State, or a State and county if possible, and also to a place whenever one can be deciphered.

At this point a final monthly tape is prepared that recombines all the matched records, sorts them together, and produces a "clean" data tape containing geocodes on every truck stop. This tape is next recombined with the original questionnaire data regarding the truck and the commodities it carried, and is ready for final summarization and printing.

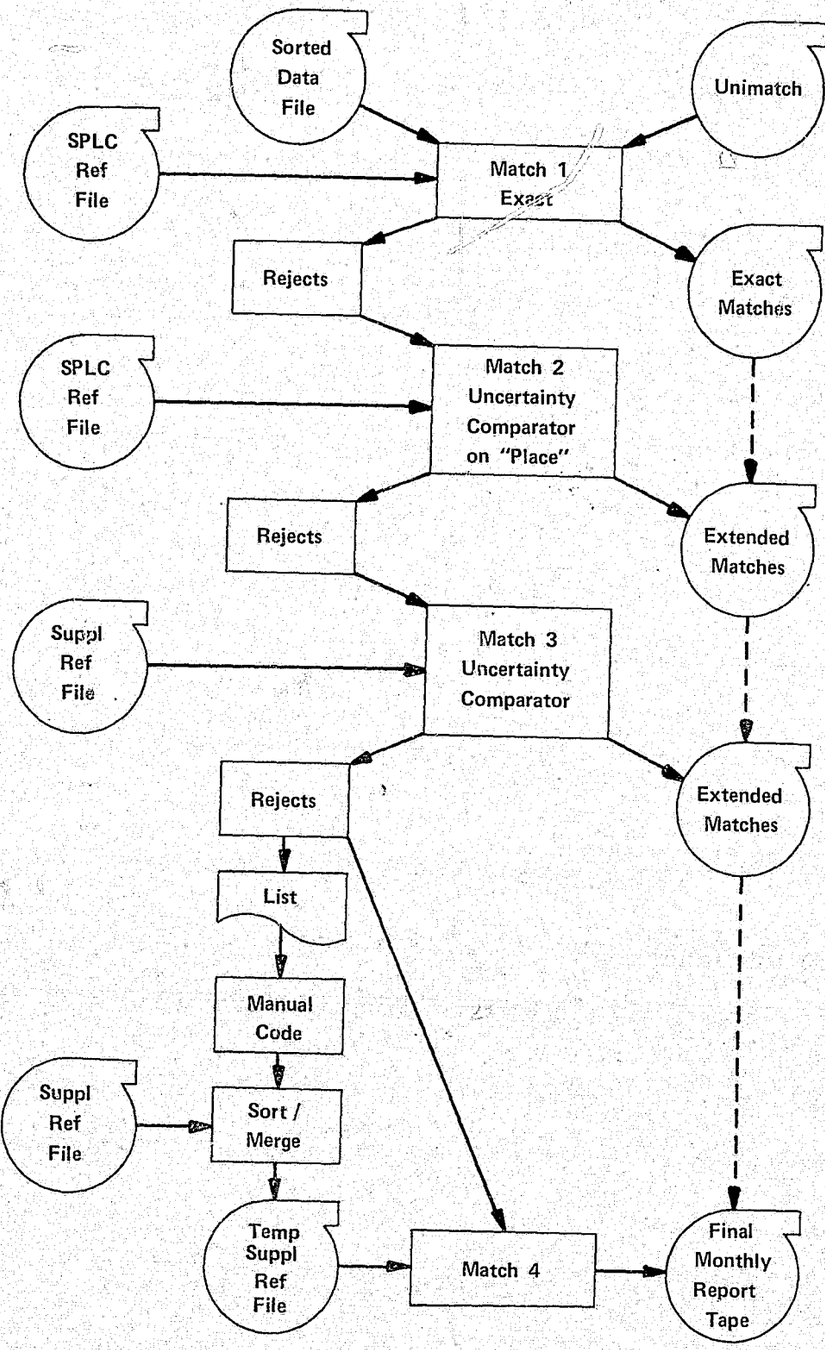
The flow of this entire process is shown by the chart on the following page.

The four matching steps are broken down as follows with respect to the requirement for exact character-for-character matches, as opposed to the use of the uncertainty comparator:

Data Item	Match 1	Match 2	Match 3	Match 4
State (2 characters)	Exact	Exact	Exact	Exact
Place (36 characters)	Exact	Uncertain	Uncertain	Exact
County (20 characters)	Uncertain	Uncertain	Uncertain	(NA)
SPLC/GACI Reference File	Yes	Yes	No	No
Permanent Supplementary Reference File	No	No	Yes	Yes
Temporary Supplementary Reference File	No	No	No	Yes

Note that the State code must always be an exact match. With a two-letter code it would be hard to do otherwise! However, this code is standardized at the time the questionnaires are prepared for keypunching; we don't depend on the truck owners' spelling. Thus, the only cases of State mismatch will be clerical or keypunching errors on the part of the data encoders, and such errors are quite rare.

PACE Matching Flow



CONCLUSIONS

While final results are not yet in, our experiences with UNIMATCH have been good. The PACE subset has been running reliably. As may be seen from the overall match rate list below, the matching has been consistently above 90 percent.

Lately, due to the constant additions being made to the supplementary reference list, the match rates have been exceeding 98 percent.

The task is not finished. More data is yet to come, and the match rates should continue to improve. If our experiences with PACE are any criterion, UNIMATCH is destined to become a major new tool for the geocoder.

NTCFS Geocoding

Overall Match Rate

Tape	No. of records	Match rate	Rejects
1	TEST		
2	4,049	94.8	212
3	4,096	90.6	385
4	3,804	95.9	281
5	7,235	95.8	305
6	13,030	99.2	105
7	1,600	99.8	3
8	20,000	97.2	560
9	23,169	98.1	494
10	20,263	(rerun)	1
11	23,973	98.9	249
12	20,708	98.7	351

SAMPLE REFERENCE FILE

```
AK
AK
AK
AKADAK
AKADAK STATION
AKAFOGNAK
AKAKHIOK
AKAKIACHAK
AKAKIAK
AKAKOLMIUT
AKAKUTAN
AKALAKANUK
AKALATNA
AKALEKNAGIK
AKALEKNAGIK LAKE
AKALEKNAGIK MISSION
AKALEXANDER
AKALLAKAKET
AKAMBLER
AKANAKTUVUK PASS
AKANCHOR POINT
AKANCHORAGE
AKANDERSON
AKANOREAFSKI
AKANDREAFSKY
AKANGOON
AKANIAK
AKANNETTE
AKANVIK
```

STATE / PLACE

```
WADE HAMPTON              C227C
WRANGELL PETERSBURG       02280
YUKON KOYUKUK             C225C
ALEUTIAN ISLANDS          02010          814280
ALEUTIAN ISLANDS          02010     0010505
KODIAK                    C215C          813C10
KODIAK                    02150     003050081 3C65
BETHEL                    02050     CC4C5C1816C50
BETHEL                    C205C     CC5C50CE16C4C
BETHEL                    02050     0060402
ALEUTIAN ISLANDS          C2C1C     CC7C50C81422C
WADE HAMPTON              0227C     CCEC4C1824C40
YUKON KOYUKUK             02290          818245
BRISTOL BAY DIVISION      C2C7C     CC5C5C0815230
BRISTOL BAY DIVISION      02070     CC505CC815230
BRISTOL BAY DIVISION      C2070     C0905C0815230
ANCHORAGE                 C202CC38C      81CC8C
YUKON-KOYUKUK             02290     01C05C0E1E25C
KOBUK                     0214C     C11C5CCE2212C
BARROW                    02C4C     C12C4CC81E2C5
KENAI-COOK INLET          02120     0145500812140
ANCHORAGE                 C2C2C     C14C41C81C1C0
YUKON-KOYUKUK             C229C     C15C4C1
WADE HAMPTON              C2270          824120
WADE HAMPTON              C227C          824120
ANGOON                    02030     01604C18C4065
KUSKOKWIM                 C216C     0170501817210
OUTER KETCHIKAN           C219C     C1EC5CCEC219C
KUSKOKWIM                 02160     0190400817260
```

COUNTY / STATE CODE / COUNTY CODE / SMSA / PLACE CODE / PLACE DESCRIPTION / PLACE SIZE / ATA-SPLC CODE

GACI

MATCH ONE OUTPUT—EXACT MATCH

TYPE CODE	MONTH	BATCH	SERIAL	COMPUTER REPLACEMENT CODE	STOP NO.	STATE NAME	PLACE NAME	STATE CODE	SPLC-STATE & SECTION	SPLC POINT CODE	PLACE CODE	PLACE DESCRIPTION	COUNTY	PLACE SIZE	SMSA	STATE & COUNTY CODE
21	11	01	01400037	26	80	AZ	BUCKEYE	04	79	71	69	00304	MARICOPA		06620	0013
21	11	01	02000038	07	80	AZ	CAMP VERDE	04	79	41	32	00335	YAVAPAI			025
21	12	02	04900052	15	30	AZ	CAMP VERDE	04	79	41	32	00335	YAVAPAI			025
21	12	01	01800048	22	69	AZ	CAREFREE	04	79	71	16		7MARICOPA		6200	0013
21	11	02	07600045	26	72	AZ	CASA GRANDE	04	79	65	62	00354	PINAL	08		021
21	11	02	07600045	26	71	AZ	CASA GRANDE	04	79	65	62	00354	PINAL	08		021
21	11	02	07600045	26	70	AZ	CASA GRANDE	04	79	65	62	00354	PINAL	08		021
21	11	02	07400045	16	61	AZ	CASA GRANDE	04	79	65	62	00354	PINAL	08		021
21	11	02	07400045	16	60	AZ	CASA GRANDE	04	79	65	62	00354	PINAL	08		021
21	11	02	07400045	16	66	AZ	CASA GRANDE	04	79	65	62	00354	PINAL	08		021
21	11	01	02400038	25	81	AZ	CASA GRANDE	04	79	65	62	00354	PINAL	08		021
21	12	02	07000055	18	90	AZ	CHANDLER	04	79	72	24	00504	MARICOPA	08	6200	0013
21	12	01	02000049	03	65	AZ	CHANDLER	04	79	72	24	00504	MARICOPA	08	6200	0013
21	12	01	02000049	03	60	AZ	CHANDLER	04	79	72	24	00504	MARICOPA	08	6200	0013
21	12	02	07000055	18	99	AZ	CHANDLER	04	79	72	24	00504	MARICOPA	08	6200	0013
21	11	01	02400038	25	85	AZ	COLLIDGE						PINAL			AZ
21	11	01	02400038	25	86	AZ	COOLIDGE	04	79	65	54	00704	PINAL	06		021
21	11	01	02400038	25	87	AZ	COOLIDGE	04	79	65	54	00704		06		021
21	11	01	02200038	20	31	AZ	DOUGLAS	04	79	57	99	00804	COCHISE	08		003
21	11	01	00300036	23	43	AZ	DOUGLAS	04	79	57	99	00804	COCHISE	08		003
21	11	01	00300036	23	42	AZ	DOUGLAS	04	79	57	99	00804	COCHISE	08		003
21	11	01	00300036	23	41	AZ	DOUGLAS	04	79	57	99	00804	COCHISE	08		003
21	11	01	00200036	21	20	AZ	DOUGLAS	04	79	57	99	00804	COCHISE	08		003
21	12	02	05900054	02	10	AZ	EAGEN						APACHE			AZ
21	11	01	03000039	22	60	AZ	EHRENBURG						YUMA			AZ
21	11	01	02400038	25	84	AZ	ELOY	04	79	65	82	01054	PINAL	07		021
21	11	01	02400038	25	83	AZ	ELOY	04	79	65	82	01054	PINAL	07		021
21	11	01	02400038	25	82	AZ	ELOY	04	79	65	82	01054	PINAL	07		021
21	11	01	05600043	14	20	AZ	FLAGSTAFF	04	79	21	40	01154	COCONINO	10		005
21	11	01	05600043	14	22	AZ	FLAGSTAFF	04	79	21	40	01154		10		005
21	11	01	05600043	14	21	AZ	FLAGSTAFF	04	79	21	40	01154	COCONINO	10		005
21	11	01	04200041	12	80	AZ	GILA BEND	04	79	72	52	01294	MARICOPA	04	6200	0013

RESULTS—FIRST MATCH: REFERENCE FILE AGAINST DATA FILE—EXACT MATCH

RUN STATISTICS

20,708 INPUT RECORDS READ--(DATAIN)

INPUT REJECT RECORDS READ--(REJTIN)

20,708 INPUT RECORDS SELECTED FOR PROCESSING

103,090 REFERENCE RECORDS READ--(REFERIN)

3,715 UNIQUE REFERENCE FILE KEYS

6 TEMPORARY OUTPUT BLOCKS WRITTEN-(WKFILE)

1,116 TEMPORARY INPUT BLOCKS READ--(WKFILE)

254 REFERENCE CANDIDATES IN MAXIMUM DOMAIN

23 REFERENCE CANDIDATES IN AVERAGE DOMAIN

22 REFERENCE CANDIDATES FOR AVG DATA RECORD

15,881 OUTPUT RECORDS WRITTEN--(MTCHOUT)

4,827 OUTPUT RECORDS WRITTEN--(REJTOUT)

53 CRITICAL FIELD REJECTS (CR)

1,069 SEARCH FIELD REJECTS (SR)

GROUP REJECTS (GR)

3,705 SELECT REJECTS (SE)

DATA FILE SEQUENCE CHECKS

15,881 RECORDS MATCHED (MT)

12,913 EXACT MATCHES

MATCH RATE = 76.69 PER CENT

SECOND MATCH OUTPUT—UNCERTAINTY COMPARATOR ON PLACE AND COUNTY

```
211202054001082562CTBARK HAMSTED                                  09164131    7LITCHFORD          005
211302062001270224CTBREDGEPORT ——ACCEPT                           0916828002003                121160001
211201017001040831CTCHESIRE                                       09167134    7                    8880009
211403058001343232CTDARIAN                                        09168440    7FAIRFIELD           8040001
211402034001310722CTDEERBY ——————ACCEPT                           0916764005904NEW HAVEN      09   009
211201018001041291CTEAST GRAMBY                                   09163174    7HARTFORD            3280003
211403070001362240CTGLASTONBURT                                   09163276    7HARTFORD            3280003
211402043001330744CTHIGGAMUN                                      09166175    7MIDDLESEX            007
211402043001330741CTHIGGMIN  }——REJECT                                        MIDDLESEX             CT
211402043001330742CTHIGGOMUN }                                                 MIDDLESEX             CT
211403063001352892CTHUNTINGTON                                    09168186    7                    9999001
211502045001430860CTMARLBOROUGH                                   09163290    7HARTFORD             003
211203066001111821CTMIDDLESEX <————————————————————————————————————————————> MIDDLETOWN          CT
211202038001061536CTNAGATUCK  }——ACCEPT                           0916732013704NEW HAVEN       098880009
211202059001100132CTNANGATUCK }                                   0916732013704NEW HAVEN       098880009
211402046001332650CTNEW BRITAINY }                                0916345013903HARTFORD        115440003
211202062001102265CTNEW BRITIAN  }                                0916345013903HARTFORD        115440003
211401026001300824CTNEW BRITIAN  }                                0916345013903HARTFORD        115440003
211401026001300822CTNEW BRITIAN  } ACCEPT                         0916345013903HARTFORD        115440003
211403056001342882CTNEW BRITIAN  }                                0916345013903HARTFORD        115440003
211401026001300820CTNEW BRITIAN  }                                0916345013903HARTFORD        115440003
211403058001343231CTNEW CANAUN                                    09168410    7FAIRFIELD           8040001
211401013001283151CTNOUGHATUCK }                                               NEW HAVEN             CT
211501025001401172CTOID GREENWICH } REJECT                                     FAIRFIELD             CT
211303030003420760MAREIDVILLE    }                                             CT
211403072001370907CTSEYMORE                                       09167620    7NEW HAVEN           009
211502037001412080CTSOUTH NORWICH                                                                      CT
211401003001280340CTWALLINFORD                                    09167120    7NEW HAVEN           009
211202053001081333CTWARERBURY ——REJECT                                         NEW HAVEN             CT
211402047001333009CTWETHERFIELD                                   09163410    7HARTFORD            3280003
211502036001411184CTWETHS                                                      HARTFORD              CT
211502036001411185CTWETHS                                                      HARTFORD              CT
211101015000922364CTWILLAMANTIC }—ACCEPT                          0916129025904TOLLAND         08   015
211403070001362243CTWILLIAMANTIC}                                 0916129025904WINDHAM         08   015
211202054001082560CTWINDSOH LOCKS—ACCEPT                          09163140    7HARTFORD            3280003
211107091002612013ID                                                           CANYON                  0
```

RESULTS—SECOND MATCH: UNCERTAINTY COMPARATOR ON PLACE AND COUNTY

RUN STATISTICS

```
4,827 INPUT RECORDS READ--(DATAIN)
      INPUT REJECT RECORDS READ--(REJTIN)
4,827 INPUT RECORDS SELECTED FOR PROCESSING
103,090 REFERENCE RECORDS READ--(REFERIN)
  606 UNIQUE REFERENCE FILE KEYS
   24 TEMPORARY OUTPUT BLOCKS WRITTEN-(WKFILE)
  514 TEMPORARY INPUT BLOCKS READ--(WKFILE)
  254 REFERENCE CANDIDATES IN MAXIMUM DOMAIN
   55 REFERENCE CANDIDATES IN AVERAGE DOMAIN
   54 REFERENCE CANDIDATES FOR AVG DATA RECORD
  650 OUTPUT RECORDS WRITTEN--(MTCHOUT)
4,177 OUTPUT RECORDS WRITTEN--(REJTOUT)
   53 CRITICAL FIELD REJECTS  (CR)
   37 SEARCH FIELD REJECTS    (SR)
  382 GROUP REJECTS           (GR)
3,705 SELECT REJECTS          (SE)
      DATA FILE SEQUENCE CHECKS
  650 RECORDS MATCHED         (MT)
    5 EXACT MATCHES

                    MATCH RATE =   13.46 PER CENT
```

SAMPLE SUPPLEMENTAL REFERENCE FILE

STATE / PLACE	COUNTY	STATE / COUNTY / SMSA / PLACE / PLACE DESCRIPTION / PLACE SIZE / ATA-SPLC CODE
AKMENDELTNA		02 7CC999999
AKNELCHINA		02999999999999999999999
AKNELCHINS		029999999999999999999999
AKNORHT KENAI	KENAI	02120 7CC999999
AKNORTH KENAI	KENAI	02120 700999999
AKPRIMROSE		029999999999999999999999
AKRENAIE		02120 1E3C4CE812C6C
AL		01999999999999999999999
AL	CALLAS	01047 999999
AL	COTTONDALE	011258ECC 476164
AL	HALEYVILLE	01133 0820406473527
AL	HOLT	011258600 476174
AL		01 999999
AL	COVINGTON	01039 999999
ALCOVINGTON		01C91 C4754CC476933
ALDLAYTON		01C035160C4654C5479645
ALDOPHNE	COFFEE	01031 053046647763C
ALELBO	MONTGOMERY	011C1524C 999999
ALFLETA	WASHINGTON	01129 999999
ALFOLR PARTS	WASHINGTON	01129 999999
ALFOUR POINTS	ETOWAH	01C5528ECC68531147175C
ALGODSDEN	MARENGO	01091 999999
ALHALF ACRE		01 999999
ALMONPGOIY	WILCOX	01131 999999
ALOAKMULGEE	RUSSELL	011131800136541047471C
ALPHOENIX		

RESULTS—THIRD MATCH: UNCERTAINTY COMPARATOR AGAINST PERMANENT SUPPLEMENTARY FILE

RUN STATISTICS

4,177	INPUT RECORDS READ--(DATAIN)
	INPUT REJECT RECORDS READ--(REJTIN)
4,177	INPUT RECORDS SELECTED FOR PROCESSING
2,547	REFERENCE RECORDS READ--(REFERIN)
136	UNIQUE REFERENCE FILE KEYS
	TEMPORARY OUTPUT BLOCKS WRITTEN-(WKFILE)
	TEMPORARY INPUT BLOCKS READ--(WKFILE)
25	REFERENCE CANDIDATES IN MAXIMUM DOMAIN
7	REFERENCE CANDIDATES IN AVERAGE DOMAIN
7	REFERENCE CANDIDATES FOR AVG DATA RECORD
3,820	OUTPUT RECORDS WRITTEN--(MTCHOUT)
357	OUTPUT RECORDS WRITTEN--(REJTOUT)
13	CRITICAL FIELD REJECTS (CR)
205	SEARCH FIELD REJECTS (SR)
93	GROUP REJECTS (GR)
46	SELECT REJECTS (SE)
	DATA FILE SEQUENCE CHECKS
3,820	RECORDS MATCHED (MT)
29	EXACT MATCHES

MATCH RATE = 91.45 PER CENT

MANUAL CODING OF TEMPORARY SUPPLEMENTAL REFERENCE FILE

INDIANA - NTCFS GEOCODING SYSTEM - REJECT LISTING

ST	PLACE-NAME	COUNTY NAME	FIPS STATE	FIPS COUNTY	CENSUS SMSA	CENSUS PLACE	CENSUS DESC	CENSUS SIZE	SPLC STATE	SPLC COUNTY	SPLC PLACE
IN		BEDFORD	18	093	----	0145	4	08	37	53	50
IN	Morgan	MARGAN	18	109	3480	----	-	--	37	--	--
IN	(IL + KS) IN ELGIN		18	9	----	----	-	--	--	--	--
IN	(KS) IN PAGE CITY	LOGAN	20	109	----	----	-	--	59	62	35
IN	YORKTOWN IN TOUKTOWN		18	035	5280	2840	4	04	36	73	68
IN	SOUTH IN WHITLEY		18	183	----	2385	4	03	36	69	22
IN	VIN IN WINCENNES	KNOX	18	083	----	2610	4	08	32	69	20
IN	ELK IN WLKHART		18	039	----	0740	4	10	36	21	20
IN	WLKHART	WLKHART	18	039	----	0740	4	10	36	21	20
IN	5	PORTER	18	127	2960	----	-	--	36	--	--

MANUAL CODING OF TEMPORARY SUPPLEMENTAL REFERENCE FILE (CONTINUED)

ALABAMA - NTCFS GEOCODING SYSTEM - REJECT LISTING

ST	PLACE-NAME	COUNTY NAME	FIPS STATE	FIPS COUNTY	CENSUS SMSA	CENSUS PLACE	CENSUS DESC	CENSUS SIZE	SPLC STATE	SPLC COUNTY	SPLC PLACE
AL		HALEYVILLE	01	133	----	0838	4	06	47	35	27
AL	FOUR PARTS	WASHINGTON	01	129	----	----	-	--	47	--	--
AL	FOUR POINTS	WASHINGTON	01	129	----	----	-	--	47	--	--
AL	(OK) OAKMULGEE	WILCOX	01	131	----	----	-	--	47	--	--
AL	PRATTV PRAYYVILLE	AUTAGUA	01	001	----	1415	4	08	47	54	70

TEMPORARY REFERENCE FILE FOR FOURTH MATCH

INDIANA -NTCFS GEOCODING SYSTEM ------FIPS------- ----CENSUS----- -------SPLC------- PAGE

STATE	PLACE NAME	COUNTY NAME	STATE	COUNTY	SMSA	PLACE	DESC	SIZE	STATE	COUNTY	PLACE	16
IN		BEDFORD	18	093		0145	4	08	37	53	50	
IN		MARGAN	18	109	3480				37			
IN	ELGIN		18	999	9999	9999	9	99	99	99	99	
IN	PAGE CITY	LOGAN	20	109					59	62	25	
IN	TOUKTOWN		18	035	5280	2840	4	04	36	73	68	
IN	WHITLEY		18	183		2385	4	03	36	19	92	
IN	WINCENNES	KNOX	18	083		2610	4	08	37	69	70	
IN	WLKHART		18	039		0740	4	10	36	21	20	
IN	WLKHART	WLKHART	18	039		0740	4	10	36	21	20	
IN	5	PORTER	18	127	2960				36			

ALABAMA - NTCFS GEOCODING SYSTEM ------FIPS------- ----CENSUS----- -------SPLC------- PAGE

STATE	PLACE NAME	COUNTY NAME	STATE	COUNTY	SMSA	PLACE	DESC	SIZE	STATE	COUNTY	PLACE	1
AL		HALEYVILLE	01	133		0830	4	06	47	35	27	
AL	FOUR PARTS	WASHINGTON	01	129					47			
AL	FOUR POINTS	WASHINGTON	01	129					47			
AL	OAKMULGEE	WILCOX	01	131					47			
AL	PRAYYVILLE	AUTAUGA	01	001		1415	4	08	47	54	70	

SOURCE CODE/JCL SAMPLE

```
//TCF003JP JOB (DS,1000010000,008270,A99),TERRY,MSGLEVEL=1,
//   PRTY=13,NOTIFY=DXG004
//     EXEC GEOMONTH,OUTSER=W03194
XXGEOMONTH PROC TRK=10,              SORT ALLOC FOR SUPPL REF FILE
XX             BLK=80,               OUTPUT BLKSIZE
XX             OUTSER=               SERIAL NO OF 7TRK TAPE FOR CDC
***
***  PLACE AND COUNTY ENCODER   --PACE--          CENSUS USE STUDY
***
***  NATIONAL TRUCK COMMODITY FLOW STUDY
***  GEOCODING SYSTEM
***
***  END OF MONTH PROCEDURE
***     CODES REMAINING REJECTS BY MATCHING TO SUPPLEMENTARY
***     REFERENCE FILE OPTIONALLY AUGMENTED TEMPORARILY WITH
***     REFERENCE RECORDS TO CODE EVEN GIBBERISH DATA RECORDS
***          CREATES 7 TRK TAPE FOR CDC
***
XXNTEM4  EXEC  PGM=NTEM
***  NTEM4 - TEST FOR REJECT FILE NOT EMPTY
XXSTEPLIB  DD  DSN=TCF001.UNILOAD,DISP=SHR
XXFILE     DD  DSN=TCF001.REJECTS,DISP=SHR
IEF236I ALLOC. FOR TCF003JP NTEM4
IEF237I 654   ALLOCATED TO STEPLIB
XXGIBBERSH EXEC PGM=IEBGENER,COND=(0,NE,NTEM4)
***
***  GET SUPPLEMENTARY REFERENCE FILE RECORDS TO CODE DATA WITH
***          FAULTY STATE ABBREVIATIONS
***
XXSYSPRINT DD  SYSOUT=A
XXSYSUT1   DD  DDNAME=FIX
XXSYSUT2   DD  DSN=&&GIBBERSH,DISP=(NEW,PASS),UNIT=SYSDA,
XX             SPACE=(TRK,(2,5),RLSE),DCB=TCF001.SUPREF
XXSYSIN    DD  DSN=TCF001.SOURCE(GEC80),DISP=SHR
```

SOURCE CODE/JCL SAMPLE

```
***
XXSYSIN      DD  DSN=TCF001.SOURCE(GEOSTATE),DISP=SHR
XXSYSPRINT   DD  SYSOUT=A,DCB=BLKSIZE=1210
XXREPORT     DD  DUMMY,DCB=BLKSIZE=1210
XXINFILE     DD  DSN=&&GIBBERSH,DISP=(OLD,PASS)
IEF202I - STEP - LIST    , WAS NOT RUN BECAUSE OF CONDITION CODES.
IEF236I ALLOC. FOR TCF003JP LIST
  74008.1701
  74008.17C1 CPU    0MIN 00.00SEC MAIN    OK LCS    OK
  000000.98  RUT    000007.42  SEC  ID  L168
IEF373I STEP /LIST  / START 740
IEF374I STEP /LIST  / STOP  740
BCS37XX STEP /LIST  / CRU   000
XXSORTGIBB  EXEC PGM=SCRT,REGION=100K,PARM='CCRE=MAX',
XX          COND=(0,NE,NTEM4)
***
***  SORTGIBB - SORT SUPPLEM REF FILE WITH TEMPORARY REF RECORDS
***             TO CODE ALL REMAINING REJECTS. THIS FILE IS
***             TEMPORARY TO AVOID INTRODUCING GARBAGE INTO
***             SUPPLEMENTARY REFERENCE FILE
***
XXSORTIN     DD  DCB=TCF001.SUPREF,DISP=(OLD,DELETE),DCB=TCF001.SUPREF
XX           DD  DSN=&&GIBBERSH,DISP=(NEW,PASS),UNIT=SYSDA,
XXSORTOUT    DD  DCB=TCF001.SUPREF,SPACE=(TRK,(50,10),RLSE)
XX
XXSYSOUT     DD  SYSOUT=A
XXMATCH4  EXEC   PGM=XEQ,CCND=(0,NE,NTEM4),REGICN=150K
***
***  MATCH4 - SAME AS MATCH 3 OF GEOCODE
***           CODES REMAINING REJECTS FROM LAST GEOCODE CYCLE
***           CORRECTED IN LAST GEOFIX CYCLE
***
XXSTEPLIB    DD  DSN=TCF001.UNTLOAD,DISP=SHR
XXSYSIN      DD  DSN=TCF001.SOURCE(GEOMTCH3),DISP=SHR
XXSYSTEMP    DD  UNIT=SYSDA,SPACE=(TRK,(15,10))
XXSYSPRINT   DD  SYSOUT=A,DCB=BLKSIZE=3458
XXDATAIN     DD  DSN=TCF001.REJECTS,DISP=OLD
```


Question Period

Mr. DeVorkin—Since this is perhaps the first major application of UNIMATCH, I am certain that there will be a tendency to be careful about loosening up the uncertainties, the uncertainty windows. I would be interested if you have any plans for some reverse direction experiments. That is, how loosely can one set the uncertainty, to determine perhaps a totally machine process where there is very little done by hand excepting perhaps a small percentage of incorrect matches, so as not to effect the statistics of the final deductions.

