



**Economic and Social
Council**

Distr.
General

ECE/CES/2006/36
3 April 2006

Original: ENGLISH

ECONOMIC COMMISSION FOR EUROPE

STATISTICAL COMMISSION

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-fourth plenary session
Paris, 13-15 June 2006
Item 6 of the provisional agenda

**SEMINAR ON POPULATION AND HOUSING CENSUSES
SESSION II**

Reengineering the Census of Population and Housing¹

Submitted by United States Census Bureau

I. ABSTRACT

1. The United States Decennial Census of Population and Housing is an essential part of the American political, economic, and social systems. Census data are critically important in achieving equitable political representation, in the distribution of over \$200 billion in Federal funds annually, and in a variety of other public and private sector uses.

2. The reengineered 2010 Census consists of three highly integrated activities designed to dramatically improve upon Census 2000. We will accomplish this by taking advantage of opportunities for innovation – made possible through the expanded use of technology – in order to: 1) increase the coverage, accuracy, and quality of census data; 2) reduce operational risk; 3) increase the relevance and timeliness of census long form data; and 4) contain costs. The three highly integrated activities we have embarked upon to meet these goals are: 1) the American Community Survey; 2) the Master Address File/Topologically Integrated Geographic Encoding

¹ This paper has been prepared at the invitation of the secretariat.

and Referencing (MAF/TIGER) Enhancements Program; and 3) a multi-year program of comprehensive planning, development, and testing for a short form only 2010 Census.

3. We expect that the cost reductions in the last component will be sufficient to offset the costs of all three components of the reengineered census. That is, all three components can be carried out at a cost that is no greater – and probably somewhat less – than the cost of repeating the Census 2000 process.

II. THE UNITED STATES DECENNIAL CENSUS OF POPULATION AND HOUSING

4. In every nation, there are unique legal, historical, operational, and political factors that shape the way a census is conducted. The United States Decennial Census is no different.

(a) Our Constitution requires that a census of the Nation's population be taken every ten years to decide how many seats in the House of Representatives will be allocated to each state.

(b) Also by Federal law, state and local governments use census data to draw legislative (including Congressional) districts of approximately equal population to comply with the constitutional "one-person-one-vote" mandate and the statutory requirements of the Federal Voting Rights Act.

(c) The Federal government also distributes over \$200 billion annually using formulae that mandate the use of, or must rely on, Decennial Census data.

(d) Federal, state, local and tribal officials use Decennial Census and other data in deciding the location of new hospitals, highways, bridges, and schools.

(e) Businesses, large and small, have come to depend on the U.S. Census Bureau's population, income, education, economic, housing, and other data to make informed decisions about locating new offices, shops, and factories, and in identifying markets for new products and services.

(f) Decennial Census results also serve as benchmarks or controls for many of the household surveys conducted by the U.S. Census Bureau and others.

5. In short, census data are critically important in achieving equitable political representation, a fair allocation of resources, and for a myriad of other public and private sector uses. These factors in turn dictate three over-arching concerns that shape the way the United States takes a census:

(a) accuracy of the population count at all geographic levels and for all demographic groups is of paramount importance to virtually all stakeholders. Given the amount of political power and funding that are affected by census counts, the U.S. Census Bureau has been involved in numerous lawsuits on this topic over the last four decades. This issue also was at the foundation of a major confrontation about Census 2000 between the U.S. Congress and the President, and a landmark decision by the U.S. Supreme Court

concerning the use of statistical sampling for determining the total population figures; (b) consistency of the data also is a key requirement for most stakeholders. To most, the unique value of decennial census counts is that they are produced at the same time for all areas using the same methods and rules everywhere. Another key element is that the census, by law, is conducted at the Federal level, and cannot be delegated to the individual States. This again relates to the fact that the census results are used to determine how many Congressional representatives, and how much Federal funding, should be allocated

to each state;

(c) content also is determined and controlled at the Federal level. By law, the Office of Management and Budget (within the Executive Office of the President) must approve all data collection efforts, and the questions to be used, for any Federal statistical effort. Also by law, the U.S. Congress must be notified of the planned topics and questions to be used.

III. CENSUS 2000 BACKGROUND

6. During Census 2000, the U.S. Census Bureau implemented multiple nation-wide operations that included:

- (a) hiring 860,000 people for data collection and data capture;
- (b) answering about 5.8 million telephone calls;
- (c) printing over 20 million paper maps;
- (d) printing 398 million questionnaires (including English and 5 other languages);
- (e) providing Questionnaire Assistance Guides in 49 languages;
- (f) opening 27,000 local Questionnaire Assistance Centers and 520 temporary field offices;
- (g) capturing data from 1.5 billion pieces of paper from March through August 2000; and
- (h) tabulating data for 9 million census blocks, and 39,000 governmental units.

