Census 2000 School District Special Tabulation: Effective Access and Use

Warren G. Glimpse and Catherine D. Wightman

Presented at the

American Educational Research Association
2004 Annual Meeting

San Diego, CA
April 13, 2004

Proximity ... The mission of Proximity is to design, develop, implement, facilitate access to, and deploy component-based information resources that empower education leadership, researchers, and stakeholders to improve the state of education and be better able to analyze and assess children's demographics and their living environments.

Proximity areas of specialization include:

- providing guidance in development of decision making resources,
- developing demographic estimates and projections,
- conducting demographic studies,
- providing comprehensive mapping and GIS solutions,
- conducting training and workshops, and
- developing custom software and data analysis tools.

The Proximity Modeler software and database enables districts to develop and revise their own demographic estimates and projections, update estimates and projections as new/updated historical data become available, apply alternative assumptions and produce estimates and projections under different scenarios, and perform what-if, impact type analyses.

This document has been prepared by Warren Glimpse and Cathy Wightman.

Contact Information:

Warren G. Glimpse
Proximity
PO Box 7511
6121 Lincolnia Road
Alexandria, VA 22307-7511
703-721-0826
wglimpse@proximityone.com

Catherine D. Wightman
15711 Greencourt Drive
Houston, TX 77062
281-286-3432
cwightman@swbell.net

Abstract

The Census 2000 School District Special Tabulation contains the most extensive and current set of data on children's demographics and their living environment summarized by school district, county, and state. Closely paralleling the 1990 Census School District Special Tabulation, these data provide insights into the state of education and are essential to analyze long-term trends on issues pertaining to children and PK-12 education.

This paper reviews the structure, means of access, and ways to effectively use the Census 2000 School District Special Tabulation (SDST) completed in August 2003. The SDST is the most comprehensive source of data on children's demographics and their living environment by school district. It provides unique data for the analysis of the population enrolled in public and private schools and not enrolled.

One focus of the paper is on methods of integrating data from the SDST with other data resources (school, attendance area, administrative, achievement, and other geographic data). Another focus of the paper is on the use of GIS tools and geospatial analysis of SDST data integrated with school (building), attendance areas, neighborhoods, and street segment level geography.

Warren Glimpse was developer of the 1990 Census SDST and has worked with wide ranging users in academia, research organizations, and education agencies in integrating these data with other data. Glimpse has been a principal architect of the Census 2000 SDST and was developer of the National Center for Education Statistics Web site providing access to the data. Perspectives on longitudinal uses of the SDST are also covered in the paper.

Contents

1 Introduction 5
2 Organization of the SDST and Data Access and Use 6
2.1 Overview 6
2.2 Importance: Enabling Abilities, Benefits, and Application Examples 7
2.3 Geography 8
2.4 Subject Matter 8
2.4.1 Record Type by Enrollment Type Iteration 9
2.4.2 Subject Matter Tables and Items 10
2.5 Suppression and Disclosure Avoidance 10
2.6 Data Access and Use 10
2.6.1 NCES Facilities for Access 10
2.6.2 Proximity Data Access & Analysis Resources 13
3 Effective SDST Data Access and Use & Decision Making 14
3.1 Tools, Options and Procedures 14
3.2 Data Integration 14
3.2.1 Federal Statistical Data Integration 14
3.2.2 Federal and non-Federal Data Integration 15
3.3 Advancing the Role of Visual and Trend Analysis in Education Decision Making 16

1 Introduction

As of Census 2000, children make up more than 53 million, or 24%, of the U.S. population. What do we know about this population for smaller area geography such as school districts? At the individual school district level, what is known about the characteristics of children enrolled in public school as compared to those enrolled in private school? How many school age children are not enrolled in school by school district? What are the characteristics of the "school district community"? How have these attributes of school districts changed between 1990 and 2000? What can we really say about the absolute and relative (district to district) state of education without access to data providing answers to these questions?

While there are extensive data from wide-ranging sources about children, only the School District Special Tabulation (SDST) provides a comprehensive set of "Census 2000 Summary File 3" like demographic data for children tabulated by school district, county and state. The SDST data provide the baseline and benchmark data for analysis of children's characteristics that are unique and of continuing use throughout the intercensal period. The SDST data uniquely empower us to answer questions such as those posed above.