Mr. Carlberg—Let me say first that the parameters that were set into UNIMATCH, that is the match parameters, were set by the Bureau of the Census and DOT and not by us. Any such experimentation would be their experimentation and not ours. We could process it, if they wanted to loosen the reins. It would probably be something that DOT should do. In terms of the State code, you couldn't really do much since it is only two characters long. The Census Bureau was only looking at the first 15 characters of the place code. They had to have an exact match on the first two characters and then beyond that they used what they called default penalties. I don't want to get into too much detail, but the penalties that they used had been experimented with and determined to be reasonable. I think that it might be real interesting to try loosening the reins and to see if we would still get a good match. You would want to look at it very closely for a while, but the 98 percent achievement level isn't bad.

Mr. Kondo—What was the logic in rejecting only the first two letters wrong as opposed to any errors in any other location?

Mr. Cooke—I think I can offer a tentative explanation. If you allow the uncertainty comparator to look at all characters in a place name, you are going to be looking at a huge domain. If you say that it's got to match on the first two characters of a place name, then you considerably reduce your domain. You reduce it by two orders of magnitude.

Mr. Kondo—Is that already wired into UNIMATCH?

Mr. Carlberg—No, it is strictly optional. You specify exactly what you want to match. You decide what you will allow, and you specify the weights and penalties. Every hit gets a weight, and you can decide what it is. Every penalty gets a weight, you decide what it is. This is all variable. You can decide how many characters that you want to look at and how you want to look at them. It is very fexible.

Mr. Kondo—You said you had 75 percent match on the first run. Was this an average across the country? Was there a wide range with a greater match in some areas and less in other areas of the county?

Mr. Carlberg—I have no statistics on that at all. I don't know. I don't think we even tried to find out. I would only guess that is is pretty uniform. I doubt if there is anything geographical about it at all.

Census Data User Services Activities

ROBERT B. VOIGHT

INTRODUCTION

As the title of this discussion implies, the Bureau of Census Data User Services Office (DUSO) was established in the fall of 1972 to serve as a focal point for improved access to census information and techniques on the part of experienced and potential users. Its mission includes publications, seminars, conferences, training sessions, and research and development activities to achieve the maximum exploitation of census data and methodology by government agencies at all levels, nonprofit organizations, business and industry, and the general public.

DUSO devises, tests, and applies techniques of improving access to the census data base, extends the uses of census data through surveys and studies of the uses made, the needs for census data and the gaps in data which may be appropriate for the Census Bureau to fill. It researches new techniques for incorporation into regular Bureau operations and procedures to improve services to all data users. It serves as the single point to which a customer may come to purchase summary data tapes, published and unpublished data, census maps, computer programs such as ADMATCH, GRIDS, CARPOL, the GBF/DIME files and the like, or obtain special tabulations or compilations of data not available on summary tapes or in published form on a cost reimbursable basis. It prepares general purpose statistical compendia such as the *U.S. Statistical Abstract*, the *Pocket Data Book*, data guides by subject and area for both census and other federally collected statistics, and nonfederal data available for States and local areas; and reports on the methodological and procedural aspects of major censuses and other statistical programs of the Bureau which serve as guides or "cook books." A recent development which has generated widespread interest has been the distribution of computer-generated narrative profiles for States, central cities, SMSA's, and congressional districts providing the socioeconomic characteristics from the 1970 Census of Population and Housing, and we plan to expand this to other subject areas in the Bureau.

DUSO stresses assistance to data users in the access to use of published and unpublished census data through information provided in newsletters, data access descriptions, user guides, face-to-face exchanges in conferences and workshops. Through its research and development activities and surveys of users and nonusers it seeks to develop better ways of meeting their needs through regular Bureau operations and procedures.

ORGANIZATION

The Data User Services Office has a staff of about 125 people, the majority of whom are statisticians, social science analysts, editors, and urban specialists and a small unit of highly qualified computer programers. There are six staffs within the office:

The Data Access and Use Laboratory—Michael Garland, Chief

Special Tabulations Programing Staff—Richard Hornseth, Chief

Statistical Compendia Staff—William Lerner, Chief

Census History Staff—Phyllis Carter, Chief

Users' Services Staff—Larry Carbaugh, Acting Chief

Census Use Study—Caby Smith, Chief

Although these staffs are under my general supervision, our general philosophy of organization might best be thought of as a federation in which each staff chief is responsible for the activities and programmatic structure of his unit but the common mission of all units is service to users and potential users.

It may be of interest to note that the bulk of the funding for these activities comes from the products and services which are provided on a cost reimbursable basis. This has a very salutary effect on our efforts to provide the best service to the customers.

CURRENT ACTIVITIES

Since most of you are either actual or potential users of census data and/or techniques at either the operational, planning, administrative, or granting quantum of the total data spectrum, I would like to review briefly some of the current DUSO activities which may be of interest to you. In several of these you will note the vital necessity for having a good updated GBF/DIME file which is the basic theme of this conference.

The following projects were recently completed in the Indianapolis Unified Statistical Evaluation Study where a small staff from the Census Use Study is resident and working with the various city organizations in practical applications of census data and techniques.

We were asked by the Manpower Administration of the U.S. Department of Labor to determine whether it is

technically feasible to generate neighborhood profiles of job seekers. They were hoping for development of a data system to help local decision makers determine manpower priorities for the manpower revenue sharing program recently enacted.

We used the Indiana Employment Security Automated Reporting System (ESARS file) as a basic source since it is an accumulation of demographic information about job seekers. We extracted the home address and social security number of persons in the active files of the five Employment Security field offices that serve Marion County. Using the social security number as a file link we added addresses of the job seekers to the ESARS file. The Marion County portion of the file was then geocoded using the ADMATCH program and the GBF/DIME file. The socioeconomic profiles were then constructed by computers. These reports were generated for each of the 9 townships of Marion County, for each of the 25 city-county council districts, and for 9 selected poverty areas in which a multiplicity of agencies have an interest.

Our next interest was to learn whether there was any additional value for this type of profile. So sample reports were distributed to a number of local groups in Indianapolis. Thus far, we have been informed that the Manpower Area Planning Council Secretariat is using the data extensively in preparing their list of priorities and expenditures for the coming year.

We have also been told that the Indiana Employment Security Division and the Indianapolis Metropolitan Manpower Commission are considering opening branch offices in the neighborhoods with the largest number of job seekers so as to be more accessible to the people they are trying to serve.

The Director of the Community Service Program's Multi-Service Centers has used the data to determine which Multi-Service Centers should be staffed with an employment counselor. Incidentially the decisions of Employment Security, the Indianapolis Metropolitan Manpower Commission and the Multi-Service Center manager of the Community Services Program were made jointly, so as to establish services in a maximum number of neighborhoods of need, while avoiding setting up rival employment offices next door to one another.

The planning staff of the Indianapolis Public Schools indicates that the data will be used in conjunction with other information that will be mentioned later in this report, to determine the locations at which to present basic adult education courses.

So although the data was produced for a specific need, a number of additional concrete decisions have already been made from the reports and this will continue to be true for many types of data applications.

The following examples have been provided by local users of the GBF/DIME files. In each of these the value of an up-to-date file such as is achievable through cooperation in the Census Bureau CUE program should be self-evident.

Incidence Mapping

As an extension of an agency's information system, data maps were needed. By geocoding local data to the GBF/DIME file, computer maps could be presented showing geographic trends of housing deterioration, ethnic population shifts, patterns of specific disease incidence or specific project service areas. Geocoding allows all data-budgetary, demographic and project output measures to be filed within the same system. Virtually any data category can be mapped and compared with any other. (Boston Model City Research and Program Assistance Office).

Land-Use Information System

The objective was to create a system for area profiles and simulation. The land-use file containing zoning, lot size, land-use, condition, and address data for each parcel was geocoded. Public safety records and other local data were integrated into the GBF/DIME file as well. An area profile of local data cross-classified by census socioeconomic data was then possible. (Planning Department, Stamford, Conn.)

Housing—Code Inspections

The object was to provide housing inspectors with profile information of the areas they were inspecting. A real estate file of assessments, housing inspections and building permits was matched to the GBF/DIME file. Other local data (school, police, health) as well as census socioeconomic data were integrated with the DIME file. A "state of the neighborhood" profile resulted informing the inspectors about what to look for in the neighborhood to be surveyed. (City Planning Department, Springfield, Mass.)

Market Analysis - Economic Surveys

The objective was to determine potential sites for industrial and commercial establishments. This was accomplished by using the GBF/DIME file as a reference file for the assignment of coordinates to addresses of commercial and industrial establishments. The addresses of the establishments were obtained from the yellow pages, Polk City directory, Dunn and Bradstreet, national business lists and special devices. The resulting establishment file with coordinates was used in the production of automated maps of the incidence of establishments by standard industrial classification. (Urban Data Processing, Cambridge, Mass.)

CARPOL Program

The computerized CARPOL recently issued by the Bureau to help alleviate the gasoline shortage is being established in a number of major metopolitan areas. More than 3,000 copies of the documentation have been distributed to date. This is a real life case demonstration of the value of an updated GBF/DIME file available in all of the SMSA's.

We have been assembling local experiences with the use of the GBF/DIME files and expect to issue a Census Use Study report on these in early June. It may give local people clues to additonal uses to which they can put their files.

FUTURE ACTIVITIES

After many months of testing and the incorporation of significant improvements, the record linkage program for application to urban data files, called UNIMATCH, developed by the Census Use Study has been published and copies of the documentation manual and the computer program tape will be available from the Users' Services Staff of DUSO at the usual fee of $70 by the end of May 1974.

UNIMATCH is a much more powerful and faster record linkage than its predecessor, ADMATCH, which has been in local use for several years. Previous matching systems have been restrictive in the sense that they have been tied to fixed linkage algorithms. With UNIMATCH a user can define a matching algorithm suitable to his task, and then automatically execute his algorithm to complete his application. The system can be used for almost any conceivable linkage application by defining, through the UNIMATCH language, the nature of the record linkage task.

Over the next several months DUSO will be making available computer summary tapes of the 1972 censuses of wholesale, retail, and services trades; manufacturers and mineral industries; construction industries, transportation, and State and local governments similar to those made available from the 1970 decennial census. In the very near future you will be able to purchase from DUSO summary tapes containing the data used for the allocation of general revenue sharing funds to some 38,000 local governments for the first three entitlement periods. These may be of some interest since they contain data for very small governmental units. Summary tapes of the data contained in the *1972 County and City Data Book* have been available for several months, containing over 190 cells of statistical information for States, counties, SMSA's, and cities of 2,500 or more containing data from the various censuses and other Federal data collected on a nationwide basis.

The Director of the Census, who spoke to you yesterday is on his way to Madrid with a small cadre of DUSO staff members to discuss the establishment of the GBF/DIME system in Spain. It is already in use in Israel and France and a number of other countries have expressed interest in adopting it for their use.

In the planning, but not yet budgeted, stage is an extensive census user training program which we hope to conduct at the Bureau where users and potential users may come in to take concentrated 3 and 4 week courses in various census methods and techniques which should assist them in the application of statistical data geocoding systems to their own needs.

Within the next 60 days we expect the *Index to Selected 1970 Census Reports.* This companion document to the *Index to 1970 Census Summary Tapes* should prove to be an extremely valuable guide to what is available in detail from all the cross-classified data published from the 1970 census. I can testify personally as to its value for I have referred to it many times in the last few weeks to answer inquiries from users while it was being readied for the printer.

SUMMARY

I would like to sum up by telling you that the DUSO organization stands ready to assist you in gaining access to census data and techniques. We want to improve the data services of the Bureau. It is equally important for you to let us know about the shortcomings and gaps of census products as well as the uses you make of them. It is only through a continuous dialogue with users that we can improve our services.

Finally, there is a little data below the national level that does not require geographic orientation to be useful. The degree to which the Census Bureau and the local users can maintain up-to-date GBF/DIME files through cooperative efforts in the CUE program, can greatly enhance the value of Federal, State, and local information for a wide spectrum of administrative, operational and planning efforts.

Question Period

Mr. Etienne—We asked for a special tabulation on some neighborhood units and found the price very, very high. Is there any way we could get around something like this? The price was $7,000 for a neighborhood tabulation.

Mr. Voight—I agree that is a rough one. One of the difficulties, perhaps, is the fact that we haven't yet gotten ourselves to the stage where we can handle smaller jobs as efficiently as we do the larger jobs. Part of this is simply due to the configuration of our hardware. In the future, of course everybody talks about the future, but with the hardware improvements that we hope to achieve over the next couple of years at the Bureau, we will be able to reduce some of those costs substantially.

We have compared some of the costs on reimbursable jobs that the Bureau does with some of the summary tape processing centers that are around the country providing data from census summary tapes. We stand in about the middle. Those that are university based and can use slave labor among the graduate students can give you a pretty good price. Those that are operated by private business have prices that are somewhat higher. At the moment, we are sort of in the middle range. The cost also depends to some extent on the complexity of the job, of course, even though it may be a small job.

I don't know about your job in particular, but we are not the cheapest, I grant you that. I don't know if we will ever be the cheapest because, like all bureaucratic things, there are overheads that we automatically have to apply. We simply can't bypass those.

Mr. Hearle—Any other experience with special tabulation requests that anyone would like to describe or raise questions about? It's clear that there is going to be more and more that as people seek to match local data with the data which the Census Bureau cannot, for confidentiality reasons, release. Was this a matching process or just a tabulation of census data?

Mr. Etienne—It was a special tab on neighborhood unit block data.

Mr. Hearle—So it was just census data in effect. Were you going to do any matching of it yourself with data that was locally generated when you got it?

Mr. Etienne—No, our policy planning unit, which was newly established in New Orleans, did a survey and developed 25 neighborhood units for the city. They wanted a neighborhood profile to work on policy planning for the city.

Mr. Hearle—Taking the example that Mr. Etienne raised, would the Bureau staff, your staff, counsel with the local agency about how to achieve their objectives in ways that might be less costly than the way that the local agency actually presented it to you.

Mr. Voight—Yes, if we have not in your particular case, we perhaps should have. As a matter of fact, for the majority of requests for special tabulations that come in, when we look at either the table outlines that the customer sends in or the specifications that he provides, one of the first things that we do is review it to see if it can be streamlined and thus reduce the cost. Both the cost and time factor is a crucial one because inevitably there is queuing up at the Bureau on special tabulations. Many times we work out with the customer a satisfactory approach that would be far less costly than that which he had originally assumed.

You also have to remember that the people come in, I won't call it with a request, I'll call it a "wish list." They want a terrific amount of data; they really aren't appreciative of the costs involved in getting it. We try to gently let them down to the critical items that they are really interested in. We try to reach an amicable solution to their data needs. I would urge any of you who have cause for requesting a special tabulation to first write and let us know that you are interested, describing it as well as you can. Then let us come back to you for the specifics and make a pretty thorough review before we actually get to the question of providing you a cost estimate for the work.

Mr. Hearle—I think if those letters included not only the specifications of the data but also the policy objectives that are sought in the analysis, the Census staff, from experiencing similar requests elsewhere, could often think of a different way to get at the same answers through some other array of data that might either be less costly or more interesting or useful.

Mr. Etienne—I must say that the Census Bureau did cooperate with us and made several suggestions. As a result, due to the cost and some of the suggestions from the Census Bureau, the boundaries of the neighborhoods were realigned to match tract boundaries. We got the same information, but it wasn't as specific as we wanted in the beginning. They chose to get a more generalized neighborhood.

Mr. Kondo—In my contacts with the Census Bureau, particularly with the User staff, my experience has been very good. They have helped us, making suggestions, redefining what we wanted and providing information on available data of which we were previously not aware. I guess the only comment that I do have is of the time that it takes to get some of the material. There may not be a real good answer to that problem because of the great volume of work.

Mr. Voight—We are, if I may put it bluntly, literally ashamed of some of the time estimates that we have to provide. Part of it is the nature of the animal. Whenever you go to the basic records it means you have to come to the Census. There is to some degree a queuing up simply waiting your turn to be serviced. I think our handicap at the moment is not so much an inability to handle the programing end and to get ready to tabulate, as much as it is the way in which we have to do the tabulations. We have to go to the basic tape file. That means lots of manual tape handling and making sure the right tapes are mounted on the machines. There is a large ray of hope though. As soon as we are able to achieve a status of random access on our computers, our turn-around time will be much, much better.

Mr. Thyagarajan—The 1960 census was used by a private company to produce county-level cross-tabulations that proved extremely valuable in terms of general trends especially in regional planning. I had occasion to use the data in the Detroit region where I was working. I find that that kind of gap exists now in spite of the fact that the Census Bureau can do all these fancy tabulations. Just the access to a publication that is relatively cheap and inexpensive, that is readily available, I think is missing now. I don't think any amount of cost cutting and so forth is going to replace that kind of standardized cross-tabulated data that was available for the 1960 census. Could you comment on whether something like that is likely to come out, or is it too late? The 1970 census data is sort of getting old.

Mr. Voight—I have to agree with you to a certain extent. On the other hand, it's the only data that is available in many instances. I am aware of that activity in 1960. I think the man had a good idea, but he went broke simply because he didn't have enough business to continue it. We would have been very happy if someone else had picked up the ball this round and repeated that type of activity. On the publication side, what we have attempted to do and how well we have succeeded remains to be seen a little later down the decade.

The Bureau is limited in terms of overall funding so that the amount of data published can't be open ended. We have to be pretty selective in the amount of data that we publish. That is the reason for the extensive effort to produce summary tapes in the hope that people could make extensive use of the tapes through the summary tape processing centers or on their own equipment and thus derive some of the data, perhaps even more cheaply, than would have been possible had it been published. One other factor that is discouraging, at least to us, is the general cost of getting data published between 1960 and 1970 and particularly in the last couple of years when GPO has raised its prices on government publications as high as 300 percent on some of the items.

Mr. Thyagarajan—A suggestion, if it is possible to do so. You are going to publish the *Urban Atlas*. You are going to pursue that for several SMSA's as I understand it. Is it possible to include in the *Atlas* sort of a simplified summary table using the public use sample and attach maybe half a dozen key cross-tabulations for SMSA-level purposes? As I understand, populations of 250,000 is the criterion for public use sample stratification.

Mr. Voight—Well, it could be done. It's a little late in the game, I am afraid, to get that type of information into the *Urban Atlas* which will already carry about 12 general categories of data. Mr. Meyer, do you want to speak on that point in more detail?

Mr. Meyer—The *Urban Atlas* has been designed only to provide a graphic summary of selected existing census data on a tract by tract basis to facilitate comparisons and relate differences between tracts spatially. It was not designed as a vehicle to provide additional cross-classifications of data not already available from the census. There are no possibilities at this time of restructuring the program.

Mr. Voight—How familiar are you with the *County and City Data Book*?

Mr. Thyagarajan—Pretty well.

Mr. Voight—You don't think that would provide enough information?

Mr. Thyagarajan—No, not enough for the purposes that I'm interested in. We had to go through a special tab to do this.

Geographic Base File Development in the Minneapolis-St. Paul Area

REGIS VOYE

The Metropolitan Council is the regional coordinating and planning agency for the Twin Cities area covering seven counties that make up the Minneapolis/St. Paul metropolitan area. This area consists of 189 minor civil divisions, a large part of which is unincorporated area. The Council was created in 1967 by the Minnesota State Legislature. It is supported financially by a .7 mill levy, augmented by State and Federal funds. Areas of planning for the Council are health, housing, aging, cable TV, criminal justice, transportation, open space, sewer, water, solid waste disposal, air quality and physical development.

The scope of our planning and the problem of coordinating large masses of data for planning led us to conclude that the GBF was the best tool in organizing and analyzing geographic data as a part of a larger data base project. Due to lack of funds to develop the GBF, nothing was done until the spring of 1973. In 1973, the Council budgeted local funds for the development of the GBF. During that year we spent a total of $34,500. This expenditure breaks down as follows:

	Dollars	Hours
1. Coding limit/nonmatched	$2,098	672
2. Segment name consistency	2,955	936
3. Address edit	7,488	2,266
4. Coordinate change	3,254	1,050
5. Extension map work	2,679	796
6. Computer services	3,201	
7. Programing and supervision	12,825	
	34,500	

We are presently into the update phase of the Census Bureau's CUE program and hope to complete the extension to the seven-county area by January 1975. Our staff is made up of four coders, one demographer, one part-time programer and me.

CUE

The CUE program has been moving along with a few minor delays. The first delay was for the running of the ADDEDIT report by the Census Bureau. This resulted in a delay of a month until the computer processing was completed by the Census Bureau. The second delay was due to getting corrections keypunched and processed with FIXDIME C. FIXDIME C was not delivered until January 1974.

The file for our area contains 84,042 records. The coding limit/segment name phase of CUE produced 27,964 segment corrections. The address edit change/coordinate phase of CUE, which we are winding up now, has produced 40,000 segment corrections. Of these, 9,000 were for the x-y coordinates. The major areas for errors in our file are ZIP codes followed by address range. Another problem is duplication of node numbers. A great deal of staff time was devoted to tracking down and correcting this type of error. Duplication of nodes occurred on census tract boundaries for the most part. This probably occurred during the original coding for the urbanized area. A large staff was employed to do a rush job, with inadequate supervision and quality control procedures. For this reason, we now employ a smaller crew (four coders) who work in teams and are assigned to one task, on one map, at a time. In this way, supervision is kept close to the work performed.

During Phase I of CUE we processed all corrections on keypunch cards. This became a handling problem due to the volume of cards. To overcome this, I developed a system for storing and handling all corrections. The system required two programs written in COBOL. (See exhibit A.)

Program 1 lists the corrections in the sequence in which they were written. The heading for the list shows card columns 1 through 80. A record count is also shown on each detail line for control.

Program 2 has five parts—

1. Input void records are loaded into a table of sequence numbers. The table is used to drop matching sequence numbers in part 3.
2. Input current corrections from card.
3. Input previous corrections from tape. Each record is checked against the table generated in part 1 and matches listed and dropped.
4. Sort all input records by record sequence numbers, card types and the from/to, left/right indicators.
5. The output is a listing of all corrections and a magnetic tape file. Each record is numbered and this shows on the listing and is inserted into the tape record. This number is then used for voiding records on the next run of the program. The listing also shows error diagnostics such as duplicate corrections, blanks and input/output totals.

The final version of the file is used as input to FIXDIME C for updating of the GBF/DIME file.

PROBLEMS

We faced many problems during the coding and mapping of the seven-county area. There are a few we have not yet overcome.

