

DOCUMENT RESUME

ED 418 150

TM 028 248

TITLE Census Mapping Project School District Data Book. Working Paper Series.

INSTITUTION MESA Group, Arlington, VA.

SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.

REPORT NO NCES-WP-96-04

PUB DATE 1996-02-00

NOTE 75p.

CONTRACT RN92161001

AVAILABLE FROM U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Avenue, N.W., Room 400, Washington, DC 20208-5652.

PUB TYPE Reports - Evaluative (142)

EDRS PRICE MF01/PC03 Plus Postage.

DESCRIPTORS Cartography; *Census Figures; Children; Comparative Analysis; Computer Software; *Databases; Elementary Secondary Education; Geographic Regions; *Maps; *School Districts

ABSTRACT

The School District Data Book (SDDB) is a database and information system housed on 51 CD-ROMs containing the most extensive set of data ever developed on children, their households, and the nation's school systems. The Data Book allows comparisons among school districts and permits the extraction of data about districts with particular characteristics. The database provides up to 200,000 data items for each school district or county, and a mapping feature enables users to view maps of all individual school districts in the nation for the first time. Approximately 15,000 school districts have been mapped. These are usually the same districts included in the Common Core of Data. This briefing document is organized into the following sections: (1) development of the SDDB; (2) features of operation and software; (3) database content and applications; (4) mapping features and applications; and (5) school districts and the mapping project. (SLD)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

NATIONAL CENTER FOR EDUCATION STATISTICS

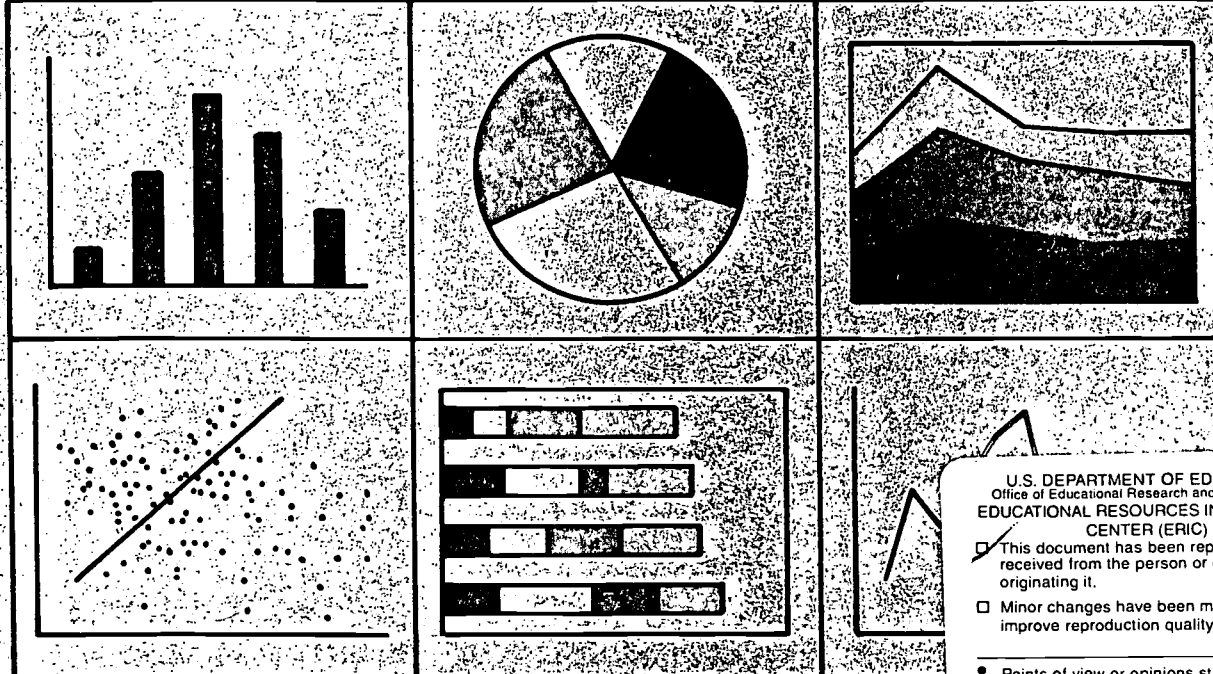
Working Paper Series

ED 418 150

Census Mapping Project School District Data Book

Working Paper No. 96-04

February 1996



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

**U.S. Department of Education
Office of Educational Research and Improvement**

TMO28248

Census Mapping Project
School District Data Book

Working Paper No. 96-04

February 1996

Contact: Tai Phan
Projections and Special Programs Branch
(202) 219-1627

U.S. Department of Education

Richard W. Riley

Secretary

Office of Educational Research and Improvement

Sharon P. Robinson

Assistant Secretary

National Center for Education Statistics

Jeanne E. Griffith

Acting Commissioner

Elementary/Secondary Education Statistics Division

Paul D. Planchon

Associate Commissioner

National Center for Education Statistics

The purpose of the Center is to collect and report "statistics and information showing the condition and progress of education in the United States and other nations in order to promote and accelerate the improvement of American education."—Section 402(b) of the National Education Statistics Act of 1994 (20 U.S.C. 9001).

February 1996

Foreword

Each year a large number of written documents are generated by NCES staff and individuals commissioned by NCES which provide preliminary analyses of survey results and address technical, methodological, and evaluation issues. Even though they are not formally published, these documents reflect a tremendous amount of unique expertise, knowledge, and experience.

The *Working Paper Series* was created in order to preserve the information contained in these documents and to promote the sharing of valuable work experience and knowledge. However, these documents were prepared under different formats and did not undergo vigorous NCES publication review and editing prior to their inclusion in the series. Consequently, we encourage users of the series to consult the individual authors for citations.

To receive information about submitting manuscripts or obtaining copies of the series, please contact Suellen Mauchamer at (202) 219-1828 or U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, 555 New Jersey Ave., N.W., Room 400, Washington, D.C. 20208-5652.

Susan Ahmed
Acting Associate Commissioner
Statistical Standards and
Methodology Division

Samuel S. Peng
Statistical Services and
Methodological Research

Census Mapping Project

School District Data Book

The MESA Group

February 1996

**2775 South Quincy Street, #620
Arlington, VA 22206**

Table of Contents

	Page
Foreword	iii
Table of Contents	vii
1. School District Data Book Product Summary	2
1.1 School District Data Book--An Overview	2
1.2 SDDB Development	3
1.3 Features of Operation and Software	7
1.4 School District Data Book Contents	8
1.4.1 Demographic CD-ROMs	8
1.4.2 Cartographic CD-ROMs	11
1.5 SDDB Mapping Features	12
1.6 School Districts	14
2. Recommendations	16
2.1 Year 2000 Census Planning	18
2.2 Effective Use of the SDDB Products	21
3. Distribution of CDs	23
4. Chronological Overview	24

Census Mapping Project/School District Data Book

This document is the final report for the Census Mapping Project and School District Data Book Contract No. RN92161001. The purpose of the report is to summarize activities and products and services resulting from the project under the purview of this contract.

Subjects covered in this document include:

- (1) A summary of the products and services developed and delivered to the U.S. Department of Education
- (2) Recommendations
- (3) Distribution of CDS
- (4) Chronological summary of events

For more information about this report, contact:

The MESA Group
2775 South Quincy St., #620
Arlington, VA 22206

1. School District Data Book Product Summary

1.1. School District Data Book -- An Overview

The School District Data Book is a database and information system housed on CD-ROM containing the most extensive set of data on children, their households and the nation's school systems ever developed.

The School District Data Book enables the user to:

- o Obtain unique insights about any school district
- o Compare one school district to another
- o Locate districts having certain characteristics
- o Draw school district maps depicting patterns
- o Extract data for use with other software
- o Use reference features as a handy electronic library

The School District Data Book has been distributed on a set of 51 CD-ROMs. 44 of the CD-ROMs are referred to as *demographic* CD-ROMs. 7 of the CD-ROMs are referred to as *cartographic* CD-ROMs. Using a conventional microcomputer equipped with a CD-ROM reader, immediate access is provided to data for every school district, county and state and the United States as a whole. This section provides a summary of what is contained in these CD-ROMs and how the information can be used.

The immense SDDB database contains approximately 20 billion characters of data, the equivalent of 14,000 high-density diskettes. The database provides up to 200,000 data items for each school district or county. The mapping feature enables users to view maps of all individual school districts in the nation for the first time.

The School District Data Book is an information resource of the National Center for Education Statistics of the U.S. Department of Education. It has been developed by The MESA Group in cooperation with the U.S. Census Bureau and involvement of all state education agencies. The School District Data Book is the only source of the 1990 Census School District Special Tabulation. The School District Data Book has been developed to facilitate effective access to these and related data for planning and evaluation of the nation's education delivery system and related issues concerning children and their households.

This briefing is organized into the following sections:

- o SDDB Development
- o Features of Operation and Software
- o Database Content and Applications
- o Mapping Features and Applications
- o School Districts and the Mapping Project

1.2. SDDB Development

Census Mapping Project. Development of the School District Data Book started in 1988 with the Census Mapping Project. Under this initiative, sponsored by the National Center for Education Statistics and coordinated by the Council of Chief State School Officers, all states participated in a program to develop school district maps. The maps, the first complete set ever to be developed for the nation, were the critical first step in the development of the database.

A public school district is an area whose public schools are administratively affiliated with a local education agency recognized by the state education agency as responsible for implementing the state's elementary and secondary public education program. Through the Census Mapping Project, approximately 15,000 school districts were mapped.

School districts delineated by the Census Mapping Project are usually the same as those referenced in the NCES Common Core of Data Program. The Census Mapping Project used names and codes from the 1989-90 Common Core of Data as a means of identification.

Most areas of the United States are covered by one or more school districts. However, there are parts of some states that are not covered by any school district. These 60 areas are referred to as "balance of county" areas and treated as "pseudo" school districts in the SDDB. As a result, all areas of the United States are accounted for through the Census Mapping Project.

Paper maps developed by individual states were sent to the U.S. Bureau of the Census. The Census Bureau digitized the maps and transferred the resulting data into the Census Bureau's TIGER System. The TIGER (Topologically Integrated Geographic Encoding and Referencing) System is used by the Census Bureau as a way of tabulating address-oriented data. Once the school district maps were a part of the TIGER system, each of the nation's 6.5 million census blocks could be uniquely associated with their respective school districts and the census data could be assembled for all these districts.

MESA Group and SDDB Development. In 1992, the National Center for Education Statistics contracted with The MESA Group of Arlington, Virginia to conceptualize and develop the School District Data Book (SDDB). It would be MESA's responsibility to assemble the raw data into the databases that became a part of the SDDB and to design and develop the software to meet the goals of the Department of Education for utility and ease of use. MESA continues to update and extend the software and database system and provide assistance to SDDB users.

The MESA Group developed the School District Data Book Demographic and Cartographic CD-ROMs as described in this report working with the Census Bureau and the National Center for Education Statistics. One product developed under the broad umbrella of Census Mapping Project/School District Data Book was the *School District Analysis Book*. The MESA Group had limited responsibility for development of the School District Analysis Book (SDAB). The MESA Group was charged with responsibility to produce certain extract files from the Census Bureau special tabulation files and provide these files to Synectics for Management Decisions. Synectics

then took these data files, reorganized the data, and developed separate retrieval and display software for the School District Analysis Book. Once the SDAB software and data files had been completed by Synectics, NCES requested that MESA produce the final SDAB public use CD-ROM which it did.

There is one SDAB CD-ROM which contains U.S. and state-level (only) data for the nation. All demographic data on the SDAB CD is a small subset of the demographic data on the SDDDB CD-ROMs.

The original plan for the SDAB CD-ROM was that it would pre-date release of the SDDDB CD-ROM. As it turned out, the SDDDB CD-ROMs were all released before the SDAB CD-ROM.

The original purpose for the SDAB CD-ROM was for it to serve as a partial electronic extension of the tables provided in a report to Congress based on the 1990 Census school district special tabulation. As it has turned out, as of this date, no report to Congress has been prepared based on these data. The MESA was never asked to plan nor prepare such a report.

These few paragraphs have been focused on the SDAB CD-ROM to clarify what is, why it was produced and how it was produced. The idea that the SDAB would contain state-level spreadsheets which could convey the same types of data summarized at the national level and contained in the report to Congress was correctly conceived. This concept should be applied with the corresponding year 2000 project and report.

1990 Census School District Special Tabulation. In 1993, under the sponsorship of NCES, the Census Bureau produced the 1990 Census School District Special Tabulation files that comprise approximately 95 percent of the SDDDB's data. MESA and Census Bureau staff worked together to develop data compression techniques to transfer the data files from a mainframe computing environment into microcomputer databases.

The Census Bureau delivered the school district special tabulation files ("File D") to MESA on approximately 200 high density magnetic tape reels. MESA transformed the census special tabulation data into a database structure suitable for CD-ROM and microcomputer use.

The following information may be useful for readers unfamiliar with the term "special tabulation" as it applies to decennial census files. A special tabulation refers to a tabulation of data prepared using the basic record file of the decennial census. The basic record file contains individual respondent data. As these data are confidential, only the Census Bureau has access to the basic record file. Thus, only the Census Bureau can prepare "special tabulations."

The 1990 Census school district special tabulations are summary data. It contains only aggregate data describing attributes of groups of persons and households in school districts, counties, states and the nation. *There are no data about individuals in the special tabulation.*

The 1990 Census school district special tabulation is the largest special tabulation ever compiled by the Census Bureau. It is also the largest demographic database ever developed from a decennial census.

The two features of the 1990 Census school district special tabulation that set it apart from other special tabulations are (1) the way the data were estimated for school districts (hence the geography) and (2) the nature of the record structures and universe of tabulation (hence the subject matter specifications).

Geography. Only the 1990 Census school district special tabulation has used a process of splitting census blocks to develop estimates for the target area of tabulation (school district). In this regard it is unique both in complexity and methodology.

Subject Matter. Conventional Census tabulation records are developed for persons or households. In the 1990 Census school district special tabulation, the scope of person and household record types were extended. Of primary interest was where the persons record was developed specifically for children (persons 3-to-19 years of age, not high school graduate). Only the 1990 Census school district special tabulation provides detailed demographic data on children--for any point in time or for any type of geographic area below the national level (where it is possible to develop limited corresponding estimates using the Current Population Survey).

The Census Bureau delivered two types of special tabulation data for this project: File A and File D. Other short-hand file names were applied to other intermediate datasets, but those datasets were never distributed for NCES or used outside the Census Bureau.