The Census 2000 School District Special Tabulation (SDST) is a special tabulation of decennial census basic record data for school district boundaries. The Census 2000 SDST closely follows the scope of subject matter and types of geography found in the 1990 Census SDST. The SDST was developed to meet data and geographic specifications of the National Center for Education Statistics (NCES) by the Census Bureau under the sponsorship of the U.S. Department of Education.

The Census 2000 SDST became available on a flow basis starting in March 2003 and was mostly complete in the fall of 2003. All of the data will be accessible before mid-2004. The 1990 Census dates of data availability paralleled those of Census 2000.

As of Census 2000 14,286 "regular" school districts existed for which complete count data were tabulated. While there are no Census Bureau Census 2000 data summarized at the school district level, the Summary File 1 census block level tabulations carry the school district codes in each block geographic identifier segment enabling the user to tabulate school district complete count data for all summary data items included in Summary File 1.

The Census 2000 school district special tabulation product includes Summary File - like data for 13,559 school districts. Summary data were not released for school districts with less than 50 unweighted respondents. As a result, all of the Census 2000 SDST data are subject to sampling variability as well as other errors of collection and tabulation.

Looking at enrollment and assessment data only, it is not possible to fully understand the state of education. It is essential to use the SDST in analyzing the state of education for local education agencies as well as in regional, statewide and national scope analyses. In many states, where school district boundaries do not align to city or county boundaries, the only source of "school district community" demographics is the SDST. In those instances where PK-12 education is administered through combinations of elementary, secondary, and unified school districts, only the SDST provides data on "relevant children" for those district's educational programs.

History and Impact. The first decennial census school district data were developed with the 1970 census and involved a process for aggregating census block group equivalent data to approximate school district boundaries. With the 1980 census, the first effort to develop a special tabulation for school districts was undertaken but yielded mixed results due to inadequate school district boundary data and the fact that the number of children were "double counted" for areas having more than one school district on the same earth surface area.

For the 1990 Census, a "school district mapping project" was undertaken. An important product from 1990 Census SDST was the first set of maps for all U.S. school districts. School district codes were added to the Census Bureau TIGER/Line files which have made it possible to develop map boundary files for use in mapping and GIS applications. The availability of school district map files makes it possible to view many other types of subject matter (such as enrollment and test data) for school districts. Also, the addition of right- and left-segment school district codes in the TIGER/Line "intersection to intersection" line segments makes it possible to associate address-oriented data with corresponding school district.

The Census 2000 SDST development mirrored that of the 1990 Census SDST in many respects. An important difference is that for the 1990 Census, census blocks could be split by school district boundaries (even though the census block geographic area was not split for any other purpose). For Census 2000. census blocks were defined so that all school districts could be defined as a set of whole census blocks.

The real impact of the SDST is not fully known. A technically complex and somewhat specialized data resource, the SDST is not as widely in use by state and local education agencies as it might be. The SDST is used in determination of the U.S. Department of Education Title I funding program distributions. In related work it is used as the basis for developing updates of children in poverty developed by the Census Bureau to meet needs of the Department of Education.

SDST school district boundary files, and related GIS enabling applications, are used by many state departments of education and are available on state education agency web sites for downloading. These map resources often form the basis for updates to school district boundary changes which occur irregularly in many states.

The SDST data are used in wide-ranging research and analysis applications to meet the needs of Federal, state, and local government, associations, higher education institutions, and businesses. These data provide unique insights into issues such as fiscal capacity and other economic measures for school districts as well as the almost sole source of demographic data for the school district community, children enrolled in public school, children enrolled in private school, and children not enrolled in school.

Effective Access and Use. There are at least two dimensions to the matter of effective access and use. One has to deal with the more operational features – how can I get the specific subject matter for geography of interest to me in a form that I can best use in my analyses. The other dimension has to do with the more "strategic" dimension – determining which data are most useful in the scope of analysis of interest.

This paper first reviews organizational features of the SDST followed with a brief look at the dimension of effective use, more strategic aspects. Operational features and alternative methods of accessing, integrating, and using the data are reviewed.

2 Organization of the SDST and Data Access and Use

2.1 Overview

To properly frame other aspects covered in the paper, a few basic overview items are briefly reviewed first.

The Census 2000 SDST was developed by Census Bureau from decennial census basic record files under National Census for Education Statistics (NCES) sponsorship and specifications. The purpose of the SDST, from the standpoint of NCES, was to provide data from the decennial census to assess the state of education.