1. No street name and house numbering systems, especially in the unincorporated areas.

   The metropolitan area is presently planning for the installation of an emergency telephone number system (911). We hope that our rural areas will see the benefits of this system and be encouraged to develop street names and numbering systems. In most cases, our local communities in the rural areas are apprehensive about State and Federal data systems. Convincing them of the need for street names and house numbering for GBF is an impossible task.

2. The geopolitical makeup of our area presents another major problem. Seven counties, 189 minor civil divisions, 7 rivers, and 927 lakes over 10 acres in size, create mammoth mapping and coding problems. The State highway department has developed a set of maps taken from aerial photos which assist in coding, but many lakes are not shown and many rural streets are unnamed. We have had to go to the local communities to solve many of these problems.

3. The slowness of the Census Bureau in delivering maps, programs and procedures in some cases has made it necessary for us to go out looking for work for the GBF crew in order to retain them as a working unit. I believe this is due to staffing and budget problems at the Bureau.

4. The Office of Management and Budget has expanded the SMSA to 10 counties. The Metropolitan Council was created by State legislation for seven counties and has no legal responsibility outside that area. One of the three new counties is in the State of Wisconsin. This is potentially an even larger problem. We know little about these counties and do not know the local officials. The size of the 10-county area is a large one, 4,845 square miles, roughly the size of the State of Connecticut. An area of this size would present many and varied mapping and coding problems.

USES OF GBF

The Council is planning to integrate the file into our geopolitical component of a data base. (See exhibit B.) The data base is presently in the testing stage using locally developed correspondence tables. The GBF and its coordinates will be added when coding is completed for the SMSA.

Presently, we have a number of files with coordinate values. We were using arbitrary coordinates for all files. These files cover data collection zones, traffic analysis zones, health planning areas, and urban planning districts. With GBF becoming available we will use its state plane x-y coordinates as the coordinate system which will be a part of our data base. A single coordinate system will then be used by all planning and operating units within the Council.

We are using the file in its present form for ADMATCH processing. I wrote a splitter program in COBOL to prepare the file for the preprocessor ADMATCH program.

The St. Paul Police Department is using the GBF in their ALERT system. The file is used to code incidents for later reporting by geographic area. They have used the file to build an intersection name file with address ranges. This is used for automobile accident reporting and analysis.

The St. Paul Housing and Health Departments are using GBF for geocoding all data in their reporting systems. Work is then assigned by census tract and reporting to the Federal Government is done by census tract.

The city is using the Ramsey County numeric street number system. All input from the two departments is coded with the numeric equivalent for the street name. The code and house number are then processed against a segment side version of the GBF and the census tract is assigned to the record.

SCHOOL DISTRICTS

At present several school districts in the metropolitan area are looking into the feasibility of using GBF for busing students. GBF gives them the ability to determine if a student lives over a mile from the school to which he or she is assigned. This will enable the districts to meet the State law busing program requirements and also minimize their costs by using automated methods.

HENNEPIN COUNTY HEALTH COALITION

This group is looking into the feasibility of using GBF for ADMATCH for processing health records.

CITY OF MINNEAPOLIS

The Planning Department has a need for ADMATCH but due to computer hardware differences is unable to utilize the program and GBF for its processing. They have developed their own address-matching program.

UNITED FUND

We have several funds in the metro area that have shown interest in GBF. They are waiting for the completion of CUE to start utilizing GBF for address processing.

SUMMARY

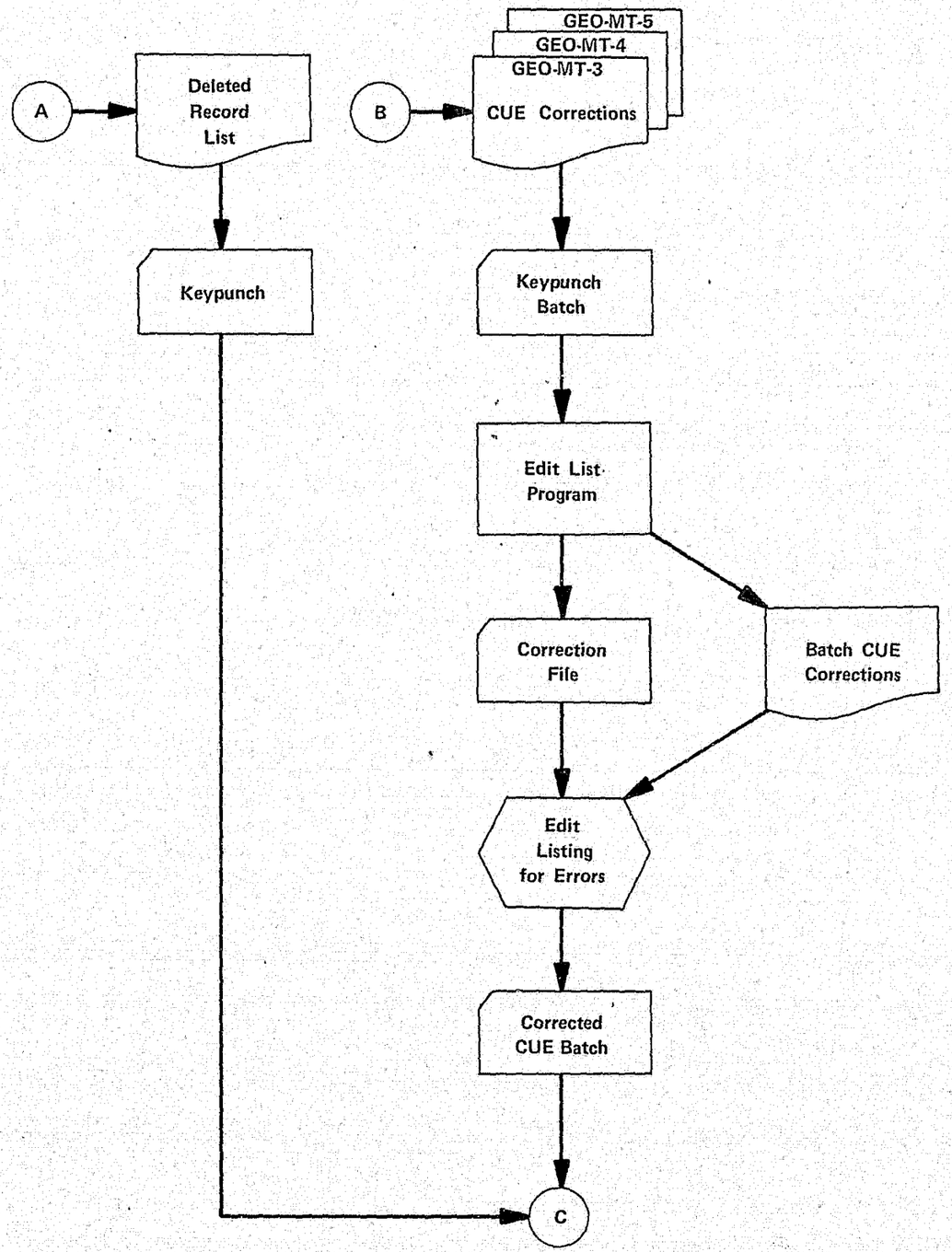
The key to local governmental units using GBF is in the Metropolitan Council developing and maintaining the GBF

and a master street name numeric code. Additionally, we must provide for numbering new streets by local agencies who need to record local data in local data files as part of their responsibility as operating agencies.

Presently, street name numeric coding is fragmented with each unit maintaining their own code structure. In order for us to utilize local data in a planning mode we must have easy access to this data. Regionwide numeric coding gives us a means of accomplishing this. With our coordinating, maintaining and checking of the GBF and street name number coding, we have the ability to control the accuracy of the GBF. Using CUE as a continuing program, we will continue to provide the GBF to the local users.

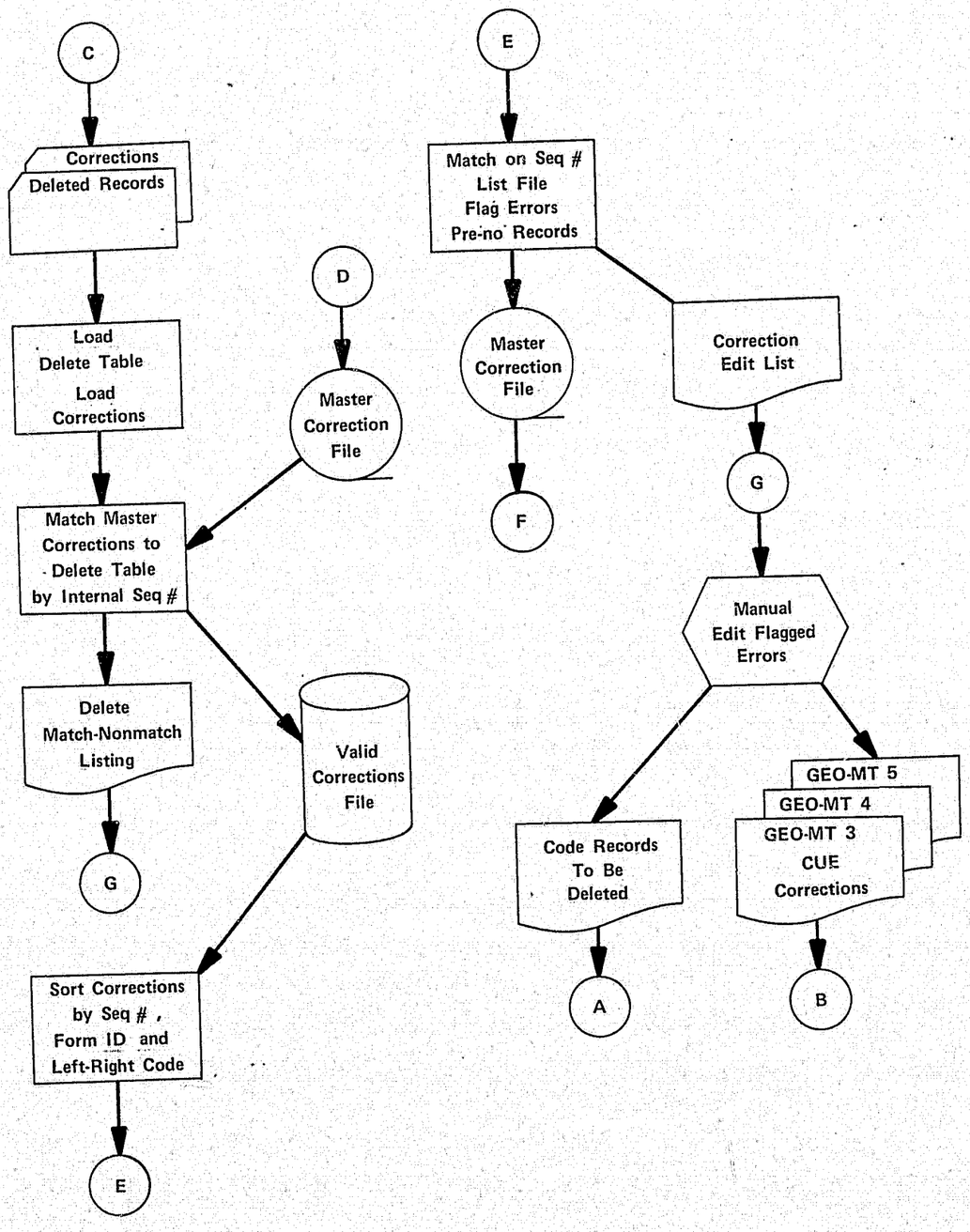
Pre FIXDIME Edit

Exhibit A-1



Pre FIXDIME Edit—Continued

Exhibit A-2



Pre FIXDIME Edit—Continued

Exhibit A-3

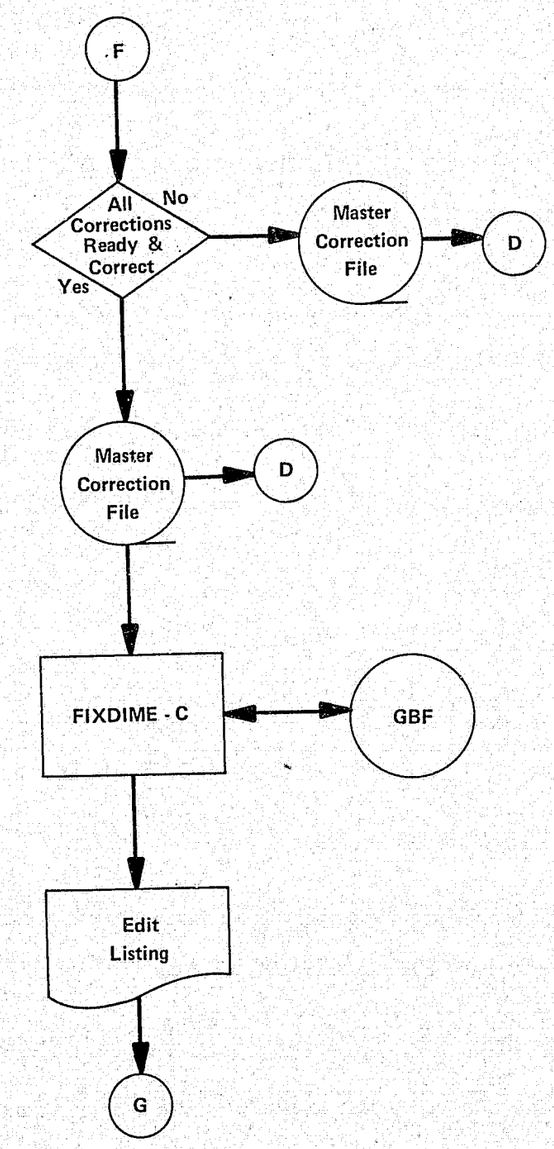


Exhibit A-4

DATE 03/22/74 C-U-E C O R R E C T I O N S G-B-F EQ - 0125 PAGE NO 0677

CARD TYPE	FILE CODE	RECORD NUMBER	C K	CLF	COLUMNS 13 THRU 80 IN SEQUENCE 3456789012345678901234567890123456789012345678901234567890123456789012345678901234567890 2 3 4 5 6 7 8	ERROR DIAGNOSTICS	SEQ NO.
4	5120	079997-4		R	55343		36681
5	5120	079997-4			026002025502900F4602140200999744857409342630678086214864 1		36682
3	5120	079998-2			51200799982000001W 78TH ST *		36683
4	5120	079998-2		L	55343		36684
4	5120	079998-2		R	55343		36685
3	5120	079999-0			51200799990000001W 78TH ST *		36686
4	5120	079999-0		L	55343		36687
4	5120	079999-0		R	55343		36688
4	5120	080000-4		L 308	00240200247655113		36689
4	5120	080000-4		R 307	00240300247755113		36690
4	5120	080009-5		L 106	00242200244655113		36691
4	5120	080009-5		R 103	00242100244755113		36692
4	5120	080015-2		R	002301002361		36693
4	5120	080021-0		L	55102		36694
4	5120	080021-0		R	55102		36695
4	5120	080024-4		L	55102		36696
4	5120	080024-4		R	55102		36697
4	5120	080025-1		L	55102		36698
4	5120	080025-1		R	55102		36699
4	5120	080038-4		L	000629000631		36700
3	5120	080056-6			51200800590000004S HABASHA ST		36701
4	5120	080056-6		L	55107		36702
4	5120	080056-6		R	55107		36703
5	5120	080056-6			036100015501500F46037815015829449418093091107092212235326		36704
4	5120	080057-4		L	55107		36705
4	5120	080057-4		R	55107		36706
5	5120	080057-4			036100015501500T46037815015829449418093091107092212235326		36707
4	5120	080058-2		L	55107		36708
4	5120	080058-2		R	55107		36709
4	5120	080059-0		L	55107		36710
4	5120	080059-0		R	55107		36711
3	5120	080062-4			51200800673000006 WACHTLER RD		36712
5	5120	080064-0			060601009502400T46036109013110449024093126006947882226422		36713
5	5120	080065-7			060601009502400F46036109013110449024093126006947882226422		36714
4	5120	080140-9		L	001745001831		36715
4	5120	080141-6			3		36716
4	5120	080142-4		L	001833001923		36717
4	5120	080149-9		L	55418		36718
4	5120	080149-9		R	55418		36719
4	5120	080152-3		L	55391		36720
4	5120	080152-3		R	55391		36721
5	5120	080152-3			026602016701800F46016976019877450009093516507297612125038		36722
4	5120	080153-1		L	55391		36723
4	5120	080153-1		R	55391		36724
4	5120	080154-9		L	55391		36725
4	5120	080154-9		R	55391		36726
4	5120	080155-6		L	55391		36727
4	5120	080155-6		R	55391		36728
4	5120	080157-2		L	55427		36729
4	5120	080157-2		R	55427		36730
3	5120	080160-6			51200801606000001* WALNUT GROVE LA 9		36731
4	5120	080188-7		L	55106		36732
4	5120	080188-7		R	55106		36733
4	5120	080189-5		L	55106		36734
4	5120	080189-5		R	55106		36735

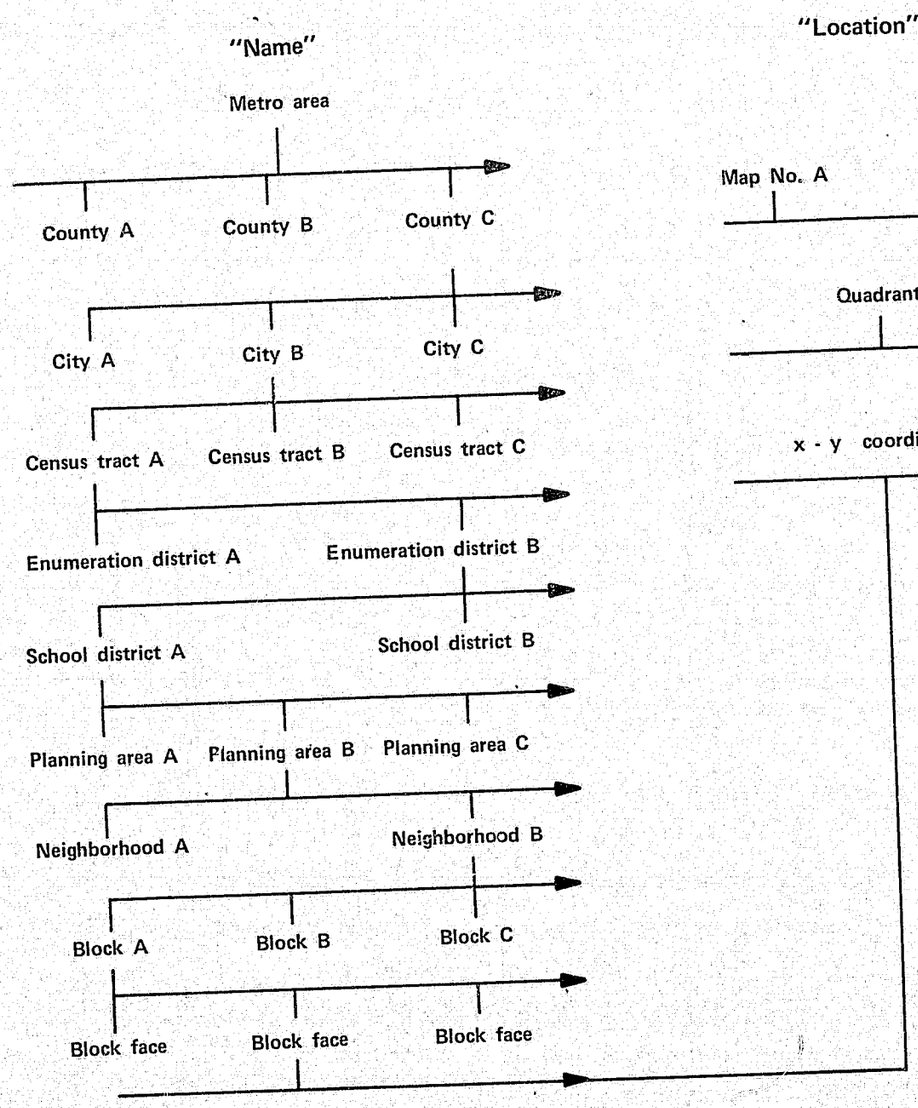
Exhibit A-5

DATE 03/22/74 C-U-E C O R R E C T I O N S G-B-F EQ - 0125 PAGE NO 0705

CARD TYPE	FILE CODE	RECORD NUMBER	C K	CLF	COLUMNS 13 THRU 80 IN SEQUENCE 3456789012345678901234567890123456789012345678901234567890123456789012345678901234567890 2 3 4 5 6 7 8	ERROR DIAGNOSTICS	SEQ NO.
3	5120	082485-5			51200824855000001 WINNETKA AV N		38221
4	5120	082505-0		L	00013000015055107		38222
4	5120	082505-0		R	00012900016355107		38223
4	5120	082506-8		L	00015200019855107		38224
4	5120	082506-8		R	00016500019955107		38225
4	5120	082548-0		L 215	000824000850		38226
4	5120	082548-0		R 216	000823000851		38227
4	5120	082549-8		L 214	000852000880		38228
4	5120	082549-8		R 216	000853000881		38229
3	5120	082554-9			51200825548000001* WINSLOW AV		38230
3	5120	082555-5			51200825555000001* WINSLOW AV *		38231
5	5120	082558-9			030100014601600T46033993018200449762093169107215802215034		38232
5	5120	082559-7			030100014601600F46033993018200449762093169107215802215034		38233
4	5120	082560-5		L	000000000030		38234
4	5120	082561-3		L	000032000062		38235
5	5120	082619-9			000000142601700F46025667018960449874093339107252512170996		38236
4	5120	082625-6		L	55109		38237
4	5120	082625-6		R	55109		38238
4	5120	082626-4		L	55109		38239
4	5120	082626-4		R	55109		38240
4	5120	082627-2		L	55109		38241
4	5120	082627-2		R	55109		38242
4	5120	082628-0		L	55109		38243
4	5120	082628-0		R	55109		38244
4	5120	082629-8		L	55109		38245
4	5120	082629-8		R	55109		38246
5	5120	082629-8			070903004101300T46043165021131450185092981707374382263377		38247
4	5120	082630-6		L	55109		38248
4	5120	082630-6		R	55109		38249
5	5120	082630-6			070903004101300F46043165021131450185092981707374382263377		38250
4	5120	082631-4		L	55109		38251
4	5120	082631-4		R	55109		38252
5	5120	082631-4			070903004601300T46043165021193450194092981707377722263377		38253
4	5120	082632-2		L	00257700259955109		38254
4	5120	082632-2		R	55109		38255
5	5120	082632-2			070903004401300T46043165021192450194092981707377722263377		38256
5	5120	082632-2			070903004601300F46043165021193450194092981707377722263377		38257
4	5120	082654-5		L	004901004 4755428		38258
4	5120	082654-5		L	4901 49475542R	DUP-RECORD ERR	38259
4	5120	082654-5		R	55428		38260
4	5120	082655-3		R	55428	DUP-RECORD ERR	38261
4	5120	082655-3		L	55428		38262
4	5120	082660-3		R	55428		38263
4	5120	082660-3		L	55428		38264
3	5120	082661-1		R	55428		38265
3	5120	082662-9			51200826611000001 WISCONSIN CIR * 9		38266
4	5120	082669-4			51200826629000001* WISCONSIN AV N		38267
4	5120	082669-4		L042001903	00162400164855113		38268
4	5120	082670-2		R041900305	00162300164755113		38269
4	5120	082670-2		L042001902	00171800178455113		38270
5	5120	082670-2		R041900901	00170100185555113		38271
5	5120	082688-4			000000210101600F46033964019267469917093171707272122214291		38272
4	5120	082700-7			040601018000500T46038399026432450955093078907651782237992		38273
4	5120	082700-7		L	55391		38274
				R	55391		38275

Geopolitical Reference File Concept

Exhibit B



Question Period

Mr. Conboy—I notice that you mentioned in your problem area, ZIP codes. I seem to get the inference that it was basically a key punch problem, not a definition problem. Is that correct?

Mr. Voye—Key punch or coding, I don't know which. I haven't dug back into the records to determine exactly which one created the problem.

Mr. Conboy—I am going to throw this out for discussion. In Rhode Island the ZIP codes are almost nondefined. We have cases in which if you live within a half a mile of a post office you will have one ZIP, beyond that another. If you wish to pick up your mail on the way home, you have one ZIP; if you want to have it delivered, you have another.

The area in which I live is a suburban area. There are no defined boundaries. It has its own ZIP code within the city where I live. No one has been able to show me where one ZIP area ends and the other begins. Is anybody else having this problem? Is Rhode Island unique? Secondly, has anybody found a solution because, quite frankly, we are having a lot of trouble trying to put these in the file with any kind of confidence. Basically, ZIP code depends on where you live and whose word you take.

Mr. Hearle—Yes, that is correct. It is helpful to remember that the ZIP codes were designed solely for the purpose of delivering the mail.

Mr. Conboy—I might ask then, if we all know that to be the case, why is it so necessary in this file? If we know that it has problems, and it is basically a tool for sorting mail in general areas, is it as necessary as people seem to say it is, considering the time, money, and the frustration involved?

Mr. Kunz—The gentleman asked my question. I was going to ask the same thing. I am very sensitive to this since we have had over 1,000 pages of corrections for ZIP codes. We have been working on this for about 3 or 4 months. This is our real albatross. The only way that we solved it, if this is any help to you, is that we sent staff people out to every post office within the two-country region. We sat down with the postmaster or superintendent of mails, or whoever was most knowledgeable, and made them lay out an exact service area on the metropolitan maps.

Mr. Conboy—We have done that to the extent of providing the maps to the Postmaster at the Turnkey Post Office.

Mr. Kunz—You have to go with the maps, and you have to check two districts against each other. When you go to an adjacent one, sometimes they don't agree. Then you have to go back to the other one. Finally, they agree.

Mr. Conboy—I don't know what the situation is in New York, but in Rhode Island we have some relatively large areas which are exclusively rural. You literally have a one- or two-man post office department. It is almost impossible.

Mr. Kunz—There is no solution for that.

Mr. Conboy—No, there isn't any. I am wondering, is anybody here from Census who would like to give me some kind of an answer because I am going into this again this summer. It took quite a bit of time with very little results last summer.

