

**USDA Service Center Agencies
Geospatial Data Management Team
Data Management Plan For**

Census Data

**Randy Frosh
May 2005**

I. Purpose and Scope (business case)

A. Purpose

The Census Bureau conducts many important censuses and surveys. The most well-known is the official population census of the United States, called the decennial census. It is conducted every ten years, most recently in April 2000. During each decennial census, the Census Bureau collects data from every household in the U.S. and its territories.

The Census Bureau, American FactFinder displays the most commonly requested results of a census. Information and characteristics from the demographic profiles is provided for fact sheets related to:

- ❑ General demographic - Population, race, ancestry, income, disability, education, employment, language, marital status, occupation, poverty status, and more.
- ❑ Social - Population, race, ancestry, income, disability, education, employment, language, marital status, occupation, poverty, housing characteristics and more
- ❑ Economic - Industry, establishments, sales, receipts, revenues, shipments, expenses, products, payroll, employees and more
- ❑ Housing - housing characteristics

Factfinder does not have map data per se. Instead, American FactFinder data is joined to the TIGER map data to produce a map with attribute data.

Factfinder is a specialized census web site that provides information on demographic profiles. The demographic profiles provide key, most widely used, Census 2000 data for cities/places, counties, and larger geographic areas.

The Demographic Profile contains the 100-percent and sample data. 100-percent (complete count) include data on sex, age, race, Hispanic or Latino, household relationship, household type, group quarters population, housing occupancy, and housing tenure. Sample items include population data on school enrollment, educational attainment, marital status, grandparents, veteran status, disability status, residence, nativity and place of birth, foreign born, language spoken at home, ancestry, employment status, commuting to work, occupation, industry, class of worker, income, and poverty status. Sample items include housing data on units in structure, year structure built, rooms, year householder moved into unit, vehicles available, house heating fuel, occupants per room, value, mortgage status and monthly owner costs, and rent-related items.

B. Scope

The scope of the dataset is the United States and Puerto Rico. It does not include the island areas of: American Samoa, Commonwealth of the Northern Mariana Islands, Guam and the U.S. Virgin Islands.

The default geography for the Fact Sheet is the United States. Fact Sheet data is available for states, counties, cities and villages, most towns, and most ZIP Codes. It is available by county, census block group, census block and census tract.

The Demographic Profiles are available for the following types of geographic areas:

- Nation, Region, Division
- Metropolitan Area (whole and by state part)
- American Indian Area/Alaska Native Area/Hawaiian Home Land (whole/state part)
- State
- County and County Subdivision
- Place
- Consolidated City
- Alaska Native Regional Corporation
- Congressional District

II. Acquisition

A. Data Source

1. Producer Information

a. Name

U.S. Census Bureau

b. Location of Headquarters

U.S. Census Bureau
4700 Silver Hill Road
Washington DC 20233-0001

c. Internet Address

<http://www.census.gov>

2. Publisher Information

a. Name

U.S. Census Bureau

b. Location of Headquarters

U.S. Census Bureau
4700 Silver Hill Road
Washington DC 20233-0001

c. Internet Address

<http://www.census.gov>

3. Acquisition Information

a. Delivery Media

FTP. The online copy of the data may be accessed without charge

b. Download URL

The Demographic Profile technical documentation (PDF) may be accessed at <http://www.census.gov/prod/cen2000/doc/ProfileTD.pdf>

Access the CSV structured demographic profiles at the URL:

http://www2.census.gov/census_2000/datasets/100_and_sample_profile/0_All_State/

c. Projected Data Availability Schedule

The data is currently available.

B. Standards Information

1. Geospatial Data Standard

a. Standard Name and Steward Information

There is no recognized standard for the attribute data from the U.S. Census Bureau. The Service Center Agencies provide map, metadata and symbology standards that are referenced in the TIGER data management plan.

b. Standard Version

SCI Std 003-02
October 15, 2003

c. Standard URL

<http://www.itc.nrcs.usda.gov/scdm/docs/SPG-GeospatialDataStandard.pdf>

2. Metadata Standard

a. Standard Name and Steward Information

Metadata are compliant with:
Federal Geographic Data Committee (FGDC)
Content Standard for Digital Geographic Metadata FGDC
STD-001-1998 Version 2 revised June 1998

And:

United States Department of Agriculture (USDA) Service Center Agencies (SCA)
Standard For Geospatial Dataset Metadata
SCI Std 003-02 October 15, 2003

<http://www.itc.nrcs.usda.gov/scdm/docs/SPG-GeospatialDatasetFileMetadata.pdf>

b. Description of Metadata Captured

None

c. Metadata Accuracy and Completeness Assessment

None

C. Acquired Data Structure

1. Geospatial Data Format

- a. Format (raster, vector, etc.)

