

ACS Initial Summary File Specification

**American Community Survey Office
U.S. Census Bureau**

Table of Contents

1	INTRODUCTION	3
2	ASSUMPTIONS.....	4
3	GENERAL PROPOSAL	4
3.1	<i>CONTENT AND RELEASE OF ACS SUMMARY FILE</i>	5
3.2	<i>ORGANIZATION AND FORMAT</i>	5
3.3	<i>FILE IDENTIFICATION.....</i>	6
3.3.1	<i>File naming convention.....</i>	6
3.3.2	<i>Identification Fields Within Summary File.....</i>	8
4	NOTES(N) AND QUESTIONS(Q) – OPEN ISSUES FOR FEEDBACK.....	8

1 Introduction

The American Community Survey (ACS) is a new approach to producing critical information about the characteristics of local communities. It will eliminate the need for a long form in the 2010 Census and is a key part of the Census Bureau's Decennial Census Program. The ACS publishes social, housing, and economic characteristics for demographic groups covering a broad spectrum of geographic areas in the United States and Puerto Rico. The ACS shifted from a demonstration program with a different sample design and sample size to the full sample size and design in 2005. It became the largest household survey in the United States, with an annual sample size of about 3 million addresses. Every year the ACS can support the release of single-year estimates for geographic areas with populations of 65,000 or more.

The ACS will accumulate sample over 3-year and 5-year intervals to produce estimates for smaller geographic areas including census tracts and block groups. The data product based on the five-year time period is designed to be similar in scope and geography to the data products released from the traditional long form.

In order to ensure that the data products from the ACS are meaningful and usable for all data users, the Census Bureau releases many different kinds of tabular products on the Census Bureau's American FactFinder website. We are now proposing to develop and release an ACS "Summary File" (ACS-SF) that will contain the complete set of ACS detailed tables released each year in a computer file format. One of the data products supported by the long form in Census 2000 was the Summary File 3 (SF3). SF3 is basically a set of matrices (or tables) on a variety of subjects drawn from the long form that are repeated for all the supported geography down to the block group level. The proposed design of the ACS-SF is very similar to the SF3.

(For more information on SF3, go to <http://www.census.gov/Press-Release/www/2002/sumfile3.html> .) The final goal for the ACS-SF will be to provide a set of files (i.e., a separate ACS-SF) for estimates from each time period the ACS will cover: one-year; three-year; and, five-year.

In 2006, the ACS will implement, as a test, a prototype of the ACS-SF based on the 2005 ACS detailed tables (aka "base tables"). The Census Bureau hopes that this prototype ACS-SF will give data users the chance to work with this set of files and provide the Census Bureau with useful comments. The Census Bureau will consider comments it receives to the prototype release into the design of the final ACS-SF. All comments must be received no later than October 31, 2006. The Census Bureau plans to release the prototype ACS-SF by the end of 2006.

2 Assumptions

For the initial product, only base/detailed tables will be included in the 2005 ACS-SF. Tables suppressed because they did not meet the criteria of the ACS data release rules will be included in the file and will contain “null” cells.

The basic structure of the file for the ACS-SF should be similar to the segmented format used for the Census 2000. (See the documentation given in the hyperlink referenced above).

Eventually, both the ACS and the Puerto Rico Community Survey (PRCS) will release separate ACS-SF products for the 1-, 3-, and 5-year estimates.

The Census Bureau will provide definition of the ACS detailed tables. These tables will be mapped to file segments once they’ve been defined. The technical documentation accompanying the ACS-SF will spell out this mapping in detail.

By the end of 2007, there will be complete technical documentation of ACS-SF.

3 General Proposal

This proposal describes a general scheme that does not depend on the number or size of the data tables for ACS.

The 2005 ACS-SF will be a compact, portable compilation of all Base/Detailed Tables released for 2005 ACS, similar to the 2000 Census summary files. However, there will be differences that will reflect the new and unique product initiatives that are a hallmark of the ACS program.

Summary file users are typically large institutions like governmental agencies and universities, which have an established IT infrastructure and staff that are experienced with using Census Bureau data products in their own environment. The needs and preferences of summary file users are relatively unknown; their number is indeterminate.

3.1 Content and Release of ACS Summary File

General Concept: Eventually there will be six National-level Summary File Products, three for ACS (one-year, three-year, and five-year estimates), and three for the PRCS(one-year, three-year, and five-year estimates).

Assumption: The release of ACS summary file will be synchronized with the release of the same data in AFF. The initial layout of the file will account for all of the detailed tables.

3.2 Organization and Format

The summary files will be very dense, very portable, and very highly structured. Their contents will be ASCII, they will contain only data, an ASCII header, file label, and each data element will be identified by its file position. Included with this file will be an ACSII Data Dictionary.