The File A data were not carried forward into the School District Data Book Project. Only the File D data were used in the School District Data Book.

With exception of two data items (total population and total housing units) the File D data are all sample-based estimates. That is, the data items are subject to sampling variability as well as others sources of estimation error.

Despite the existence of the sampling and nonsampling error sources in the data, MESA finds the more aggregate File D data to be impressively accurate. User may examine this fact themselves by comparing the complete population item with the sample-based estimate of total population for a school district of interest. Indeed, the accuracy of the estimates of draw into question the accuracy of the administratively reported data from other sources.

The File A data were used by budget at the Department of Education for purposes of administering distribution of funds to states that are distributed on the basis of population or other demographic data. The File A data are in the public domain and have been distributed by the Department of Education to individual states.

It is also noted that the File A data were re-processed by the Census Bureau, at the request of the Department of Education, after errors in geography were found and subsequently corrected. As a result, it is possible that, in a few cases, the total population data in the final File A tabulations will differ from the total population in the Filed, School District Data Book, database. The number of districts affected is small and the size of the differences, when they do exist, is small.

The Census Bureau supplied most of these 200 reels to MESA several times. In each case, MESA would process the files, sometimes most of the tapes, before locating a systematic error in the tabulations. The Census Bureau would then re-run the tabulations, place the data on tape and MESA would pick up the tapes from the Census Bureau, each time returning the tapes with errors. There were at least six versions of the special tabulation tapes supplied to MESA.

Additional Statistical Data Sources. In 1993-94, The MESA Group acquired two non-decennial census data files and integrated these into the SDDB CD-ROM framework. Described in more detail below, these files include:

- o NCES Common Core of Data
- o Census of Governments School District Finances

School District Boundary Files. In 1994, using TIGER/Line files from the Census Bureau, The MESA Group developed the first set digital boundary files ever produced for all school districts in the United States. The TIGER/Line files are a digital representation of earth surface features, such as streets and streams, each recorded as line segments in the TIGER/line files. These line segments, such as a street segment between two intersections are coded on the left- and right-sides as to school district, census block, etc.

The MESA Group developed software to process the TIGER/Line files. Using this software with the TIGER/Line files, MESA developed boundary files for all school districts in the United States. These boundary files are used by the SDDB software to draw maps of school districts. The boundary files are an integral part of the SDDB CD-ROM series.

1.3. Features of Operation and Software

Standardized software is provided on each SDDB Demographic CD-ROM. When the system is started, the main menu appears as shown in the display presented below. The SDDB software design is critical to meeting the goals of making the data not only easily accessible but also highly usable -- by users with varying interests and technical backgrounds.

Specific Operations

Profiles and Tables:

- o select geography through menu-driven operations
- o select profiles providing highlight data
- o select tables providing more detailed census data
- o view or print the data

Database Operations:

- o extract data for use with other software
- o locate districts meeting certain criteria
- o prepare reports with custom selected items
- o obtain basic distributional statistics

Mapping Operations:

- o display maps of individual states by district or county
- o display thematic maps showing subject matter items for states, districts and counties

Electronic References:

- o online user's manual for guidance
- o glossary of definitions and terms
- o subject matter index to locate data

Extensive menus enable the user to make guide operations using "pick-from-list" selections as shown in the Profiles menu displayed below. There is no need to memorize complex keystrokes or make choices based on vague abbreviations.

Once the user has selected geography and subject matter of interest, the data are presented in a comparative display manner as shown below.

1.4. School District Data Book Contents

The School District Data Book is organized into two types of CD-ROM titles: Demographic CD-ROMs and Cartographic CD-ROMs. The content of each type of CD-ROM is described below.

1.4.1. Demographic CD-ROMs

The 44-volume SDDDB Demographic CD-ROM set includes one volume for the U.S. by State and a volume for each state by school district and county.

Basic content of each Demographic CD-ROM includes:

- o SDDDB software and reference files
- o For all districts, counties, states and the U.S.:
 - "Top 100" database of key demographic items
 - Administrative database (no county data)
 - Financial database (no county data)
- o Boundary files for maps
 - U.S. by State
 - State by county (all states)

The U.S. by State CD-ROM contains, in addition to the basic content:

- o Detailed 1990 Census school district special tabulation data for the U.S. and all states.

Each of the State CD-ROMs contain, in addition to the basic content:

- o Detailed 1990 Census school district special tabulation data for the state and each of its districts and counties.

Several states require two or more CD-ROMs, while in other cases two or more states are contained on one disc.

The content and structure of four types of subject matter files (1990 Census, Top 100, Administrative and Financial) on each CD-ROM are described below.

1990 Census School District Special Tabulation. The 1990 Census School District Special Tabulation data are provided for each school district, county, state and the U.S.

Data are organized by 7 types of tabulation records (discussed in more detail in the Reference Manual):

Type	Data	# Items
1	Characteristics of All Households	981
2	Characteristics of All Persons	5,688
3	Characteristics of Households with Children	808
4	Characteristics of Parents with Children	3,187
5	Children's Households Characteristics	808
6	Children's Parents Characteristics	2,813
7	Children's Own Characteristics	2,271

As shown above, there are some 2,271 data items about children themselves (Children's Own Characteristics) for each geographic area of tabulation.

Roughly 70 percent of the data items in each record correspond to the Census Bureau subject matter tables used in the 1990 Census Summary Tape File 3. The additional tables follow similar numbering/reference nomenclature but have been defined by NCES to serve more specific data uses; e.g., dropout population and at-risk populations.

For record types 3 through 7, the basic tabulation universe is children 3 to 19 years of age who are not high school graduates. Tabulation categories are further detailed by type of enrollment:

Total Enrolled & Not Enrolled
 Total Enrolled (Public & Private)
 Enrolled in Public School
 Enrolled in Private School
 Not Enrolled

This means that the 2,271 data items for children's own characteristics in a district are available for each of enrollment categories. As shown above, each of the data items are equally available for those enrolled in public school as well as those enrolled in private school.

For each type of enrollment category, as applicable for a school district age/grade coverage, in record types 3 through 7, the data are further broken down by the following age/grade categories:

By Grade:	By Age:	By Grade:	By Age:
Total Relevant	Age 0- 2 years	Grade 5- 8	Age 18-19 years
Pre-Kindergarten	Age 3- 4 years	Grade 9-12	Age 3-19 years
Kindergarten	Age 5-13 years		Age 5-17 years
Grade 1- 4	Age 14-17 years		

This means that the 2,271 data items for children's own characteristics in a district are available both for all children and for each of these age/grade categories. As shown above, each of the data items are equally available for those enrolled in grades 1-4, for example, in public school as well as those enrolled in grades 1-4 in private school.

Top 100 Database. The "Top 100" database was developed to provide a compact file of key data items to be provided on each CD-ROM for each district, county, state and the U.S. These data have been drawn mainly from the Census school district special tabulation. They include:

- Persons by Sex, Race and Other Selected Attributes
- Families/Households by Selected Attributes
- Housing Units by Selected Attributes
- Economic Characteristics (selected items)
- Dropouts and At-Risk Children
- Attributes of Children
 - Total Children (selected items)
 - Children Enrolled in School (selected items)
 - Children Enrolled in Public School (selected items)
- Selected Administrative Items (Common Core of Data)
- Selected Financial Items (Census of Governments)

Financial Data. The financial data, from the 1989-90 Survey of School District Finances conducted by the Census Bureau, includes data on the following types of subjects.

- Revenue by Sources, by Category
 - Local
 - State
 - Federal

- Expenditures by Function
 - Current
 - Non-Current

Administrative Data. The administrative data have been derived from the 1989-90 Common Core of Data - School Level File. Using the school level data, school district level aggregates were prepared for various cross-categories involving the number of students by type, teachers by type and schools by type.

Readers familiar with the NCES Common Core of Data (CCD) may be aware that data from the basic file are released in a public use form annual. The CCD file used in this project was custom developed by the Census Bureau. The SDDDB CCD file presents similar subject matter data as the conventional annual file but the data have been organized differently.

1.4.2. Cartographic CD-ROMs

Seven Cartographic CD-ROM contain geographic files (no subject matter data) for mapping. Geographic files on these CD-ROMs are designed to work with subject matter data provided on the Demographic CD-ROM and user supplied data.

Basic content of each Cartographic CD-ROM includes:

- o Boundary files for mapping
 - State by county by school district
 - State by county by census tract
 - Census tract by block group
- o Street overlay files for mapping organized by census tract

The above files have been developed by The MESA Group using the TIGER/Line files prepared by the Census Bureau.

The cartographic CD-ROMs may be acquired with or without the IMAGE System geographic information system (GIS) software. However, to use the cartographic CD-ROM files, a user must have some type of GIS.

Users of the cartographic CD-ROMs who elect to use GIS(s) other than the IMAGE System should note that the files on the cartographic CD-ROMs will first have to be converted into the users preferred GIS internal file format.

1.5. SDDB Mapping Features

An important feature of the School District Data Book is that electronic maps are provided for all public school districts--the first time that a complete set of school district maps has ever been assembled for the U.S. in any form.

The mapping feature enables the user to prepare thematic or orienteering maps for districts and counties organized by state. Thematic maps present the characteristics of a data item, for example children in poverty, as hatch patterns for each district in a map. Different hatch patterns correspond to different levels of poverty.

The maps can also be used for orienteering. In the simplest case, district boundaries can be seen relative to county boundaries.

There are two types of map support for the School District Data Book. The SDDB Demographic CD-ROM contains the mapping software enabling the user to draw U.S. by state, state by county or state by district maps. The SDDB Cartographic CD-ROMs provide extended boundary and overlay files.

Mapping with the Demographic CD-ROMs. The Demographic CD mapping software is distributed as a component of each demographic CD-ROM. This software enables the user to view a U.S. by state, state by district or state by county thematic map. Subject matter data used in the thematic map may be from the demographic CD-ROM census database or from a user supplied file (subject matter data of any origin).

Due to limiting features [by design to make it easy to operate] of the demographic CD mapping software, users may desire to use the IMAGE System GIS to display thematic maps. Limiting features of the demographic CD mapping software architecture are that, by the intent to have its operations simple, there are very few options to select from and a reduced number of user controllable features. The process of drawing a state by district thematic map can be exceedingly slow because of the arrangement and number of polygons. Also, the state by district boundary files used in MapView to depict the maps integrate all types of districts that exist in the state--elementary, secondary and unified as applicable. For many states this means that the map display will offer very limited ability to examine relations between, say, elementary and secondary districts. Finally, for states with a very large number of small districts, the display screen is filled with a statewide by district map with no zoom-in ability.

IMAGE System GIS and the Cartographic CD-ROMs. By using the full IMAGE System GIS, all of these limitations are removed. But with the additional features and flexibility, come added technical requirements and the need for user technical proficiency.

The IMAGE (Integrated Mapping and Geographic Encoding) System is a proprietary software and database resource developed by MESA. While IMAGE System operations easily adapt themselves

to school and school district applications, IMAGE applications support virtually any type of geography and any type of subject matter data.

The SDDDB cartographic CD-ROMs, together with the MESA IMAGE System GIS, comprise dynamic electronic atlases. The cartographic CD-ROMs include only mapping and geographic data such as data for street overlays and supplementary geography such as county by census tract boundary files. The cartographic CD-ROM may be optionally used with the IMAGE System GIS software. The IMAGE System is a full-scale geographic information system for drawing and manipulating maps files. Features of the IMAGE System and more details on mapping applications are described in more detail a separately available IMAGE demonstration kit.

With the IMAGE System, the user may focus analysis on, for example, a set of three counties by school districts. Analysis may be further restricted to only elementary or secondary districts in those areas where such districts co-exist. The ability to put-together seamless maps using IMAGE, means a user can control the geographic breadth of analysis as well as the number and type of vertical layers of geography.

1.6. School Districts

Boundary definitions used for school districts in the School District Data Book originate with the Census Mapping Project beginning in 1988. Under this initiative, sponsored by the National Center for Education Statistics and coordinated by the Council of Chief State School Officers, all states participated in a program to develop school district maps. The maps, the first complete set ever to be developed for the nation, were the critical first step in the development of the database.

A public school district is an area whose public schools are administratively affiliated with a local education agency recognized by the state education agency as responsible for implementing the state's elementary and secondary public education program. Through the Census Mapping Project, 15,304 areas were mapped: 14,985 school districts and 319 "pseudo" school districts.

School districts delineated by the Census Mapping Project are usually the same as those referenced in the NCES Common Core of Data Program. Districts mapped in the Census Mapping Project were assigned names and codes from the 1989-90 Common Core of Data where possible.

Pseudo Districts. In addition to the 14,985 school districts mapped, 30 pseudo districts, referred to as "balance-of-county" areas, were also mapped. Additionally, 268 pseudo districts corresponding to sub-district areas were mapped. Also, in California, the only state not to fully participate in the Census Mapping Project, 21 pseudo districts exist corresponding to county equivalent areas.

Balance-of-County Pseudo Districts. Most areas of the U.S. are covered by one or more school districts. However, there are parts of some states that are not covered by any school district. These 30 areas are referred to as "balance-of-county" areas and treated as "pseudo" school districts in the SDDB. Balance-of-county areas, as the name suggests, are the residual part(s) of a county not assigned to any school district. Although these areas are treated as one area within a county for data tabulation purposes, in most cases a balance-of-county area is actually comprised of several non-contiguous parts.

While balance-of-county areas are not recognized by the State or Federal Government as true school districts, data have been tabulated for these area in the 1990 census school district special tabulation. Due to the inclusion of these balance-of-county areas, all areas of the U.S. are accounted for through the Census Mapping Project and 1990 census school district tabulations. Balance-of-county pseudo districts have district codes that begin with the two characters 81 (no other district codes in the U.S. have 81 as the starting digits).

Sub-District Pseudo Districts. In Hawaii, which is collectively one school district, under the Census Mapping Project, 231 sub-districts were mapped. In New York, 37 sub-districts, within the New York City Public School District were mapped (32 community school districts and 5 borough secondary school districts). As both true district and sub-district data are presented in the SDDB database, users may need to take special precautionary measures not to double count the data for these areas.

California Pseudo Districts. The State of California is the only state not to fully participate in the Census Mapping Project. As a result, some school districts in California were not mapped, nor were census data tabulated for these areas as true school districts. Twelve California counties, which are comprised of two or more true districts, were not mapped at the district level (Butte, El Dorado, Humboldt, Kings, Madera, Mariposa, Monterey, Napa, San Benito, Santa Barbara, Siskiyou, Tehama and Trinity). In these cases, the county itself was used as a pseudo district for census data tabulation purposes. Since data were tabulated for these county-wide pseudo district areas, the demographic data are available for California on a statewide basis even though data are not separately provided for many true districts. The number of true districts in California as conveyed by the Census Mapping Project is approximately 20-percent fewer than the true number of districts in the State.