Mr. Meyer—The problems that you speak about are all too real. We are aware of them on a daily basis from areas all over the country. Every area has problems involving ZIP codes; some admittedly much worse than others. Unfortunately, ZIP codes are necessary for geocoding operations. Let me give one illustration: Commonwealth Avenue, right here in Boston, extends through many different communities, and since each community operates its own house numbering system, not only do the address ranges keep on repeating but sometimes one side of the street has a different address range than the other. The only available discrimination is ZIP code. We recognize the problem and the impact upon the local area and are exploring ways of conducting at least a partial ZIP-code correction process by computer before sending the file to the local agency.

Mr. Kondo—I notice in the last chart that you had, there was hierarchy of areas. In the Massachusetts program, there was mention of the small-area analysis zone which is an attempt to try to get a common unit by which we can interrelate different areas that different people use. I can't remember the exact sequence of the hierarchy in your chart but you had a census tract, an enumeration district, on down to a neighborhood. Was the neighborhood the common basic unit of the hierarchy with each higher level unit area composed of multiples of the next lower level?

Mr. Voye—No, it is not. The block is the determining factor, the lowest level; the enumeration district, school district, planning area, neighborhood, and census tract are not. There is no hierarchical structure as far as a planning area being within a school district, or a school district being within an enumeration district, or an enumeration district being within a census tract. That is why we had to use the block face. The block is the determining factor in which one of these it lays. They are not homogeneous.

CUE Program Development in the Northeast Ohio Area

ANTHONY MA

INTRODUCTION

While I do not have a completely successful story to tell, I believe our experience in the preparation of the CUE program would contribute to the total development of GBF/DIME. This paper attempts to analyze the development of the CUE program from a systems approach which describes the interaction between the organism (NOACA's CUE program) and its environment. Our focus is particularly on the immediate environment.

The Northeast Ohio Areawide Coordinating Agency (NOACA) was formed in 1968 for the purpose of coordinating, chiefly through review of local general and special purpose governmental units' applications for Federal grant-in-aid, affected comprehensive and functional planning activities in northeastern Ohio. The Demonstration Cities and Metropolitan Development Act of 1966 and later the Intergovernmental Cooperation Act of 1968, provided the impetus for the agency's establishment.

In 1969 NOACA absorbed the Seven County Transportation/Land Use Study (SCOTS) which had been established in 1964 to fulfill the requirement of the 1962 Federal Highway Act that a geographic area receiving Federal highway funds prepare a comprehensive transportation/land use plan.

Since that time NOACA's obligations to coordinate and review Federal- and State-funded planning activities and proposals in metropolitan northeastern Ohio have grown. Now, the agency serves as the areawide program coordinator for the following Federal and State agencies: Executive Office of Management and Budget (OMB), designated metropolitan clearinghouse (OMB Circular A-95 Revised); HUD, metropolitan regional council; FHWA, UMTA, and Ohio Department of Transportation (ODOT), conducting a defined "continuing phase" transportation study.

Presently, NOACA does not operate under one particular Ohio enabling statute. Chiefly, its authority and structure are determined by several Federal and State laws, and their managerial requirements and guidelines. A change in the agency's status, hence its authority and responsibilities, is anticipated as the State of Ohio establishes Regional Planning and Development Organizations (RPDOs).

The NOACA area encompasses seven northeastern Ohio counties—Cuyahoga, Geauga, Lake, Lorain, Medina, Portage, and Summit—which total 2,931 square miles. The largest is Portage, 506 square miles; the smallest, Lake, 236 square miles. The landforms range from coastal along Lake Erie to, in the interior, relatively level woodland plateaus that are drained primarily through six river basins. The character of the area varies from major metropolitan to rural agricultural.

The 1970 census records a total area population of 3,000,276. The most populous county is Cuyahoga at 1,721,300, which includes Cleveland (750,903) and four other cities with populations exceeding 50,000. Two other populous counties are Lorain, 256,843, which includes the cities of Lorain (78,185) and Elyria (53,427); and Summit (553,371), which includes the city of Akron, 275,425.

For various demographic and economic purposes, the Bureau of the Census aggregates counties and large cities into standard metropolitan statistical areas (SMSA's). Presently, the NOACA area is made up of three SMSA groups: Cleveland SMSA—Cuyahoga, Geauga, Lake, and Medina Counties; Akron SMSA—Summit and Portage Counties; Lorain-Elyria SMSA—Lorain County. A map of the NOACA area is attached.

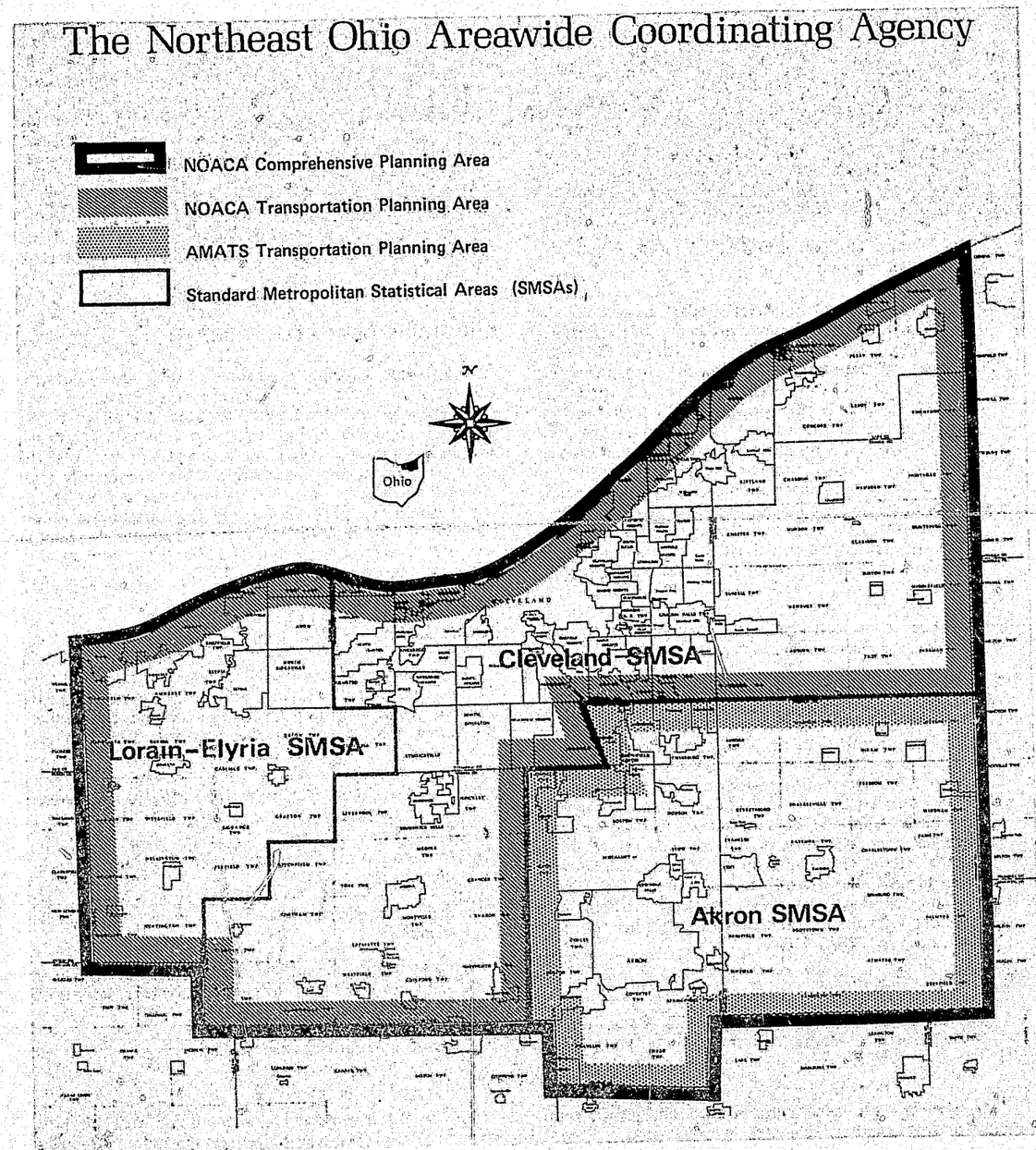
COMPONENTS OF THE IMMEDIATE ENVIRONMENT

A. Bureau of the Census

This brief description of our organization and the area would indicate the complexity of the preparation of the CUE program. In 1973, we were encouraged by the Bureau of the Census to begin the preparation of the CUE program. Mr. Jake Silver has been most helpful in assisting us to initiate the program. His assistance gave us a good start and his continuous support keeps us persistently pursuing the program under a very difficult and complex environment. It is very exciting to work with people who are cooperative. This working relationship with the Bureau of the Census is one component of the immediate environment.

B. The Nature of Our Organization

NOACA at this time does not collect primary data for comprehensive land use plan and correction of base maps. Much of this work is to be done by the Regional and County Planning Commissions within our area. The



immediate requirement is to establish an effective mechanism to coordinate the input data from these RPCs and CPCs. In addition, there are many functional and single-purpose planning agencies in the area which could provide much of the primary data to the areawide data base file and they are potential users of the GBF/DIME. Hence, much of the effort in the preparation of the CUE program is directed to understand and evaluate the components and to identify an appropriate mechanism to coordinate the work of these agencies.

C. The Financial Difficulties in the Preparation of the CUE Program

Initially, we sought financial support from both the private companies and public programs. Unfortunately, due to the lack of knowledge in utilizing GBF and oversold by the staff, this approach failed. When we began to develop the CUE program this time, we decided to finance the project by Federal program (one-third HUD, one-third UMTA and one-third FHWA). There was no difficulty in the appropriation of funds from the HUD program for this work but it was unclear whether transportation planning could finance part of this effort. We failed to receive concurrence from the Department of Transportation until the energy crisis came. The agency needed the GBF for the car-pooling program and CUE may be undertaken as eligible project cost.

D. The Lack of Knowledge in the Applying of GBF in Planning and Operational Program

There was a lack of understanding in the application of GBF/DIME on the part of the planner and the public. Consequently, little or no enthusiasm has been generated to proceed with the development of CUE, even though the Bureau of the Census continuously attempted to communicate with planners through workshops, conferences, and individual contacts. However, Jake Silver from the Bureau, persistently works with areawide planning agencies and provides them with materials on CUE. Because of his effort, our program began to take shape. In an attempt to promote the CUE program, we have designed a training program and established localized conferences to demonstrate the application of GBF/DIME.

E. The State Government

The State of Ohio proposed to divide the State into 11 substate districts for the delivery of State services and 15 districts for planning and coordination. This concept basically is a good one; however, the problem in northeast Ohio is that the proposed boundary cuts through the SMSA line. At the present time, much of the data on economics, health, and transportation are on the SMSA basis. This proposal creates two fundamental issues: (a) The structure and function of the organization will be changed and the role between the areawide agency and its county, regional and city planning commissions will have to be redefined, and (b) the imposition of the proposed boundary will create many technical problems in reestablishing the base data file.

THE STATUS OF OUR CUE PROGRAM

The Census Bureau gave us some of their survey data on 80 areas in doing the CUE program. Based on this experience, we have estimated our total project cost. We also have applied to various Federal governmental agencies to jointly fund the program and it was part of the FY 1975 work program activity. Negotiation with the county and regional planning commissions and consultants have begun and it would be contracted by July 1974. In summary, our program is on its way to begin. It is anticipated that part 1 and part 2 correction phases and the first phase of the update will be completed by July 1975.

CONCLUSION

The success of this program depends on how well the agency (organism) interacts with its immediate environment which is dynamic and continuously changing. Essentially, the agency who will be involved in the CUE program should—

(a) establish a close working relationship with the Bureau of the Census,

(b) understand the conditions imposed by the nature of this agency,

(c) secure adequate financial as well as technical resources,

(d) demonstrate the utility and application of GBF/DIME in order to provide a cost/benefit analysis of CUE, and

(e) adapt to externally uncontrolled changes.

Question Period

Mr. Harward—My experience in Massachusetts is that many of the functional planners in other than transportation seem to think that this is not going to do them any good because the data will be too fine grained. They are interested in the eastern half of the State, western half of the State, four districts, statewide districts, or something like that. Personally, I feel they are being very shortsighted. Have you run into that problem?

Mr. Ma—We haven't really encountered that kind of problem, as such. But we do have the question continually asked in terms of jurisdiction. Regional Planning Commissions ask, "What is the areawide agency doing in my territory collecting this microdata? You only deal with area-wide issues, why are you coming down here to collect this detailed kind of information?" Our basic answer to that is, "If you want to get an overall picture you have got to set up some building blocks. The best building block is from the GBF." I think that is probably the only answer. You just can't have an overall picture and look at it that way. You have got to somehow build it up. Am I answering your question?

Mr. Harward—To a certain extent. I hope after we are all done and I listen to all these people with their tales of how they tried to sell the program that we will end up with everybody saying, "Give us the program," but I doubt it.

Mr. Ma—As I said before, we have an experience of overselling ourselves; therefore,we are holding back so that we do not oversell ourselves. We want to really put some effort into demonstrating this as a workable and operational program. I don't know how we can do it because we haven't reached this point yet. I am glad that Mr. Voye, from Minneapolis, was sharing with us the kinds of frustrations and problems that he had. We haven't really jumped into it yet, but we are preparing to jump into it. When I see the kind of things that are ahead of us, we hope that we are better prepared than before.

County Boundary Coordinate File for Graphic Display of Data

GERARD BOETJE, CHRISTIAN DAVIS, AND
DONALD DeVORKIN

(Presented by Dr. Donald DeVorkin)

1. INTRODUCTION

Several software systems have been developed for the display of data. Certain of these systems such as SYMAP, CALFORM, DPS, etc. are capable of displaying data values associated with a geographic area. This is accomplished by density shading the geographic area based on the relative magnitude of the data values.

These software systems implicitly require certain capabilities and/or files. Requirements are—

A name or code associated with each geographic area.

A machine-readable file which contains the boundary coordinates for the labeled geographic areas.

A machine-readable file of the data to be displayed. This data must have the same name or code as used in the geographic area definitions.

2. GEOCODED DATA

Many Federal agencies and industrial organizations collect or associate data by geographic area. Although there is an increasing tendency toward standardization, in practice, much useful data can be found geocoded in a wide variety of systems. Therefore, the association of data coded to different systems requires a conversion capability.

3. GEOGRAPHIC CODE CONVERSION CAPABILITY

Because of the conversion problem, a Geographic Converter File (File 1) has been developed. This file provides conversion capability at the county or county equivalent level. It encompasses those codes listed in table 1 for each county in the United States.

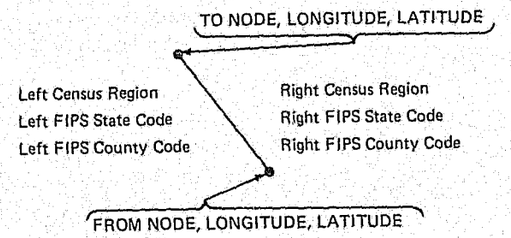
File 1, therefore, provides the capability to associate data under geocoding systems.

4. BOUNDARY COORDINATE FILE

While File 1 allows data associations at the county level, and the systems listed above provide for machine display, a county boundary coordinate file is necessary.

The Bureau of the Census has developed a machine-readable file containing the coordinates of the county and State boundaries of the continental United States. This file is in DIME format. Each record in this file represents a county boundary segment, and contains information as shown below.[1]

DIME FILE County Boundary Segment Information



5. BOUNDARY COORDINATE FILE FOR DISPLAY PURPOSES

The software display systems generally require that the boundary coordinate file be in a sequenced format. The Charles Stark Draper Laboratory, Inc. (CSDL), under contract with the Transportation Systems Center (TSC), obtained a copy of the DIME file. Initial use at CSDL demonstrated that a sufficient sequence for ordering the county boundary segments was not available in the DIME file. Consequently, the DIME file could not be used with display software existing at CSDL.

A design and development effort was sponsored by TSC in order to provide a county boundary coordinate file compatible with existing display software. The effort had as its objective, a compressed file which would facilitate—

[1] The digitizing work of the county coordinate file was done by the Bureau of Public Roads (now FHWA) during 1969-1970. The digitizing was performed individually on State clusters which were then integrated up to a total county boundary file. Accuracy of the digitizing was claimed as ±0.025 degrees. The Bureau of the Census assumed the job of reorganizing the file into DIME format.

TABLE 1. Elements of the National Geocoding Converter File 1

REGION	STATE	AREA	COUNTY CLUSTER	COUNTY
An areal unit smaller than the United States and larger than a single State. The number of regions within a File 1 geocoding system ranges from 8 units to 20 units.	An areal unit which is the first order political subdivision of the United States. The number of States within a File 1 geocoding system ranges from 48 units to 51 units.	An areal unit not necessarily smaller than a State but larger than the average county cluster. The number of areas within a File 1 geocoding system ranges from 120 units to 200 units.	An areal unit smaller than a State, generally consisting of two or more counties. The number of county clusters within a File 1 geocoding system ranges from 430 units to 550 units.	An areal unit which is the second order political subdivision of the United States. The number of counties and county equivalents within a File 1 geocoding system is approximately 3,142 units.
9 Region Codes	*11 State Codes*	*5 Area Codes*	*9 County Cluster Codes*	*11 County Codes*
Standard Federal Region	Federal Information Processing Standard State	Office of Business Economics Region	State Economic Area	Federal Information Processing Standard County
Bureau of the Census Division	Interstate Commerce Commission State	ZIP Code Market Area	Interstate Commerce Commission Area	Bureau of Public Roads County
Standard Point Location Code Region	International Business Machine State	Economic Sub-Region	Office of Emergency Preparedness Area	International Business Machine County
Bureau of Public Roads Region	Bureau of Public Roads State	Water Resources Sub-Area	National Location Code Area	Dun and Bradstreet County
National Location Code Region	Dun and Bradstreet State	Land Resources Area	Bureau of Public Roads Area	General Services Administration County
Office of Emergency Preparedness Region	General Services Administration State		Standard Point Location Code County Cluster	Office of Business Economics County
Water Resources Region	Office of Business Economics State		Transportation Zones	Interstate Commerce Commission County
Land Resources Region	Bureau of the Census State		Standard Metropolitan Statistical Areas	Office of Emergency Preparedness County
Freight Rate Territory	Office of Emergency Preparedness State		Market and Production Areas	National Location Code County
	National Location Code State			Standard Point Location Code County
	Standard Point Location Code State			Bureau of the Census County

Use with available systems: CALFORM, SYMAP, etc.

Efficient information storage.

Efficient machine processing.

Areal aggregation of data.

The design of the DOT compressed DIME file calls for a single record to contain all relevant information about the entire boundary between two counties. In this format, all information contained in the original DIME file is preserved.

Because no relevant information is lost, the records and contents of the original file could be reproduced from the compressed DIME file. However, the exact ordering of the records of the original file could not be duplicated. Since the ordering was somewhat arbitrary to begin with, this is not considered a significant issue.

The compressed DIME file focuses on the basic connectivity features of the boundary segments when viewed in a network context. Intermediate points on a network (boundary) link have boundary shape detail as their only purpose.

An explanation of the DOT compressed DIME file concept, along with technical details on file content and structure can be found in appendix A.

Considerable sorting, sequencing and information rearrangement were required in the development of the DOT compressed DIME file. The necessary machine processing steps are described in appendix B.

6. GBF DATA MANAGEMENT SYSTEM

The approach described above, makes possible a geographic base file (GBF) data management system under development at CSDL.

The goal of the system is to provide a means for flexible and efficient areal aggregation.

System features include—

The GBF is disk resident.

Completely independent of the GBF data file, pointer lists are held in core.

Chaining can be carried out in core. This means that if boundaries internal to a region are to be dropped, it can be accomplished in core before the GBF file must be accesssed.

CSDL approaches the DIME file and the resultant GBF organization as a network structured data management system.

This will be of particular relevance as geocoding work proceeds below the county level. In this situation, both the number of records and the aggregation possibilities will increase drastically. The compressed DIME file approach will then become indispensable to the task of interrelating and aggregating data for tabular and geographical display.

7. MAP DIGITIZING

The format of the CSDL compressed DIME file is suggestive of an approach to map digitizing. A development program at CSDL is presently under investigation for accomplishing mapping at the lowest possible cost to the user.

The approach blocks into three major steps—

1. Define the primary node network (on the county level, there are 8,947 primary links) in the county-level DIME file.

2. Define the clockwise-directed closed polygon relationships among the primary nodes.

3. To each primary link, simply add intermediate nodes as necessary to give further shape detail.

ACKNOWLEDGMENTS

This document was prepared under CSDL Project 56-60100 sponsored by the Transportation Systems Center of the Department of Transportation through Contract TSC-698.

The authors would like to express their appreciation to Robert Tap of the Transportation System Center for his guidance and assistance in the development of the material contained in this document.

The authors would also like to express their appreciation to Adele Volta who is responsible for the presentation format design and for editing, and to Rafaela Rodriguez for clerical assistance.

The publication of this paper does not constitute approval by the Department of Transportation of the findings or the conclusions contained therein. It is published for the exchange and stimulation of ideas.

APPENDIX A. DOT Compressed DIME File

Number of records:	17,490
Tape:	9 track, IBM 360
Density:	1600 BPI
Logical record length:	Variable, maximum 1272
Block size:	7294
Record format:	VB (variable blocked)
Label:	None

FILE DESCRIPTION

The DOT compressed DIME file is composed of records which contain the data for all boundary segments common to two adjacent counties. Figure A-1 and the following text describes the file and record structure.

Figure A-1 illustrates a hypothetical county (County A), its boundary segments labeled with small letters, the nodes for each segment numbered from 1 to 14, and adjacent counties. Since there are five counties which are adjacent to County A, there will be only five records in the DOT compressed DIME file for County A. The format and content of each of these records is shown in figure A2.

As illustrated in table A-1 each record can be thought to have two sections. the fixed section contains that information defining the right and left counties, the first node common to both counties and the last node common to both counties. A variable number of nodes defining intermediate boundary segments common to two counties is contained in the variable section of the record.

The number of nodes field is the sum of the number of nodes in the fixed section of the record, which is always two (2), and the number of nodes in the variable section of the record.

Each field occupies four bytes (characters). The record format of the DOT compressed DIME file is listed in the table.

The file is organized such that all records necessary to form a closed county boundary are grouped together in a sequence. The sequence itself gives a clockwise tranverse around the entire county boundary. Therefore, the county being bounded by the closed traverse is toward the right.

It should be noted that, except for records defining county boundaries which are also a country boundary, a record appears once for a county, and then again for the neighbor county but with right and left identifications reversed.

Record Format for DOT Compressed DIME FILE

Field discription	Begin position	End position	Length	Format
Segment number	0	3	4	Fixed binary
From node	4	7	4	Fixed binary
To node	8	11	4	Fixed binary
Left census region	12	15	4	Fixed binary
Left State	16	19	4	Fixed binary
Left county	20	23	4	Fixed binary
Right census region	24	27	4	Fixed binary
Right State	28	31	4	Fixed binary
Right county	32	35	4	Fixed binary
From longitude	36	39	4	Float binary
From latitude	40	43	4	Float binary
To longitude	44	47	4	Float binary
To latitude	48	51	4	Float binary
Number of nodes[1]	52	55	4	Fixed binary
Segment number[2]			4	Fixed binary
Node[2]			4	Fixed binary
Longitude[2]			4	Float binary
Latitude[2]			4	Float binary

[1] This field indicates the number of nodes contained in the record.
[2] A series of this field sequence representing intermediate nodes will be appended to the record. The number of these field sequences is equal to number of nodes field less two.