IV. THE 2010 CENSUS REENGINEERING PLAN

7. In response to the lessons learned during the implementation of Census 2000, and in striving to better meet this nation's ever-expanding needs for social, demographic, and geographic information, the Department of Commerce and the U.S. Census Bureau have developed a multi-year effort to completely modernize and reengineer the 2010 Decennial Census Program. This reengineering effort has four major goals:

- (a) improve the relevance and timeliness of census long-form data;
- (b) reduce operational risk;
- (c) improve the accuracy of census coverage; and
- (d) contain costs.

8. Our plan to achieve these goals consists of three highly integrated activities designed to dramatically improve upon what was a very good census in 2000. We also will accomplish this by taking advantage of opportunities for innovation – made possible through the expanded use of technology – that will enable us to meet our above-stated goals for the 2010 Decennial Census Program.

9. The three integrated components are:

- (a) collect and tabulate detailed characteristics data every year throughout the decade through a large survey called the American Community Survey;
- (b) enhance and improve our existing Master Address File (MAF) and Topologically Integrated Geographic Encoding and Referencing (TIGER) system by bringing them into alignment with true global positioning system coordinates and converting our TIGER data base system to a commercial off-the-shelf data base environment; and
- (c) implement a program of early and comprehensive planning, development and testing designed to completely restructure the management and conduct of a short form only

(basic characteristics) census in 2010. This component also will provide the savings needed to support the reengineering initiative.

V. THE AMERICAN COMMUNITY SURVEY

10. Given the rapid demographic changes experienced in recent years, and the strong expectation that such changes will continue and accelerate, the once-a-decade data collection approach of a decennial census is no longer acceptable for producing much of the data required by the Federal government, states, municipalities and tribal governments. To meet the needs and expectations of the nation, one of the Census Bureau's approaches has been to develop the American Community Survey.

11. This survey collects decennial census long-form data every month instead of once every ten years, and the Census Bureau provides tabulations of these data on a yearly basis rather than only once each decade. After years of development and testing, the Census Bureau in 2005 implemented the American Community Survey at the full nationwide sample of three million housing units per year, or about 250,000 housing units per month. The American Community Survey will allow Federal agencies, state governments, tribal officials, and local customers to make decisions based on current information, not based on data that could be six or more years old. The Census Bureau plans to produce and release block group and tract-level² data comparable in detail to the Census 2000 long form in August of 2010, in advance of the 2010 Decennial Census short form data. The Census Bureau will release the estimates from the American Community Survey program each year, based on the previous year's data collection.

12. Just like the census long form, the American Community Survey will provide data on the following subjects: families, children, and the elderly; income and poverty; educational attainment and school enrollment; work and unemployment; disability; immigration and language ability; housing; and many more.

13. Starting in 2006, the American Community Survey will provide single-year estimates for all states, as well as for specific areas and population groups with more than 65,000 people. For smaller areas, it will take three to five years to accumulate sufficient samples to produce appropriate estimates. For example, for areas of approximately 20,000 to 65,000 people, three-year estimates will be produced starting in 2008. These multi-year estimates will be updated annually thereafter. For rural areas and city neighborhoods (census tracts and block groups) and other areas of less than 20,000 people, we will produce five-year estimates, starting in 2010. These estimates will also be updated annually thereafter.

14. The Census Bureau collects American Community Survey data in continuous, 3-month cycles using a combination of mailout/mailback, Computer Assisted Telephone Interviewing, and Computer Assisted Personal Interviewing data collection modes. Optimal use of these three modes of data collection results in cost-efficient, high-quality statistics.

15. In the first phase of the American Community Survey, mailout/mailback, the Census Bureau sends out a pre-notice letter, the initial mailing package (which includes the American Community Survey questionnaire, an instruction booklet, and other materials), and a reminder card. A replacement mailing package with a second questionnaire is mailed about three weeks after the first mailing to those who do not respond. This strategy of four mailings is designed to maximize response, and the inclusion of a replacement mailing to those who do not initially respond constitutes a significant operational improvement over Census 2000. Currently, English questionnaires are mailed in the United States and Spanish language questionnaires are mailed in Puerto Rico.