Database import compatible test files.

- b. Format Name

Comma separated values (CSV)

- c. Data Extent

The United States and the island areas, which consist of: American Samoa, Commonwealth of the Northern Mariana Islands, Guam and the U.S. Virgin Islands.

- d. Horizontal and Vertical Resolution

None. The data is attribute data. See TIGER for map resolution.

- e. Absolute Horizontal and Vertical Accuracy

None. The data is attribute data. See TIGER for map accuracy.

- f. Nominal Scale

1:100,000 for TIGER data

- g. Horizontal and Vertical Datum

The TIGER datum is North American Datum 1983 for all appropriate areas areas (UTM zone 3 through 22) and World Geodetic System 1984 elsewhere. The vertical datum is mean sea level.

- h. Projection

See TIGER for map accuracy.

- i. Coordinate Units

See TIGER for map coordinate units.

- j. Average Data Set Size

Each of the four datasets: demographic, economic, social and housing are approximately 60 megabytes as a map of all counties in the U.S.

- k. Symbology

None

2. Attribute Data Format

- a. Format Name

Comma separated values (CSV)

b. Database Size

15 megabytes average for each of the four datasets.

3. Data Model

a. Geospatial Data Structure

TIGER® database <http://www.gistools.com/ftp/Manual6.pdf>

b. Attribute Data Structure

Comma separated values (CSV)

c. Database Table Definition

<http://www.census.gov/prod/cen2000/doc/ProfileTD.pdf>

d. Data Relationship Definition

Each demographic profile is a self contained table that can be joined to TIGER data by FIPS code.

e. Data Dictionary

<http://factfinder.census.gov>

D. Policies

1. Restrictions

a. Use Constraints

The Census Bureau asks that they be cited as the source.

b. Access Constraints

None

c. Certification Issues

None

2. Maintenance

a. Temporal Information

Decennial census

b. Average Update Cycle

10 years.

E. Acquisition Cost

1. Cooperative Agreement

a. Description of Agreement

None

- b. Status of Agreement

None

2. Cost to Acquire Data

FTP download is free.

III. Integration

A. Value Added Process

1. Benefit to the Service Center

A centralized location to acquire to most commonly sought demographic, economic, social and housing statistics.

A continuous database for the US is created so that any area can be extracted as a shape file. A consistently named FIPS code field is created so that counties and states can easily be extracted.

2. Process Model

- a. Flow Diagram

None. See description below.

- b. Process Description

- 1. ftp www2.census.gov

- Login anonymous and supply your email address.
- Use the ftp Dir command to go to the site
census_2000/datasets/100_and_sample_profile/0_All_State/
- Use the ftp command prompt n
- Use the ftp command get ProfileData*.ZIP
- Use the ftp command quit

- 2. Unzip all the ProfileDataAA.ZIP files (where AA is the state postal code).

- 3. Separate the files into directories where:

- 2kh##.csv into DEMOGRPAHIC
- 2ks##t2.csv into SOCIAL
- 2ks##t3..csv into ECONOMIC
- 2ks##t4.csv into HOUSING

- 4. use MSAccess to define tables for each of the four layers. The first fields are the same for all tables. The type is text.

Name	Description	Len
RECTYP	Record Type	2
SUMLEV	Summary Level	3
GEOCOMP	Geographic Component	2
FIPSST	State code	2
FIPSCO	County code	3
COUSUB	County Subdivision code	5
PLACE	Place	5
CONCIT	Consolidated City code	5

MSACMSA	Metropolitan Statistical Area/Consolidated Metro. Stat. Area	4
PMSA	Primary Metro. Stat. Area	4
AIANHH	American Indian Area/Alaska Native Area/Hawaiian Home Land	4
ANRC	Alaska Native regional Corporation	2
CD106	Congressional District 106th	2
FUNCSTAT	Functional Status Code	1
AREANAME	Area Description	95