Types of Districts. School districts are classified as elementary, secondary and consolidated (unified) based on the grade range administered by the district. Each school district's grade range in the 1989-90 Common Core of Data Public Education Agency master list represents the lowest and highest grades with non-zero student counts in the schools operated by the agency. Grades recognized for inclusion in the universe of elementary and secondary agencies range from prekindergarten (PK) through grade twelve (12).

The 15,006 school districts covered by the Census Mapping Project are categorized as follows:

- o 11,269 consolidated districts
- o 3,175 elementary districts
- o 562 secondary districts

Comparing District Data from Other Sources. For the various reasons reviewed in this section, the set of school districts included in the School District Data Book will differ some from those in other statistical programs. Explanations and distinctions cited above will help you with analyses involving multiple sources of data.

2. Recommendations

2. Recommendations

A few of the introductory remarks in this section are repeated from section 1. It is anticipated that this section on recommendations may be distributed separate and apart from the larger document.

The School District Data Book is a database and information system housed on CD-ROM containing the most extensive set of data on children, their households and the nation's school systems ever developed.

The School District Data Book enables the user to:

- o Obtain unique insights about any school district
- o Compare one school district to another
- o Locate districts having certain characteristics
- o Draw school district maps depicting patterns
- o Extract data for use with other software
- o Use reference features as a handy electronic library

The School District Data Book has been distributed on a set of 51 CD-ROMs. 44 of the CD-ROMs are referred to as *demographic* CD-ROMs. 7 of the CD-ROMs are referred to as *cartographic* CD-ROMs. Using a conventional microcomputer equipped with a CD-ROM reader, immediate access is provided to data for every school district, county and state and the United States as a whole.

The immense SDDB database contains approximately 20 billion characters of data, the equivalent of 14,000 high-density diskettes. The database provides up to 200,000 data items for each school district or county. The mapping feature enables users to view maps of all individual school districts in the nation for the first time.

The School District Data Book is an information resource of the National Center for Education Statistics of the U.S. Department of Education. It has been developed by The MESA Group in cooperation with the U.S. Census Bureau and involvement of all state education agencies.

Important and Unique Features

The School District Data Book:

- (1) is the single most important statistical database of the 1990's with respect to the analysis of characteristics of children, their households and participation in the American education system;
- (2) is the only source of the 1990 Census School District Special Tabulation prepared by the Census Bureau, and is the largest single public use demographic database ever developed by the Census Bureau;

- (3) contains the only definitive demographic database about children at the county and school district levels of geography;
- (4) is the largest public use statistical database ever developed under sponsorship of the U.S. Department of Education;
- (5) contains the first electronic mapping system for school districts ever developed;
- (6) offers an expandable platform for the continuing analyses of children, school districts and education throughout the 1990's; and
- (7) offers easy access in profiles for use by persons who are not experts in computer and statistical analyses.

Recommendations

Recommendations provided in this section are organized into two areas:

- (1) Suggestions based on experiences of this project for improvements relating to a possible Census 2000 school district special tabulation
- (2) Suggestions regarding mechanisms to provide for widest and most effective use of the information resources developed in this project

In some cases, recommendations presented here may assume knowledge about the background, scope and structure of the School District Data Book. Refer to section 1 of the report for additional information

2.1. Year 2000 Census Planning

- o The geographic and subject matter specifications should be repeated in a manner similar to that used in the 1990 census adjusting for changes resulting from sampling and questionnaire specifications.

The one most important missing tabulation needed is count of total relevant children by single years of age.

- o School districts should be a "regular" geographic area of tabulation carried on all of the Census standard machine-readable products and treated as if, for example, the school district is a place. In this respect, subject to confidentiality guidelines, there should be tabulations for the entire school district. Where school districts are split across county boundaries data should be presented for the portion in each county. Further subdivisions are not recommended unless attendance area level summary statistics are developed.
- o The Census public use release dates for key files should be, at the latest:
 - (1) June 2002 -- first regular tabulation estimates files
(incorporating districts as a regular area of tabulation)
 - (2) January 2003 -- completed special tabulation files
- o Public use TIGER/Line files with the 1999-2000 school district boundaries should be available to NCES no later than mid-year 2000.
- o The basic record files structure used by Census to output the special tabulation data was adequate other than these considerations:
 - (1) A data validation plan should be implemented from the outset rather than counting on the producers of the final product finding errors as the product is developed. The data validation plan should begin with the techniques applied by MESA in the data validation starting with Version 4 of the File D data. This step will save enormously on the cost of production of the public use files and possibly make the full data available a year earlier than experienced with the 1990 version.
 - (2) All data records developed by Census should be in the same file in sequence (delivery of separate 2B and 6F records resulted in last "minute" (that resulted in weeks to months delay) processing and long-term organization problems in producing the final SDDB demographic CD-ROMs.
 - (3) There should not be two separate total persons records; the total persons record should all be in one record.

- (4) The preferred output media would be CD-ROM (the records would still likely need to be restructured for the final public use CD). The concept of publishing data only in an Internet or counterpart online media is not practical for the school district data.
 - (5) Census should produce a summary record for the U.S. summary (produced by MESA for the 1990 project). Consideration should be given to producing summary records for metropolitan areas (not provided in the 1990 data).
 - (6) There should be a better understanding between Census and NCES as to the exact scope of documentation that Census is to deliver. In many respects the documentation was inadequate and had to be pieced together.
 - (7) A step can be saved if the data are output initially using a compression algorithm (before the mainframe output goes to tape or alternative media).
 - (8) Some record types were never moved from the Census tabulation output tapes to the final CD-ROM structures. This was a subjective decision made by Roger Herriot, the consideration being that the public use CD-ROM volume had already gotten unexpectedly large (44 CD-ROM compared the originally estimated 9 CD-ROM). These record iterations should be removed from the development process, thus saving volume and reducing time and cost.
- o The report to Congress should be delivered in three phases. The outline for all of these reports, to the degree feasible, should be completed before the final NCES tabulation specifications go to Census.
 - (1) The first report should be delivered in mid 2002 and be comprised simple summary profiles for the U.S. and states based on data released in the initial sample files (these would be the Census "regular" tabulations for whole districts as discussed above).
 - (2) The second report should be delivered to Congress in January 2003 and based on the same data, possibly extended, as (1). The second report should be more narrative-oriented and address relevancy of these data to current issues in educating America's children. This report should contain a CD-ROM supplement (possibly with the entirety of the report) containing spreadsheets, and possibly graphics, providing data for individual states as summarized for the nation in this report.
 - (3) The third report should be delivered to Congress in mid-2003. This report should make use of special tabulation data analyzing data from the "children's own" and, if produced, "households with children" records. This report should be an extension of the report (2).
 - o School district codes should be fully standardized on an annual basis across all NCES-related statistical programs. The year 2000 tabulation codes should be made a part of the TIGER system based on the 1999-2000 school year.

- o Extended subject matter tabulations are an essential part of analyzing characteristics of school districts. That is, it is essential to have data records describing attributes of the children, the children's households and the children's parents.
- o If it is necessary to reduce the scope of tabulations, here is a suggested priority:
 - (1) the least used categories of data would be the iterations of record types 3 through 7 by age and grade of child. While it does appear that these data are important, most users stopped with analysis of total relevant children in a district. Also, the augmentation to get grade detail based on age and the district grade coverage both increased cost and may have contributed to error in these specific estimates unnecessarily.
 - (2) the number of subject matter cells for total persons was excessive (5,688). This number could be narrowed to approximately 3,500 cells by reducing the number of iterations for some type of race/ethnic categories.
- o It is not recommended that the number of cells tabulated in existing record types be reduced as there is little financial gain from that action.
- o Split census blocks should no longer be used for several reasons.

It is recommended that under the suggested annual boundary change program that districts be re-aligned to census block boundaries. There are technical issues associated with this recommendation, having to do with the treatment of housing units by block face in the tabulating process, that are beyond the scope of these recommendations. These recommendations may be submitted at a later time. In some cases the re-assignment may require a permanent split in census blocks rather than redrawing the boundary of the school district.

2.2. Effective Use of the SDDB Products

- o Most generally, the data access and use infrastructure for education information, in particular K-12 data resources and related systems, is so limited that even existence of useful, highly, relevant data does not mean it will be used.

State education agencies typically do not have staff nor equipment to support their own agencies requirements to analyze needed information. An equivalent, or redirected, program similar to the Census Bureau State data center program should be considered. State units, possibly just one person, could both perform analyses and assist within state regional service units and local education agencies.

While many of the OERI labs are both well staffed and have adequate equipment to provide data access and use services to operations within their states, the interests, motivations and actual support mechanisms vary widely--with many providing no outreach or support for access to the SDDB database.

- o The expectation that SDDB training and support would be provided by the OERI labs and State Education Agencies did not develop as hoped. There may be several reasons for this. In any event, access to the basic data has not been adequately promoted. We have frequently encountered prospective users interested in the data who had never known of its existence.

It is suggested that an effort be made to publicize availability of the data working closely with trade and professional associations which relate to the data.

- o Beyond creating an awareness about the data existence and usefulness, there is the need for user support. While many questions could be answered by an analyst reading the manual, it is not realistic to expect this to happen. Users need someone to call to gain assistance with use of these data.

It is suggested that NCES provide an ongoing Census data user support process for the next two years. This support should not be limited to the 1990 decennial data. There should be increasing attention given to the boundary file development and maintenance as we move toward the year 2000 census.

- o The School District Data Book should become the annual publishing media for K-12 data released by the Government when data for all or most districts are released. The resulting product would be easy to use and become well-known for its form of operation. Access to year 2000 census data would thus be easier, quicker and less costly to implement. End user training should be a minimum. Problems with the existing SDDB operation could be eliminated. New important features could be added over time rather than at possibly the last minute in the year 2003.

It is suggested that the Government transform the existing SDDB demographic system into a 1996 edition and subsequent annual editions. The 1996 edition would include an expanded subset of subject matter over the present "top 100" file and be integrated with the longitudinal time series data from F-33 and CCD.

3. Cartographic CD-ROMs

There are seven CD-ROMs containing the files for all states organized in the following manner:

SDDBMAP1	Alaska Arizona California Hawaii Idaho Montana Nevada New Mexico Oregon Utah Washington Wyoming	SDDBMAP5	Connecticut Delaware D.C. Maine Maryland Massachusetts New Hampshire Rhode Island Vermont Virginia West Virginia
SDDBMAP2	Indiana New Jersey New York Ohio Pennsylvania	SDDBMAP6	Louisiana Oklahoma Texas
SDDBMAP3	Arkansas Colorado Iowa Kansas Missouri	SDDBMAP7	Alabama Florida Georgia Kentucky Mississippi North Carolina South Carolina Tennessee
SDDBMAP4	Illinois Michigan Minnesota Nebraska North Dakota South Dakota Wisconsin		

4. Chronological Overview

The contract was awarded to The MESA Group on July 3, 1992. The contract ended on November 30, 1995.

Key milestones and developments are summarized below.

July 1992

- o Project starts
- o Developed plans around expected data delivery from Census to begin in September
- o Developed plans to produce 9 CD-ROMs

August 1992

- o Determined that the dates contained in the proposal for the Census Bureau to provide prototype files would be later than planned. Prototype TIGER files would not be available until 2/93. Prototype "File D" data would not be available until 11/95.
- o MESA proceeds with software development for SDDB demographic CD-ROMs

September 1992

- o MESA continues with software development for SDDB demographic CD-ROMs

October 1992

- o MESA continues with software development for SDDB demographic CD-ROMs

November 1992

- o MESA continues with software development for SDDB demographic CD-ROMs

December 1992

- o First File D tapes received, for Kansas
- o Data compression software and file transfer evaluation
- o Based on data to date, estimated that 25-to-30 CD's would be mastered (ultimately to determine that 44 would be required)

- o MESA continues with software development for SDDB demographic CD-ROMs

January 1993

- o Received more File D test files from Census
- o Finalized a data compression method working with Census
- o MESA continues with software development for SDDB demographic CD-ROMs

February 1993

- o MESA received first production (Version 1) File D files from Census (for most states)
- o MESA continues with software development for SDDB demographic CD-ROMs

March 1993

- o MESA finds systematic error found with record type 2B in File D data in files supplied to MESA by Census
- o MESA notifies Contracts that due to
 - (1) Census data delivery delays and
 - (2) underestimated size of File D database on part of Governmentthat contract scope and time frame would need modification
- o Census advises MESA that first set of TIGER files on tape would be available in April (5 months beyond the original project planned date)
- o MESA continues with software development for SDDB demographic CD-ROMs using known erroneous Census supplied data for development purposes. A focus is placed on data locator and ways for user to find needed data.
- o MESA reports geographic-oriented errors with the financial (F-33) data file as supplied by NCES to MESA (problems with file development that are responsibilities of the Census Bureau). Meeting scheduled at the Census Bureau for April to address these issues and develop a plan for amending the geographic coding.
- o MESA reports that no record layout was provided for the final Common Core of Data (CCD) file.
- o A problem with the Census Mapping Project district code correspondence with the CCD was identified by MESA.

- o Much of MESA staff time is concerned with conversion and processing of erroneous data supplied by Census.

April 1993

- o Census corrects File D 2B record error. Census starts supplying MESA with Version 2 File D data.
- o Census supplies MESA with Version 2 File D data for 42 states.
- o MESA provides Synectics first test files for developing School District Analysis Book.
- o MESA identifies a problem with Version 2 File D Data.
- o Census File D data and starts to supply MESA with Version 3 of the File D data
- o Much of MESA staff time concerned with conversion and processing of erroneous data supplied by Census
- o MESA continues with software development for SDDDB demographic CD-ROMs using known erroneous Census supplied data for development purposes.
- o MESA starts receiving TIGER files on a flow basis from Census. MESA receives approximately 80 reels of expected 200 reels.
- o MESA starts processing TIGER files and finds that county have been randomly supplied and that most states have files for part or all of many counties. No whole states.
- o MESA develops retrieval software for table data access in with NCES desired feature comparative data display for multiple geographic areas.
- o MESA completes development of the first hard copy version of the seven "record type" index files for circulation.
- o MESA completes a demonstration version of the system.
- o MESA conducts two orientation sessions at the Census Bureau.