16. An ongoing question about this approach concerns the reference period for the resulting estimates. The U.S. National Academy of Sciences, advisory groups, data users, and Census Bureau staff have agreed that American Community Survey multi-year estimates should be conceptualized as “period” estimates.

Table 1:
Sets of Sample Cases Used in Producing American Community Survey Estimates

Data Product	Population Threshold	Year of Data Release							
		2006	2007	2008	2009	2010	2011	2012	2013
1-year Estimates	65,000+	2005	2006	2007	2008	2009	2010	2011	2012
3-year Estimates	20,000+			2005-2007	2006-2008	2007-2009	2008-2010	2009-2011	2010-2012
5-year Estimates	All Areas*					2005-2009	2006-2010	2007-2011	2008-2012

*All legal, administrative and statistical geographic areas down to the tract and block group level.

17. To illustrate, the first American Community Survey 5-year estimate is based on interviews conducted from January 2005 through December 2009. The 5-year estimates reflect the characteristics of the interviews conducted in each of the 60 months equally. No month is given preference over any other month nor is more weight given to any particular month (neither middle nor last). In essence, 5-year estimates represent the 60-month “period.” So, for a small town with a population of 10,000, the 5-year estimate will reflect the characteristics of the population as collected at the month of interview during the 60-month period.

18. Therefore, multi-year estimates will be conceptualized as period estimates that are meant to reflect the characteristics of the area over the entire data collection period: 60 months for 5-year estimates, 36 months for 3-year estimates, and 12 months for 1-year estimates. With this conceptualization, the American Community Survey estimates will be labeled through the use of the period of sample years that comprise the estimates, as shown in Table 1.

19. Each year the entire set of the American Community Survey data is weighted to produce 1-year estimates. The weighting process includes several factors, such as the probability of selection, computer-assisted personal interviewing subsampling, the monthly weighting

adjustment, and noninterview adjustment. Finally, the 1-year estimates are controlled to population estimates released for that year by the Census Bureau's longstanding official Population Estimates Program. Using these 1-year weights, the Census Bureau will release tabulated data products for all entities with a population of 65,000 or more.

20. In keeping with the conceptualization of multi-year estimates as "period" estimates, the sample cases from the entire period will be combined (or pooled) for the application of the final weighting factors. For this purpose, the estimates are controlled to the simple average of the population estimates. Using the first 5-year period estimate (2005-2009) to illustrate, the American Community Survey estimates would be controlled to a simple average of the population estimates from 2005 through 2009 released in 2010. The American Community Survey weighted population totals for a multi-year estimate will not align with any particular year of official population estimates.

21. However, citing concerns about possible negative impacts of using county level population estimates (by race/ethnicity, sex, age) in the weighting of American Community Survey data, some of our advisory committees and data users have called for the American Community Survey to discontinue the use of these population estimates as controls. This recommendation comes with a call for empirical justification that the use of these county level population estimates as controls improves the accuracy of the American Community Survey estimates.

22. Therefore, the Census Bureau has determined to undertake a comprehensive evaluation of the quality of the official population estimates we release annually. This review will include an assessment of the population estimates at various levels of geography (nation, state, county, and place) by various levels of demographic detail (total population, and populations by race/ethnicity, sex, and age groups). This research will include a comprehensive review of our population estimation models and the use of administrative records data.

23. In addition, we plan to conduct a research project to produce multiple sets of period estimates using different assumptions about the geographic level and demographic detail level of the population controls used in the weighting process, such as:

- (a) most detailed: Use the county level population estimates by race/ethnicity, sex, and age and the place level total population estimates;
- (b) roll up one level of geography: Use the state level population estimates by race/ethnicity, sex, and age and the county level total population estimates;
- (c) remove component(s) demographic detail: Use the county level population estimates by sex and age and the place level total population estimates; and
- (d) collapse geographic entities: Use the population estimates for groups of counties (meeting some minimum population threshold) by race/ethnicity, sex, and age.

24. Our goal is to determine which alternative produces the most accurate American Community Survey period estimates. Such research must be concluded by mid-2009 at the latest in order to implement any changes for our planned 2010 data release.

25. Additional improvements for the American Community Survey are unfolding at the writing of this paper. These include: the implementation of a partnership program to educate the public about the importance of American Community Survey data and to improve response; an

expanded language program to provide additional assistance to those for households with limited English speaking abilities; and, as discussed earlier, the expansion of the American Community Survey research and evaluation program. The American Community Survey is now producing data profiles for areas with population greater than 65,000.