5. Define the rest of the fields for each table. The type is Number as a double. These fields are defined in the data base table definition in <http://www.census.gov/prod/cen2000/doc/ProfileTD.pdf>
 - V01 – V96 for DEMOGRPAHIC
 - DP2001 – DP2102 for SOCIAL
 - DP3001 – DP3107 for ECONOMIC
 - DP4001 - DP4100 for HOUSING
6. Use MSAccess File/Get External Data and import each of the four demographic files for each state into the tables.
7. Create a FIPS_C field that is text with length five. This field is used to join to the FIPS_C field in the TIGER county file.
8. For each table concatenate the FIPSST and FIPSCO filed into the FIPS_C field with a statement like Update CN00DEMG set FIPS_C = FIPSST + FIPSCO
9. Separate out the county records into a county table for each of the four tables. Delete CN00DEMG where RECTYP <> '07' and get rid of the county subdivisions with Delete CN00DEMG where COUSUB <> ''
10. Delete attributes. RECTYP all are '07', SUMLEV all are '50', GEOCOMP all are '00', and COSUB, PLACE, CONCIT, MSACMSA, PMSA, AIANHH, ANRC, CD106 (all are blank for counties)
11. Join the TIGER county map to each table then save as a shape file.
12. Delete the features for Palau, Virgin Islands, Guam, Marshall Island, etc that do not have demographic profile data.
13. Load each map into SDE with a scale factor of 100,000 to provide one meter of resolution.

3. Technical Issues

a. Tiling

None

b. Compression

None

c. Scale

1:100,000

d. Tonal Matching

None

e. Edge-matching

See TIGER

4. Quality Control

a. Procedures

None. However, spot checks for completeness are performed.

b. Acceptance Criteria

None

5. Data Steward

a. Name and Organization

Natural Resources Conservation Service
National Cartography and Geospatial Center
P. O. 6567
501 Felix St., Bldg. 23
Fort Worth, TX 76115-3405

Original Data steward will remain:

U.S. Department of Commerce
Bureau of the Census
Geography Division

b. Responsibilities

NCGC will apply updates to the data set and notify the Service Centers.

B. Integrated Data Structure

1. Geospatial Data Format

a. Format (raster, vector, etc.)

Vector

b. Format Name

ESRI ArcSDE

c. Data Extent

The extent is the same as the source.

d. Horizontal and Vertical Resolution

One meter

e. Absolute Horizontal and Vertical Accuracy

The accuracy is the same as the source.

f. Nominal Scale

1:100,000

g. Horizontal and Vertical Datum

Each layer in ArcSDE has a single projection/datum code value. As a result, the datum must be defined as either NAD83 or WGS84. NAD83 is an illegal datum for the Pacific areas west of Hawaii. So, WGS84 should be the assigned datum.

h. Projection

Longitude/Latitude

i. Coordinate Units

Decimal degrees.

j. Symbology

<http://www.itc.nrcs.usda.gov/scdm/docs/SPG-StandardforGeospatialSymbology.pdf>

2. Attribute Data Format

a. Format Name

ESRI ArcSDE

b. Database Size

Varies by theme

3. Data Model

a. Geospatial Data Structure

ESRI ArcSDE

b. Attribute Data Structure

ESRI ArcSDE

c. Database Table Definition

Varies by theme

d. Data Relationship Definition

ArcSDE

e. Data Dictionary

Same as source

C. Resource Requirements

1. Hardware and Software

Unknown.

2. Staffing

Unknown.

D. Integration Cost

1. Hardware and Software

Unknown.

2. Staffing

One day for one person to download the data, get into access, join it to the TIGER shape file and load it into SDE.

IV. Delivery

A. Specifications

1. Directory Structure

a. Folder Theme Data is Stored In

F:\geodata

2. File Naming Convention

<http://www.itc.nrcs.usda.gov/scdm/docs/SPG-GeospatialDataSetFileNameStandard.pdf>

a. List of Theme Files and The File Naming Convention

Shape files:

Poly Files	
map shp	shp file
map dbf	dbf file
map shx	shx file
map sbn	sbn file
map sbx	sbx file

B. User Information

1. Accuracy Assessment

a. Alignment with Other Theme Geospatial Data

Data aligns well with other TIGER data but will not align with data at a scale of better than 1:100,000

b. Content

The content is the same as the source.

2. Appropriate Uses of the Geospatial Data

a. Display Scale

1:100,000

b. Plot Scale

1:100,000

- c. Area Calculations
Accurate only to source scale
- d. Decision Making
None

C. Maintenance and Updating

1. Recommendations and Guidelines

- a. Original data location and structure

Natural Resources Conservation Service
National Cartography and Geospatial Center
P. O. 6567
501 Felix St., Bldg. 23
Fort Worth, TX 76115-3405

The integrated database is stored in an ArcSDE database. The data is delivered to the Service Center.

- b. Update Cycle

Update data when Census releases new TIGER files

- c. Availability

Make the updates available as soon as the database is updated.

- d. Change Control

This is to be determined.