Geographic levels of tabulation include school district, county, state, and the U.S.

The scope of subject matter parallels the Census 2000 "Summary File 3" tabulations with special augmentations.

The time frame of the SDST includes the 1990 Census and Census 2000.

Products from the SDST include the following:

1. Web-based database of summary statistic data
   (downloadable and available for direct display)

2. Selected extract files developed for downloading

3. School district code embedded in Summary File 1 Geographic Identifier Segment
   (enables tabulation of school district data from SF1 block level data)

4. School district codes carried on left- and right-side TIGER/Line file segments
   (enables GIS applications, covered later)

5. School district polygon/boundary files (TIGER/Line derived)
   (for Mapping and GIS applications)

2.2 Importance: Enabling Abilities, Benefits, and Application Examples

The importance of the SDST substantially transcends the purposes and uses by the Federal government. A few of the key examples of benefits, uses, and application examples are summarized below.

1. Provides data on children's demographics based on Census
   - not generally available from any other source
   - race/ethnicity reported differently by local education agencies

2. Provides data on "school district community demographics"
   - not available elsewhere where districts not coterminous with cities/counties

3. Enables ability to compare census-based attributes of children enrolled in public school with administratively reported data

4. Provides access to the same set of subject matter tabulations for children enrolled in private school as for children enrolled in public school
   - only source of demographics of children enrolled in private school by school district

5. Enables ability to assess the scope and composition of school age children not enrolled in school

6. Enables comparison of school districts and children's demographics over time
   - 1990-to-2000
   - limited in some cases due to geography concordance and subject matter definition

7. Enables preparation of thematic maps for pattern recognition by school district

8. Enables ability to view school district boundaries in context of other types of geography

9. Enables the geocoding of address-oriented data (e.g. student residences) to school districts

10. Used as basis for developing demographic estimates and projections

11. Used as basis for computing funding distributions for the ESEA Title I and other Federal and state funding programs

12. Used as basis to assess fiscal capacity and fiscal equity

13. Provides unique tabulations on "At-Risk Children"

2.3 Geography

The geographic dimensions of the SDST seem basic, including tabulations for school districts, counties, states, and the U.S. However, there are additional geography attributes.

School district boundaries and names are defined by states. The maps of districts were developed by the Census Bureau and have been made a permanent part of the TIGER/Line system. School districts are classified as unified, secondary, and elementary.

- Unified -- serving children of all grade levels
- Secondary -- primarily serving children in secondary grades, typically 9-12
- Elementary -- primarily serving children in elementary grades, typically PK-8

Some small geographic areas are not in school districts. Such areas have special codes and have data tabulated for as undefined areas.

In addition to the regular school districts and undefined area quasi school district areas, pseudo districts have been designated for community and borough 'school districts' in New York City and for middle and high school complex areas in Hawaii.

In several states, including Arizona, California, Illinois, and others, multiple school districts can cover the same earth surface area. The most typical situation is where elementary and secondary districts cover the same area.

The concept of Relevant Children adds another geographic dimension. While the concept of relevant children is a subject matter concept, it is based on type of school district. A child is relevant to a school district if he/she lives within the territory of the district and his/her assigned grade falls within the grade range provided by a district.

School district boundaries change from year to year in some states. The Census 2000 SDST data relate to school district boundaries defined as of the 1999-2000 school year.

There are many important relationships between school district geography and other types of geography and areas for which data were tabulated from Census 2000. Here are a few key relationships.

- School district boundaries do not cross state boundaries
- School districts are comprised of a set of census blocks
- The school district to census block code to school district code relationship is contained in Summary File 1
- Geographic codes correspond to 1999-2000 school year Common Core of Data codes
- Geographic codes are 5-digit codes, unique within a state, provided separately for unified (SDUNI), secondary (SDSEC), and elementary (SDELM) districts.

2.4 Subject Matter

In this section, the basic concepts of children are reviewed as well as the tabulation record types and enrollment iterations.

Child Defined. For Census 2000 SDST, a child is defined as a person age 0 to 17, or age 18 or 19 and not a high school graduate. For the 1990 Census SDST, the definition is the same as Census 2000 but does not include persons ages 0 to 2. In the 1990 Census, children do not include high school graduates under age 18.