Figure A-1. Fictional County For DOT Compressed DIME File

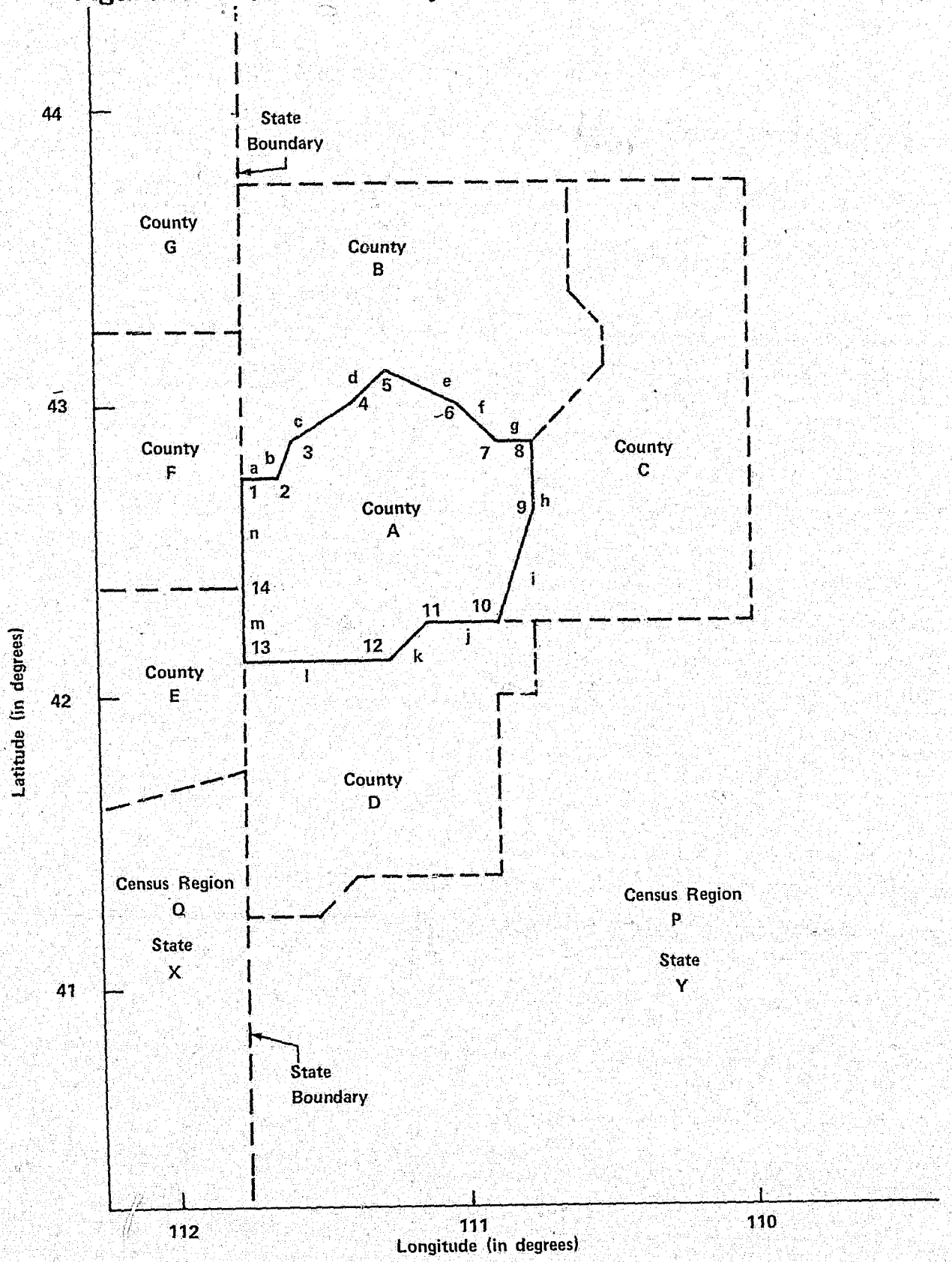


Figure A-2. Records for Fictional County in DOT Compressed DIME FILE

RECORD NO.	FIXED SECTION OF RECORD													VARIABLE SECTION OF RECORD																									
	SEGMENT	FROM NODE	TO NODE	LEFT CENSUS	LEFT STATE	LEFT COUNTY	RIGHT CENSUS	RIGHT STATE	RIGHT COUNTY	FROM LONG	FROM LAT	TO LONG	TO LAT	NO OF NODES	INTERMEDIATE SECTION 1				INTERMEDIATE SEGMENT 2				INTERMEDIATE SEGMENT 3				INTERMEDIATE SEGMENT 4				INTERMEDIATE SEGMENT 5				INTERMEDIATE SEGMENT 6			
															SEG	NODE	LONG	LAT	SEG	NODE	LONG	LAT	SEG	NODE	LONG	LAT	SEG	NODE	LONG	LAT	SEG	NODE	LONG	LAT	SEG	NODE	LONG	LAT
1	g	1	8	P	Y	B	P	Y	A	111.75	42.75	110.75	42.875	8	d	2	111.625	42.75	b	3	111.575	42.75	c	4	111.375	42.875	d	5	111.25	43.125	e	6	111.00	43.00	f	7	110.875	42.875
2	f	8	10	P	Y	C	P	Y	A	110.75	42.875	110.875	42.25	3	h	9	110.75	42.625																				
3	f	10	13	P	Y	D	P	Y	A	110.875	42.25	111.75	42.125	4	j	11	111.125	42.25	k	12	111.25	42.125																
4	m	13	14	Q	X	E	P	Y	A	111.75	42.125	111.75	42.375	2																								
5	n	14	1	Q	X	F	P	Y	A	111.75	42.375	111.75	42.75	2																								

APPENDIX B. Machine Processing Steps for DOT Compressed DIME File

The following machine processing steps were required to produce the DOT compressed DIME file.

1. Record (segment) numbers were added to each Census DIME file record in order to uniquely identify each segment.

2. Each Census record was duplicated with the TO and FROM and the RIGHT and LEFT information inverted. This procedure permits sorting on RIGHT state and county and LEFT state and county FIPS code positions to form a file which has all boundary segments for a given county sorted into adjacent records.

3. The adjacent records (segments) representing a complete county boundary were sequenced in clockwise order.

4. It was observed that the actual boundaries for Park and Teton Counties in Wyoming and Fremont County in Idaho were not in the file. In lieu of these boundaries, the Yellowstone National Park (YNP) boundary had been included. These segments of the YNP boundary, which were not actual county boundary segments, were deleted from the file. The coordinates for the above county boundaries were manually obtained (+ .01 degree accuracy) and inserted into the file.

   The file resulting from the steps 1 through 4 is referred to as the DOT SEQUENCED DIME File. This file contains all of the boundary segments for any given county in adjacent, clockwise-sequenced, segment-oriented records. A technical description of the tape file is listed in appendix C.

5. Those records in the DOT SEQUENCED DIME File common to two adjacent counties are compressed into one record in the DOT COMPRESSED DIME File. A more detailed description of the file is contained in appendix A.

APPENDIX C. DOT Sequenced DIME File

Number of records:	85,054
Tape:	9 track, IBM 360
Density:	1600 BPI
Logical record length:	52
Block size:	7280
Record format:	FB (fixed blocked)
Label:	None

Record Description

Field description	Begin position	End position	Length	Format
Segment number	0	3	4	Fixed binary
From node	4	7	4	Fixed binary
To node	8	11	4	Fixed binary
Left census region	12	15	4	Fixed binary
Left State	16	19	4	Fixed binary
Left county	10	23	4	Fixed binary
Right census region	24	27	4	Fixed binary
Right State	28	31	4	Fixed binary
Right county	32	35	4	Fixed binary
From longitude	36	39	4	Float binary
From latitude	40	43	4	Float binary
To longitude	44	47	4	Float binary
To latitude	48	51	4	Float binary

Question Period

Mr. Weaver—I would like to agree with one comment you made and that is the correction, update and extension of geographic base (DIME) files could be simplified. However, I am afraid that when you take your concept of compressing DIME files down to the street segment level DIME parts, you are going to find that the vast majority of street segments in the GBF/DIME file in urban areas don't have intermediate nodes and I don't think you will find that your techniques are applicable.

Mr. Boetje—We have found that DIME file compression aids you greatly when you are "DIMEing" a geographic area which includes or is bounded by terrain features which are not part of the connected network under consideration—for example, Minneapolis-St. Paul which has a large number of lakes around.

Admittedly, certain of the work in which we are engaged is applicable to only those who would have access to large computing facilities which have cathode ray tubes (CRT or Scope) with light pen graphics. It would then be possible to create, update, and edit node information, address ranges, block face numbers, census information, etc., quickly and efficiently. The DIME file creation, update, and edit capability using such a system could be much faster than current methods.

Mr. Kondo—I didn't fully understand the function of the primary "from" and "to" nodes.

Mr. Davis—The primary nodes are those nodes which form the initial and terminal points of a boundary common to two adjacent counties. The cosmetic or intermediate nodes are those nodes which cartographically define the shape of the boundary between the primary nodes. For example, if a boundary common to 2 counties is defined by 40 nodes, the first and the 40th nodes are the primary nodes and the 2d through the 39th are the cosmetic or intermediate nodes.

Mr. Boetje—A word about the numbers in the county boundary DIME file. The DIME file which we obtained from the Bureau of the Census has approximately 50,000 records, each record representing a county boundary segment. The county boundary segments, as they existed on the file, did not represent closed polygons, i.e., contiguous, sequenced boundary segments which form a complete county boundary.

After compression, we had approximately 8,900 primary nodes, and the entire file can be resident on a small amount of direct access space.

Mr. Weaver—You made mention that the file management techniques are enhanced. I wonder if you would take a moment or two just to give me an idea. The one concern that I have is that I see this thing applicable to the larger areas, but I am really concerned how it would be applicable down to the street level where you are trying to associate individual data at block face.

Mr. DeVorkin—I hadn't really intended to speak much about that here because I am not sure that the machine aspect of things is of interest to the group. Very briefly, since entire county boundaries are represented in a single record, it is possible to set up pointer systems entirely in core, drop internal boundaries for State or regional data aggregation purposes, do this entirely in core and seek only the external boundary records. It can be a very efficient approach.

Back to your question about the block face level. As long as one deals with a single square block or rectangular block, I think you are quite correct. I think we simply have a block record and that's the end of it. In terms of network of points, we are hoping to extend the approach from block level with a certain amount of automatic intersection capability on the display scope display screen; that is, put a "from" node up here, a "to" node up there, something up here and something down there, and have automatic intersection. Where those intersections are not true, simply with a light pen take them out. We are not there yet. That is the next stage.

Mr. Hearle—Could you say a word about your schedule to produce these products, and what availability you will expect to have?

Mr. DeVorkin—It depends on whether it is on the schedule to our sponsor. I feel, and they may say, "No," which changes the whole thing around, I feel that this is a fairly high priority kind of work and subject to someone supplying us with an SMSA file. I would think that over the next 4 or 5 months we will begin to have a pretty good look at it. We may have anything other than a laboratory in-house plug-together system; that is, probably nothing within 6 months directly useable to our outside users. I think, somehow, we would know with this experience where we are.

Mr. Hearle—Did you mention which computer this uses?

Mr. Boetje—We have developed the compression software on an IBM 360/75, which is a fair-sized machine, at the Charles Stark Draper Laboratory. Much of the existing and developed software can run on a smaller machine. For example, SYMAP and CALFORM can run on any IBM 360 with 200K available core. In addition, some of the data management techniques which we are considering can run on an IBM 360/25 or equivalent machine.

If one is interested in going on-line, interactive with a scope or CRT, fairly sophisticated hardware and software would be required. Our work is being conducted on an IBM 360/75, as I previously mentioned, with an IBM 2250 CRT. One could tailor the software to work with other CRT's fairly easily. It is important to note that once the files have been developed with such a sophisticated system, they could be used readily on much smaller hardware.

Mr. Hearle—You would not object of the appropriate comments are made to the sponsoring agency about encouraging you to continue along the path of designing this for computers that somebody has?

Mr. DeVorkin—No.

Mr. Etienne—Have you looked into any of the Tectronics display scopes which are considerably less expensive than the IBM variety? Nonrefreshable screens, smaller programs?

Mr. Boetje—The answer is no, because we have just started the development work and it is still in a preliminary stage. I do not know any reason why the programs could not be adapted for use with Tectronics equipment. We have been using one on-line, interactive display system which is tailored to the IBM 2250. It would require a major effort to convert this system for use with other equipment. The other software could be readily converted and used with other equipment.

Mr. Pisarski—I think maybe there is some confusion about what it is we are up to. One of the elements of it is simply that the Department of Transportation has data files of massive scale—railway statistics running to a quarter of a million individual records; the truck survey is going to go several hundred thousand records, many of them at the place and county levels and higher levels; and of course, census data, etc. So our goal in this whole process is number one, to steal as much of the large software development work that has been done by the military that we can not even come close to touching financially, and two, to grab that stuff and see how applicable it is to our kinds of programs—and so we are utilitizing a large portion of multi-million dollar software. In addition to that, our concern is in the display area and that is primarily at the county level. County file one, shown in this slide is available now and the DIME-like system to display information at the county and aggregate levels of county, dropping out the internal county boundaries, is now available. The base files are the crucial tools that we need to display our statistics. I guess that at this stage we are presenting what we are doing for its potential utility, conceptually and logically, to local applications. I think we are well aware that the scale at which we are approaching it is quite different.

Mr. Weaver—At Charlotte we have looked at the Tectronics 4014 display device, and I think you will find that there is a fundamental design philosophy difference with the storage tube approach as opposed to the vector generation approach. If you are heavily committed to the 2250 the transferability to the local level, from the cost effective standpoint and just system changes, will be significant. If you have not looked at the 4014 in detail take a good look at it. It's a very cost-effective device.

Mr. Etienne—There is fully interactive graphic system designed by M and S Computing out of Nashville, Tenn. It runs off PDP form and uses the Tectronics, I think it is a 4014, which is a 19-inch screen. You can purchase the whole thing for about 50K, computer software, terminal and all.

Mr. Davis—I would like to further address the comments by Mike Weaver and Don Cooke with respect to the applicability of the DIME compression to urban area geographic base files (GBF). Basically, you are correct in stating that there probably would not be very many cosmetic nodes in a street segment-oriented GBF. Consequently, the primary node orientation of DIME file compression would not be very beneficial to urban area GBF. This, however, is only one aspect of DIME file compression.

As brought out in several presentations, a major effort is required to create, edit, and update GBF. Let us consider, for an instant, an urban area which has street segments primarily laid out in a grid network. It would be very feasible, with interactive graphics and compression technology, to work only with the ends of the streets which fall on the urban area boundary. Once all end points had been identified, a computer program could calculate the intersection coordinates for the streets within the urban area boundary. The resulting grid network could then be displayed for purposes of editing out nonexistent street segments, editing in skewed street segments, and editing in address ranges, block numbers, census information, etc. The GBF creation, edit, and update could be accomplished quickly and efficiently.

Another aspect, which could be made available through DIME file or GBF compression, is a real aggregation of data associated with the DIME file or GBF. In this sense the compressed file would serve as a computerized index between the user and the data file. For example, one might ask for a display of the distribution of the journey-to-work origin and destinations by enumeration district (ED). Such aggregation capability could also help in school district planning, health and planning, etc.

Mr. Soller—Much of the work we have been discussing is, the kind of technology that is used to support the New York City geographic base file. We do not include any cosmetic nodes in our basic data file, but we do have a coordinate description of all line-at-map features. We also have a coded segment description of all streets and boundaries so that any intersection can be calculated.

Mr. Cooke—One technical postscript to that: Calculating intersections with the method that you described is basically the guts of Bob Dial's thesis on the SAC system which is 10 years old now. I think you will find it also in the written Arithmicon system of Mr. Corbett. Mr. Corbett uses that type of idea to insert missing coordinates in DIME files and to check consistency.

Mr. DeVorkin—I don't think we can represent this to be entirely new thinking. Some of it has just come out in informal discussions with Alan Pisarski about a highway system. I can not even remember the name of it. That system was much the same as this, but didn't get tied down.

Let's Get Off The DIME

THOMAS O'BRIEN

One of the great-great grandfathers of the DIME file is with us today. There may even be more than one progenitor present, but I would like to recognize Mr. Donald Cooke. Mr. Cooke and his partner, Mr. Jack Sweeny, suckered me out of some money when I was with the city of Boston to install a GBF/DIME file which they had prepared on speculation after they left the Census Use Study in New Haven.

Mr. Cooke initiated me into URISA and USAC. The city of Boston responded to the famous USAC request for proposals as did almost every other city in the country. I had a chance to follow USAC when I went to work for Secretary Romney at HUD and as a consultant to Westinghouse on their USAC project in Dayton. I have seen the limited progress that has been made in data base development and data analysis, and I would like to talk briefly about that progress today. I would like to share with you some of my thoughts on data development and on the need for structures for analysis other than the DIME file. That tool, a marvelous structure, is just now getting to be used.

I note with some chagrin that the city of Boston which has had a magnificent DIME in its bag of tricks for the last 4 years had never used it, not once. It also has a very sophisticated group of analysts that could use it, and that gives me pause.

I suspect that some of you have been in contact with the National Center for Higher Education Management Systems at the Western Interstate Commission for Higher Education, known affectionately as NCHEMS at WICHE. I think if those who first thought of the DIME file had thought of a catchy title like that we might never have seen any progress. NCHEMS at WICHE is a fascinating group trying to develop higher education management systems that would be transportable from university to university. In many ways it is like a USAC trying to develop urban information systems and urban management systems. Although, if you have been close to USAC you realize that the management side was pretty much dismissed at an early stage and that neglect is coming back to haunt USAC. This national center, because of the leadership of the Office of Education in HEW, has coerced all the major colleges and universities in the country to participate in an elaborate committee structure and highly participatory development led by a strong central staff. This consortium is developing higher education management systems in spite of the very diverse interests of the participants, achieving cooperation between the Office of Education, State boards for higher education, private universities, public universities, State colleges, and community colleges, all very strange bedfellows. These diverse interests are working together despite their mutual distrust. If there is one group that could never get along, it would be that bunch. State and local governments, certainly, have many diverse interests, but no where near the kind of paranoia about sharing information or developing program budgets that exists in higher education.

There is a great deal of interest among legislatures and in the executive branch among city counselors and city managers in developing program budgets and having more accountability. The thrust for accountability in the public sector overall, I think, is much greater than at universities. Yet, somehow this NCHEMS at WICHE has succeeded in co-opting everyone and proceeding down a path which many of the participants are very scared of, but they say, "If I don't participate, it is going to be worse than if I do. Why can't we get ourselves together?"

We are lucky in Massachusetts. We have got a great leader in Mr. Alan Altshuler, Secretary of Transportation. Mr. Altshuler, and Mr. Tom Humphrey who has spoken to you already, have gone a long way to share the wealth that exists in their transportation budget—not to win friends and influence people in the sense of buying them off, but in recognizing that transportation planning was really too centralized and was not sensitive to the needs of the local areas. Transportation planning has determined much of the future of the State, determining many things outside of transportation that transportation planners did not have the mandate to do. So, under Mr. Altshuler's leadership and Mr. Humphrey's direction the transportation pot has been invaded for the benefit of the State, but certainly in a nonbureaucratic way.

They are putting money into the cities and towns, through the regional planning agencies mainly, and into the regional planning agencies because they know that unless they have strong leadership and strong analytic capability at that level, they will not be able to develop the transportation plans that the State really needs. They have also been participating with my office, and with others who have responsibility in State planning, to try to help develop analytic capacity.

But today I would like to talk with you briefly about what we need at a broader level, since we do have representation from the Department of Transportation and the Bureau of the Census. I would like to talk about one of the initial thrusts that came out of the Census Use Study in New Haven, which secret sources report was largely resented

by the Bureau of the Census; that, in fact, the cities and towns in this country could obviate the need for a census if they were allowed to use data they already record in an intelligent fashion. If the research and development effort were put in to make that data available to people who need it for planning and program design, they would really do away with the need for that great event every 10 years and have a much more alive, a much more responsive way of analyzing their problems and in thinking up new approaches and new solutions.

Naturally, when I had my latest incarnation as a State planning official, I began to look around and lo and behold, I saw the most marvelous economic data base in the country, totally untapped. That is the State's Division of Employment Security Records, which includes a terrific amount of information about employed and unemployed people. I saw information in our Tax Department which could be sensitively treated with no threat to any individual's privacy to give us current information on the income distribution of the population. And I looked over to the Department of Education, which receives reports on the enrollment of students, and based on my experience directing a program in Boston, I knew that between census years, from 1960 to 1970 or 1950 to 1960, that the proper manipulation of school enrollment data could provide very accurate estimates of population, because the ratio of whatever it is to school-age children changes very slowly and very predictably. I thought about two other data bases. My work with IMIS in Dayton and the work I had been doing at HUD and the city of Boston convinced me that we did not know anything about the intergovernmental flow of funds. We still have not addressed the issue of how State government, local government, and Federal Government relate in terms of the flow of funds and resource allocation. What kind of structure would you need to translate information about Federal money and State money into meaningful information at whatever level decision makers find necessary?

The fourth data base we need is one which should be readily available from property assessment. In the city of Boston we have a unique assessing practice; we don't. Boston has not been reassessed for at least 30 years, and there are those who think much longer. The disparity in the sales and the sales assessment ratio throughout the city is notorious. It has been written up in academic journals, in popular magazines, in court journals, and nothing happens. I examined the potential of the assessing data base in Boston and said, "Wow!" With the GBF/DIME file and accurate assessment information we would know where all the new construction is, the number of dwelling units by section of the city, the location of all industrial property, etc. We would know the greatest things in the world—what a planning tool! So we spent hundreds of thousands of dollars on mechanizing the records of the Assessing Department and during that process, in order to get the job done, we sold out completely and guaranteed that nobody would ever look at it. We did the technical job and lost the data base as a research and policy tool.

The need for land-use planning information and other information is a natural tie-in with assessment records. An assessor needs to know the number of dwelling units, the size of a building, the type of construction, the location of the land, the number of square feet in a particular parcel, and he needs to know who the owner is. Lawyers need such data for property transfers, and geographers are frantic for a way of getting good assessment records plugged into remapping the State or particular subdivisions of the State. Those four data bases, economic, demographic, fiscal, and physical land use are fundamental data bases which don't exist, or exist in a form which is now inaccessible.

The State has a responsibility to develop these data bases. I would like to see a Census Use Study at the State level. I would like to see an attempt by the States' Regional Planning Agencies and State and local government join together in establishing a national center for State management systems. I would like to see an attempt to go beyond the structural developments which have been so successfully demonstrated with the GBF/DIME file, to talk about the other structures that we need.

Mr. John Keller, Program Budget Consultant to the State of Hawaii, recently said, and I share his views, that if he went into any State and had $5 million, the place he would spend it is not on staff analysis, but on information development. It was his experience that inquisitive analysts experience monumental frustrations in trying to obtain relevant information and are soon driven out of the system. The State must become a planning services delivery agency. We must deliver the tools to the Regional Planning Agencies. We have the potential data bases and we have the power to require that those data bases be maintained. We must service needs other than transportation. Because we don't have the bricks and mortar that follow behind transportation planning, we are unlikely to get the kind of State money we need to do data base development. I recommend that you consider the leadership that has been shown by the Department of Transportation. They have dipped into the till, not just demonstrating their good faith by coming to our meetings, but sharing their resources because they want these problems to be addressed collegially. We need to work together to develop these resources, and I hope over the next few years that the Federal Government will point to the DIME file and the work of the Census Bureau to show the progress that has been made, the return that we are getting for the investment. Meanwhile, we must work to make the other investments which are necessary for better fiscal, economic, and social planning.

Question Period

Mr. Voye—You apparently feel pretty strongly about the DIME file for Boston. What is your opinion of the concept of extending the DIME file outward from Boston to the entire State?

Mr. O'Brien—Transportation planners are enthusiastic about a statewide DIME file. Its usefulness would be limited, but it is unlikely that anyone will agree on an alternative expenditure of funds. From my discussion, you can see I would spend the money differently.

Mr. Voye—What do you see as an alternative?

Mr. O'Brien—The development of a data base to produce intercensal year estimates or the development of an economic data base. We also need better structures for manipulating data, like the Consistent System. The Defense Department has invested a great deal at MIT on the Cambridge project's attempt to reformat data quickly and allow for highly flexible retrieval. The result, the Consistent System, provides a structure which could be very useful in the development and manipulation of the planning data bases we need.

The GBF/DIME is a structure which is necessary, but it is time to develop other structures appropriate for an economic or demographic data base. Just as the work of Mr. Simon Kuznets in developing basic economic data was essential to converting the theories of Keynes into practical economic policies, we need small-area economic and demographic data to understand what is going on and to test hypotheses before accepting them.

For example, Boston was operating on the totally erroneous assumption that population decline in the city had continued at the same rate during the 1960's as in the 1950's. The perception that the decline was continuing affected everyone's thinking. The Mayor was a pessimist. Good economic data turned that perception around and turned the Mayor around. Boston's ambitious capital outlay program is the direct result of good analysis and good planning. We need better data, we need better structures to manipulate that data, and we need better analysis.

Mr. Kondo—You indicated that the city of Boston has not utilized their GBF/DIME file since it was developed during your tenure with the city. You also suggested the use instead of many other existing data bases and systems—health statistics, employment security files, etc. I think they are a complement to the existing data bases. The GBF can also be a considerable aid in combining these files. However, there are still considerable problems remaining with differences in definition, unit coverage, date data was collected, etc.

Mr. O'Brien—I agree with you. Interactive data processing is the tool we are going to use to make the technology accessable. Data processing has been oversold, but the day must come when you will be able to sit at a terminal and interrogate the Division of Employment Security files for the detailed economic information you need. We should have the capability to interrogate a variety of data bases, developed, maintained, and properly safeguarded for confidentiality by the agency responsible for data collection. They can make sure that the information cannot be disaggregated to jeopardize individual rights. However, the system must be interactive, as expensive as that is going to be and as long as it is going to be before analytical skills are at the operational level.

Mr. Silver—Is there anything specific happening right now which you can mention?

Mr. O'Brien—Yes, the work now going on with the Division of Employment Security files and the Consistent System. We have committed State funds to develop a very strong analytic staff and are working jointly with MIT on the software structures which allow the staff to manipulate the data.

Mr. Kondo—What kind of geographic referencing are you using with the employment security data?

Mr. O'Brien—Mainly city and town, but ZIP code is possible, too. They can give you employment by SIC, by ZIP code, anyway you want it.

Mr. Cooke—I must say that in working with the city of Boston and specifically with Messrs. Tom O'Brien and Bob Hanson, I think that is where I learned pretty much all I know about using the GBF/DIME files and what it takes in order to put a GBF/DIME file to work.

It came down to identifying about 6 or 7 items which you have got to have in order to put GBF/DIME to work. For one thing you have a GBF/DIME file. We knew this in 1968, when we started working with you. We came up with that—a DIME file with all the coordinates and all blocks bounding, pretty good addresses from a very good set of source maps—from the Boston City Redevelopment Authority. You need census data, that is the second element. We got that eventually.

Third, You need local data. And, as you pointed out, there were some terrible problems just developing the tax-assessor file. Now we do have some data bases we can work with, and we should work with—that is your key point here.

So you get the geographic base file, local data, and census data. Now you need a computer to run on and that was a terrible stumbling block in Boston. Boston at that time has an IBM system 360 model 20, I believe. You could not run anything on it. I believe only now that they are finally up to the point that they could actually run some advanced software.

We also worked with Boston Model Cities. They had access to a very good computer at the Urban Systems Lab at MIT, and we did develop some software for them. The software was at the point where we were generating incidence maps of data in the model cities area on cathode ray tube plotters. This was being run by very low-level technicians in the model cities agency—not programers at all.

So we had a little bit of success there. It is unfortunate that because of the machine problems in the city of Boston, we never transferred that software to the city officially.

Another thing you need is people. You need two kinds of people: one, technicians to actually do the work and run the programs. In Boston this was basically Mr. Bob Hanson. That was it, and Mr. Hanson is not with the Redevelopment Authority anymore. That is a very, very key problem. The other kinds of people, besides technicians, are understanding, forward looking, and patient management—which we certainly had while Mr. O'Brien was in charge of the Boston Redevelopment Authority.

In summary, you need a GBF/DIME, census data, local data, a good computer with good software to run it, creative technicians, understanding managers and, of course, money to keep the whole thing running. We were missing several key elements in our experience back then.

Geographic Reference Files in New Hampshire

JAMES M. LANGLEY, JR.

INTRODUCTION

The Planning and Economics Division of the New Hampshire Department of Public Works and Highways maintains data files which record information on or pertaining to the highway system in New Hampshire. Some of these files contain data on the physical characteristics and status of the highway system itself while others store information such as traffic recorder counts which indicates how traffic flows through the highway network. Another type of file records various data, for example, traffic accident details, which may be related to specific points along the highway system.