May 1993

- o Much of MESA staff time concerned with conversion and processing of erroneous data supplied by Census

- o Still More Errors with File D. In April and May, MESA reported the existence of previously unknown tabulation errors in the File D data. New errors were found in May.
- o MESA proceeded to process all of the File D data available under the assumption that a caveat statement or explanation to be forthcoming by the Census Bureau.
- o MESA processed most of the "version 3" File D tape files. Much of these data had been placed on a prototype CD-ROM.
- o Unreadable File D Tapes. Census was notified by MESA that one or more tape reels for seven states could not be read by MESA equipment. (This matter is totally different from the erroneous tabulations and has to do with the manner in which data are recorded onto and read from the magnetic tapes.) By late May the tapes for all but one of these states had been re-supplied to MESA.
- o The problem of unreadable tapes occurred with reels provided earlier and Census replaced them just as with this more recent group. It should be noted that not one TIGER reel was identified as defective or could not be read by MESA (though they are plagued by their own processing problems). This led MESA to conclude that the problem was totally with the way that "end-of-reel" marks were placed on the Census supplied File D tapes.
- o No File D for Some States. At this time, the "first iteration" of File D data (for any "version") had not been received for three states, including Illinois, one of the largest.
- o By the end of May, MESA had received TIGER tape files for approximately 30 states though many states were incomplete.

In addition to the missing county file sets (TIGER files unlike File D data, are provided on a county-by-county file basis. For each county there are approximately 12 types of files -- the number varies.) there were also whole tapes missing. In two instances, tapes were shown on the shipment manifest but were not contained in the boxes. The nature of the omission appears to be clearly a matter of packaging rather than a loss of tapes in shipment.

To complicate the matter of TIGER file accountability, the tapes were shipped to MESA on a flow basis. As a result, a shipment might include all Michigan reels but one which might be scheduled to ship three days later. Since MESA had no advance schedule of when or what was being shipped, MESA had to wait a few days to see if missing reels/counties were a problem or simply a "spread-out" shipment.

- o As to quality control, like the File D data, MESA had no distinct charge nor complete ability to examine the completeness and accuracy of the files. It should be noted that the Census Bureau cannot prepare a school district map for any school district in the U.S. MESA checks in three states indicate complete technical accuracy. However, MESA determined in the case of Idaho that at least one district was incorrectly drawn. The source of this problem is with

the district mapping coordinators -- not the state coordinator nor the Census Bureau. Alas, the boundary remains in error.

- o F-33 Data. No change from April, this file remained under development by MESA (this cannot be fully completed until all of the CCD is completed.)
- o CCD Data. Some progress occurred with the Common Core of Data files in May. MESA received the record layout and possibly final files for two states. MESA was advised the remaining states to be delivered in June.
- o MESA continued to develop software to handle production versions of the File D data. This month's developments featured ability to process larger, more complex files. Also, the profiles reporting feature was further developed.
- o Easier forms of systematized data file extraction processing was implemented at the request of NCES so that data files can be extracted in Level 1 without learning features of Level 2 -- for more elementary applications.

Digression. Originally, there was a level 1 and a level 2 software. Level 1 was intended to be an easier to use software. Level 2 was intended for more technical applications. For a time, these program were developed and operated separately. Later in the project, the programs were put together into a single software product. Level 1 became the first option in the SDDB demographic CD-ROM software and Level 2 became the second option in the SDDB demographic CD-ROM software.

- o District boundary file generation software (using the TIGER files) was tested in several states with several types of districts.
- o In the distributable software, MESA contemplates finalization of profile layouts remains contingent upon having the complete CCD files.
- o MESA proceeds with development of user documentation for the composite system.
- o Further demonstration versions of the system were provided to NCES staff to use and evaluate.

June 1993

- o Several TIGER tape processing issues continued to develop over the past three months. Some of these issues are highlighted below.

(1) Limited Scheduling Detail. MESA had not received a plan summarizing what files are being sent at what time. Instead, the boxes of tapes arrived, presumably on a flow basis as Census was ready to release them. There was no packing list, as such, describing what the

boxes contain. In many cases a set of boxes will be missing tape(s) for many counties of a state which are "mostly" contained on tapes in those boxes. MESA could not determine which states the Bureau believes that MESA should have received in total and which ones are missing in part or in whole. MESA had received some TIGER files for most states although many states remain incomplete.

(2) Documentation. There was a printout in the shipping box for each tape in the box. MESA could visually examine that printout and determine (generally) if all counties are included on those tapes. This documentation was adequate with regard to summarizing what files are believed to be on the tapes.

There was, however, a related issue of concern. There were typically 14 or so TIGER files for each county. Some counties should not have all file types. As a result, MESA could look at the number of TIGER files and know with certainty that MESA had received all of them.

(3) Validation. MESA's limited approach to validation was basic (in absence of other instructions) -- try to use them once loaded. When all of the files have been loaded for a state, staff runs a boundary file generation test against the entire state and then examine a diagnostic file. This was a very limited means of validating that MESA had received all of the correct files for each county (keeping in mind there is no master list for check-off).

(4) Missing Counties. A recurring problem was where files for some counties are missing because the tapes containing them were missing (never shipped?). The problem is exemplified with Alabama. MESA received tapes G04289, G04295, G04298 and G04304 containing Alabama counties through code 079. MESA had no tapes containing the remaining Alabama counties. It was not known if these reels have been sent and/or if MESA should have received them by what date. This situation existed for several states. MESA had to assume that the tapes containing the "other counties" were to be shipped later. For states with this "partial delivery" status MESA could not process the files until all county files are complete.

(5) File Sequencing. Given that all county files apparently exist for a state MESA could, in theory, begin processing. The order of county files is always out of sequence. The first step was to re-sequence the files -- a cost to NCES which might be avoided by having them properly sequenced as released by Census (they ultimately have to do this anyway for their own work with the TIGER files).

(6) Incorrect Tape Contents. Another type of problem was encountered with some tapes where the contents of the tape does not match the description of the tape contents based on the printout shipped with the TIGER files. One of the most extreme problems of this type was where one tape was determined to contain STF-3 data rather than TIGER data.

More typically some TIGER tapes do not contain the same set of files as described by the printout documentation sheets accompanying the tape. As an example, upon examining the

printouts for the tapes received for Nebraska, all counties appeared to be on tapes in our possession. MESA processed all five Nebraska tapes, only to find that the printout for tape G04341 (the final tape), which shows files going through county code 091, actually stops short and only goes part way through county 055. As a result, Nebraska cannot be completed and this effort is lost (most of a person day and key machine usage day).

This problem proved to be the most costly to NCES because resources are expended that then have to be re-expended when the correct tapes are supplied.

(7) Urbanized Area County Replacements. Couple with these processing problems, the matter that the Bureau announced plans to issue replacement files for 745 urbanized area counties due to coding errors found in these files. MESA found it infeasible to replace the approximate 10,400 files that Census proposes to release as replacement files. The likelihood of errors creeping into the processing is too great to consider a piecemeal file correction procedure.

- o MESA Suggested Remedies to TIGER related issues. Minimally, MESA needed all tapes/states with reported problems to be replaced as soon as possible. Without this step MESA cannot even develop the most basic district boundary files for each state. To meet the more immediate obligations of the "demographic CD's," MESA needed to have all of TIGER files for the U.S., suitable for generating the district boundaries, on a rather immediate basis. Recalling that our targeted date for the Bureau to provide all types of files is June 30, further delay is going to push the limits of our ability to meet the target deliverable schedule.

For long-term considerations, MESA suggested that the Bureau provide a complete set of replacement TIGER data files. These files could be on CD-ROM or tape but in any event should be the ones used for the cartographic CD production.

- o File D Being Revised - Again. In early June, while developing case study materials for the DC area, several anomalies were noted with the data. Upon further examination, the number of apparent errors grew and MESA suggested that further validation of the data be performed.

By this time in early June, MESA had processed most U.S. File D ("Version 4") data into the provisional final database structures on interim CD-ROM media. These data were then determined to be of no value and the files will have to be re-created when Census provides the corrected data.

- o File D schedule. As of June 30, MESA had no final File D data for any area. Census reported that they expect to have all states reprocessed for delivery to MESA (Version 5) by mid-July. It will not be clear until late July as to how this delay will affect MESA's product delivery schedule.

- o Much of MESA staff time in June was absorbed by processing, reprocessing and validating erroneous and incomplete data supplied by Census -- both the File D data and the TIGER data.
- o Census delivered to MESA the CCD file structure documentation and extended CCD for two states as planned, MESA identified problems with these files. These problems were reviewed with Census and the files are to be revised. As of June 30, none of the revised files have been provided. MESA is told by Census that the new files will arrive by mid-July.
- o A briefing was held with the Census Bureau State Data Center steering committee on June 24. Plans for dissemination and support were reviewed. The SDC group is considering whether or not they might want to augment the presently planned basic training session at the time of their annual conference to extend to a full day training session.
- o A briefing was held at NCES on June 30 for sociologists attending a data seminar. The SDDB project and product were well received and attendees reported great interest in receiving the data.

July 1993

- o Validation Processing of File D. Due to the important of the quality of the data in this project, MESA gives special attention to efforts in validation work.

After MESA found several apparent errors in data extractions from File D performed for the case Study using Washington, DC, NCES requested that MESA check the entire File D in an effort to detect other errors that could affect the validity of the data. MESA undertook this effort on a rush basis performing the initial evaluation in one week. Errors and comments on data validity were reported to Census which then prepared a revised File D for DC. This file was rechecked within a subsequent week to verify that the errors noted in the initial checks had been corrected and that no new errors could be located.

The full procedure was carried out only on the file for the District of Columbia, although on occasion checks were made to determine whether the same problems showed up in other areas. More than 200 tabulations were checked covering all seven basic record types (HT, PT, HC, PR, CH, CP, CO) almost all subject matter areas. Some tabulations could not be checked because sufficient cross-checking data were not available.

Four approaches were taken:

1. Checks for correspondence with data form Summary Tape File 3 of the Census Bureau where comparable tabulations existed;
2. Checks with consistency with other tabulations for the SDDB, both within and across file types, and with STF3 tabulations that did not correspond exactly.

3. Logic-based checks to identify illogical numbers (e.g., persons living in vacant housing units), and to locate illogical relationships between tabulations (e.g., more households than occupied housing units).

4. Knowledge-based checks were made for tabulations that were not congruent with known facts about the District. (e.g., racial composition).

Validation processing conclusions. *A number of errors were located, including some systemic errors that impacted entire data series.* Four of the seven basic record types were found to be systemically compromised. There were also errors located affecting individual tables in the remaining series.

Several different kinds of errors were identified:

1. Programming errors. In one series, all data on parents were high by several thousand persons. In another series, five different total were reported in successive tables for the number of children in housing units. In several tables of a third series, per capita income figures were several thousand dollars too high because they had been computed on only part of the population. In still other series, children were reported to reside in vacant housing units, and persons were said to reside in rural and non-urbanized areas of the District (not possible by definition). Census agreed to fix these errors and they did not appear in the revised file.

MESA recommended, and NCES concurred, that Table HC-P025 be withdrawn from the data set to be released because the number of subfamilies shown differed greatly from data in STF-3 that should be comparable.

2. Data errors. In one series, a four-year-old child was reported to have graduated from high school, and thus was not included in the "relevant" child population for the school district. This impaired the comparability of the "universe" total with the equivalent data from STF-3. While resulting discrepancy was small, it was nonetheless disturbing (the impact in other districts may be substantially larger). In other series, women as old as 85 years were reported to be mothers of school age children. Census declined to correct these errors because they did not reside in the programming but resulted instead from coding errors or inadequate editing procedures that affected the underlying data. It will be necessary to caution users about such facets of the usability of the data in the documentation. We also note that if the edit procedures were having this apparent failure to catch "outlier" codes, that there are likely other tabulations affected by failed edit procedures that could not be readily identified in our validation processing.

3. Errors resulting from weighting. Census uses different weights for persons and housing units to determine universe estimates. On occasion, this results in anomalies such as about 800 more households than occupied housing units in the District, since households are

assigned the person weight while the housing units they occupy are weighted differently. This apparent inconsistency will be confusing to most users of the data and will need to be dealt with in the documentation.

MESA noted that, even though the statistical procedures are correct, estimates for very sparsely populated districts (some as few as 50 children) will have very misleading results. In some instances there may only be two or three actual responses which are weighted to then provide estimates of the characteristics of 40-to-60 children.

The validation process also helped to identify problems with table titling and data presentation as conveyed by the SDDB software or database structure. The result of this has been to allow us to identify and correct problems that might have otherwise gone unseen until the SDDB had been released.

- o The Census Bureau supplied a few "replacement tapes" for the TIGER files known to be in error. MESA continued to process the TIGER files that have been received from Census.
- o The Census Bureau did deliver a "final" extended version CCD file for the U.S. This file also had erroneous data (unable to process missing data and no verification that the data aggregated to U.S. totals). Census will recreate the file. MESA awaits receipt of the "final extended" CCD file in August.
- o *Level 3 Software Plan.* A Level 3 Integrative Software Plan, a contract deliverable, was prepared during July. The Level 3 Integrative software is the mapping software to be used with the cartographic CD-ROMs.
- o *Demographic CD-ROM Content and Structural Plan.* A Demographic CD-ROM Content and Structural Plan, a contract deliverable, was prepared during July.
- o *Cartographic CD-ROM Content and Structural Plan.* A Cartographic CD-ROM Content and Structural Plan, a contract deliverable, was prepared during July.
- o *Training Plan.* A training plan, a contract deliverable element, was completed.
- o A training session program was conducted on July 28 in Washington. This program was well attended by state education agency staff through the nation.
- o *Training Manual.* A training manual, a contract deliverable, was prepared and distributed in draft form at the July 29th session.
- o A computer-based slide show was developed for the above mentioned training session. The slide show provides basic instruction on what the SDDB is and how it can be used. The slide show is distributed to interested parties on a diskette. The Census Bureau was given a copy for use by regional office staff.

- o *Support Training Materials.* The above mentioned slide show is a part of the support (related) training materials, a contract deliverable, and was completed in July.

August 1993

- o *Limited Demographic File.* A provisional (known to be erroneous due to Census errors) limited demographic file, a contract deliverable was supplied to NCES.
- o *File D Problems continue.* Numerous processing problems and errors were encountered with the Census supplied File D tapes. These are the data sets ("Version 6") that have been supplied after correcting for the errors identified in the validation processing. A summary of the types of problems follows.