2.4.1 Record Type by Enrollment Type Iteration

A summary of record types follows. A set of subject matter tables are provided for each type of geographic area, provided the confidentiality thresholds are met.

Record Type

1. Total – Population and Households (TT)
   - same as SF3; only difference is that tables are for school district geography
2. Children (CO)
   - Universe limited to children (age 0-17 and age 18-19 not high school graduates)
   - "How many children have a given characteristic?"
3. Households with Children (HC)
   - Subset of total households: any household that includes at least one child
   - Prominent subgroups: householders with related children,
     households with foster children,
     households with children living with siblings or grandparents,
     households where children over 15 serve as the householder, and others
   - "How many households with a given characteristic have children in them?"
4. Parents with Children (PC)
   - Subset of the total population: householders with related children,
     spouses of householders with related children,
     parents of children in subfamilies
   - Does not include householders with unrelated foster children
   - "How many parents with a given characteristic have children?"
5. Children – by Household Characteristic (CH)
   - Number of children living in households with a given characteristic
   - Relationship requirements same as HC record type,
     but result identifies units of children rather than households
   - Characteristics of households are counted for each applicable child in household
   - E.g., a household with three children would be counted three times
   - As with the HC record type, data for CH
     do not include children living in Group Quarters (a non-household arrangement)
   - "How many children live in a household that has a given characteristic?"
6. Children – by Parent Characteristic (CP)
   - Number of children living with a parent that has a given characteristic
   - Relationship requirements same as PC record type,
     but the result identifies units of children rather than parents
   - Characteristics of each qualified parent counted for each applicable child in household
   - Allows a single child to be counted twice in given table.
   - "how many children are living with parents that have a given characteristic?"

Enrollment Iterations

For each of 5 child-oriented record types listed under Record Types, each set of table/items is iterated for each of the following enrollment iterations.

- Total Children
- Relevant Children – (Enrolled and Not Enrolled)
- Relevant Children – Not Enrolled
- Relevant Children – Enrolled
- Relevant Children – Enrolled Public
- Relevant Children – Enrolled Private

2.4.2 Subject Matter Tables and Items

Data are organized in a manner similar to the summary files.

Table/Item Concept and Example
- examples of Table P88 follow

Summary File 3-like subject matter tables
- see matrices of table/items organized by record type

At-Risk Tables. SDST At-Risk tabulations are one major exception to subject matter coverage as compared to the SF3 tables. Generally, children who are at-risk have this attribute based on characteristics of the child's mother. These are children who are living with a mother who is:

- not a high school graduate,
- single, divorced, or separated, and
- whose 1999 income was below the poverty level

Some At-Risk tables differentiate children and mothers by age group. Tables are similar to those designed for the 1990 Census SDST.

2.5 Suppression and Disclosure Avoidance

Suppression and disclosure avoidance affects data access and use in two ways. First, availability of any child oriented record type (all types except TT) is subject to a population threshold of 50 children (unweighted cases) in the geography being tabulated. If school district XX has only 49 children (unweighted cases), the iteration tables will not appear for that district. Second, each cell value in the special tabulation files is independently rounded. Some tables lack additivity. In rounding, 0 remains 0, 1-7 rounds to 4, 8 or greater rounds to the nearest multiple of 5. All totals, subtotals, and derived measures were computed prior to rounding and independent cell rounding.

2.6 Data Access and Use

At present, available options to access SDST data are through the National Center for Education Statistics and Proximity. The Census Bureau provides access to the TIGER/Line files which contain school district codes and are the source of derived boundary files. The Census Bureau also provides downloadable access to the school district boundary files. The Census Bureau does not, at present, provide access to the SDST subject matter data.

2.6.1 NCES Facilities for Access

NCES SDST data access services are focused on a Web-based model.

Table query/call-up. Data may be retrieved in a tabular structure having some similarity to the Census Bureau FacfFinder.

An example of using the NCES facility to display data in a tabular manner is shown below. These sample tables show how Table P88 is retrieved for the total population (TT record) and Children's Own -- relevant children enrolled in public school (CO – RCEP).