Traditionally, files of this type have been maintained on a disjunctive basis, that is, a separate file with its own update mechanism and reporting features was initiated when the need for each file arose. As file size has grown and the number of data files has increased this approach has become quite unwieldy in terms of the effort expended per file and the difficulty encountered in attempts to correlate data from different files.

About 5 years ago, a comparison of the Division's files revealed that although their structure and data content varied widely most contained geographic descriptors of at least one and usually several types. These descriptors included point names and coordinates, various linear feature identifiers such as road and route numbers, and a great variety of area descriptors ranging from traffic zone to county codes. The consistent use of geographic classification codes in these files suggested that their reorganization around a common geographic reference framework would facilitate interfile data correlation and reduce the redundancy of stored information.

In the course of this analysis the Division learned of other organizations working along similar lines. The advantages of filing systems based on a central geographic framework had been perceived and were being applied in many ways by various Federal and State agencies as well as their counterparts in the larger urban areas. While generically similar, specific applications differed depending on the size of the area of interest, the degree of abstraction of the geographic reference framework, and the nature and relationship of subsidiary data files to this framework. These efforts were proceeding along two main paths or directions—the elementary area or grid approach to a geographic reference and the network or linear graph method of establishing this base. Of these, the latter appeared preferable to us for obvious reasons.

Among the geographic reference systems that came to our attention during this evaluation were the geographic base files of the Census Bureau. Although these files are intended for use in small area data analysis they are network oriented and organized following topological principles that are independent of the actual size or nature of the area involved. The recognition by the Census Bureau that a GBF itself is primarily a reference file which is best maintained separately from any associated data file also appeared to be significant. Accordingly, the Division began to participate in the Census Bureau GBF program to gain experience in the construction and application of files of this type.

URBAN TRANSPORTATION PLANNING AND GEOGRAPHIC BASE FILES IN NEW HAMPSHIRE

A brief review of the geography of New Hampshire will place the relationship and application of small area geographic files in proper perspective with similar files for other areas of the State. At present there are two standard metropolitan statistical areas, Manchester and Nashua, that lie entirely within New Hampshire. Manchester, with 1970 SMSA population of 132,512 and Nashua with a 1970 population of 86,280 are both located in the southeastern portion of the State and are contiguous with the Lawrence-Haverhill and Lowell SMSA's whose central cities are in Massachusetts. The total New Hampshire SMSA extent, which forms part of the northern fringe of the Boston-Washington megalopolis, contains towns (MCD's) which have more than doubled in population between the 1960 and 1970 censuses, and contrasts markedly with the remainder of New Hampshire which is essentially rural in nature. In 1970 about 35 percent of the State's population resided in towns which are now a part of SMSA's although these towns form only six percent of the State's total land area.

In New Hampshire one of the current and continuing Highway Department applications of Census Bureau GBF's is in connection with the urban transportation planning process. Metropolitan Manchester, containing the largest city in the State with a 1972 estimated population of 94,000, was the only urban area in the State that met the comprehensive urban planning requirements of the 1962 Federal-Aid Highway Act. In 1964, the Metropolitan Manchester Planning Study (MMPS) was authorized by participating local, State and Federal agencies in response to this requirement. The Southern New Hampshire Regional Planning Commission (which actually covers the Manchester

region) evolved in part as a result of the MMPS and currently is responsible for the Manchester GBF which was coded by local planning agencies in 1969. At present, the New Hampshire Department of Public Works and Highways acts as coordinating agency for continuing phase of the MMPS and provides data-processing services for GBF maintenance and application.

The Manchester GBF is now in the correction phase of the Census Bureau CUE program. In 1969 the coding limit line was established as a combination urban-limit/useable-address boundary that encompasses Manchester but is generally well within the limits of the SMSA as now defined. The GBF has been used as a coordinate source for MMPS base and future year networks in the area that is common to both files. Computer-produced plots of traffic assignments may thus be overlayed on a plot of the GBF itself for study purposes.

Traffic zone definitions used in the MMPS predated the construction of the Manchester GBF and several arbitrary boundaries used for zone definition were not subsequently included in GBF coding. This fact coupled with the limited extent of the GBF in comparison to the total MMPS area has so far precluded the use of programs such as ADMATCH for the aggregation of socioeconomic data in the continuing phase of the transportation study although this could certainly be accomplished, in the central city at least, with minimal effort. Although standard FHWA procedures permit socioeconomic forecasting from base year data, a check on these forecasts in the form of actual data aggregation for selected traffic zones would appear to be of value.

The 1970 census indicated that Nashua, N.H. had passed the 50,000 population mark, and this city and its environs became the subject of an urban transportation study and was defined as the second SMSA entirely in New Hampshire. The Nashua GBF was coded in the summer of 1973 by Highway Department personnel with the cooperation of the Nashua Planning, Engineering and Assessor's Departments and town officials in Hudson. In this case, GBF coding preceded initiation of the Nashua Area Transportation Study and the Census Bureau Metropolitan Map Series sheets used in coding the GBF were available for reference when the NATS traffic zone definition was established. The initial phase of the transportation study is now being completed by a consulting firm but the utility of the Nashua GBF in the continuing phase of the study will not become apparent until the consultants' procedures and preliminary results are fully evaluated. The initial version of the Nashua GBF itself, presently at the address-edit stage of construction, will be completed with coordinates by the summer of 1974.

LARGE-AREA GEOGRAPHIC REFERENCE FILES

The primary benefit of Highway Department work to date with Census Bureau GBF's has been the experience gained in the organization, construction, editing, and maintenance of geographic reference files (GRF's). The active support of the Census Bureau which has included the provision of operational computer programs such as FIXDIME, TOPOEDIT, and the forthcoming FIXCORD has been of great value in accelerating this learning process. Two large-area GRF's, one covering the entire State and the other covering Sullivan County, are operational and under construction respectively. Both of these files, which are the direct result of experience with Census Bureau GBF's, are being used to explore the possible construction and application of a statewide GRF.

The first of these files, the Town Boundary File, was developed from single-line tracings of an 8 foot (2 mile per inch) three-section map of New Hampshire prepared some years ago from a mosaic of U.S. Geological Survey 15 minute topographic quadrangles. Construction of this file was analogous to that of a standard Census Bureau GBF except for the replacement of MCD code by county code and of block code by MCD code. Obviously, street addresses were not involved. Coordinates for this file were measured by overlaying graph paper on the noded single-line maps and added to the file using a specially written computer program. Except as previously noted, the file is in standard Census Bureau GBF format with unused fields blanked and at present includes map mile coordinates only. The file, which is primarily intended for the display of town-aggregated statistics, contains 3,656 records and is also used for experiments in GBF/cell-file equivalencies.

In New Hampshire State primary and secondary highway systems are posted with accident reference markers that are used in reporting the location of a motor vehicle accident. Other roads in the State not on these systems (local city streets, town roads, etc.) are not so posted and the reporting of accurate accident location data for roads of this type has proved troublesome. In an attempt to improve this situation a second large-area GRF is being constructed for evaluation as a position reference file (PRF) for Sullivan County, an essentially rural area of the State.

The Sullivan County PRF is based on a set of 15 single-line maps prepared by tracing 4-inch per mile enlargements of the New Hampshire General Highway Series maps covering this area. In addition, two urban area insets are involved. All single-line maps are block numbered and noded in a manner analogous to the preparation of MMS sheets for GBF coding. A special class of nodes, represented by a unique symbol, denotes the position of accident reference markers. The Highway Department road inventory identification code is used for basic segment identification although this is supplemented by street name in urban areas.

For field use in reporting accident location a second set of Sullivan County maps combining features from the General Highway Series and single-line maps is in preparation. These maps consist of duplicates of the original General Highway Series maps reproduced as a screened background on which all node symbols and numbers representing landmarks that may be identified in the field appear for contrast in a darker tone. At present the PRF is being coded and a preliminary evaluation of the map set by Highway Department field personnel has produced favorable response.

As a result of this work, it is now possible to estimate with reasonable precision the size of a Census Bureau type GRF that would cover the entire State of New Hampshire. The estimate is based on the use of quarter-mile spaced section-delineating or curvature nodes along featureless stretches of rural highway. In addition, relative county land area, population, and linear feature density plus the current sizes of the existing SMSA GBF's have been considered. The estimated size of a GRF or data index of this type—150,000 records—indicates that a complete reappraisal of current file management responsibilities and techniques must precede the actual construction and use of a statewide GRF.

SOME PROBLEMS CONCERNING THE DEVELOPMENT AND APPLICATION OF GEOGRAPHIC REFERENCE FILES

Adequate base maps are necessary for the construction of a GRF. For New Hampshire USGS quadrangles (generally 15 minute topographic plus some 7.5 minute topographic and as yet no 7.5 minute orthophoto), town maps prepared from uncontrolled natural resource photogrammetry by the State Department of Resources and Economic Development, plus the General Highway Series maps are currently available. None of these sources is by itself completely satisfactory as a base for the preparation of single-line maps for GRF coding. Both USGS coverage and the DRED maps are apt to be out of date with respect to the current status of the highway network in certain regions. General Highway Series maps reflect yearly changes in the highway system but these maps have not been carefully controlled so that the positional accuracy of features shown is a present rather low.

In 1971, the New Hampshire Legislature enacted RSA 31:95-a which requires that a properly indexed and up-to-date tax map shall be available in every city and town of the State by 1980. Although funding was not provided by the Legislature, many towns in which tax mapping is now inadequate have started to appropriate funds or build escrow accounts for this purpose. Regional Planning Commissions will in some cases coordinate the purchase of controlled photogrammetry which will serve as a base for the preparation of these tax or parcel maps. These photos plus the tax maps and files that are already in use in the cities will provide an up-to-date base for GRF development that is superior in accuracy and resolution to many of the present sources of this information. In addition, the prospect of the future widespread availability of parcel maps creates the possibility of the downward extension of GRF's to the subblock level or, alternatively, of the compilation of subsidiary data files containing this information that are keyed or indexed to a GRF.

Street addressing in the fringe areas of New Hampshire SMSA's remains a problem. It appears likely that the extension of current SMSA Census Bureau type GBF's to the limits of these areas will take place before satisfactory addressing schemes exist in some localities. While this will limit the use of these files for the aggregation of address-related data, the Regional Planning Commissions foresee applications of the expanded files such as the maintenance of land-use data that do not involve addresses. In time the fringe areas will become more heavily settled and regular addressing schemes will evolve. As this happens address data will be added to the file records.

A GRF may be regarded as an index to other data files containing information on the points, lines, and areas which it describes. The size of a wide-area GRF itself plus the variety and amount of data that may be related or keyed to this index calls for careful evaluation and application of file management techniques. Although the situation may be alleviated in some cases by constructing parallel files of similar structure for a series of disjoint geographic regions, this approach avoids the issue of consolidation which may be an objective of file construction. File management or data base systems such as IMS or TOTAL, on the other hand, impose considerable overhead in terms of initial and/or continuing cost, manpower and procedural mechanics. At present, the question of file centralization and, therefore, the utility of a data base system are unresolved issues in New Hampshire.

A basic problem with GRF's of the type under discussion is their relative inflexibility with respect to zonal definitions specified by boundaries that do not appear in the file. This is of concern to most planners since a major aspect of their work involves the interrelation of data referred to different zonal subdivisions of the same area. The merging of two zonal definitions of an area as a single cathode ray tube display is common practice. The production of a new GRF (i.e., data index) restructured to represent the overlay is apparently less common as is the extension of this process to include a disaggregation capability for subsidiary (i.e., index keyed) data files. It is apparent, however, that operations of this type might best be approached on a level which permits the concise specification of file definitions and manipulation procedures. Once again the use of an information management system seems to be indicated.

CONCLUSION

The next few years will bring a quantum increase in both the amount and precision of geographic and geographic-related data that is in a form suitable for computer assimilation and analysis. One of the greatest challenges currently facing planning authorities in New Hampshire is the definition and establishment of responsibilities, priorities, and procedures that will permit the effective selection and use of this type of information in abstract or in detail. Questions of interagency cooperation at several levels of government must be resolved so that duplication of effort will be avoided and the exchange of complementary data will be facilitated.

The combined efforts of Federal, State and regional and local government agencies in the creation and application of the Census Bureau GBF's have clearly demonstrated that a unified technical approach to the management of geographic and demographic data is effective in urban areas. In New Hampshire the Highway Department looks forward to the development of similar techniques for other geographic regions and additional classes and stratifications of data. Participation in the Census Bureau GBF program will continue to be an invaluable guide to these efforts.

Question Period

Ms. Liddle—Are you familiar with the statewide GBF/DIME being prepared by the Rhode Island Highway Department which will be used in their accident location studies? Will your use of the GBF/DIME for accident location and study purposes in rural areas differ with anything that they are proposing to do?

Mr. Langley—Differences between GBF development and the use of GBF type files for accident reporting and recording in the cases of Rhode Island and New Hampshire arise from the disparate sizes of the two States and the fact that a much greater portion of Rhode Island is urbanized—about 50 percent versus perhaps 10 percent for New Hampshire. Because of this, street names are relatively well established and used less ambiguously in Rhode Island than in New Hampshire and will, I believe, be used directly in reporting accident location in Rhode Island. In New Hampshire, noded maps will be provided for use in reporting rural accident location in order to avoid variations in street or road name usage.

Mr. Post—How will information be reported by persons filling out the accident report in the field in order to get the appropriate information to relate back to your geographic reference file?

Mr. Langley—The position reference file is coded from noded, single-line maps developed by tracing enlargements of general highway series maps. A set of maps intended specifically for use in reporting the location of accidents is constructed by preparing node symbol and number overlays which are then overprinted on copies of the original maps. An accident in a given town is reported as occurring at a particular node—which will usually represent a highway/highway or highway/feature intersection—or at a stated distance between two nodes.

Mr. Koch—Two questions. First, you said that legislation in 1971 now "requires that a properly indexed and up-to-date tax map shall be available in every city and town of the State by 1980." Will these maps be compatible and consistent from one area to the next so that they can be used without a major number of adjustment factors? Second, what is the estimated cost of creating a GBF for the entire State of New Hampshire?

Mr. Langley—The existing tax map legislation establishes no coordinating agency or standard for tax mapping and therefore statewide tax map compatibility is not assured at the present. It appears, however, that the more active Regional Planning Commissions, which are located in the southern, most heavily populated portion of the State, will attempt to secure controlled photogrammetry at a consistent scale to serve as a base for tax mapping and will to a certain extent coordinate parcel mapping activities in these regions.

Based on progress to date on the Sullivan County PRF/GBF plus a careful estimate of the cost of constructing a similar file for Merrimack County, it appears that the files for these counties will cost about $40,000 or approximately $1.15 per record. This figure does not include addressing but does include correlation of highway station in feet with node number in the case of segments representing roads. As experience in coding and editing files of this type accumulates, the cost per record should decrease somewhat so that an estimate of $150,000 as the cost of a statewide PRF appears reasonable.

Mr. Halterman—You spoke of traffic zone definitions used in the Manchester transportation study which included several arbitrary boundaries used for zone definition, but which were not subsequently included in GBF coding. Are you adjusting these transportation zones to overcome this problem?

Mr. Langley—The Manchester transportation study is currently producing results which are increasingly difficult to reconcile with observed data. This is apparently due to the use of trip tables that are based on O/D data obtained about 10 years ago and defined by traffic zones which on the fringes of the study area are larger than they should be considering the current state of development in these areas. Starting in 1976 the study will undergo extensive modification which will include new O/D surveys, the subdivision of peripheral traffic zones, and the correlation of traffic zone definition with the Manchester GBF by adding, when necessary, traffic zone boundary defining segments to a copy of the then current GBF.

Computer Mapping With Micrographics

RICHARD H. SCHWEITZER, JR.

Computer mapping, or to be more exact, the creation of maps through the use of automated data processing on computer-driven output devices, is not new. Apparently the first computer maps were produced by a group of Swedish meteorologists in 1952. They were contour maps produced on the line printers. Since that time the development of newer output devices have resulted in computer maps with much better quality. However, the end products were still recognizable as what they were—facsimiles of maps produced which happened to have been produced by a computer. A major breakthrough would have to be achieved before a computer map and a cartographer's map could be judged by the same standards of quality. The most critical limitation was the hardware used to produce the maps.

The earliest computer maps were produced on a typewriter driven by the computer. Even today I will venture to say that most computer maps are still produced on a high-speed line printer where various symbols or combinations of symbols are used to differentiate the different data classes or zones. The software packages that are most generally available, SYMAP, CHORO, GRIDS, CMAP, and LINMAP, all use the printer as their output device. These systems suffer from the low resolution and coarseness of the resulting maps.

A more precise means of producing a map is through the use of a computer-driven drum or flatbed plotter. These plotters have a resolution up to 1,000 lines per inch and a speed up to 40 inches per seconds. The resulting line maps are very useful for depicting networks or street patterns. This technique is not used as often as the line printer to produce thematic maps.

As cathode ray tubes become more common, many of the mapping software packages will be adapted to output their displays on the various types of CRT's refreshed, storage, or raster scan tubes. This approach is particularly attractive when several optional map forms are being evaluated for a graphics display. Similarly, the CRT offers many benefits for interactive editing of the mapping files or the GBF/DIME files, the latter being essentially line files.

The geospace plotter is not a generally available or widely known computer output device. It is a large (40-by-60 inch) raster scan device that draws its pictures directly on photographic film. Its high rate of plotting speed makes it an attractive device. However, the resulting pictures are continuous-tone images made up of a series of discrete points which lose some of their sharpeness in any reproduction process.

COMPUTER MAPPING AT THE CENSUS BUREAU

The Census Bureau has experimented with all of these approaches to automated cartography during the last 6 years (1). Each technique has its advantages and applications. However, none of these techniques produced high quality, publication-ready maps which could be printed either in color or black and white without the expense of going back to the computer for additional processing or expensive secondary manual enhancement. The Geography Division pioneered a technique using computer micrographics to achieve these goals. Surprisingly, the cost of producing computer maps by this new approach was even lower than producing maps on a line printer.

The heart of this new approach to computer cartography is a computer output on microfilm or COM unit. These devices plot a picture or write alpha-numerics on a high precision CRT at a very fast rate of speed. This image is then photographed on microfilm, usually 35, 70 or 105mm film. A new image can be created and photographed as fast as every 3 seconds. More complex images may take a minute or longer. The COM unit we use has 16,384 addressable points in both axis across a 3-inch CRT. Thus, a 35mm film image can record an image of approximately 16,00 by 12,000 units or rasters. Normally we use only one-fourth this level of resolution to produce page-size maps. The device is capable of uniquely addressing each raster with up to 64 levels of intensity; however, we have chosen, for many reasons discussed later, to have the beam sweep across the screen in a vector mode and uniformly expose the film in designated zones. This capability is really the secret of the quality of the maps since screening to get various shades can be done later and at a lower cost. However, this approach necessitated the development of a computer mapping system designed to optimize the capabilities of the COM unit in respect to the desired end products.

PHILOSOPHY AND SOFTWARE

Every computer mapping technique requires a machine-readable file of descriptive information about each area that is to be mapped. Assuming for the moment that the final map is to be a choropleth map where the individual areas are to be uniformly shaded corresponding to varying data values, a file of the boundary segments describing each area or polygon must exist in a machine-readable format. This data might be in a master file such as a GBF/DIME file or in a polygon boundary file. In any case the interior area and/or the perimeter for each polygon must be described in very

specific terms before the computer can correctly and uniformly assign the desired symbols to them.

Individual areas or polygons can be described in one of two ways. First, a polygon can be defined in terms of the individual boundary line segments required to enclose the polygon. This might be called a perimeter description file. A second technique would describe the polygon in terms of the individual strips required to completely cover or "paint" the surface of the polygon. This can be considered as a surface description file. These strips are defined by the end points where the strips would cross the boundary of the polygon. These two methods of describing a polygon for computer mapping are shown in figure 1. Every computer mapping program depends upon one or both of these techniques to describe the individual areas that are to be shaded in the actual mapping of real data.

The widths of the individual strips required to describe each polygon is really a function of the computer output device. For instance, maps produced by SYMAP or CHORO or CMAP on a high-speed line printer describe the polygons in strips of 1/8 inch in width or 8 strips per inch. Since the COM unit we use is capable of addressing over 16,000 lines per inch on a strip of 35mm film, the width of the individual line is very narrow. This fine line width means that even after a 35mm image was enlarged to a 7- by 9-inch image, there would still be approximately 1,800 lines per inch.

The conversion of a polygon perimeter description file into a set of strips describing each polygon is accomplished by a program that is called SCAN. This can easily be extracted from a GBF/DIME file by use of a simple computer program such as DACS. The description of each polygon is converted from a set of coordinates describing the perimeter of the area into a series of horizontal line segments each of which is described by its starting and ending coordinates.

The resulting coordinate pairs are the beginning and end points for each sweep of the COM unit's electron beam. The beam is turned on at the coordinate location defined as the beginning and is swept across the COM unit's CRT (and through optical techniques, across the film) to the next pair of coordinates where the beam is turned off. As the COM unit reacts to the sets of plotting instructions, each polygon is rapidly painted on the CRT.

This conversion of a polygon boundary file into sets of plotting commands must be able to properly describe donuts, concave shapes, spirals, and also polygons with more than one area (i.e., islands). Fortunately, this is not a very difficult task for a computer. For example, it takes the Bureau's Univac 1108 computer 2.8 minutes to create a SCAN file for the Nation's 3,076 counties with over 180,000 vectors; the 38 census tracts in Des Moines took only 20.2 seconds. Also this operation is done only once for each area to be mapped, not each time a map for the same area is desired. The complete SCAN file is stored on an intermediate file for later use. This is particularly useful when a file like the U.S. county file is to be used many times.

It is, of course, possible for the polygon data to be extracted from a master data file along with each polygon's geographic area codes or entered into a data file from cards.[1]

The last computer operation before the production of a plot tape is the merging of the classed data. I assume that the data values to be mapped with the SCAN file exist in a systematically formatted data file. In this operation the program matches, on a random basis, the geographic area code of each data value with the appropriate geographic area code for the correct set of plotting instructions. The geographic area code for each polygon on the SCAN file is simply replaced by class interval of the data. The non-matches and multiple-matches are printed out on a diagnostic listing. The file is then sorted by the class intervals so that all of the polygons for the first class interval and grouped together and so on for as many classes as have been defined. The appropriate plotting command functions are then written on the plot tape so that each set of polygons that have been grouped by class intervals is plotted as a separate frame of microfilm. The resulting output tape is a properly formatted tape containing the required commands so that the COM unit plots each map one class interval per frame with descriptive titles and class identifications. This operation takes about 1 minute on our 1108 computer for the Nation's 3,076 counties and only a few seconds for the 38 census tracts in Des Moines. This whole process is accomplished in a single computer program run.

Our tapes are plotted on an FR-80 COM unit which is built by Information International Incorporated and operated by the National Oceanic and Atmospheric Administration. Similar COM units are manufactured by several other firms. You will find that there are service bureaus in most large cities which have COM units. Also many major corporations are turning to COM units as a means of outputting their records for easy storage. Almost any vector-type COM can be used to produce quality computer maps.

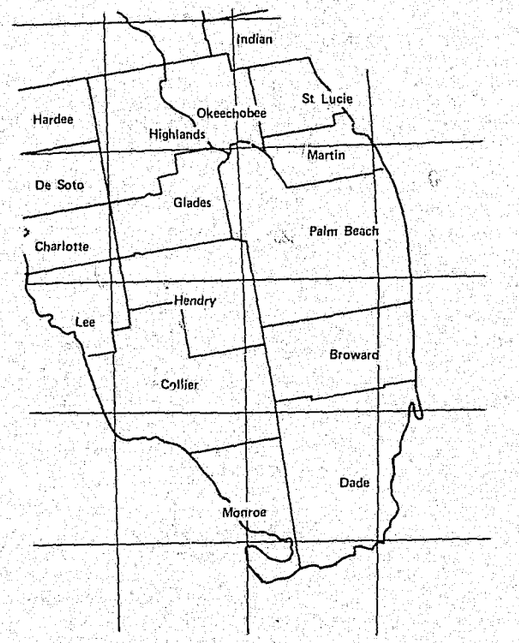
Because of the technique of sweeping across each line so that the entire polygon is covered, the film is completely exposed where the electron beam was instructed to draw and is left unexposed on the remainder of the image. After the tape is plotted, the film is processed as full reversal negatives in an automatic film processing unit. This produces a black negative with clear windows where the selected polygons were located on the film image. The cost of the microfilm is approximately 25 cents per frame. Figure 2 is an enlargement of a single frame of film for one of the classes of an agricultural data map of the United States. Note the sharpness of the contrast at each county boundary and the lack of any "stairstepping" which is common with most inexpensive hard copy line plotters or maps produced on a line printer.