Tape Numbering Errors/Incorrect Tapes. There are eight New York state tape reels. Each set of four of tapes is targeted to go to a separate CD-ROM. The first CD was created requiring the better portion of one person-day and one computer operations-day. The second set of four tapes was under processing on a second day when MESA learned that the eighth tape contained no data.

Since these data are in one large 8-reel file, there is no way to simply pick up a corrected eighth reel and add that data to the processing from the previous seven. All eight reels are being replaced. MESA is again at the loss and expense of a couple of days time.

MESA speculates that the source of this problem was operator error. The same problem has been found for Arizona and New Jersey. These are not programming problems -- the wrong tape is being mounted and/or labeled incorrectly.

More Errors Found in Development of "Limited Demographic File". As reported in the documentation for the dropout analyses file delivered yesterday, the state total records had a zero value for the 100-percent housing units field (H003 cell).

The H003 cell was added to some other cell in the record. This situation occurs with a random pattern. MESA cannot tell which cell will be affected. These problems affect only the state total records for record types 1 and 2.

MESA cannot simply reaggregate the data from the county records and produce the same data values as Census due to the distributional statistics (e.g., medians). Distributional statistics fields are not computed algorithmically by Census from the county median cell values so MESA could not replicate their processing.

MESA suggested having Census prepare a U.S. state file containing record types 1 and 2. While this will be the most expeditious way to get the data from Census, it should be noted that MESA will have to (1) write a special program to handle input of the exceptional records

and (2) devise a means of splicing in the "new" state records and removing the erroneous ones. While this might appear simple and not require much time in a mainframe environment, once the data are on CD's, complexity and time requirements grow dramatically.

MESA would also have to provide special handling for the dropout analyses file, since that processing cannot wait for the CD file augmentations.

MESA did not "see" this problem in the validation processing, since by our mutually devised plan, analysts were looking only one district - DC - and not state summary records. This raises the specter of other such possible errors yet to be uncovered.

Status of File D Tapes. MESA has now processed all states and D.C. File D data successfully (moving the files from tape to the provisional master CD-ROM). While MESA reports this in a summary manner, there were several problems where MESA needed to acquire replacement states tapes.

Digressing on the apparent nature of the unreadable File D tapes, close examination of a Minnesota reel (that was unreadable) revealed that a crimp in the tape had rendered the tape unreadable. When this occurs, all tapes for a state must be replaced rather than just one. The reason for this is that Census is supplying the state file (one file) on a spanned-reel set. Since the length of tape varies from reel to reel, each reel will almost never start and stop with the same physical record.

It should also be noted that the conversion of the files from tape onto the provisional CD's does not necessarily mean that all of the data are known to be "good." As MESA subsequently process the files on the provisional CD's for analytical purposes, errors may be found. For example, in July MESA found that a tape that had been labeled as Arizona actually contained data for New Jersey yet this file was transferred onto the provisional CD as MESA had assumed the tapes had been labeled correctly. MESA expects to be through with an initial analytical processing of the final files by mid-September.

- o TIGER Files. At this date MESA has not received all TIGER tapes - for any version of the files. The Census Bureau liaison reported to MESA that she thought MESA had the TIGER files for all states! It remains that some files identified as missing in MESA memos as far back as May are yet to be supplied.

MESA has received diskette files showing which files Census believes that MESA should have received. That information is still in the process of being examined.

Both according to the Census supplied inventory list and MESA records, no TIGER files have been shipped/received for DC, IL, NH, NJ, NM, NC, TX or VT.

Digression. The Census "TIGER-supplied" list appears internally inconsistent which is leading to confusion in determining the reliability of the list itself. Here is one example for the state of Alabama. On the tapes that MESA has, MESA could not locate all of the files for this state. Alabama county FIPS codes go through 133. The tapes that MESA has for AL, based on hard copy printouts showing DSN's with the reels (G04289, G04295, G04298 and G04304) have as the highest FIPS code provided to be 079 (located on tape reel G04304). This finding is consistent with the Census supplied diskette documentation file named "__AL.TXT." However, a different Census supplied diskette documentation file (DBF) shows that tape G04304 contains a totally different set of files starting with Alabama county 073 and proceeding through 133 (actually 133 is followed by 125 as the final county shown in that listing). We believe that the Census documentation file DBF is incorrect, relative to the tapes that MESA received. The short of this, aside from the confusion, is that MESA does not have all TIGER files for all counties in Alabama - many are missing (all above 079).

The foregoing situation of some (a few or many) county files not present for a state is repeated for several states. It remains that there has been no files provided based on our memos stating that certain county files are missing, dating as far back as May.

The previous TIGER file review only relates to the TIGER files that are known to be in error. Census apparently proposes to provide the NCES with only the TIGER files with the UA and related coding errors. MESA has never received any communication regarding plans (though various options have been discussed) for the manner in which Census proposes to provide NCES with the final corrected TIGER files. The Census liaison reported to MESA that those files are not yet available. At this time, it is unclear when they will be available. While MESA can meet NCES basic needs (demographic CD) with the version of the TIGER files that are in error, there should be a plan communicated by Census to NCES in writing as to how and when the corrected TIGER files will be provided (as the files will be used throughout the 1990's).

MESA proposed that Census provide the TIGER files on once-off CD-ROM's which would accelerate our elapsed processing time. As far as MESA knows, this is not considered as an active possibility by Census. However, apparently Census is generating once-off TIGER CD's for some internal purpose.

Because of the various uncertainties that surround the TIGER files, as set forth above, it is not clear when MESA will have all of these files.

- o The Census Bureau supplied MESA what is believed to be the final extended CCD file. This file has been supplied with documentation on August 26, 1994, and will be reconstructed for use in the SDDB demographic CD in September.
- o *Prototype Demographic CD-ROM.* A prototype demographic CD-ROM, a contract deliverable, was prepared and distributed for use at the OERI training sessions (see below).

- o *Training Plan.* A final training plan, a contract deliverable element, was completed.
- o Training sessions were conducted in Charleston, WV and Portland, OR for the OERI Labs and affiliated organizations.
- o *Training Manual.* Revised training materials, a contract deliverable, was prepared and distributed at the OERI sessions.

September 1993

- o Due to continuing problems with processing of TIGER files supplied by Census on tape, MESA requests that the Government supply the files on CD-ROM as the corrected files are now being distributed to the public in that form.
- o MESA receives additional TIGER replacement and original tape files.
- o MESA identifies problems with processing TIGER files in several states and requests replacement files.
- o MESA locates yet another File D error (Record Type 6 Error). In the process of performing final checks of the latest File D, MESA could not locate the total relevant children record (F) for the children's parents characteristics (record type 6) for any state. In consultation with Census Bureau staff, MESA confirmed that the record was never placed on the tapes that MESA was sent by Census.

An examination of the previous File D files (pre-August), that MESA had processed, revealed that the record 6-F did exist at that time. The record was a part of the validation processing that MESA performed. The record 6-F is absent only in the files received by MESA from Census after validation processing.

October 1993

- o MESA identifies more File D errors. MESA found additional errors requiring Census to replace File D tapes for three states. MESA also found that the state level summaries were incorrect, resulting in (1) Census re-supply the state level 2B records and (2) the requirement for MESA to splice these data into the master database. MESA awaits Census to supply the state 2B corrected data files and the New York City 2B record data.
- o MESA processed TIGER files for several states. Most of these processing applications involved states where MESA had [previously received the state files (e.g., Alabama) but had reported errors with the files. For two of these states, MESA again encountered problems with the replacement files.
- o All F-33 data has been received from Census and is believed to be complete.

- o All CCD data has been received from Census and is believed to be complete. Processing of the final files supplied by Census in September continued into October.
- o User software: Further development of the primary distributable SDDB software (demographic CD) was placed in a hold pending (1) completion of the major extract demographic CD-ROM and (2) receipt of final File D data from Census.
- o The composite software for the demographic CD is now largely considered complete. MESA cannot complete the files until the final demographic CD databases can be constructed from the provisional CD versions. This process requires development of index files and other tasks that require knowledge of exactly which final data are on which final CD's.
- o Progress was made on development of a demonstration version for the cartographic CD software.
- o Development of the Summary File Set I CD-ROM has been a major focal point for October. The Summary File Set I (SFS-I) is the extended version of the limited demographic file. SFS-I contains more than 200 specially developed extract files produced from the provisional File D CD's. The SFS-I CD also contains a special program enabling users of the CD to extract data for use. Development of the SFS-I required preparation of several previously unplanned programs to extract the files as required to meet NCES requirements.

NCES requested MESA to develop the SFS-I to expedite deliver of the basic demographic data in any form due to the many errors in the data supplied by Census slowing up the total completion of the preplanned SDDB demographic CD-ROMs.

Preparation of the SFS-I has been a large scale mini-version of the demographic CD-ROM. In October MESA developed an integrated user's guide and technical documentation. The printed table data dictionary is more than 130 pages detailing the various file's contents.

Development of the SFS-I has been the first occasion where MESA has received sufficiently correct data to aggregate to U.S. state totals (the Census Bureau is not constructing U.S. total data). This process has been completed for the total households records but awaits corrected Census data before completing this step for the total persons record. We believe that the record types 3-7 aggregates to U.S. totals completed for the SFS-I are also correct and final.

- o The SDDB Demographic CD system is substantially completed less the development of the revised File D data files and associated indexes. Having completed development of the provisional demographic CD's, MESA continues to expect that the entire U.S. will be contained on 40 CD's.
- o Documentation for the SDDB Demographic CD is as complete as possible pending inclusion of the final File D data and TIGER file boundary data.

- o Order forms and distribution plans for the SFS-I CD were developed and MESA planned to implement this process in November.

November 1993

- o File D Nears Completion. During November, Census supplied what appears to be the final set of File D data. MESA used the latest set of File D revisions to develop the Summary File Set - I extract data (this CD-ROM contains the "limited demographic file" as specified in the contract).
- o Summary File Set I CD-ROM. The Summary File Set I (SFS-I) has been completed as a provisional database on CD-ROM. Twenty copies were delivered to NCES on November 17. The SFS-I is the "limited demographic file" specified in the contract. The limited demographic file has grown in scope due to the delays encountered in Census delivery of the correct File D data and the urgent need within the Department for certain data expected to be available in the final SDDDB form by this time.
- o Some File D Errors Remain. Since Census did not provide U.S. total records, MESA constructed a portion of the U.S. total records during November as a part of developing the SFS-I database. This work will continue into December for record types not included in the SFS-I data. During the U.S. summary processing, MESA aggregated the state level records to U.S. level and verified the accuracy of the results, insofar as possible with respect to time and comparable tallies, and found that the tabulations are correct -- again, as far as the checking could be completed.

In the process of checking the totals for each state (verifying that the district records summed to the state record totals), MESA found that most states checks exactly. In several states, however, as noted in other memos, there were several minor discrepancies. For example, one district had two record type 1 data records in one state. MESA believes that the existing problems can be remedied by MESA without further re-supply of the data by Census (this would be the most cost-effective solution and generally no impair the quality of the database).

During December, a major task will be the re-assembly of the parts of files that Census has supplied during the data delivery process and re-supplied in piecemeal fashion to correct errors that MESA located.

- o Census Backup File D Data. Census has advised NCES and MESA that no copies of the special tabulation files will be maintained at Census any longer. This means that if anything happens to the files now in the possession of MESA, or if subsequent problems are identified, that the entire project could be in jeopardy. MESA recommends that the Census Bureau not delete all copies of such files (these files apparently include software and many intermediate files that would be nearly impossible to regenerate). For the scope and size of

this special tabulation, MESA believes that a responsibility lies with Census to retain these materials for at least six months after the delivery of the final data (May 1994).

- o TIGER Files. During November MESA processed TIGER files for several states. Much of the processing has involved development of boundary and overlay files for both the demographic and the cartographic CD-ROM deliverables.

Several missing or erroneous TIGER files have been reported in previous monthly reports or other memos for which Census has still not provided replacements. Two of the most significant problems are Texas and California. No version of any TIGER file has been supplied for most Texas counties. While the California TIGER files have been supplied, reel G04299 contains an unreadable section on the tape. As a result, MESA was unable to process California counties 083 and 091. Typical of the process of encountering errors in the tape processing, MESA is delayed to (1) encounter the error, (2) try to fix it and/or work around it, and then (3) wait for the replacement files to be supplied.

Regarding TIGER files to be re-supplied, it has been noted that "For two of these states, MESA again encountered problems with the replacement files (the primary problem is that the files to be corrected and supplied were incomplete)." It remains that no replacement files for the erroneous files have been supplied.

More generally, MESA again requests that the Census Bureau provide the TIGER files on CD-ROM as released. For example, MESA is aware that the TIGER files for California have been released by Census on CD-ROM. By the Census Bureau supplying these files on CD-ROM rather than tape, it will cost the Census Bureau less than supplying the erroneous two county file sets noted above for California and minimize the re-work on MESA's part of processing the tape version.

- o All F-33 data has been received from Census and is believed to be complete. Processing of these files continued in November as it will until these data have been integrated one-for-one with the File D data.
- o All CCD data has been received from Census and is believed to be complete. Processing of these files continued in November as it will until these data have been integrated one-for-one with the File D data.
- o SFS-I Extract Module. A data extraction program, actually a version of the Level 2 integrative software, was completed and delivered to NCES. This software is contained on the SFS-I CD-ROM.
- o Development of the Summary File Set I CD-ROM continued as a major focal point for November. Originally expected to be completed by mid-October, the CD was completed and delivered on November 17.

- o The Summary File Set I (SFS-I) is the extended version of the limited demographic file. SFS-I contains more than 200 specially developed extract files produced from the provisional File D CD's. The SFS-I CD also contains a special program enabling users of the CD to extract data for use. Development of the SFS-I required preparation of several previously unplanned programs to extract the files as required to meet NCES requirements.
- o The SFS-I integrated user's guide and technical documentation was completed and delivered to NCES on November 17. This documentation is an extension of the documentation for the more comprehensive SDDB System.
- o A training program was conducted by MESA for Federal agency and Congressional staff on November 30 in Washington. This training session had been delayed from the earlier scheduled date due to the delays in Census data delivery.
- o A second training program was conducted by MESA for NCES staff on November 30 in Washington. This training session had been delayed from the earlier scheduled date due to the delays in Census data delivery.
- o Planning discussions were carried on between NCES and MESA regarding prospective changes in the training schedule for the months ahead.
- o Order forms and distribution plans for the SFS-I CD were developed and MESA planned to implement this process in early December. This schedule is dependent on when MESA could complete "sizing" processing of the File D data and determine which areas will be on which CD's.