Using NCES Facility to perform a query
-- application example related to GIS tutorial
http://proximityone.com/sdmaps/sdmaps07.htm

Sample conclusions: 171,375 total population (P88 tab universe) compares to 31,535 RCEP population

1 - Objective:

display Table P88 Population by Ratio of Income in 1999 to Poverty level for Spring Branch ISD

(total population: "school district community")

- http://www.nces.ed.gov/surveys/sdds/selectgeo.asp
- Choose School District>Texas>Spring Branch ISD
- Choose Total -- Population and Households (TT)
- Choose Select A Table P88
- Click Continue -- view this table

Geography: School District -SPRING BRANCH INDEPENDENT SCHOOL DISTRICT export to excel file | format for printing

Universe: Total – Population and Households Characteristics

Table: P88 : RATIO OF INCOME IN 1999 TO POVERTY LEVEL [10]

P88. Ratio of Income in 1999 To Poverty Level [10]

Universe: Population for whom poverty status is determined

Total:	171,375
Under .50	10,865
50 to .74	6,715
75 to .99	7,680
1.00 to 1.24	8,780
1.25 to 1.49	9,045
1.50 to 1.74	9,010
1.75 to 1.84	3,210
1.85 to 1.99	4,515
2.00 and over	111,555

All "Race Alone" tables (i.e. White Alone, Black or African American Alone, American Indian and Alaska Native Alone, Asian Alone, Native Hawaiian and Other Pacific Islander Alone or Some Other Race Alone) include Hispanic and non-Hispanic origins.
Note on independent cell rounding:
Each cell in the Census 2000 special tabulation matrix has been independently rounded. As a result of the cell rounding process:
0 remains 0, 1-7 rounds to 4, 8 or greater rounds to the nearest multiple of 5
Totals, subtotals, and derived measures computed prior to rounding; independent cell rounding process disrupts table additivity.
Source: The Census 2000 School District Tabulation (STP2) is a special tabulation prepared by the U.S. Census Bureau's Population Division and sponsored by the National Center for Education Statistics.

2 - Objective:

display Table P88 Population by Ratio of Income in 1999 to Poverty level for Spring Branch ISD

(relevant children enrolled public)

- http://www.nces.ed.gov/surveys/sdds/selectgeo.asp
- Choose School District>Texas>Spring Branch ISD
- Choose Children (CO)
- Choose Total Relevant Children - Enrolled Public
- Choose Select A Table P88
- Click Continue -- view this table

Geography: School District -SPRING BRANCH INDEPENDENT SCHOOL DISTRICT export to excel file | format for printing

Universe: Children Characteristics /**Enrollment:** Total Relevant Children – Enrolled Public

Table: P88 : RATIO OF INCOME IN 1999 TO POVERTY LEVEL [10]

P88. Ratio of Income in 1999 To Poverty Level [10]

Universe: Population for whom poverty status is determined

Total:	31,535
Under .50	2,550
50 to .74	1,965
75 to .99	2,210
1.00 to 1.24	2,275
1.25 to 1.49	2,335
1.50 to 1.74	1,985
1.75 to 1.84	740
1.85 to 1.99	1,090
2.00 and over	16,390

All "Race Alone" tables (i.e. White Alone, Black or African American Alone, American Indian and Alaska Native Alone, Asian Alone, Native Hawaiian and Other Pacific Islander Alone or Some Other Race Alone) include Hispanic and non-Hispanic origins.
Note on independent cell rounding:
Each cell in the Census 2000 special tabulation matrix has been independently rounded. As a result of the cell rounding process:
0 remains 0, 1-7 rounds to 4, 8 or greater rounds to the nearest multiple of 5
Totals, subtotals, and derived measures computed prior to rounding and independent cell rounding process disrupts table additivity.
Source: The Census 2000 School District Tabulation (STP2) is a special tabulation prepared by the U.S. Census Bureau's Population Division and sponsored by the National Center for Education Statistics.

Downloading Data. Data are downloaded from the same NCES website

Mapping Resources. The NCES map server is accessible via http://www.nces.ed.gov/surveys/sdds/. The mapserver provides access to basic thematic maps for a selected set of SDST data and basic orienteering features

2.6.2 Proximity Data Access & Analysis Resources

Proximity has developed a range of tools more geared to analysis, data integration, and GIS applications. Several of these resources are summarized below.