The detailed table ACS-SF release will contain one file of geographic information (the “geoheader”), containing one record for each geographic entity, and many files of detailed table data. Logical Record numbers will be assigned to the geoheaders sequentially, beginning with ‘0000001,’ and the same numbers will be assigned to data records so that the data and geographic information can be linked together correctly.

The geoheader file will be released only once for a data set, with the detailed table data of the first release. The geographic header file will be fixed field format. The logical record numbers in the files for subsequent releases will link to those geoheaders.

There will be three sets of data files in identical format. One set will contain the estimates, one will contain the standard errors, and one will contain the margin of error. There will be one field per estimate.

The format of the files in each of the four sets will be similar to that of the Census 2000 summary data files. These files will be positional and comma delimited. Each file will contain approximately 7 fields that will enable users to identify and combine the files, and up to 245 comma-delimited data items. (This limit is chosen with Excel limitations in mind.) Numeric data items will not have embedded commas. The order of the data items will be documented in the Technical Documentation and the ACSII Data Dictionary.

3.3 File Identification

Summary files' names, and certain fields within summary files, will enable concrete identification of their contents (consistent with the 2000 Census but not identical).

3.3.1 File naming convention

This convention will enable a user and/or a computer program to determine the contents of a file from its name.

The format of geoheader file names will be *st geo r . datayear - x yr*

<i>st</i>	'us' for ACS; 'pr' for the Puerto Rico Community Survey
<i>geo</i>	constant – always 'geo'
<i>r</i>	data revision indicator, only present if needed. 'a' 'b' etc
<i>datayear</i>	data year
-	constant – always '-'
<i>x</i>	number of years of data used to compute the estimates (1, 3, or 5)
<i>yr</i>	constant – always 'yr'

Example: usgeo.2006-1yr
name of the geoheader file of the original release of data for ACS for the 2006 data year, the 1-year estimate

Example: prgeob.2007-3yr
name of the geoheader file of the second revision ('b') of the Puerto Rico Community Survey data for the 2007 data year, the 3-year estimate

The format of data file names will be *st iter seq t r . datayear - x yr*

st SS, 'us' or 'pr' where "SS" is the postal abbreviation for the state. The 'us' for the national file which will contain geographic entities that can cross state boundaries (eg., metro areas), and 'pr' stands for Puerto Rico.

iter '000'

seq file sequence number, '0001' – '9999' a four-digit code.

This sequence number will be used in a manner similar to the approach taken for the SF3. Each sequence number will correspond to a series of detailed tables. Each record of this file will be for a unique geographic area published by the ACS. All the data cells contained in the detailed tables for that geographic area would be on this record. There will be a file for the estimates and a separate one for the standard errors.

t type of data:

'e' for estimates

's' for standard errors

r data revision indicator, only present if needed. 'a' 'b' 'c' etc.

datayear data year

- constant – always '-'

x number of years of data used to compute the estimates (1 or 3 or 5)

yr constant – always 'yr'

Example: us00014e.2006-1yr

--name of the 14th estimates data file of the original release of ACS data for the 2006 data year, the 1-year estimate

Example: pr00062sc.2007-3yr

--name of the 62nd standard error data file of the third revision ('c') of the Puerto Rico Community Survey data for the 2007 data year, the 3-year estimate

3.3.2 Identification Fields Within Summary File

The following fields will be populated within the summary files, so that the data and geographic information can be identified and linked together correctly by computer programs.

CENSUS GEOID: 1 field unique identifier

FILEID: File identification - 6 characters
Constant and always "ACSSF"

STUSAB: State/US-Abbreviation (USPS) - 2 characters
Ex. 'tx', 'us' or 'pr'

CHARITER: Iteration number - 3 characters
Ex. '001'

SEGMENT/IFSN: Characteristic Iteration File Sequence Number - 2 characters
Range: '01' - '??' (SF3 had 76 data files)

LOGRECNO: Logical Record Number - 7 characters
Range: '0000001' - '9999999'

4 Notes(N) and Questions(Q) – open issues for feedback

N: The detailed tables defined for ACS include some tables that are iterated. This proposal handles them like the A-I iterations within SF3 – simply as more detailed tables.

N: The new summary file layout could be split up in alternative ways. For example: (1) State Based files (the SF3 method); or, (2) State-based files plus a file for all counties in the U.S. **Other suggestions?**

Q: This proposal maintains the Census 2000 fields CHARITER and IFSN as separate fields of 3 and 2 characters respectively. Should they be merged into one 5-character field? Will future ACS products need more than 99 data segments? Or will future years of ACS products be more iterated?

Q: ACS includes alphabets in their data, whereas Census 2000 used numeric jam values in similar situations, primarily for users' convenience. Should ACS alphabets be replaced with numerics?

Q: Will the ACS-SF be made available in alternate formats such as SAS, or csv?

Q: How will we treat yearly table changes or dropped tables?