A. The American Community Survey and Hurricanes Katrina and Rita

26. Beyond the massive operations required to address short and long-term rescue and recovery operations related to Hurricanes Katrina and Rita in August and September of 2005, a major challenge exists in developing statistics to describe the changes affecting the local population and economies so that emerging recovery needs of affected areas can be identified. Because the American Community Survey collects detailed information on the characteristics of population and housing in the counties in Alabama, Louisiana, Mississippi, and Texas affected by the hurricanes, it can be a useful tool in the recovery effort.

27. To address the impacts of Hurricanes Katrina and Rita, the Census Bureau intends to develop a special data product designed to reflect characteristics of areas in which the hurricanes had a major impact. This product will have two components, one representing the characteristics of these areas for the first eight months of 2005 (from January through August), and the other representing the characteristics of these areas for the last four months of 2005 (from September through December). The American Community Survey sample is sufficiently large to provide “before” and “after” profiles of affected areas; however, producing 8-month and 4-month estimates means that the sample sizes and the reliability of the estimates will be reduced. Other limitations of this special product include possible higher levels of nonresponse in some areas where it was very difficult to conduct interviews immediately after the hurricanes. These tabulations will be available in early summer 2006, before the official release of annual American Community Survey data.

VI. THE MAF/TIGER ENHANCEMENTS PROGRAM

28. The MAF/TIGER Enhancements Program is multifaceted -- taking advantage of well-established technology to improve continually on the methodologies currently in use, while substantially expanding geographic partnerships with state, local, and tribal governments to maintain the address and geographic information essential for a successful 2010 Census and American Community Survey. Ongoing geographic partnership programs, coupled with technological improvements (such as the use of global positioning system technology) will help to reduce or eliminate the address duplication and incorrect housing unit and group quarters location problems that hampered Census 2000. Procedures will be streamlined so that many of the duplicates that may arise during address list updating and validation can be resolved more quickly. The American Community Survey and 2010 Census field staff will be equipped with a more comprehensive, accurate, timely, and spatially-correct geographic database—one of the best predictors of a successful data collection effort. They will be able to use geographic information systems technology (with global positioning system capability) to guide them to the correct units and to use in recording the locations of both new addresses and new streets. Use of this technology will increase enumerator efficiency, facilitate identification of duplicate addresses, eliminate location errors, and reduce fieldwork—the most expensive component of any census. Coupled with opportunities made possible by American Community Survey, the

MAF/TIGER Enhancements Program must be completed to allow the Census Bureau to fully benefit from the use of global positioning system-equipped Hand Held Computing devices to find, interview, and update data about living quarters and their occupants for the short-form only census in 2010.

29. An updated and enhanced TIGER database with global positioning system coordinates will allow the Census Bureau to maintain a more complete and accurate inventory and location of addresses and features. In addition, it will greatly expand the Census Bureau's ability to improve the accuracy and completeness of the statistical and geographic information provided to data users. An updated TIGER database also will enhance data users' ability to use and further process American Community Survey and 2010 Census data using their own geographic information systems.

VII. THE 2010 SHORT FORM CENSUS

30. A sustained, multi-year, integrated program for planning, testing, and development of the Constitutionally mandated enumeration in 2010 is the third key component of our reengineering effort. Without it, we are left with a census that improves data timeliness and relevance (through the American Community Survey) and geographic accuracy (through the MAF/TIGER efforts), but at a greatly expanded cost and with no serious reductions in operational risk or improvements in coverage accuracy. With it, the data collection effort for 2010 can take advantage of and build on these other improvements to contain costs and improve accuracy while keeping operational risk to a minimum. This will be accomplished through things such as:

- (a) using global positioning system-equipped Hand Held Computing devices for data collection. Use of these devices will allow the program to make major improvements to the data collection operation—the largest and most expensive component of any census. For example, their use will significantly reduce the need for paper forms and maps, the huge staff and space required to handle that paper, and the printing, postage, and data capture costs associated with data collection using paper forms. These devices also will provide better information to field staff as they conduct their work, resulting in improved productivity and fewer errors;
- (b) mailing a second questionnaire to households that do not respond to the initial mailout. Research by the program has shown this to significantly increase mail response rates, thus lowering field followup workloads and costs by roughly 75 million dollars for each percentage point improvement in mail response;
- (c) finding ways to increase data quality for all population groups by improving questionnaire wording and instructions when collecting data about race and Hispanic Origin;
- (d) exploring ways to increase within-household coverage for all groups and areas by improving questionnaire wording and instructions regarding our residence rules;
- (e) making methodological improvements in the way data is collected for persons who live in group quarters. These will build on the knowledge gained from American Community Survey data collection in group quarters.