December 1993

- o File D Nears Completion. MESA has proceeded to process the File D data throughout December with no new errors noted. Much of the project effort in December has been placed on merger of the corrected files that were supplied by Census in mid-November. This process involves taking the "parts" of the corrected files and merging them together into the distributable database structure. This process has required the development of a new set of provisional master CD-ROMs. Approximately half of the states were completed in December with the remainder scheduled for completion in January. From the set of final provisional File D data, MESA will be able to proceed with development of the final distributable CD's.
- o Summary File Set I CD-ROM. The Summary File Set I (SFS-I), delivered by MESA in November, was further distributed by NCES. No errors or problems in using these files have been reported. MESA staff has started to provide assistance to users of the 1990 census special tabulation files (contained on the SFS-I CD), an activity which will progressively increase as more files are released.

- o U.S. Aggregates. Since Census did not provide U.S. total records, MESA must construct the U.S. aggregate records. This process was successfully completed in November for that portion of File D required for the SFS-I CD. To accelerate completion of the SFS-I CD, only those records requiring aggregation for that product were included in the aggregation process.

As the final File D provisional CD-ROM is developed, MESA is also developing the U.S. aggregates files. These files must all be developed and verified before the first SDDB CD-ROM can be released (since each SDDB volume must contain the same set of U.S. aggregate data).

- o TIGER Files. MESA has repeatedly reported errors associated with missing TIGER files supplied by Census. As a result of a meeting in early December with the Census Bureau, it is now believed that this problem has been resolved.

Census has agreed to supply all files on CD-ROM. TIGER files for several states were supplied by Census on CD-ROM to MESA. Census has stated that they expect to deliver TIGER files for all states on CD-ROM to MESA by mid-February.

Census is able to process the TIGER tape files and place them CD-ROM, whereas as MESA has not been able to, because they are able to identify tape file problems, solve those problems (having to do with the internal source files), and then build the corrected files on CD-ROM.

- o All F-33 data has been received from Census and is believed to be complete. Processing of these files continued in December as it will until these data have been integrated one-for-one with the File D data.
- o While all district level CCD data has been received from Census, MESA has found that there are no state or U.S. total records. As a result the scope of the geography would be incomparable with the 1990 census data and the financial data.

MESA advised NCES of this problem and requested the state and national summary data. Due to problems of "missing data," the data that were developed from the CCD school level file cannot be supplied for all fields at the state and national levels.

During December, MESA received a file from NCES containing a limited set of CCD items at the state and national level. MESA is now in the process of evaluating these data to determine the feasibility of integrating these data with the district level data. The CCD files must be reconstructed and new data fields estimated based on the state and national data supplied.

- o Materials were developed for a training program scheduled for GIS users of the mapping software in January to be held in Washington.

- o Planning discussions were carried on between NCES and MESA regarding prospective changes in the training schedule for the months ahead.
- o Order forms and distribution plans for the all SDDDB products were further developed. This process will be implemented during the first week of January.

January 1994

- o Final File D CD-ROM Remains Under Development. MESA has proceeded to process the File D data throughout January. Some new errors have been found as summarized below.

As in December, much of the project effort in January has been placed on merger of the corrected files that were supplied by Census between project start-up and November 1993. This process involves taking the "parts" of the corrected files and merging them together into the distributable database structure. This process has required the development of a new set of provisional master CD-ROMs.

MESA expected to complete all states during January but encountered some problems and expended relatively more effort on the level 3 integrative software and cartographic CD-ROM to meet deliverable and training requirements associated with the GIS training (also described in this report).

MESA has now completed preparation of the interim File D CD-ROM for 42 states and D.C. While only 8 states remain to be prepared, these are the larger states (multi-CD). Unless unforeseen errors are encountered MESA expects this phase to be completed in February.

New York Errors. In developing the interim File D CD-ROM two errors have been encountered with New York state. There may be other errors in the New York records that are now being examined.

New York Error 1. The population by race data for the pseudo districts 60??? have been found to erroneously contain zeros (data on the tapes supplied by Census). These areas are the secondary county equivalent districts for New York City PSD. Other parts of the data records for these areas may also be in error. We have not performed an study of the extent of the error. The suggested alternative procedure for dealing with this problem are to, since the data are for pseudo districts, (1) not release data for these areas at all or (2) release the data as they exist and annotate the tabulation problem in the documentation. We await direction as to the procedures for handling this problem.

New York Error 2. The revision data tapes supplied by Census are not all readable. To review, the File D 2B data record for New York City PSD was found to be in error. The Census Bureau responded by providing the corrected data for the record in November. Rather than supplying the corrected data records for NYC PSD, the entire set of tapes were re-run for the state (9 reels provided to MESA). In November, MESA processed only one

of those reels to obtain the data needed for the SFS-I CD for the NYC PSD. Thus the whole set of 9 reels was not processed until January 1994. In processing the most-recently-supplied NYS 9 reel set, an unresolvable read error was encountered on the second reel. At this time, at the objection of MESA, Census no longer has the ability to replicate nor re-tabulate the data. Our best final data file production option appears to be to splice in the known erroneous NYC PSD 2B record similar to the process used with SFS-I. This is not wholly satisfactory since MESA believes the programmed changes made to correct the previously erroneous NYC PSD 2B problem may have also corrected other district data errors which have not been uncovered.

No other state files are known to have errors. From the set of final provisional File D data, MESA will be able to proceed with development of the final distributable CD's.

- o Summary File Set I CD-ROM. The Summary File Set I (SFS-I), delivered by MESA in November, was further distributed by NCES. Two errata were issued and made a part of the User's Guide. No significant problems in using these files have been reported although some users did require further instruction than was contained in the manual. MESA staff has continued to provide assistance to users of the 1990 census special tabulation files (contained on the SFS-I CD), an activity which will progressively increase as more files are released.
- o U.S. Aggregates. It is emphasized that MESA must aggregate the state level File D data from the state records to U.S. totals before the first SDDB demographic CD can be fully assembled (as these data were not supplied by Census). The U.S. total records are required before the U.S. by district summary extract file can be prepared that is to be placed on each SDDB demographic CD-ROM. The process of completing preparation of the final provisional CD's as described above is an essential step in this process.
- o File A-File D Data. Users examining the SFS-I data have inquired as to the existence of the data for Chapter I funding distributions. These data are only in File A. MESA has raised the question with NCES of whether or not selected File A items should be included in the U.S.-wide district file to be contained on each SDDB demographic CD.
- o TIGER Files. The first TIGER files for several states were supplied by Census on CD-ROM to MESA. Census has stated that they expect to deliver TIGER files for all states on CD-ROM to MESA by mid-February.
- o F-33 Data. All F-33 data has been received from Census and is believed to be complete. Processing of these files continued in January as it will until these data have been integrated one-for-one with the File D data. See related note under "Code Correspondence Issue" below.
- o CCD Data. MESA was provided with no state or U.S. total records. As a result the scope of the geography would be incomparable with the 1990 census data and the financial data.

During December, MESA received a file from NCES containing a limited set of CCD items at the state and national level. MESA has now processed this file and has determined that many data fields which need to be estimated are of questionable value. MESA is now attempting to summarize these findings and propose how to best handle access to these data. MESA expects to resolve this problem into a working system component during February.

- o Code Correspondence Issue. In January the question was raised by others doing research for the Department of how the "universe" of districts covered by the Census Mapping Project related to the "universe" of districts covered by the Common Core of Data. That is, what is the existence/code correspondence between the Census Mapping Project (CMP) codes, the Common Core of Data (CCD) codes and School District Finances (F33) codes. These are all 5-digit codes which start with the 1989-90 CCD master set.

NCES requested that MESA develop a cross-code existence file. MESA prepared this file during January. Along with this file, MESA prepared documentation and commentary about the implications on the "non-alignment" of codes from the three sources.

MESA noted that many districts are referred to as "unnamed areas" in the SDDB demographic CD master list because the CMP code does not have a corresponding CCD code and thus no source name exists. As a result, with no name for a district, the data may be meaningless.

MESA noted that this cross-code data had been requested by MESA many months ago and that it was agreed by NCES that the information would be forthcoming. This expectation was confirmed by NCES staff who were responsible for obtaining the information.

MESA has taken the following actions to see that the relation between CCD and CMP codes is resolved. First, at the GIS training session all attendees were given a list of districts for which no names existed. They were asked to annotate this list with names and return it. Second, NCES was provided with the same list so that NCES could distribute it directly to the CCD coordinator. MESA requested that NCES fax this list to the NCES coordinator to accelerate completion.

The F33-CMP code have a higher non-match rate than the CCD-CMP codes. Also several months ago, MESA received information as to how F33 codes should be change to bring them into alignment with the CCD-CMP codes. Although code changes were made to many F33 district record codes, many remain unresolved. MESA has no means of annotating, documenting nor further correcting the code misalignment. The source of the coding errors lie with incomplete processing on the part of Census (Census uses another code and therefore considers a standard code not pertinent to their work).

For the SDDB CD's to be of optimal use, a fully accounted for code cross-reference list must exist as a core part of the SDDB demographic CD system. MESA is now developing a process that create the best means available to access data from all three sources.

- o A training program was conducted for GIS users of the IMAGE Level 3A/B software in late January in Washington. This training program required more specialized training materials and development time than previous training sessions. Attendees at the function included mapping specialists from most individual state departments of education across the U.S.

To implement the GIS training program, several cartographic prototype CD-ROM's were developed by MESA. The Cartographic CD-ROM were used on 17 workstations by the GIS session participants. Attendees received hands-on training that instructed them on:

- o understand operations of the system well enough to explain the operations and benefits to colleagues,
 - o take away instructional materials that enabled them to repeat the application steps independently, and
 - o learn about what boundary, overlay and data files exist on the CD-ROM for their use and how these resources can be used to meet school district mapping applications for orienteering and analysis.
- o The "Guide to Mapping Applications" was developed as a combined training and reference resource for users of the Geographic Information System software and the Cartographic CD-ROM files. Attendees at the GIS training session were provided instruction on how to use the Guide. Attendees then made use of the Guide in the training session by following step by step instructions for preparing maps and performing related cartographic applications.
 - o Order forms and distribution plans for the all SDDDB products were further developed. This process will be implemented during the first week of February.

February 1994

- o File D. MESA has now completed processing of the File D data as supplied by Census. MESA has prepared the interim File D CD-ROM for all states and D.C. Errors previously reported for the state of New York have been resolved through special programming steps.

The primary remaining task with regard to the File D data is to (1) complete development of the U.S. summary file and (2) develop the U.S. all district, all county and all state extract file.

- o Interim File D CD-ROM's. MESA has completed development of the interim File D CD-ROM's. A memo has been transmitted to NCES showing the proposed layout of all of the state files on CD-ROM's. The problems with the New York data are believed to have been corrected through special processing procedures implemented in February.

- o U.S. Summary Data. From the interim File D CD-ROM's MESA stripped off the state-level data records. Special software was developed in February to aggregate the state data records to the U.S. level.

An internal validation CD-ROM has been prepared containing the U.S. aggregate and state data. During March, staff analyzed the U.S. aggregate data checking to confirm, where possible, that the aggregates contained in the MESA generated U.S. aggregates file correspond exactly to the previously published Census tabulations at the national level.

In the final generation of the U.S. summary file, MESA will insert estimates of the median and per capita distributions derived from the interim validation CD-ROM. This process has involved (1) the extraction of interval oriented data from the interim U.S. aggregates file, (2) development of special software to estimate the distributional statistics and (3) the addition of logic to the U.S. aggregate generation program to insert these derived statistics into the U.S. summary file.

Completion of the U.S. summary file with distributional statistics is scheduled for March. At this time the U.S. summary file will be joined with the state level database. This step will complete development of the File D data files. Merger of the File D decennial data can then proceed with the F33 and CCD data.

- o U.S. All District File. During February, NCES and MESA completed specifications for the subject matter in the "top 100" file. The top 100 files include one file for each: all districts in the U.S., all counties in the U.S. and all states in the U.S. The subject matter and record layout in each file are the same. The items included in the file are approximately 100 items expected to be of most use to the SDDB users. Each of these three files will be included on each SDDB demographic CD-ROM.
- o File A-File D Data. A decision has been made by NCES not to include File A data on the final SDDB CD-ROM.
- o TIGER Files. Census agreed to supply all files on CD-ROM. Census stated plans to deliver TIGER files for all states on CD-ROM to MESA by mid-February. At this time, MESA has received TIGER files for less than half of the U.S. There has been no update from Census as to when the remainder of the TIGER files will be supplied.

MESA now has enough TIGER files to proceed with development of the individual SDDB CD-ROM. This is because MESA now has TIGER files for many states in their entirety and the full set is not required before some of the SDDB CD's can be released.

- o F-33 Data. All F-33 data has been received from Census and is considered complete. Processing of these files continued in February and are slated to be placed in final form during March.

- o CCD Data. MESA has implemented steps to place all CCD data in final form during March. Problems previously reported have been resolved.
- o Code Correspondence Issue. The matter of code correspondence reported as an issue in the previous report has now been resolved. This has been accomplished by obtaining a new 1989-90 CCD master file from NCES, building a composite master file by using Census, F-33 and CCD file and working with individual state census mapping representatives to resolve remaining questions.
- o Distribution and User Access. MESA and NCES staff met in February to develop plans for the mailer for use in distributing the SDDDB CD-ROM. Plans call for the development of the mailer and related materials during March so that the U.S. demographic CD can be distributed using these materials and procedures starting in mid-April.

March 1994

- o Census File D-New Errors Found. MESA has now completed processing of the File D data as supplied by Census. MESA has prepared the interim File D CD-ROM for all states and D.C. Errors previously reported for the state of New York have been resolved through special programming steps.

MESA found a new error in the file D data that appears to be systemic throughout the U.S. under certain conditions. The problem, as documented elsewhere in detail, is that the Census Bureau did not properly set the lower end grade cut-off properly when tabulating "relevant children" for a school district. The result of this is that, for unified districts, the total number of children in a school district can be less than the number of children in the corresponding county area--in record type 4. The treatment of data is different in record type 4 from treatment of the tabulations in the other record types. NCES has advised MESA to document this problem, treat as an known problem in the user notes, and proceed with development.

- o Remaining File D Tasks. The primary remaining task with regard to the File D data is to (1) complete development of the U.S. summary file and (2) develop the U.S. all district, all county and all state extract file.

Due to the continuing problems with the File D data, and to make it clear that problems in processing are the result of Census Bureau errors and not MESA's, MESA prepared a memo for NCES to establish acceptance of a certified set of File D data giving both aggregate subject matter data summaries and control counts for districts.