We generally plot the boundaries of the polygons at the last frame of microfilm so that we have a plotted boundary

[1] For a discussion of the problems involved in the use of GBF/DIME files in computer mapping see "Use of Geographic Base Files in Computer Mapping" by Richard H. Schweitzer, Jr., in *Geographic Base File System - Establishing A Continuing Program*, U.S. Bureau of the Census, Report GE 60, No. 4: Washington, D.C. 1973

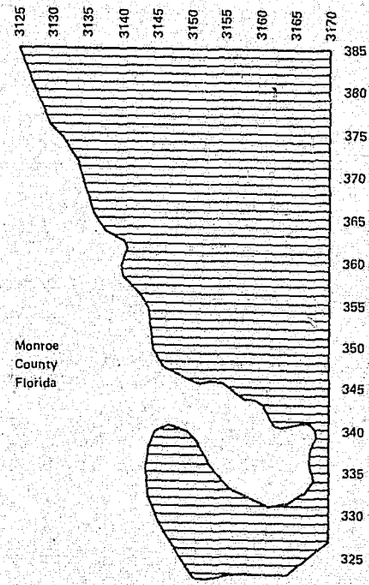
Figure 1 Two Methods of Describing Polygons

Description by Perimeter Coordinates



Longitude	Latitude
80.8975	25.1796
80.9474	25.1466
80.9629	25.1387
80.9961	25.1385
81.0458	25.1400
81.0840	25.1326
81.1087	25.1353
81.1283	25.1604
81.1625	25.2044
81.1790	25.2387
81.1815	25.2697
81.1753	25.3036
81.1671	25.3199
81.1474	25.3275
81.1192	25.3206
81.1005	25.3033
81.0791	25.2741
81.0513	25.2462
81.0216	25.2353
80.9917	25.2331
80.9598	25.2305
80.9353	25.2432
80.9232	25.2573
80.9285	25.2809
80.9264	25.2979
.	.
.	.
.	.
80.8929	25.7967
80.8976	25.1796
80.8975	25.1796

Description by Surface Area Scan Lines



Scan Line Coordinates			
Starting		Ending	
X	Y	X	Y
3158	323	3154	323
3159	324	3153	324
3167	325	3152	325
3168	326	3151	326
3169	327	3150	327
3170	328	3149	328
3171	329	3148	329
3171	330	3147	330
3171	331	3147	331
3171	332	3163	332
3161	332	3147	332
3170	333	3166	333
3157	333	3147	333
3170	334	3166	334
3156	334	3147	334
3170	335	3167	335
3155	335	3147	335
3170	336	3167	336
3154	336	3147	336
3170	337	3166	337
3153	337	3147	337
3170	338	3166	338
3152	338	3148	338
3170	339	3166	339
3169	340	3166	340
3169	341	3166	341
3164	341	3161	341
3169	342	3160	342
3169	343	3155	343
3169	344	3148	344
3169	345	3146	345
3168	346	3145	346
3168	347	3145	347
3168	348	3145	348
3168	349	3144	349
3168	350	3144	350
3168	351	3144	351
3167	352	3142	352
3167	353	3141	353
3167	354	3140	354
3167	355	3141	355

image at the same scale and projection as the individual class interval images. Figure 3 shows the boundaries for each of the Nation's counties as they were drawn on a single frame of 35mm film.

The 35mm negatives are taken to a standard photo lab for the additional processing required to produce the desired maps. All of the following steps utilize only conventional film processing procedures to enlarge the film images and produce screened publication negatives ready for printing. First, each frame of film is enlarged as a negative to the desired size. We use negative to negative film in the enlargement process; however, the photographer may desire to use an intermediate positive film. Since all of the negatives are plotted from the same mapping file and are enlarged at the time, there is no problem with the later registration of the individual images with each other. Once precise registration of all of the negatives has been achieved, they are punched so that pins can be used to insure precise registration during the subsequent composing operations.

It is appropriate at this point to digress for a minute, to explain why we do not have the COM unit screen the film negatives or produce continuous tone negatives. The COM unit we have access to can produce 64 shades of gray and could, in theory, prescreen a selected class interval with a 50-percent (or any other percentage) screen. However, if either of these were done by the COM unit, it would automatically limit our options with the forms that the images may later take. For instance, a 50-percent, 130-line screen could be visually correct for only one size of image. Thus, the final size of the image would be restricted. Likewise, a continuous tone image would have to be screened before it could be used in any printing process. This would reduce the sharpness of the final prints. Lastly, we did not want to restrict in any way the final printing to monochromatic prints. Overall, the production of window negatives allows the images to be enlarged to suit the form of presentation, or screened to the desires of the user for color or black and white. Preliminary decisions can be changed at the photo lab without ever going back for another computer run if the specifications for the final maps have been altered.

Even though the photo processing is a manual process, it is still one that is, comparatively speaking, inexpensive. The tasks outlined above are standard procedures that are identical to those that have been used for years to photographically prepare traditional manual cartographic efforts for publication. This approach to automated cartography has simply automated the tedious, time-consuming activities required to produce window negatives and left the other time-tested techniques unchanged. This blending of the old and the new has resulted in fantastic cost reductions. A page-size computer map of the Nation produced by micrographics costs about $150 or less depending on the number of maps produced in a single computer run. The same map produced by traditional cartographic techniques would cost at least $1,200. From start to finish the computer map can be ready for printing in days; the manual map may take weeks.

MAPPING OTHER DATA

Computer output on microfilm units also can be used to quickly and inexpensively produce plots of nearly any other type of readable data. For instance, figure 4 shows a micrographics plot of the census tract outline file for the Seattle-Everett, Wash. SMSA. In the same manner, a GBF/DIME file can be plotted in its entirety or by individual map sheets. These latter plots can be mounted in cardboard slide frames and enlarged with the aid of an ordinary slide projector for review during the correction of the source files.

A COM unit which operates in a vector mode can also produce graphics-art quality alpha-numerics. Thus, a map can be produced in the center of the frame and the titles, legends, and credits can be drawn in their correct locations around the map in one operation. This obviously eliminates the necessity of preparing separate titles and other descriptive information and later having to cut them into the final publication negatives.

We have been very impressed with the demonstrated ability of computer micrographics to produce quality maps at a very low cost. As the COM and Electron Beam Recorder technology continues to advance, we believe that additional increases in map quality will be realized at no increase in map production costs. In conclusion, I think it is fair to say that automated cartography has in 22 years almost equaled the standards of the traditional cartographer in the production of choropleth statistical maps.

REFERENCES

(1) For discussions of computer mapping experiences see the following articles or publications:

George Layland, "Computer Mapping—Here and Now," a paper presented at the Conference on Small-Area Statistics, American Statistical Association, 1963.

Southern California Regional Information Study, "Computer Graphics, SCRIS Report No. 1", Los Angeles, Calif., 1970.

U.S. Bureau of the Census, *Census Use Study: Computer Mapping,* Report No. 2, Washington, D.C., 1969.

U.S. Bureau of the Census, *Census Use Study: GRIDS, A Computer Mapping System,* Washington, D.C., 1972.

U.S. Bureau of the Census, *Geographic Base File System - Establishing A Continuing Program,* Report GE 60 No. 4: Washington, D.C. 1973.

FIGURE 2

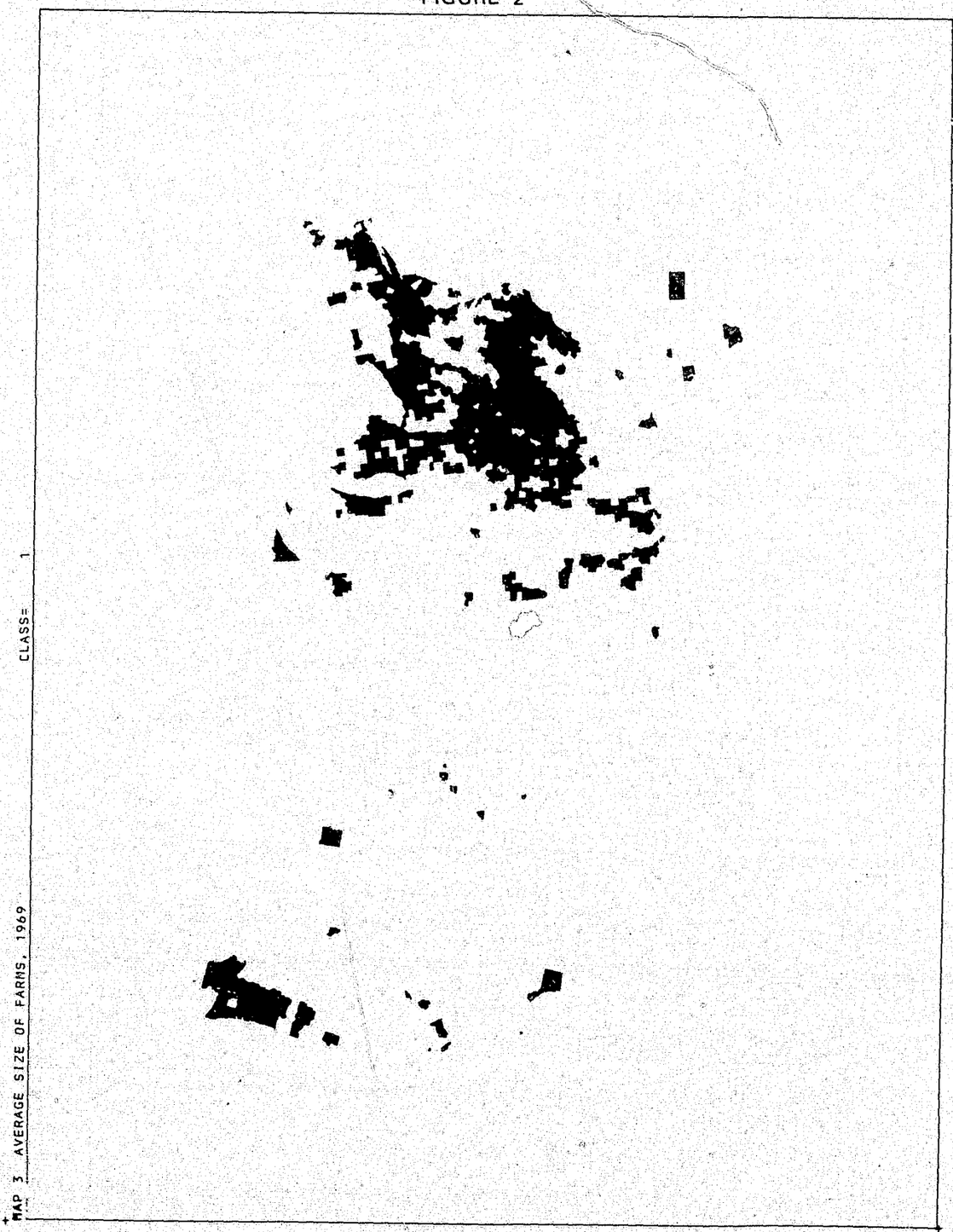


FIGURE 3

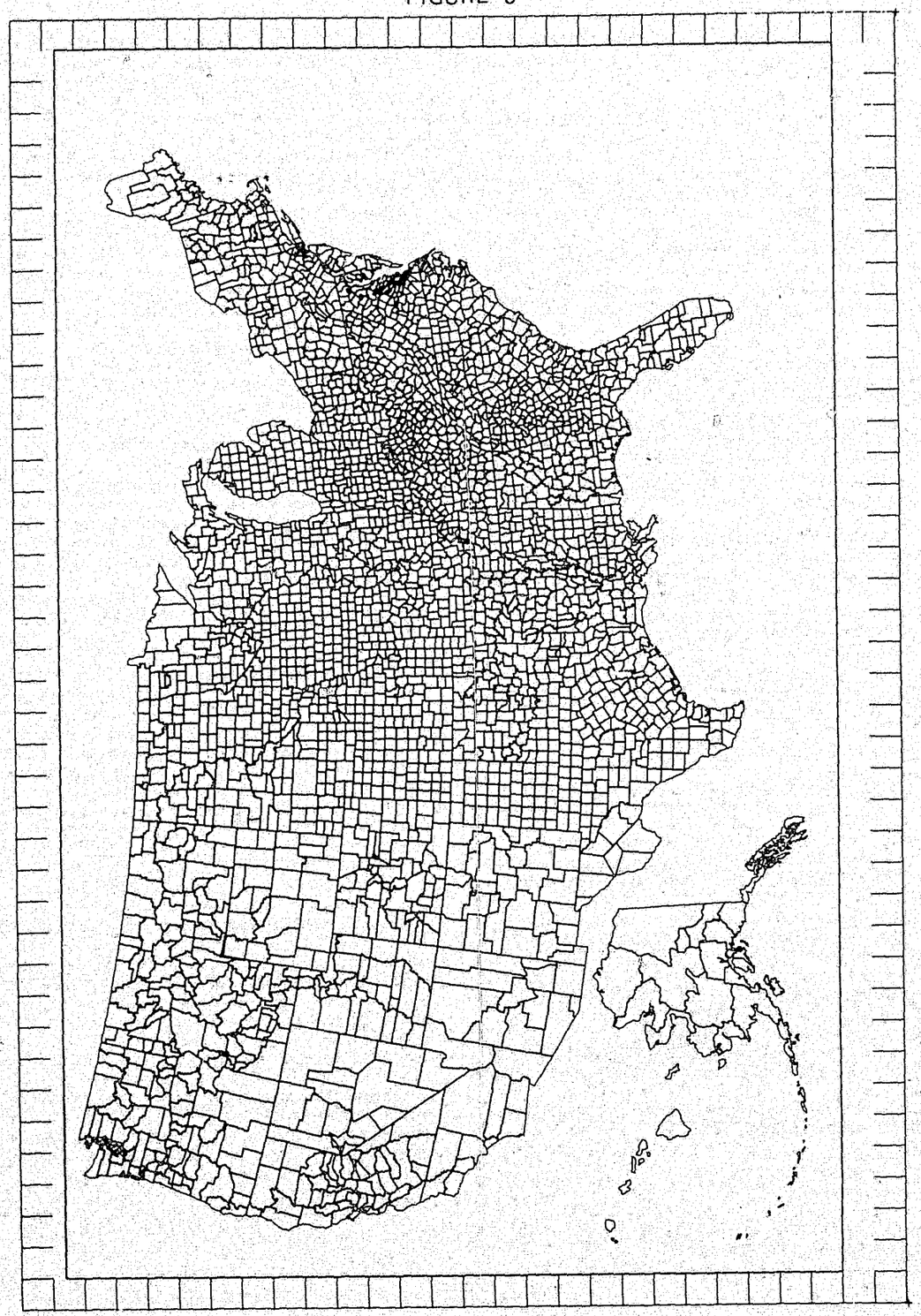
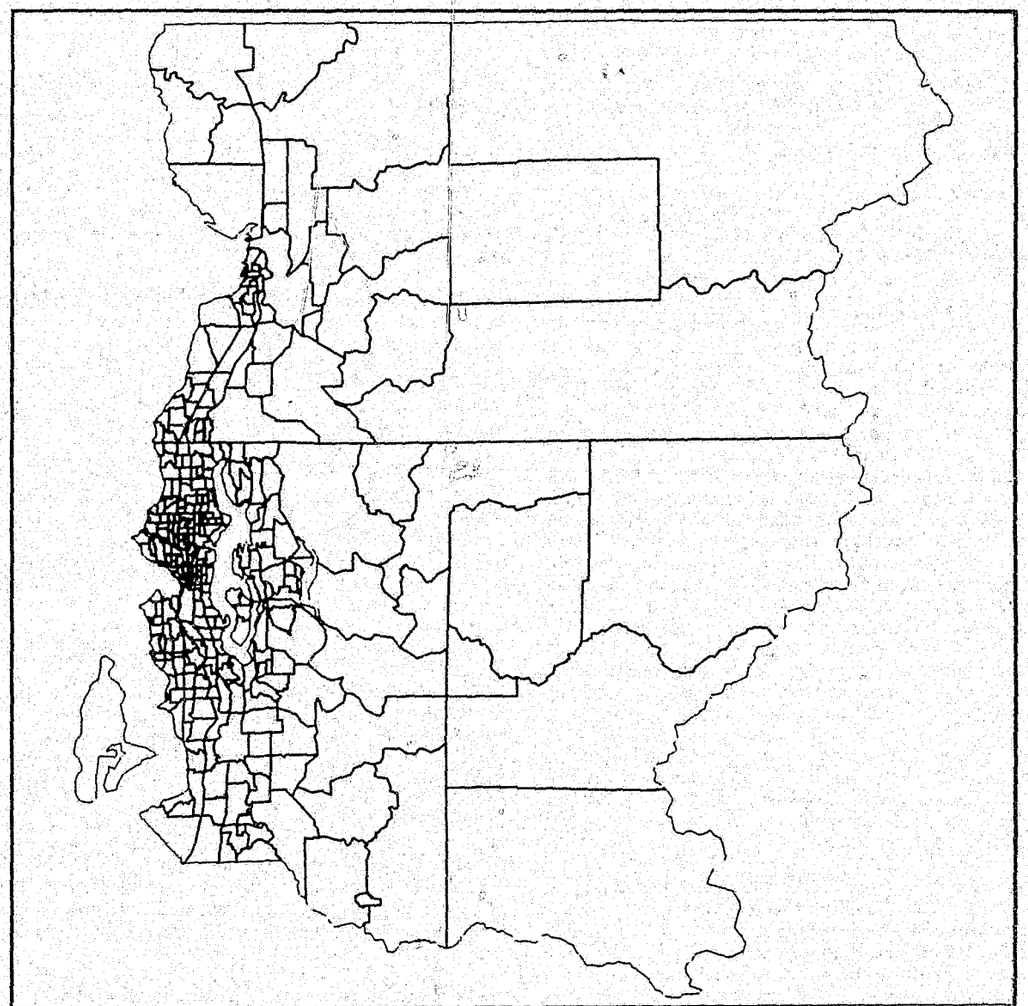


FIGURE 4



AN FR-80 PLOT OF CENSUS TRACT OUTLINES FOR SEATTLE-EVERETT, WASH. SMSA

Question Period

Mr. Kondo—The costs that you gave were, I assume, production costs. What is the general order of magnitude or range of capital cost that someone would have to incur before you can start this process?

Mr. Schweitzer—Yes, however, there are two types of capital costs. One is a function of the digitized boundary file that you are referencing. In the Urban Atlas project we are creating files basically from scratch. The capital costs are comparably high. In terms of using a GBF/DIME file, and using a program like DACS to extract tract boundaries as a polygon string, the costs are not great because most of the file developmental costs have been incurred already in the development of the GBF file. If, however, you start from scratch, you can use manual digitizing procedures and produce an acceptable file. The second type of capital cost is the equipment, the capital equipment that is involved, such as a COM unit. We fortunately have COM units available to us through the close proximity of NOAA, the National Oceanic and Atmospheric Administration. We use their COM unit and they charge us a flat rate for the operation of the machine. These costs are roughly 25 cents per frame. Private service Bureau costs, I think, would be comparable. If you plan on buying the machine, it is approximately a $150,000 to $300,000 device. Consequently, not many planning agencies or local governments would be able to afford it. Most agencies would have to purchase time from a local service bureau or corporation. You have to find access to it. So that is a major capital item. Certainly more than the line printer on the computer.

Mr. Cooke—Two questions. One, in talking about the micrographics you didn't tell us how you made the map showing population density, the one that looked like everyone held a flashlight up in the air at night. Was that with the MEDlist coordinates?

Mr. Schweitzer—That map is not a computer map. That was a manual cartographic effort produced from reverse negatives from a very large GE-50 population distribution map that we produced about a year ago. The black or very deep blue, the United States with a lighter blue background for the water areas, and Canada and Mexico were overprinted to add the nighttime effect. However, this map could have been done on COM. It probably will be done on a COM in the future. For instance, we have seen excellent plots made of oceanographic soundings where the COM, since it can uniquely address 16,000 by 16,000 points over a 6-inch glass plate, would have the resolution that would be required to produce such a map. We obviously would use the MEDlist-X tapes for the coordinates of every ED and block group. I believe that with micrographics we could create a dot map of that magnitude, probably with that level of accuracy.

Mr. Cooke—I'd really like to see it done. My second question is that you talked about the tremendous capital expense of generating the tract boundary coordinate files. Obviously, if these files are in reasonably good shape and could be used, that would be a great asset to people who would like to use their own plotting software to make tract maps. Can you tell us anything about the availability of those files?

Mr. Meyer—I asked Mr. Schweitzer to let me answer this question because I wanted to emphasize that we are not developing an in-house program. We are very much interested in advancing the state of the art of computer graphics. The digitized tract boundary files (and there will be a tract boundary file for each of the 241 SMSA's even though we can't publish an urban atlas for each one) will be available to any one who is interested in purchasing them at the cost of a tape copy, currently $70 per reel. The program tapes and documentation, including the Interactive Graphics Correction System developed by Lawrence Berkeley Laboratory to edit the new digitized output will also be available. And we hope the files and programs will be widely used.

Mr. Schweitzer—To answer Mr. Cooke's question on the technical side, the coordinates will be polygon strings not in a DIME format per se. Obviously it can be converted into a DIME-style format where adjacencies are preserved. Most mapping systems we have discovered rely upon polygon strings. Consequently, the file is being produced as polygon strings.

Mr. Boetje—I had a discussion with you about a month ago on whether this laser system was going to produce a DIME file directly. There seemed to be some question as to how difficult that could be. I have no idea whether you will be working up a system whereby you can tie this with your interactive system and actually produce a DIME file with address editing capability, block face numbering capability, etc., that way. Secondly, what would be the cost reduction if you have a program that produces your SCAN file? what would the cost reduction be if you had a boundary that already had closed polygon sequence, in terms of producing a SCAN file? You have to use your NICKEL program to produce those polygons out of your GBF right now. If your GBF contained the closed polygons directly, what kind of cost reduction are you talking about?

Mr. Schweitzer—To answer your last question first, we did not use a DIME format to produce the U.S. county outline that you saw today. So we don't have that cost to incur, the file is in a polygon string format to begin with. So I can't answer that one. However, I don't think the cost would be very large. In response to your first question, we are using DIME methodology as a part of the editing procedures. Adjacent line segments are matched. This insures that only one line is defined between each of the adjacent tracts or polygons. The editing procedures in editing work on the CRT's are designed to maximize the logic of the DIME technology interactively in the editing. This is similar to some of the methods that you described this morning. The final output files are still in polygon strings so that they have the widest utility in terms of general computer mapping. One caution on this, the accuracy of these files is really in terms of the spatial pattern that is created. They will be aesthetically correct. If you overlayed it on top of a very precise map, you may find individual streets are misplaced in terms of where they really ought to be. Since there are some outright errors in the original drafting of the tract outline maps, some of the boundaries are not quite as accurate as they ought to be. We do not have the money or time to go back to an accurate base to check to see where a line is 100 feet or a fraction of a mile left or right, of where it ought to be. The files are general purpose mapping files, not high-precision quality files for precision editing or other types of work.

Mr. Molski—I am not sure whether Mr. Meyer answered this question or not. I am going to ask it anyway. Is the Bureau offering this mapping capability as a service—as a user's service?

Mr. Schweitzer—Yes, we will do this as a special tabulation service for outside requests. We have some from other government agencies pending at the moment where we will use this mapping capability and produce maps to their speicifications. Local areas that want additional maps—more than the 12 maps that are being produced in terms of the Urban Atlas project for their particular area—can contract with us to produce additional maps at cost. We will produce the maps or publication-ready negatives to match their specifications. Likewise, areas that are not included in the *Urban Atlas* project can also have maps produced for their use at cost.

Mr. Meyer—Will you describe the data that we already have available on file and the data format that we would have to have to produce a special set of maps?

Mr. Schweitzer—For the *Urban Atlas* project, we are accessing a census data file that was produced about 2½ years ago as a special tabulation at the Bureau for an outside user. It is data extracted from the fourth count file for census tracts. The data has been put into relative terms. Most of the things are in percentages, ratios, so that you have an automatic comparison between different areas. This data file may not provide all the data that might be desired. Consequently there may be additional costs in a special tabulation cost estimate from Mr. Voight's office to create the data that you need. Alternately, we might ask you to supply us the data, if it is local data that we don't have. This would have to be in a format that we could directly map.

Mr. Cooke—The previous gentleman just tagged something onto my question. I want to tag something onto his question. To really get down to specifics, if I gave you 500 punchcards with 10 numbers in each—1 card for each tract in the Boston SMSA—and I told you the cutting intervals for each map that I wanted, what would it cost me to get 50 copies each of the 10 different variables delivered in final form?

Mr. Schweitzer—Until we get more experience in terms of areas such as Boston, the cost would be approximately $200.

Mr. Cooke—OK, let's say it's 100 percent. Two hundred dollars per map for 10 maps would be $2,000. Now for $2,000 would you deliver me the 10 piles of 50 maps each, or not?

Mr. Schweitzer—You want them printed and everything?

Mr. Cooke—Do I have to go to the printer? That is the question.

Mr. Schweitzer—Again, we would be matching your specifications.

Mr. Meyer—To answer your last question, first, we would not undertake to do the printing for you. You could do that just as easily yourself. But if you needed to print 50 different maps, we would produce the color separation negatives at a cost of, assuming the conditions that you specified, roughly $200 per negative.

General Discussion

Mr. Hearle—There are four subjects that I mentioned earlier. One is the issue of addresses in rural areas, or more precisely, the problems of geographic base files in rural areas where there are no addresses; second, is the ZIP code issue; third, confidentiality-privacy issue; fourth, any kind of perspective comments that Messrs. Pisarski or Meyer might wish to make with respect to the overall subject of the conference. Mr. Cooke, why don't you speak first on this issue of geographic base files in rural areas where no addresses exist.

Mr. Cooke—I guess the first thing to discuss when you are talking about expanding urban addresses into the rural areas is, "Why do you want to do this at all?" and asking around I came up with a few things. For one thing, it enhances the usefulness of the files if you do have urban-type addresses in as much of the area as possible. Primarily because geocoding—the translation of the street addressing to census tract, census block, or grid coordinates—is a major GBF/DIME function. It is worth it to make every effort to expand a geocodeable area or the area where there is address coverage.

Now there is an inverse question to why should we do this, and that is, "Can we get away **without** doing this?" The answer to that is also yes, technically. There are provisions in the CUE system for "null" address ranges. Obviously, there would have to be in the cases of streams and artificial political boundaries and that sort of thing. If you don't have urban-type street addresses, you can still use CUE to extend your file into rural areas.

Why would you want to extend into rural areas if you cannot geocode? Well, for several reasons. First of all, you can do computer mapping using the geographic base file for the geographic coordinate source. Also, you can encode rural traffic accidents, for instance, which occur at intersections, that is translate the intersection of two streets into grid coordinates for further processing. Finally, of course, you can add addresses later, but it is better to do it now.