Proximity Web-based Children's Demographics CDSQL

- http://proximityone.com/cdsql.htm
- Useful to generate "type of enrollment" profile
- Useful to easily display sorted lists of districts with selectivity
- Uses Structured Query Language
- Transportable to any server using no fee MYSQL

Proximity Web-based School District Demographics (SDD) Online

- http://proximityone.com/sdd.htm
- Useful to generate comparative analysis profiles
- Access to any table in the SDST Database
- Uses Structured Query Language
- Transportable to any server using no fee MYSQL

Proximity Web-based School District Information (SDIS)

- http://proximityone.com/sdsql.htm
- Useful to generate multi-sourced school district data queries
- Links selected SDST data with other national scope school district data files
- Uses Structured Query Language
- Transportable to any server using no fee MYSQL

Proximity Web-based Demographic Profile Comparative Analyses

- Arizona Example: http://proximityone.com/azsdd.htm
- See html display result here
- Access to subject matter standard DP1-DP4 Profiles
- DP1-DP4 profile structure for school districts only available from Proximity
- Links selected SDST data with other national scope school district data files

Proximity PC-based Demographic Profile Comparative Analyses

- http://proximityone.com/dpa.htm
- Generates subject matter standard DP1-DP4 Profiles
- Similar to Web-based version but adds other features
- Creates socioeconomic status indexes, rankings and compares with performance scores

3 Effective SDST Data Access and Use & Decision Making

3.1 Tools, Options and Procedures

The data access facilities described in the earlier section provide capabilities to view and download the SDST data. Only the SDIS and DPA facilities provide a capacity to examine the SDST data in a manner integrated with other data.

Users of the SDTS data can be functionally categorized into two classes. One set of users seeks basic demographics to document characteristics of one or a set of school districts. For the users, the data access facilities described above meet most requirements.

Another set of users seeks to use the data in an analytical context. This set of users might seldom need to view data for districts, and if they do, the very next thing to be done is to integrate the SDST data with other data.

3.2 Data Integration

Data integration, or providing a framework for such, is Achilles Heel of a prospective "most effective" Federal data use. Most Federal statistical agency operations view their responsibilities to collect or develop the data and then to make it available -- "data dissemination." Most users view the process as "data access." In any case, with some notable exceptions, data dissemination and data access operations are typically a) not funded very well, if at all, and b) they are almost always focused on specific statistical programs. Generally, data are not even integrated in data dissemination/access programs within the same statistical agency or even within the same division of the agency. This is, and has been for many years, the challenge and opportunity to make Federal statistical data more useful.

3.2.1 Federal Statistical Data Integration

Data Integration Priorities. With respect to the SDST, there are five primary areas of data integration that are higher priority.

1. Linking data from 1990 Census SDST with Census 2000 SDST
2. Linking data from Census 2000 SDST multiple record types and enrollment categories for a given district or set of districts
3. Linking Census 2000 SDST subject matter data with map files (shapefiles) for visual analysis
4. Linking Census 2000 SDST data for districts with other NCES district level data (e.g., CCD)
5. Linking Census 2000 SDST data for districts with other Census 2000 SF3 data for drill down geography (e.g. census tracts, block groups, blocks)

To address this and related needs, Proximity has developed the dBMerge software. See http://proximityone.com/dbmerge.htm for details.

A 'Federal data server interface' is now being developed that will enable the user to download a file from either the NCES webserver or the Census Bureau FactFinder. More Federal data servers will be added based on experiences with these servers.

Once the data is downloaded to the user computer, that file, and a possible set of other files, is linked with dBMerge. Knowing the attributes of the file contents based on metadata and file source data, the contents of the multiple files are connected and output into an integrated file.

Here are two examples of this process. Using the NCES SDDS server, two state by school district file downloads are developed: one for table P88 for the TT record (total population) and one for table P88 for the CO-RCPE record (relevant children enrolled in public school). dBMerge now links these files and combines the fields for matched school district records into a single output record, optionally adding a standardized district name from a dBMerge master file. The merged fields output file now makes it easy

to analyze TT level data with corresponding CO-RCPE level data in Excel or choice of software that processes dbase structured files.

A second example parallels the NCES server, but makes use of data from the Census Bureau FactFinder. In this application, the user downloads a table from Summary File 1 and a table from Summary File 3 for block groups that are partly or wholly contained in a school district of interest. . Using the downloaded files, dBMerge now links these files and combines the fields for matched block group records into a single output record, optionally adding a standardized district name and/or code from a dBMerge master file. The merged fields output file now makes it easy to integrate SF1 sourced data with SF3 sourced data.

dBMerge addresses needs/requirements of items 1, 2, and 5 in the above list. Solutions to other data integration needs are described below.