- o Final U.S. Summary File. The final form of the of the U.S. summary file was completed by MESA. This step completes development of all of the base data required from the File D data source.

- o U.S. All District File. Using results from the U.S. summary File D, the F33 file and the CCD file, MESA completed development of the U.S. all district file. This file will be further validated.
- o TIGER Files. As now reported several times, Census agreed to supply all files on CD-ROM. Census stated plans to deliver TIGER files for all states on CD-ROM to MESA by mid-February. At this time, MESA has received TIGER files for less than half of the U.S. There has been no update from Census as to when the remainder of the TIGER files will be supplied.

MESA now has enough TIGER files to proceed with development of the individual SDDB CD-ROM. This is because MESA now has TIGER files for many states in their entirety and the full set is not required before some of the SDDB CD's can be released.

- o Distribution and User Access. The plan for the release of the final demographic CD-ROMs was revised based on conversations between NCES and MESA. Based on this now expected final plan for the way the CD-ROMs and collective system are to be organized, NCES has started to develop an order form for the SDDB CDS.

April 1994

- o Death of Roger Herriot. The quite unexpected death of Roger Herriot in April impacts on (1) the leadership for the overall project and (2) increases demands on MESA for evaluation and quality control aspects of the project. Herriot's virtual weekly involvement with MESA staff on product issues at a variety of levels will now be assumed on a larger scale by the MESA staff.
- o TIGER Processing. Much of the staff time in April has been devoted to processing TIGER files which are starting to be received from the Census Bureau by MESA.

Still, Census stated plans to deliver TIGER files for all states on CD-ROM to MESA by mid-February. At this time, MESA has received TIGER files for less than half of the U.S. There has been no update from Census as to when the remainder of the TIGER files will be supplied.

MESA now has enough TIGER files to proceed with development of the individual SDDB CD-ROM. This is because MESA now has TIGER files for many states in their entirety and the full set is not required before some of the SDDB CD's can be released.

- o Census File D. NCES found no problems with the acceptance of the certification provided in March by MESA, summarizing the qualitative features of the File D data. The primary remaining task with regard to the File D data is to (1) complete development of the U.S. summary file and (2) develop the U.S. all district, all county and all state extract file.

- o File D Documentation Not Available. Despite repeated requests by MESA to the Census Bureau, the Census Bureau has supplied no documentation for the files developed as a part of the NCES special tabulation. The result of this is that (1) MESA will need to perform additional unexpected tasks which will require more time, (2) the overall data release time schedule will be affected some and (3) MESA cannot know all of the correct material to put into the documentation since this information remains internal (and likely written up by no one) at the Census Bureau.
- o Testing on the first CD-ROM to be released. MESA and NCES staff collectively and independently tested the first U.S. by State CD-ROM. A final version of this CD will be provided to NCES next month to approve.

May 1994

- o Census File D. For then first time, MESA reports that File D is believed to be complete--almost 18 months since the start of the contract.
- o TIGER Files. Census stated plans to deliver TIGER files for all states on CD-ROM to MESA by mid-February. At this time, MESA has received TIGER files for less than half of the U.S. There has been no update from Census as to when the remainder of the TIGER files will be supplied.

MESA now has enough TIGER files to continue development of the individual SDDB CD-ROM. This is because MESA now has TIGER files for many states in their entirety and the full set is not required before some of the SDDB CD's can be released.

- o Demographic CD-ROMs are now being released.

The SDDB-00, U.S. by State, demographic CD-ROM was mastered and replicated.

The SDDB-07, Iowa by District, demographic CD-ROM was completed.

- o Cartographic CD-ROMs are in production, as time permits consistent with completing the demographic CD-ROMs as a priority.
- o The first CD-ROM products of the project were distributed this month. The U.S. by State CD-ROM was distributed to the NCES complimentary list.
- o The address list for persons to receive SDDB CD-ROM on a complimentary basis was received from NCES. This list was automated into a basic record-keeping system. The resulting CD-ROM sales/distribution file will be further developed into more of a system structure next month.

June 1994

- o TIGER Files. In December 1993, Census agreed to supply all files on CD-ROM to deliver TIGER files for all states to MESA by February 1994. Several complete states remain missing despite our repeated requests to Census to provide the files.

During this month Census supplied TIGER files not previously available for Wyoming, North Carolina (part 2).

The TIGER files for Texas (4 CD-ROM) remains totally unavailable as do North Dakota, Utah and Oklahoma.

Processing problems with TIGER files consumed more than half of MESA staff time this month. It had been expected that TIGER processing would now be largely completed (for the demographic CD-ROMs) and require approximately twenty percent of our staff effort.

The reason for the additional time requirement is that problems with closure of boundary files require the development of new software on MESA's part and much more "manual" processing time. This situation often requires MESA staff to spend many hours examining boundary files in just one county of one state.

- o SDDDB User Software. Modifications were made to various parts of the software based on feedback from users and NCES. More in-depth modifications made be made next month but await further exposure in the user community.
- o Cartographic CD-ROM files are in production, as time permits consistent with completing the demographic CD-ROMs as a priority.
- o All Demographic CD-ROMs have an electronic reference manual and help system that is located on the CD-ROM. The desire of NCES has been to release the CD-ROMs with minimal paper documentation which MESA has done. MESA has encountered requests for many users, that seems well founded, to also have printed documentation. While the plan was that users interested in printed documentation could print it from the CD-ROM, and they can, it turns out that this process produces what the user perceives as unwieldy text, and perhaps more importantly, no graphics (such as screen examples) can be included.

As a result of this user feedback, MESA has started development of an optional/supplemental reference manual covering much of the same subjects as contained in the electronic version but organized in an easy-to-read format with graphics. The contract actually calls for the development of this product, though until now it has been planned as more electronic than on paper; there is no change in product specifications nor scope of work, only the form and mix of electronic versus paper material.

- o The following initial master CD-ROMs were completed this month:

SDDDB-07 Idaho and Wyoming
SDDDB-09 Kansas
SDDDB-30 Kentucky and West Virginia
SDDDB-32 Delaware, DC, Maryland and Virginia

Several other states may be possible to add to this list but are presently being delayed due to TIGER file processing issues.

- o The first set of paid orders was added to the automated distribution list. The U.S. by State CD was sent to those ordering that CD of the full U.S. set.
- o The SDDDB sales/distribution file was improved so that once an order is placed for CDS presently unavailable, software can keep track of this. Improvements will also be made to this software next month providing better reporting and updating features.
- o Despite the widespread testing on the U.S. by State CD-ROM, some users encountered a problem with the install procedure. In some instances the CD-ROM could not be installed. This problem was investigated and found to be the result of the way that some MS-DOS command interpreter process commands within the SDDDB install procedures.

Users who have already received the U.S. by State CD-ROM and report this problem, are being sent a diskette with revised install software on the diskette. There are no known problems with the statistical data files on the CDS. The installation diskette will need to accompany future orders of this version U.S. by State CD. MESA anticipates the need to develop a "final" U.S. by State CD, which will likely be the most widely used title, later this year after all know problems are fixed.

- o MESA staff found a problem with the way an "at-risk" ratio is being computed on Profile 001 in certain circumstances (on the U.S. by State CD-ROM). Since this is the most widely used profile, the software needed to be modified to present the data correctly. This has been completed. It will be possible to remedy this problem with the supplementary install diskette also.
- o MESA staff located another error in processing one of the files where in certain circumstances data for the last county in Wyoming would not be retrieved properly in mapping applications. This problem has now been resolved and can also be remedied with the supplementary diskette.
- o Due to the foregoing experiences, MESA plans to defer CD-ROM distribution other than the U.S. by State title, until mid-July, to insure that MESA knows pretty much everything that needs to be on the supplemental floppy disks. MESA expects that there will need to be versions of the update diskette as potential future errors become known (perhaps there will be none). This time period will give us a better period, though minimal, for user feedback and this possibly avoid (1) unsatisfied users or (2) increase of re-work and expense.

July 1994

- o TIGER Files. In December 1993, Census agreed to supply all files on CD-ROM to deliver TIGER files for all states to MESA by February 1994. It is noted that Census had originally agreed to supply these files by February 1993, and that all original plans for completion of this project were based on that schedule.

On July 28, 1994, the Census Bureau supplied the remainder of the TIGER files (files for all of six states, including Texas) to MESA. MESA immediately processed the files for Utah. Processing went smoothly and the Utah by district boundary files are now completed. North Dakota files were then processed and problems similar to those encountered in other states (boundary closures) were found with North Dakota. North Dakota processing will require additional time and processing, results being reported next month. MESA has not yet had an opportunity to examine the July 28 delivery of TIGER files for Texas and Oklahoma.

Processing problems with TIGER files consumed approximately one-third of MESA staff time this month. A continuation of difficulties reported by MESA previously, the reason for the additional time requirement is that problems with closure of boundary files require the development of new software on MESA's part and much more "manual" processing time.

MESA believes that all boundary files can be completed in a fully usable form to meet project goals but will continue to require more time and further delay on completing the final products.

- o Financial (F-33) Data. A user of the F-33 data from the SDDDB U.S. by State CD-ROM distributed in June reported that data Grand Rapids, MI to be in error. The total number of students was found to differ widely from that reported by Census estimates or CCD.

The more significant difference was that total revenues differed from total expenditures by \$40 million. This unexplainable difference was reported to Census Governments Division for their followup. Census reported to MESA that our profiles were correct and that the data problem is in the Census supplied files.

Census believes this problem is confined to Michigan districts, but has not identified the scope of possible further errors nor provided for any correction for the \$40 million differential. At this point, all that MESA can do is to document the problem as a known error in the Census files.

In examining the foregoing problem with Michigan data, MESA also found that a formula in the SDDDB software was not correctly computing certain revenue and expenditure fields. MESA software used data fields without clear definition for some data fields in the Census-supplied files. The result of this incorrect processing was to display incorrect data summaries. MESA has now revised the software and made this available to all users (see section of software below).

On the foregoing matter, it is noted that MESA had distributed sample profiles to both NCES and Census requesting validation as to appearance and quality of the data before finalizing the first CD. There was no identification of the processing error and the CD was finalized and released. It is further noted that the more recent versions of the displays have been provided to both NCES (Bill Fowler) and Census (Larry MacDonald) once again (mid-July, 1994) who confirm that the data displays are correct.

The impact of the F-33 related developments are that (1) MESA expended additional unforeseen effort in July to correct previously unknown problems, (2) the overall schedule in completion has been expanded some, (3) there is now a pressing need to re-master the U.S. by State demographic CD to correct these problems at a later date and (4) the release of the State by district CDS has been collectively delayed until mid-August to insure that these CDS have the correct files (see review of product release below).

- o Common Core of Data (CCD) Data. As a result of the errors described in the previous section, MESA further examined portions of the CCD file supplied by Census. It was found that the total population field that had been reported for states and the U.S. was too large--district level data had been directly summed by Census producing state and U.S. values that were incorrect. The U.S. total population number was being reported at approximately 262 million.

Since the state and U.S. figures appear prominently on the profiles using these data, both the software and files containing these data had to be revised in July. Updates to the files have now been made available to users of the CDS (see review below).

The impact of the CCD developments are that (1) MESA expended additional unforeseen effort in July to correct previously unknown problems, (2) the overall schedule in completion has been expanded some, (3) there is further pressing need to re-master the U.S. by State demographic CD to correct these problems at a later date and (4) the release of the State by district CDS has been collectively delayed until mid-August to insure that these CDS have the correct files (see review of product release below).

- o SDDB Demographic User Software. Further modifications were made to various parts of the software based on feedback from users and NCES.
 - Due to the problems noted earlier, modifications were made to the SDDB software that displays financial data in profiles (four independent programs).
 - Some users reported that the system could not be fully installed on their computers. A check by MESA revealed that some computers could not use the DOS command interpreter software which had been developed by MESA. A more generalized version of the install procedure was developed by MESA in July correcting for this situation.

- An upgrade program was developed by MESA to update existing user files for data errors reviewed in section 1.
- Software updates were provided to existing users (see distribution below).
- o The hard copy version of the SDDDB reference manual referenced in last month's report was completed in early July. The manual was distributed to users who have found it generally very useful. Based on the positive user reaction, MESA has started to update and extend the manual.
- o Demographic CDS Developed. The following initial master CD-ROMs were completed in June that will be released when the complete set is finished (in hold pending finalization of corrections noted elsewhere):
 - SDDDB-07 Idaho and Wyoming
 - SDDDB-09 Kansas
 - SDDDB-30 Kentucky and West Virginia
 - SDDDB-32 Delaware, DC, Maryland and Virginia
- o New paid orders were added to the automated distribution list. The U.S. by State CD was sent to those ordering that CD and/or the full U.S. set.
- o The SDDDB sales/distribution file was improved so that once an order is placed for CDS presently unavailable, software can keep track of this as well as reporting and updating.
- o The problem with the install procedure reported last month in this section has been successfully resolved and all users are now believed to be able to install and use the system as expected.
- o The problem reported in June with the profiles showing at-risk data has been successfully resolved.
- o All known problems within the system have corrected as of the end of July. Still, resolution of many of these problems (correcting both data and software, required the entire month of July. The actual release of the replicated CDS (listed above) will not re-start until mid-August.

August 1994

- o TIGER Files. MESA continued to have difficulty with processing of the TIGER files for many states. In summary, the boundaries for the school district polygons "close" with erratic quality. MESA has created specialized software to force closure for school district polygons which do not process in a normal manner. Despite the additional software, there is still an

inordinate, unanticipated amount of staff time required to develop school district boundary files that are useful for mapping purposes.

The most problematic states remain for processing at the end of the set of all states. Processing TIGER files accounted for a large part of staff time this month.

- o Financial (F-33) Data. Problems noted last month with the F-33 data were not resolved by the Census Bureau. The Census Bureau observed that the data provided by the SDDB profiles convey the same data as contained in the Census master files. There is no further explanation from Census as to explanation for the large differences in revenues and expenditures for some districts which should be closer to equal according to local sources. This situation has been the source of many user phone calls and unanticipated user support time to explain data errors that should be explained by the Census Bureau.
- o Demographic CD Software. Further modifications were made to various parts of the software based on feedback from users and NCES. Similar changes are expected to be made as a part of the normal course for development of a system of this magnitude. Software updates will be provided to users.

Revisions to the install procedures reported last month have apparently resolved all problems reported concerning installation.

Software updates were provided to existing users (see distribution below).