Now, how do we go about doing this? In talking to various people here I have heard about a wide variety of administrative arrangements for expanding urban addresses into rural areas—finding the proper line of responsibility, the engineering-mapping section or the public works department. This morning a gentleman mentioned a situation in Minnesota where people were given the offer to put a street address on their mail box, replacing the rural route number, to get the fire engine out there a little quicker. Now, Mr. Westerfeld in Baltimore also mentioned the private sector is interested. In fact, the local electric company has sent a couple of clerks to the public works people to help establish addresses in rural areas. However, assigning urban-type addresses is only part of the problem. It won't do any good unless people use these addresses, and I think the best summary of this is in a publication which by its scarcity appears to be a collector's item. This is the publication called *Use of Address Coding Guides in Geographic Coding*, and it is the conference proceedings of the first of this series of geographic base file meetings. I would just like to quote from what at that time was called "The Bureau of the Census Continuation Program" by Mr. William Fay who was chief of Geography Division back when the whole DIME idea was being born. According to Mr. Fay,

> . . .The Address Coding Guide, the Geographic Base File, the DIME system, all these things are dependent, very much dependent, on having information with a city type of address. I think what is important is a starting point in extending addresses of this kind; to start pushing to get city type of address systems extended into rural areas. We have been having some discussions with the staff of the Post Office Department. I expect that we will be meeting with other Federal agencies to see what can be done at the Federal level to create a better climate for this change.
>
> The city-type address systems within your areas are going to be extended out into the "boondocks," if I may use that term. This is going to happen when you feel a concern for this and when the needs of the police department, the fire department, the planning commission can be better satisfied by such an extension. I mention the Post Office Department, as I think it is the strong link in this sort of a system. Without the participation—the blessing let us say—of the Post Office Department, address extension is not going to happen.
>
> If you want to get the address system in your area extended, and to devise a system for doing this, and if the local Postmaster will say to his patrons, "All right, at the end of two years, if you are not using the new address system, you will not get your mail," then you will get the address system accepted. If the Post Office Department says, "Well, if you want to use it, all right, but if you want to continue to use Box 17, RFD 3, that is perfectly all right too," your new thoughts on address systems are going to be much more difficult to bring to fruition.

I think that is the best statement on the subject there is. To sum it up, I think that even though the Geography Division is going to keep working with the Post Office, it is also up to us out here in the "boonies" to try and work with our local Postmasters and try and get this problem solved at the local level.

Mr. Harward—There is something happening at the Post Office. I am a post office box holder, or was. Late last year, the Postmaster informed my wife that anybody asking for street service would get it if they put an application in at the Post Office. I am talking specifically about service in Massachusetts.

Mr. Meyer—We have not heard of any such program being carried out on a nationwide basis. In a real sense, the problems arise because the Post Office manages to deliver the mail regardless of address defects. We all have received letters with misspelled and/or incorrect post office names and ZIP codes. Obviously, this requires extra effort and involves extra costs. Someday the Post Office will be in the forefront in demanding a standardized nationwide address system.

Mr. Pisarski—Let me add a few points to that. Mr. Cooke, if you kept reading, I think from the Jacksonville Conference Proceedings, you will find that there was a representative from the Post Office at that session, as I recall, and he said after extensive cost-benefit analysis, they had decided that their first priority was new buildings and things like that. So I think that was a great part of their focus when they became the Postal Service.

Seriously, however, I think we have to step back a minute here, and take a good hard look at this. Number one, the Post Office processes more addresses in a day than all of us have ever processed. When you sit down with the Post Office people, you find out shortly scales of need and requirements. For instance, the American National Standards Institute began to address the issue of a standardized address for the country, a format for addressing. I think you would have to accept the fact that the Post Office's requirements and needs, in that determination, are greater than any of ours or anyone else's. It concerns me when I hear us saying, "Gee, we have got to change the world to fit our systems." I am not so sure that makes a lot of sense. When I spoke at the Jacksonville Conference I pointed out that the DIME files, which have since become geographic base files, are not complete area systems for urbanized areas or for regions. I think that is rather a crucial point to make, and I don't know that it is still really clear. What I am asking is, "Do we have to extend addresses outside of existing boundaries?" The question of what you are going to use it for is central.

In transportation planning when we begin to code origins and destinations, there is no question but that the studies will run outside the block address ranges; they always have since we first started address coding. So automated address coding, machine guide coding, has always been only a part of an overall area systems technology in transportation planning. Typically, a hierarchical set of systems was required; an automated address capability in the central cities, a guide coding capability outside that, and a map coding capability beyond that. With different perspectives, with different capabilities to respond to the real world where you have to work with it.

I think the first thing one must do is to define the area over which one has some concern, some interest. Typically it is well beyond the address coding guide range. But let's not then try to squeeze the whole world into your block coding area. There may be entirely different, quite appropriate techniques to apply. I don't think that simply saying that we have to address the world so that we can use our system is going to ring an awful lot of bells.

Mr. Sylvestre—It seems to me that the old rural route, post office box system isn't even good for the Post Office. They do have in mind taking care of that sometime, when and if they solve their other problems. It may be that they give this a relatively low priority because they are always dealing with one Federal agency at a time, like the Census Bureau. It seems to me, however, there are enough people with enough problems so there should be a way of coordinating the needs of the other types of agencies. Getting an emergency vehicle to a site outside the street address range is a pretty important thing to do.

The Census Bureau might open this issue again, and in some way try to get this raised, if necessary, to the Cabinet level. I don't know if the Post Office is part of the Cabinet anymore, but getting the weight of the different agencies that are interested, like DOT and HUD, might help. If the Post Office does not attempt to solve this problem it means that it will never get to this problem. They might as well put this problem back on the "front burner" and get an extra man to work on it.

Mr. Hearle—Mr. Pisarski, you mentioned this ANSI committee on standardized address coding. Could you describe that a little more fully?

Mr. Pisarski—Yes, that is good old X3L84. I hope that explains it for everyone. It's called Committee X3L84. They have a coding system for their committees, and this is a committee on codes for coding. We all get together every once in a while, and we talk about things like that. They are the people who, along with the National Bureau of Standards-FIPS types, developed the standard codes for counties. There is a set of standard codes for counties, notwithstanding the 30 or 40 odd county codes you saw that we displayed, that we have been working with. There is a FIPS-code set of standards that goes for States.

They have been working for nigh these many years, on a set of FIPS-ANSI standard codes for places. As I recall, it took something like 2½ years to define "place." A "place" was a thing that when you called it that, that was what it was.

At the last meeting, there was a proposal on the table for a standard address format, and the gentleman from the Post Office suggested that we delay any action, imperceptible as it is, so that the Post Office could resolve its concerns and come up with its own position. He did not specify, as I recall, any time frame in which that might happen. Clearly their concerns are major, and their needs are great. To respond to Mr. Sylvestre's comments, I certainly was not trying to defend the Post Office system, just defend their problem. I don't think ANSI is going to resolve this problem for us in the near term or the long term.

The work that you heard described by Mr. Bruck, and to some extent the people from Draper Laboratory, is coming at this from the other end. We have been working with national systems, in fact world coding systems, and you have the same problems when you start talking about country boundaries, believe it or not. We have been working at it from that end and coming down from the macro to smaller and smaller units, and of course, the DIME approach has been from the micro—that is, the block or the block face up.

I think that both approaches have now met somewhere beyond SMSA's, and we are all a little confused because we are now in a position of having to relate places—that is to say, points and areas—for a common denominator unit: a building block unit, something as convenient as block at the urban level, something as convenient as county at the national level. And from our researches to date, no such simple device exists that I can see. I think that is the problem.

Mr. Hearle—Will it be a little bit more responsive to Mr. Sylvestre's expression of concern as to who do we hang this around the neck of? Is there a place in the United States Government other than the President himself that can address this global question? Is there an assistant director of the Budget Bureau for standardized addresses, or some place in this mighty Republic? We ought to be able to get the legitimate interest of other Federal agencies that don't get to come to these meetings.

Mr. Pisarski—Clearly, I think the exercise is long overdue, and there are systems analysis people in the Post Office who recognize it and who are interested in address coding systems of the kind that we talk about, as opposed to the things that are really scan recognition kinds of systems that the Post Office now has and uses everyday, the bar coding systems, etc.

Mr. Hearle—Is there any more to discuss or should we shift into the ZIP-code issue which is very closely related?

Mr. Kunz—Before we get into the issue of addresses, I have a question related to ZIP code and addresses. I wonder if other people are running into the same problem of the proliferation of private streets and plazas and developments like Prudential Center here in Boston. Also, we have condominium developments now and almost every one has a private street system, with no addresses. They end up with their own systems not geared to any existing streets. I know these have to be "plugged in" to some methodology, and I am wondering if anybody has found a really foolproof method of checking these things back, getting them into a total geocoding system.

Mr. Meyer—Let me answer that if I may. The Geography Division has just finished geocoding some 5½ million business establishment addresses for the 1972 economic censuses. These establishment addresses are free-form, physical location, and mailing addresses. The entry to the geocoding system is ZIP code and Post Office name and we have developed techniques which recognize and correct defective inputs, among the most important of which is simply considering the defective inputs to be an acceptable variation and adding it to the file with, of course, a correct geographic classification. Minneapolis-St. Paul, for example, must exist in the file in a 127 different variations.

The same techniques were applied when coding to specific street address. All inputs that could not be matched against the GBF were "clustered" and the clustered address was added to the files with appropriate geography. Let me illustrate. Supposing the geocoding file for Boston did not include the Prudential Center. Obviously, all businesses reporting Prudential Center as their address could not be coded. Summarizing the uncoded file, however, would quickly reveal this fact and result in the address "Prudential Center" being added to the geocoding files.

We used this technique throughout. The gaps in the geocoding file were uncovered based on preliminary geocoding of approximately 3.5 million establishments nationwide. And only after these new addresses were inputed was the entire file of 5½ million coded. Our final coding rate was 99.7 percent.

Perhaps clustering also makes moot the question as to whether condominiums should be allowed to have their own internal street, and address ranges. In balance, while I guess I would agree with most of Mr. Pisarski's statement that everything need not necessarily fit within an address range scheme, I also think we ought to "bludgeon the beast" a little bit and at least knock off some of the irrationalities for the benefit of all.

Did any of this explanation answer your question?

Mr. Kunz—It shares my frustration or part of it, I am sure. It is a growing factor because of the shopping centers, plazas, condominiums and cluster developments. All this type of development which we talk of as planners, and want to see happen, are working against us as far as a good geographic base file is concerned because we find more and more of these things drop out of the existing street record files.

Mr. Meyer—Our geocoding system provides some help on that score. It will recognize and accept Prudential Center, and Prudential Shopping Center, among others as legitimate variations of the Prudential Center. But supposing the incoming address is listed as "Prudential Shopping Mall." It is Prudential Center, or is it, in fact, a different place located on the other side of town. Cluster coding solved the problems for us. If enough businesses (three or more) reported Prudential Shopping Mall as an address we went back to source material to determine whether it was a variation in name or new place.

However, as far as I can see right now, there is no technical reason why shopping centers, plazas, etc. cannot be incorporated in a GBF. Certainly we will explore the possiblity of so doing.

Mr. Kondo—I just had a brief discussion with Mr. Koch, who I think many of you may not have met because he has been working hard in the back room recording all of the discussions of this seminar. We discussed the problems of varying Post Office designations of an area. On one side are situations like Minneapolis. The Post Office address can be Minneapolis, regardless of whether you actually lived in the city or in the second or even third tier of suburbs as long as it still was in the Minneapolis Postal District. On the other hand, you have in Massachusetts people designating neighborhoods or communities within a city or town as their postal address. Jamaica Plain, West Roxbury, and South Dorchester, which are parts of the city of Boston, are accepted Post Office addresses. This practice, according to Mr. Koch, is not confined to the local area. Therefore, you get a different postal designation for people who may live right next to each other, or maybe this happens for one individual who puts it one way one time and another way the next time.

Mr. Cooke—The question raised this morning was, "How to clean up the ZIP codes on your GBF/DIME file." One person mentioned that 70 percent of them were either existing or correct, or something to that effect. In other words, it was a tremendous volume of work. Well, I would just like to say that the Census Bureau will shortly be releasing a program called ZIPMATCH. What this is, is somebody's "front end" on the UNIMATCH system which

allows UNIMATCH now to be used as a geographic directory, not the GBF/DIME file, but the *National Zip Code Directory*. So you put in a street address and out comes the street address with a ZIP code instead of a census tract and block. Well, the street address you put in could be your low address left, let's say, from your GBF/DIME file, and you would come out, hopefully, ZIP coded for the left address ranges. Then you could run it again with low address right as your data address in this ZIPMATCH run and, hopefully, come out with a ZIP code for your right range.

I don't think anyone has ever tried this. This might be a good experiment for some daring soul to try when the ZIPMATCH system comes out. Mr. Meyer mentioned that there are errors in the ZIP Code Directory, and I will add to that. There are 2,000 changes every quarter to the ZIP Code Directory.

Mr. Hearle—The question was raised with me at one point during the day as to why we had not in this conference, nor in the prior conferences on the geographic base file, ever devoted any systematic attention to the privacy-confidentiality issue. My response was that it is hard to figure out what to say about it after you have said that there is an issue. So the next set of moments is to invite anyone of you who has anything to say about how that issue should be addressed, to say it. It's an accelerating issue, and I think a real possibility exists that legislation may pass the Congress this session on that subject. Has anyone a thought?

Mr. Pisarski—This is probably kind of a noncomment item. I have been around information systems technology for quite a while, and my impression is that it is about the best way to hide information I can think of. Burying things in computers makes them terribly remote in most senses of the word. That point was made the other evening, I think, at dinner, and perhaps by Mr. Barabba, the Director.

Much of the base file technology is a mechanism for aggregating potentially sensitive information into aggregates of nonsensitive information. In that sense, geocoding is perhaps really quite conducive to the kinds of things that the policies of the Congress are concerned with. It seems to me that much of this work supports the concern for privacy as opposed to conflicting with it. I know we have had this debate with the Census Bureau, as to whether an address is, in fact, a descriptive item of information about a human being that is subject to disclosure, and we got into that when we started talking about recoding some of the addresses in the journey-to-work statistics. I guess the original ruling on that was that it was, and the second round was that it is base file information as opposed to individual information. The Census Bureau still has some thinking to do about that. That is really all the comment that I can make on it.

As to who has done something, the Bureau of Standards has done some very reasonable work for HUD as part of the USAC project on what the disclosure and privacy-confidentiality concerns might be in a large-scale integrated information system for an urban area.

Mr. Boetje—I would like to add a couple of points on what Mr. Pisarski said. There was a remark made at a data base seminar about a month ago here in town. It was said that since we have tremendous data base technology available, all of this information can be stored and worked with. One of the major problems is that most of the programers working with it are of above average intelligence and they are, unfortunately, of only average honesty. There is a good possibility for people to be able to get into any kind of very large-scale data system and pull off sensitive data. When you get up to the national level, if you have this capability, as we have seen, there is not much to stop someone from going in and pulling something out, other than some very tight legal restrictions on who can use it.

Mr. Pisarski—My comment is that on those occasions where ignorance is no excuse, I frequently found that stupidity is an excellent explanation. I think you are basically right, and I think that venality is very rarely the problem—stupidity very frequently is—so safeguard, in that sense, is quite appropriate. My own department at the moment, for instance, is guilty of stupidity as opposed to venality in the case of using social security numbers as a matching field for picking up carpool addresses. I am sure that there is a Congressman who thinks that is a very bad idea.

Mr. Rickert—Another problem in the invasion of privacy or in the protection-of-privacy area is the errors in the encoding, the errors in the files, either by keypunching or by machine error. These errors can cause problems to people with the credit bureaus, credit centers.

Mr. Meyer—There are no easy answers. Just recognizing that the information in the file is erroneous is a major problem in the first place. Perhaps the only solution is to prohibit dissemination of derogatory information without prior review for accuracy. I know that the whole problem of privacy and confidentiality is under discussion by many groups.

Census individual data is maintained under secure conditions. But if someone is determined to steal a reel of microfilm, there is no way you can prevent this except by physically searching every employee every time he leaves the premises. It would have to be an inside job. The identification of the microfilm reel doesn't state that company "X's" data is contained therein, and only someone working with the data would know what the identification codes mean. The penalties, by the way, for indulging confidential census information are very severe; a $10,000 fine and 5 years in jail if I remember correctly. These are effective deterrents. As regards publication of data, the Census Bureau carries out some very sophisticated disclosure analysis checks to make certain that individual information cannot be obtained by arithmetic or statistical manipulation of published results. Frequently, information is suppressed entirely to avoid any possiblities of disclosure—particulary where small geographic areas are involved.

Mr. Hearle—It seems to me, as often is the case of these discussions, we all end up perplexed: aware there is a problem, but not quite sure of how to proceed to resolve it. So why don't we move to another subject which Mr. Pisarksi suggested and which I am sure will get lots of attention. It has to do with the funding of the creation and particularly the maintenance of GBF/DIME files. I might ask Mr. Pisarski as part of any remarks he might care to make, to add any other comments about this conference and then Mr. Meyer can do the same.

Mr. Pisarski—I guess I was particularly interested and concerned by some of the presentations here at the conference that dealt with individual cities trying to get together the money and resources to work on their GBF files—to build them, to extend them, to maintain them—whichever of those activities is appropriate at the time. I guess that most of them fell into the catagory of horror stories of one kind or another, trying to put the things together and get something happening. I want to make a few observations on that.

First, several persons mentioned the difficulty of convincing people that they had something. I guess I can only share that. It is as difficult, perhaps, maybe more difficult, to convince an assistant secretary that you've got something as it is to convince perhaps a local planning decision maker. The way I have characterized that in the past, I guess, is to try to convince him that you have found a solution to a problem that he didn't know existed. When you start saying that I can now do this, he is astonished that you weren't able to do that all along. He didn't know that you couldn't do that, and so he is not impressed that you found a way to do it. The fact that you need a lot of money now to do it dosen't impress him either. Especially, at the Federal level when you are saying you want to do it for all of America, and it is only going to take a million dollars. You are going to match everybody's money. He says, "Is that a problem? I didn't know." I think that is a great part of our difficulty.

I was concerned that even in some of the presentations on the work that we and others are doing that our presentations were not appropriately enough oriented to applications and to functions. Perhaps it is appropriate for the technicians here to talk about the tool as a thing in itself, but I don't think that is the way to sell it. As far as the Department is concerned, I think we have taken a position since 1970: that the Department is not going to be in the business of funding the files, building of the files, and the tools per se. They are tools.

I think the Department has taken the position, and still does to my knowledge, of funding a program where the geographic base files are a very appropriate part of the activity. If you are going to be doing something where Department of Transportation funds are appropriate and where a geographic base file is appropriate to that activity—to your project—not only will DOT help to fund it, but it is my understanding that frequently they won't let you do it any other way. I am concerned now if there is local experience that is different from that, then certainly I would be very much interested in knowing about it. Mr. Ben, can you comment on that. Is that your understanding?

Mr. Ben—Yes, that is my understanding.

Mr. Pisarski—In other words, if you have an activity that DOT is funding, and a good way to do that activity is to us a GBF/DIME file, then by all means use a GBF/DIME file. Mr. Anthony Ma mentioned, "Gee, energy is a great excuse for getting your base file done," and that's fine by me. I think that exactly what you always need to build a base file is an excuse. If you don't have an excuse, we won't fund it, and I don't think we should. There are a number of urban areas that have excellent up-to-date GBF/DIME files that they keep up to date, but on a dusty shelf. Once a year they take them down, update them, and put them back on the shelf. I think that is really our concern.

Mr. Meyer—This, our fifth conference has been the best that we have had to date. We learn much each time, and I want to thank all the participants for an interesting and informative session. One final word. We are now beginning to update the GBF/DIME for the 1980 census, and expect the first updating phase to be completed for all areas by July of 1978. The geocoding system to be used for the census will be based upon the present structure and format of the geographic base (DIME) files. This means no changes until after 1980. However this does not mean that we don't want new ideas. We do. In planning terms, not only is 1990 just around the corner, but with luck we might even be working on a 1985 mid-decade census. So whatever suggestions you have for improvement of the system will be most welcome by us.

Summary of Proceedings

F. MICHAEL WEAVER

It is always nice to know that you are a welcomed speaker. Of course, the last speaker at a conference is always welcome. So I will try to make a few summary remarks and then release this group. It is only fair, since most of us wear several hats, to let your know which hat I am wearing today. As a technician and a "doer," my background in municipal information systems is heavily involved in using the GBF/DIME files and, of course, having used the GBF/DIME we need to have it corrected, updated and maintained. Hence our involvement with the CUE program. But today, I am speaking to you as a member of the Special Interest Group for Geographic Base Files known as the SIG-GBF, and a little bit of background on SIG-GBF seems to be in order.

SIG-GBF began at a meeting in Louisville in 1970, and holds annual get-togethers in cooperation with the Urban and Regional Information Systems Assocation (URISA). We have had meetings in New Orleans, San Francisco, Atlantic City, and this year we are looking forward to our meeting in Montreal, August 20 to 23. We have over 170 active members and we are always looking for new members. In Montreal, SIG-GBF will be tackling two major areas. One, we have our traditional role as information exchange meeting group during the URISA conference. But prior to the URISA conference, we are having a 2-day workshop on geoprocessing basics, applications, and technology. I would call to your attention this 2-day workshop, August 19 and 20, prior to the 3-day URISA conference as an excellent opportunity to introduce someone to the field of geocoding, to the application areas involved with geoprocessing, and basically to orient anyone from your agencies or groups.

Since 1970, there have been some notable changes in the SIG-GBF emphasis. A change, I think, that is going from the technologically orientated emphasis more to the user-application emphasis. In think this is good. This is evidenced within the SIG-GBF by debate which is now taking place and will soon be culminated in a decision as to what we are going to call ourselves, SIG-GBF or SIG-GEO Processing; SIG-GBF places more emphasis on the file structure and the technology, while SIG-GEO Processing or SIG-GEO emphasizes the user application of the technology. It will be interesting to see the final vote.

Given two points, you can define a line. Given a few more, you might even get a trend. Looking over our shoulder at past Geographic Base File Conferences, such as this, I think it would be interesting to highlight what I saw when I read through the proceedings and see if there is a trend. In 1970, it appears to me that there was a basic search to define the objectives and goals associated with GBF-DIME technology. I think Dr. Aangeenbrug defined the problem as "fuzzy" systems back then. I think the word is very appropriate. In 1971, there seemed to be a search to define the technology ACG, DIME, GBF's. I remember going through that period. It was quite confusing just to know which file was what and where was it going to take us. In 1971 and into 1972, there seemed to be a maturing of the technology, more stability. The users were having more and more experience with both the census data and the GBF technology. In 1973, there was the emphasis on the maintenance of the GBF, for having once discovered that it had local application, we suddenly began to find the hit ratios on the files being of question. Suddenly we became concerned with correcting, updating and extending the files.

There also were several common themes in the past conferences which I think are quite interesting. These same points have been emphasized at this get-together. Remarks such as "Cooperation among users," "The way to survive is to cooperate," and "Cooperation among agencies is essential if the GBF is to be maintained at the local level," have been made several times. Private sector involvement has been raised several times in past conferences and here also. I think we are going to see a growing emphasis in the private sector working with the GBF. The people-related problems have been popping up constantly as well, whether it is the clerical turn-over rate or the lack of training of system specialists. The people problem is still with us. It has been mentioned over and over again and perhaps at a future conference, we can concentrate on that subject. Of course, finding funding sources has been a problem for quite awhile. The need for ongoing funding to maintain the GBF will be with us as a long-term issue. This particular 2-day conference has been extremely interesting to me in a number of areas. One highlight was the demonstrated success of using the GBF at the local level, for example in Las Vegas. We need success stories. We need to raise the flag high and wave it to make sure people know that we have a winner, and we do have a winner. Another success in correcting and updating the GBF was presented. It is good to hear, for example, though the trials and tribulations they appeared to go through, Columbus ended up with a good, clean GBF. We need to raise the flag here and wave it high. We have a success in multiagency cooperation; again, Columbus and other areas have reported to us at this meeting. The secret to success, I think, is this multiagency cooperation, both local and Federal. I think we now are in a position, at least in the New Orleans case, to recognize that we have a capability. In the case of New Orleans, they recognized that they needed cooperation from local sources, and they put their emphasis on selling the technology. I think, indeed, we must sell this technology to the local users. Make them aware of it and promote it.

Seeing a movement from the "How to do it," to "Look what we have done," over the last series of five conferences, ours today is a more marketing-oriented task than a technological one. The technology is here. Now, we must let our light shine, if I may use that expression, and tell the user that we do have for him a cost-effective management tool. I think we can defend this, and we shouldn't be afraid to try. Again, ours is a selling job. Indeed, promoting the GBF technology is our responsibility to the user and the taxpayer. In the future, as I see it, we have hardware and software technology developments on our side. Cost ratios on the hardware side are each year becoming more and more in our favor with faster computers and more flexible systems, the actual cost per event is being brought well within the scope of local area users budgets. We have more operational user application needs which will be successfully met through geoprocessing technology. We have had here examples in the school board area. We have had local government examples. We have had private sector examples. More and more applications will be met that are user need oriented, again, as opposed to GBF/DIME oriented.

Common data-collection and maintenance needs at the local level will lead to the formation of what I will call geo-utilities which are self-supporting to GBF/DIME-related services. The GBF/DIME utility of the future will have a structure which will serve both the government and the private sectors at the local level.

In closing, I am in complete agreement with Mr. Meyer's final comment, "The best years are yet to come."

Mr. Meyer—If I may have one last word, I would like to express on behalf of the conference sponsors and the participants sincere thanks to Mr. Hearle for his distinguished direction of the conference proceedings.

Mr. Hearle—Thank you. I do appreciate it. Thank you all for coming.

Library of Congress Cataloging in Publication Data

Main entry under title:

Geographic base file system—a forward look; conference proceedings, April 16 and 17, 1974, Boston, Mass.

(Computerized geographic coding series GE 60, no. 5)

Co-sponsored by the Massachusetts Bureau of Transportation Planning and Development, the Transportation Systems Center, and the Urban Systems Laboratory of M.I.T.

Supt. of Docs. no.: C 56.237:5

1. Geographical location codes—Congresses. 2. Electronic data processing—Geography—Congresses 3. Electronic data processing—Cartography—Congresses. I. Massachusetts. Bureau of Transportation Planning and Development. II. Transportation System Center. III. Massachusetts Institute of Technology. Urban Systems Laboratory. IV. Series.

G108.7.G46 001.6'424 74-19013

UNITED STATES
GOVERNMENT PRINTING OFFICE
DIVISION OF PUBLIC DOCUMENTS
Washington, D.C. 20402

OFFICIAL BUSINESS

SPECIAL FOURTH-CLASS RATE
BOOK

POSTAGE AND
FEES PAID
U.S. GOVERNMENT
PRINTING OFFICE
375

END