Future versions of dBMerge will add features that relate to these capabilities.

1. The user will request the data needed using dBMerge and dBMerge will then send the request for the data download to the Federal data server making those steps semi-transparent to the user.
2. Rather than being restricted to dbase structured output structures, optional output structures will be available.
3. Determination of small area geography partly contained in a school district will be determined by dBMerge enabling the user to a) skip over the tedious steps of determining what these areas are (in terms of codes) and b) interface the data request for theses geographic areas to the Federal data server.

3.2.2 Federal and non-Federal Data Integration

Almost all users engaged in the analysis of SDST data in state and local education agencies, researchers in academia and elsewhere, and associations/special interest groups, almost immediately seek to integrate the SDST data with non-Federal data. State and local data that are most typically used with SDST data are test score data and enrollment data. In most states, the school district codes used internally with the state are not consistent with the Federal school district code. The process is further complicated, possibly necessarily, by the fact that the Census Bureau TIGER/Line files use three types of school district codes (unified, secondary, elementary) even though they all conform to the CCD code where possible.

In many states, school district boundaries and school locational data are structured in a manner inconsistent with the Federal data (SDST and otherwise). The boundaries might be updated and relate to different geography. The codes used in the boundary files often only make use of the state assigned district codes, requiring an interface with the state and Federal codes.

These data integration requirements are important special cases of those priorities listed as items 3 and 4 under Data Integration Priorities. Features of dBMerge have been developed to address these needs.

dBMerge has the basic capability of merging any two input/source files, based on user defined merge key fields in the source files, and creating an output file with all or a selected set of items from either source file. These capabilities provide a basic solution as to how Federal and non-Federal sourced data can be linked.

As an example, consider an application where source file 1 is from a state education agency that contains state sourced school district codes (possibly test score data) and source file 2 is from the NCES SDST download containing poverty data. The analytical objective might be to examine relationships between poverty and test scores by district. The dBMerge AddCode feature is used to add the Federal school district code to the state source file 1. Now source file 1 and source file 2 can be merged based on the common Federal code.

Although critically important, data integration sets the stage for more effective use of data resources. There is also a need to develop a framework for using these data once they become accessible in manner where the data can be integrated and updated.

3.3 Advancing the Role of Visual and Trend Analysis in Education Decision Making

Education leadership, particularly in EE-12 schools and school systems, is at a substantial disadvantage in making use of data and technology tools now widely in use in business and elsewhere. There are steps that can be taken to empower education leadership to make better use of data driven decisionmaking tools. Possibly a priority should be assigned to making use of visual analysis tools, such as maps and GIS, due to the power of visual analysis and the relative ease of communicating a conclusion or trend to a group using these tools.

Many EE-12 school systems have long been inclined to "buy demographic studies" to help determine where growth will occur, assess its impact, and help make decisions as to new school locations or assigning revised attendance area boundaries. Such studies typically entail expenditures of upwards of $20,000 and often much higher. The school district gets, if it is lucky, a one-time assessment of the "geodemographic state of the district." There is normally no way to update the study findings.

While there is nothing wrong with this approach, and it might be the best approach to planning for some districts, the movement in the "EE-12 system industry" should be to one of self empowerment. All larger school districts (possibly 10,000 ADA and over) should be on a path of developing their own standardized attendance area boundary files, address matching and geocoding (dynamically) student data records, lining these data into a Geographic Information System (GIS), and organizing the student, facility and community data into a time series database that enables development of demographic (and other) estimates and projections.

The SDST fits into this picture in many important ways. It provides the Census 2000 "historical baseline" data. It provides a "level 1" source of streets, census tract, census block, and other geography in the form of shapefiles that can be used as baseline and starting point for the evolution of a district's mapping capabilities. In many states, the SDST provides the only source of "school district community" demographics which should be a core part of the school district decision making information system.

Using this collective set of tools, the school district is now better able to develop updated demographic estimates and projections that might reflect new assumptions, updated historical data, alternative model specifications, revised independent data, revised attendance area boundaries, and other such changes. Estimates and projections could be more easily developed annually and conveyed for visual analysis using GIS capabilities. Linking other data, such as finance data, the district leadership is now better able to systematically address what-if and impact analysis questions.