- o Demographic CD-ROMs are now widely in use and all parts of the system are working properly. Production work continues on the remaining demographic CD-ROMs.

Production activity with the cartographic CD-ROMs continues with these CDS now scheduled for completion in November. Delays by the Government (Census Bureau) in providing the final TIGER files to MESA only in July makes it impossible to complete this processing until November.

The hard copy version of the SDDB reference manual referenced in last month's report was completed in early July. The manual was distributed to users who have found it generally very useful. Based on the positive user reaction, MESA has continued to update and extend the manual. A new version is planned for September.

- o Demographic CDS Developed. The following initial master CD-ROMs were completed in August:

SDDB-08 AZ-NM-NV-UT
 SDDB-10 MO
 SDDB-21 OH-1
 SDDB-22 OH-2

SDDB-23 TX-1
 SDDB-24 TX-2
 SDDB-25 TX-3
 SDDB-27 OK
 SDDB-35 PA-1
 SDDB-36 PA-2

- o New paid orders were added to the automated distribution list. The U.S. by State CD was sent to those ordering that CD and/or the full U.S. set.
- o All known problems within the system have corrected as of the end of August.

September 1994

- o TIGER Files. MESA continued to have difficulty with processing of the TIGER files for many states. The reasons are the same as previously reported: the boundaries for the school district polygons "close" with unpredictable quality. MESA has continued to enhance and work with specialized software, which MESA has had to develop, to force closure for school district polygons which do not process in a normal manner. It remains, that despite the additional software, there is still an inordinate, unanticipated amount of staff time required to develop school district boundary files that are useful for mapping purposes.
- o SDDB User Software. As with last month, further modifications were made to various parts of the software based on feedback from users and NCES. Similar changes are expected to be made as a part of the normal course for development of a system of this magnitude. Software updates will be provided to users.

Revisions to the install procedures reported last month have apparently resolved all problems reported concerning installation.

Software updates were provided to existing users (see distribution below).

- o Documentation. A revised hard copy version of the SDDB reference manual has been released and is now being supplied to users.
- o Demographic CDS Developed. The following CD-ROMs were completed in September:

SDDB-02 CA-1	SDDB-29 TN
SDDB-03 CA-2	SDDB-33 GA
SDDB-06 CO-MT	SDDB-34 NC-SC
SDDB-13 ND-SD	SDDB-39 NJ
SDDB-28 LA-MS	

- o New paid orders were added to the automated distribution list. The U.S. by State CD was sent to those ordering that CD and/or the full U.S. set.
- o All known problems within the system have corrected as of the end of September.
- o Delays in Being Able to Provide Technical Assistance. It had been anticipated that recipients of the CDS would be provided user support increasingly during the latter months of the project. Instead, due mainly to lateness and qualitative problems with the products supplied by the Census Bureau, more time has been expended on the matter of producing the boundary files and less on the matter of user support. Moreover, users have not had opportunity to receive the CDS as soon as anticipated and thus there has been little opportunity to obtain user feedback and provide user support.

There will be an increased demand for technical assistance and support during December 1994 and January 1995, and possibly February. This is partly because OERI regional educational research laboratories will be conducting training on the use of the CD-ROMS in December and January. Also, general user community exposure to the newest products, unavailable until the end of November, may result in identification of software, data file or documentation issues that MESA can provide remedies to expediently.

Before the encounter of the TIGER tape problems, it had been expected that the CD-ROM products would have been in users hands some months earlier. MESA has repeated had to operate under uncertain and adverse situations regarding access to and use of demographic and geographic files supplied by the Government (Census Bureau) throughout the tenure of this project.

Without adequate user support, users will likely become frustrated being unable to get questions answered which may occur after the presently scheduled end of contract. As a result, MESA plans to submit a request to contracts by mid-October for a no-cost extension to the contract for a period of three months. During this period, MESA would continue to receive user inquiries and provide assistance as required and MESA resources permit but no fees would be charged to the contract for any extended service during that period. This effort is in the best interests of MESA as well as the Government in order to make this collective project as successful as possible.

October 1994

- o Demographic CD-ROMs Need to be Remastered. The data files supplied by the Census Bureau and software were considered final, and the remainder of the project appeared to be concerned only with completion of the CD-ROM products, provision of support services and completion of documentation and reports. However, there MESA experienced follow-on repercussions resulting from modifications made to the data files while correcting for Census Bureau data errors. Demographic CDS developed prior to October would have to be

redeveloped due to an error which had remained in key files resulting from uncorrected errors in files supplied by Census.

Demographic CD-ROMs are remastered, a result of correction of data files (erroneous Census Bureau supplied data) and revisions to software (to correct for computation corrections identified in the way certain Census data were summarized).

Changes, resulting from the above actions, to the basic files that go onto each CD-ROM has resulted such file growth that some states slated to be contained on one CD-ROM now have to be remastered onto two CD-ROMs. This situation affects the states of New Jersey, Missouri and Oklahoma. MESA plans to restructure the files for these states so that the data are split onto two CD-ROMs. The structure of the files for these states will then be similar to other states that are split such as New York and Illinois.

This unexpected, additional, processing results in the production of the CD-ROMs being delayed for final distribution and an unanticipated increase in costs. All Demographic CD-ROMs are now expected to be available for shipping in early December. The impact of this addition time and cost, temporarily being absorbed by MESA, has slowed production of a few other Demographic CD-ROMs which will also be released in December.

November 1994

- o Demographic CD-ROMs. Demographic CD-ROMs continue to be remastered, a result of correction of data files (erroneous Census Bureau supplied data) and revisions to software (to correct for computation corrections identified in the way certain Census data were summarized).

The Demographic CD-ROMs are now being shipped to purchasers. Allowing time for replication processing and distribution, it is expected that the remainder of the demographic CDS will shipping by the first week of January 1995.

- o Cartographic CD-ROMs. File development continued for the Cartographic CD-ROMs. Focus was placed on development of the county by tract boundary files, tract by block group boundary files and census tract by street overlay files. Processing takes place on a state-by-state, county-by-county basis, requiring continuous operation of PCS and staff attention to a variety of possible errors in development.

December 1994

- o Demographic CD-ROMs. Demographic CD-ROM mastering has now been completed for all titles. The CD's are now being replicated. Demographic CD-ROMs are now being routinely shipped to purchasers.

- o Cartographic CD-ROMs. File development continued for the Cartographic CD-ROMs. Focus was placed on development of the county by tract boundary files, tract by block group boundary files and census tract by street overlay files. Processing takes place on a state-by-state, county-by-county basis, requiring continuous operation of PCS and staff attention to a variety of possible errors in development.

January 1995

- o Demographic CD-ROMs in Wide Use. All 44 demographic CD-ROMs have been mastered and replicated. 22,000 CD-ROMs have been produced. Dissemination and user support for these CD-ROMs was the major activity during January 1995.

User documentation for these CD-ROMs is considered complete. After some further user experience, it will probably be appropriate to provide existing and new users with an update to this version of the manual.

The demographic CD production and distribution has proceeded successfully. The dissemination processing and user support has been more substantial than anticipated during January. The dissemination and support requirement has produced a delay in completing the cartographic CDS. MESA proposes to extend the contract completion date.

- o Cartographic CD-ROMs. File development continued for the Cartographic CD-ROMs. Focus was placed on development of the county by tract boundary files, tract by block group boundary files and census tract by street overlay files. Processing takes place on a state-by-state, county-by-county basis, requiring continuous operation of PCS and staff attention to a variety of possible errors in development.
- o User Support. MESA has experienced a higher number and duration of telephone calls and user inquires than expected. MESA now averages 15 telephone calls per day with an average duration of 10 minutes telephone time each. On average, each telephone call requires an additional 10 minutes of staff time after the phone conversation--some callers require no further followup, others require an hour or more. Support for these calls is important to make the data useful.

In addition to the telephone inquires MESA receives inquires by letter and fax. MESA now averages average 4 faxes a day regarding SDDDB demographic CD-ROM use. The written materials tend to be more technical or supplement a telephone call.

User support fits into several broad categories:

1. Technical--These support requirements deal with issues such as basic installation of the system, matters of how to switch between CDS, how to operate the system in a network environment, etc.

2. Applications--Application type support usually follows the technical matters. This type of support involves matters such as:

- o ways to approach selections efficiency (complexity of large number of CDS)
- o how to use with other software
- o how to get updates
- o how to compare the 1990 census data with 1980 census data
- o how to prepare alternative types of profiles
- o how to get a variety of printers to work

3. Concepts

- o database structure
- o definitions
- o census/demographic concepts such as 100 percent versus sample estimates and their use/
- o differences in children counts data between the three primary database sources
- o adequacy of measures
- o data reliability

The availability of the support essential to facilitate effective use of the products. The 1980 census school district data was distributed without user support and the result was that the data were not used.

February 1995

- o Cartographic CD-ROMs. File development continued for the Cartographic CD-ROMs. Focus was placed on development of the county by tract boundary files, tract by block group boundary files and census tract by street overlay files. Processing takes place on a state-by-state, county-by-county basis, requiring continuous operation of PCS and staff attention to a variety of possible errors in development. Most boundary and overlay files now completed, MESA continues with development of school district boundary files.
- o File D Tapes Provided to National Archives. At the request of NCES, MESA provided the original, Census-supplied, File D tapes to National Archives. While it was the intent that National Archives would document and maintain an archival version of the tapes, it appears that the tapes now sit dormant at National Archives.
- o User Support. MESA continued to provide user support via telephone, fax and writing.

March 1995

- o Cartographic CD-ROMs.
 - MESA continues development of school district boundary files.

- MESA develops IMAGE System version 3S ("integrative software")

- o User Support. MESA continues to provide user support via telephone, fax and writing.

April 1995

- o Cartographic CD-ROMs.

- MESA continues development of school district boundary files.

- MESA develops IMAGE System version 3S ("integrative software")

- o User Support. MESA continues to provide user support via telephone, fax and writing.

May 1995

- o Cartographic CD-ROMs.

- MESA continues development of school district boundary files.

- MESA develops IMAGE System version 3S ("integrative software")

- o User Support. MESA continues to provide user support via telephone, fax and writing.

June 1995

- o Cartographic CD-ROMs.

- MESA continues development of school district boundary files.

- MESA continues development of IMAGE System version 3S ("integrative software")

- o User Support. MESA continues to provide user support via telephone, fax and writing.

July 1995

- o Cartographic CD-ROMs.

- MESA continues development of school district boundary files.

- MESA continues development of IMAGE System version 3S ("integrative software")

- MESA develops documentation for using IMAGE System with cartographic files.

- Testing takes place with Cartographic CD-ROMs and integrative software.

- o User Support. MESA continues to provide user support via telephone, fax and writing.

August 1995

- o Cartographic CD-ROMs.

- MESA continues development of school district boundary files.

- MESA continues development of IMAGE System version 3S ("integrative software")

- MESA develops documentation for using IMAGE System with cartographic files.

- MESA starts development of final cartographic CD-ROMs

- o MESA receives first set of final files for School District Analysis Book (from Synectics, developed by Synectics)

- o MESA prepares prototype CD-ROM containing SDAB for testing.

- o User Support. MESA continues to provide user support via telephone, fax and writing.

September 1995

- o Cartographic CD-ROMs.

- MESA continues development of school district boundary files. Delay encountered in developing final district boundary files as a scratch on Wisconsin Census TIGER CD-ROM requires replacement by Census.

- MESA continues development of IMAGE System version 3S ("integrative software").

- MESA develops documentation for using IMAGE System with cartographic files.

- MESA completes development of file structures for most cartographic CD-ROMs.

- o Errors found with School District Analysis Book (SDAB) prototype requires that Synectics to revise certain files and re-supply to MESA.

- o MESA converts SDAB files for use with CD-ROM.

- o MESA develops documentation files for SDAB and places on master CD-ROM.

- o MESA completes SDAB master CD-ROM, replicates 850 copies and provides to NCES.

October 1995

- o An error is found with SDAB cartographic CD-ROM. The CD-ROM master is re-developed by MESA and re-replicated. Only a few copies had been distributed to users which were recalled. MESA packages SDAB CDS for delivery to NCES and distribution to users.
- o Cartographic CD-ROMs.
 - MESA continues development of school district boundary files. Most school district boundary files now complete.
 - MESA completes IMAGE System version 3S ("integrative software") testing and prepares for distribution (onto diskette).
 - MESA develops export utility program for users of cartographic CD-ROM who do not use the IMAGE System.
 - MESA completes documentation for using IMAGE System with cartographic files.
 - MESA completes development of file structures for most cartographic CD-ROMs.
 - During development and testing of cartographic CD-ROMs MESA finds directory structural problems that may prevent some users from accessing the files (due to file/directory structures).
- o User Support. MESA continues to provide user support via telephone, fax and writing.

November 1995

- o Cartographic CD-ROMs.
 - MESA develops final masters for all cartographic CD-ROMs.
 - MESA replicates 500 copies of each of 7 cartographic CD-ROM.
 - MESA completes reference manual for users of cartographic CD-ROM who do not use the IMAGE System.
 - MESA organizes packages for distribution of cartographic CD-ROMs for purchasers and NCES distribution list.
- o User Support. MESA continues to provide user support via telephone, fax and writing.

- o Final Report. MESA prepares this final report.

Listing of NCES Working Papers to Date

<u>Number</u>	<u>Title</u>	<u>Contact</u>
94-01	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94-02	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94-03	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94-04	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94-05	Cost-of-Education Differentials Across the States	William Fowler
94-06	Six Papers on Teachers from the 1990-91 SASS and Other Related Surveys	Dan Kasprzyk
94-07	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
95-01	Schools and Staffing Survey: 1994 papers presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95-02	QED Estimates of the 1990-91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95-03	Schools and Staffing Survey: 1990-91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk

Listing of NCES Working Papers to Date (Continued)

<u>Number</u>	<u>Title</u>	<u>Contact</u>
95-04	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95-05	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings
95-06	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95-07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95-08	CCD Adjustments to the 1990-91 SASS: A Comparison of Estimates	Dan Kasprzyk
95-09	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95-10	The Results of the 1991-92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95-11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95-12	Rural Education Data User's Guide	Samuel Peng

Listing of NCES Working Papers to Date (Continued)

<u>Number</u>	<u>Title</u>	<u>Contact</u>
95-13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
95-14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95-15	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95-16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95-17	Estimates of Expenditures for Private K-12 Schools	Steve Broughman
95-18	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96-01	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk
96-02	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96-03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
96-04	Census Mapping Project/School District Data Book	Tai Phan



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)



NOTICE

REPRODUCTION BASIS

This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").