31. To do these things successfully, procedures must be fully tested under census-like conditions, and refined well in advance of Census Day. This requires a sustained, multi-year effort of integrated planning, development, testing, revising, and retesting of all the many

procedures needed to complete a successful census.

VIII. REVIEW OF COMPLETED AND PLANNED TESTS FOR THE 2010 CENSUS

32. Planning for the next census is ongoing throughout the decade. In 2003, the Census Bureau conducted a National Census Test, which was focused on improving the race and Hispanic Origin questions and testing ways to increase response.

33. In 2004, we conducted our first site test. In an area of New York City and in three rural counties in the State of Georgia, we tested both the mailout portion of the census, along with field data collection for nonrespondents. This was our first large-scale test using Hand Held Computers. In this test, we also tested our identification of duplicate housing units and persons. Overall, we were pleased with the test--particularly our ability to use Hand Held Computers to conduct personal visits for nonresponse followup. Evaluation reports for all of our research questions are available.

34. In 2005, we conducted a second National Census Test. In this test, we continued our efforts at improving the race and Hispanic Origin questions, as well as testing improvements to the other short form questions. We also focused on ways to improve our presentation of residence rules instructions, and for identifying households that might have complicated living situations that might require a followup interview to make sure the correct people were counted. Results from this test now are being studied, and reports will be issued later this year.

35. In 2006, we are conducting a second site test. This test is being conducted in the City of Austin, Texas, and on the Cheyenne River Indian Reservation in South Dakota. This test has many proposed objectives, including expanding the use of Hand Held Computers to our address listing and mapping operations, testing components of our coverage measurement program, and testing the enumeration of persons residing in group quarters, such as prisons and group homes.

36. The final testing stage will be the dress rehearsal census in 2008. We have selected two sites, the San Joaquin County, California, and the city of Fayetteville and nine surrounding counties in North Carolina. We will attempt to conduct a census in 2008 the way we intend to conduct it in 2010. Actual implementation of the 2010 Census will begin in 2009 when we conduct address listing and mapping operations.

IX. ADDRESSING THE ESSENTIAL FEATURES OF POPULATION AND HOUSING CENSUS

37. We address the following essential features that distinguish a population and housing census from other data collections as listed below.

(a) Individual enumeration--The United States always has obtained information on each enumerated person and living quarter so that their characteristics can be separately recorded. This allows cross-classifying the various characteristics and obtaining data by more than one characteristic.

(b) Specific reference period/Simultaneity/Unique reference period--For its decennial censuses, the United States uses a well-defined and unique reference period (Census Day) when it enumerates the population. The decennial census is designed to describe the

population and housing as of Census Day. In contrast, the American Community Survey is designed to describe more detailed population and housing characteristics for a specific time period – reflecting either a single-year, 3-years, or 5-years.

(c) Comprehensive results/Comprehensiveness/Universality/Benchmark/Full coverage--In the United States, decennial census data are collected and produced at the same time for all areas using the same methods and rules everywhere. No sampling is used to collect the data that are the basis of the apportionment and redistricting. Coverage measurement has been an integral part of the decennial census program since 1950. Also, for many decades, decennial census results have served as benchmarks or controls for many of the household surveys conducted by the U.S. Census Bureau and others.

(d) Small-area data--The United States decennial census produces counts of population and housing for all geographic levels from the Nation down to individual blocks, consistent with protecting individual confidentiality.

(e) Defined periodicity--The United States has produced decennial census data every 10 years since 1790.

² Census *tracts* are small, relatively permanent statistical subdivisions of a county or statistically equivalent entity, usually delineated by local data users. They are designed to be relatively homogenous with respect to population characteristics, and have an optimum size of 4,000 people. They generally follow relatively permanent visible features or governmental unit boundaries. A census *block*, a subdivision of a tract, is the smallest geographic entity for which the Census Bureau collects and tabulates 100 percent census data. Many blocks correspond to individual city blocks bounded by streets, but in rural areas may include many square miles. A census *block group* is the smallest geographic entity for which the Census Bureau collects and tabulates sample data. Also a subdivision of a tract, a block group is a set of generally compact and contiguous blocks.

* * * * *