United Nations ECE/CES/GE.41/2015/10

Economic and Social Council

Distr.: General
21 July 2015

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Seventeenth Meeting
Geneva, 30 September to 2 October 2015
Item 2 of the provisional agenda
Efficient approaches for the census in a time of increasing financial constraints

Reducing the Cost for the 2020 Decennial Census of the United States[1]

Note by the United States Census Bureau

Summary

The United States Congress, which appropriates funding for the decennial census program, has directed the Census Bureau to design and conduct the 2020 Census for a cost less per housing unit than the 2010 Census, while maintaining the highest levels of quality. The Census Bureau has embraced this goal and is fundamentally changing the way it will conduct the next decennial census. The Census Bureau began this effort in 2012 by identifying the major cost drivers of the decennial census. From 2013 through 2015, the Census Bureau conducted research and testing related to major innovations that showed promise of significant cost savings. Across four key innovation areas, the Census Bureau believes it can produce up to $5 billion in savings relative to repeating the 2010 Census design and operations in 2020. The Census Bureau will use the results of the research and testing, other key information (such as results from past evaluations, experiments, and tests), and input from a wide variety of stakeholders to make major design decisions by the end of September 2015. Based on these decisions, the Census Bureau will document and publish its 2020 Census Operational Plan, accompanied by revised estimates of cost savings and the overall 2020 Census lifecycle budget. This paper provides an in depth look at the main innovations and test census results.

[1] Disclaimer: This report is released to inform interested parties of research and to encourage discussion. Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

I. Introduction

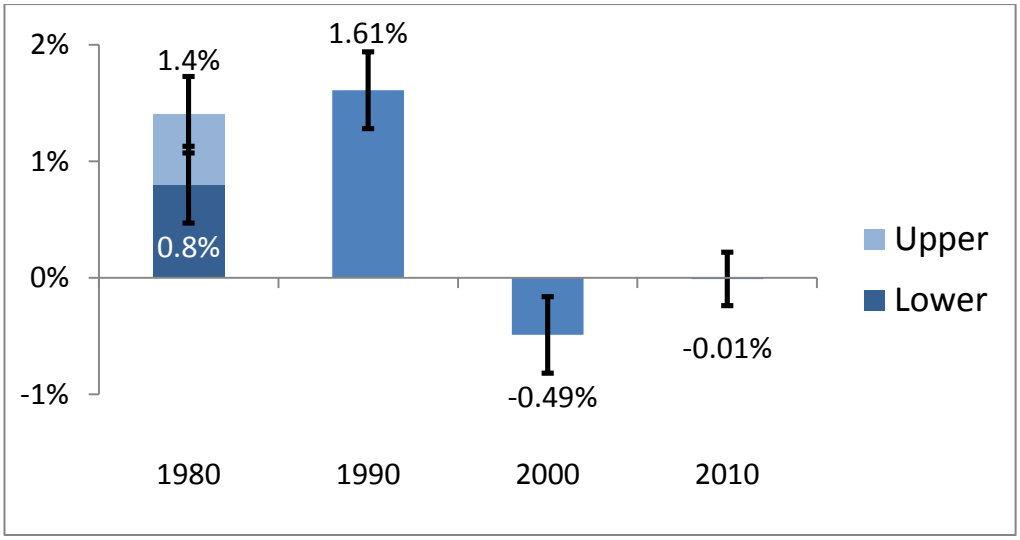
1. The United States Constitution requires that a census be conducted every 10 years to apportion the United States House of Representatives amongst the states. Census data also are used by the states to determine the boundaries of each of these House districts. In addition, census results (including those from the American Community Survey) directly affect how more than US$ 400 billion per year in federal funding is allocated to state, local, and tribal governments.

II. Background about the Plans for 2020

2. In the next census of the United States (U.S.), we will be reaching out to an increasingly diverse and growing population of around 330 million people in more than 140 million housing units. The U.S. Census Bureau is conducting research to inform key decisions about the design and operational plan for the 2020 Census. These decisions will be made by the end of this fiscal year, September 2015.

3. Over the past four decades, reducing the undercount has been a major goal for the Decennial Census. As Chart 1 shows, we have made great progress. By 2010, the estimated net undercount was nearly zero--less than ½ of 1 percent. We want to maintain quality results like this for the 2020 Census.

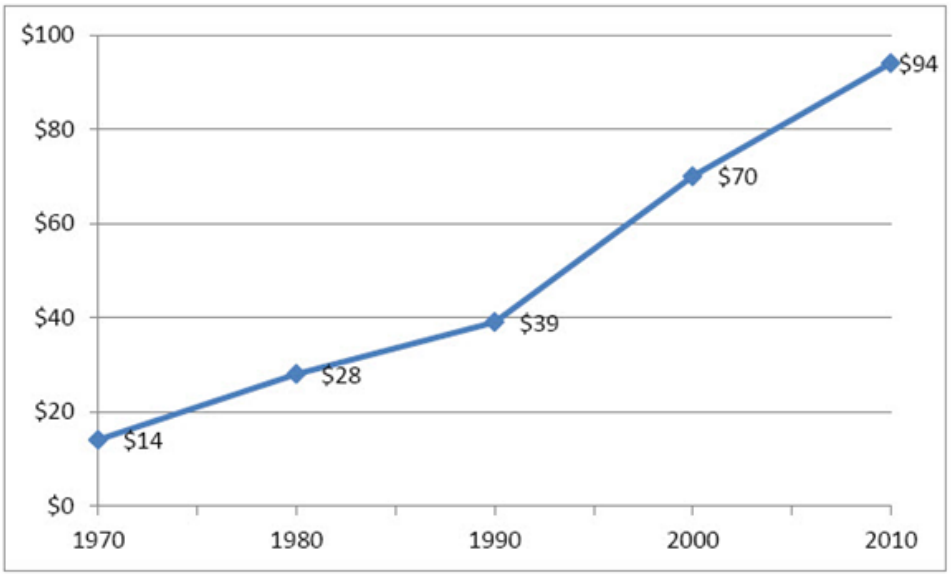
Chart 1
Percentage Net Undercount, U.S. Total, 1980 to 2010



4. The lines above and below the point estimates represent upper and lower bounds of a 90% confidence interval. For 1980, we had a series of estimates, and the upper and lower numbers shown here represent the highest and lowest point estimates of the series.

5. Over the same time, however, the cost of the decennial census has increased. As Chart 2 shows, after factoring out the effects of inflation and growth in the population, the cost of conducting censuses has steadily increased over time.

Chart 2
Cost per Housing Unit (1970 -2010) in 2010 dollars



6. For 1970-2000, the unit costs reflect a combined short-form and long-form census. The unit cost for the 2010 Census includes (1) cost for the short-form only Census, (2) costs for American Community Survey, which is conducted more frequently and replaced the long form, and (3) costs for geographic improvement programs over the same decade. After backing these costs out, the total lifecycle cost (in 2010 constant dollars) for the 2010 Census by itself was $10.3 billion.

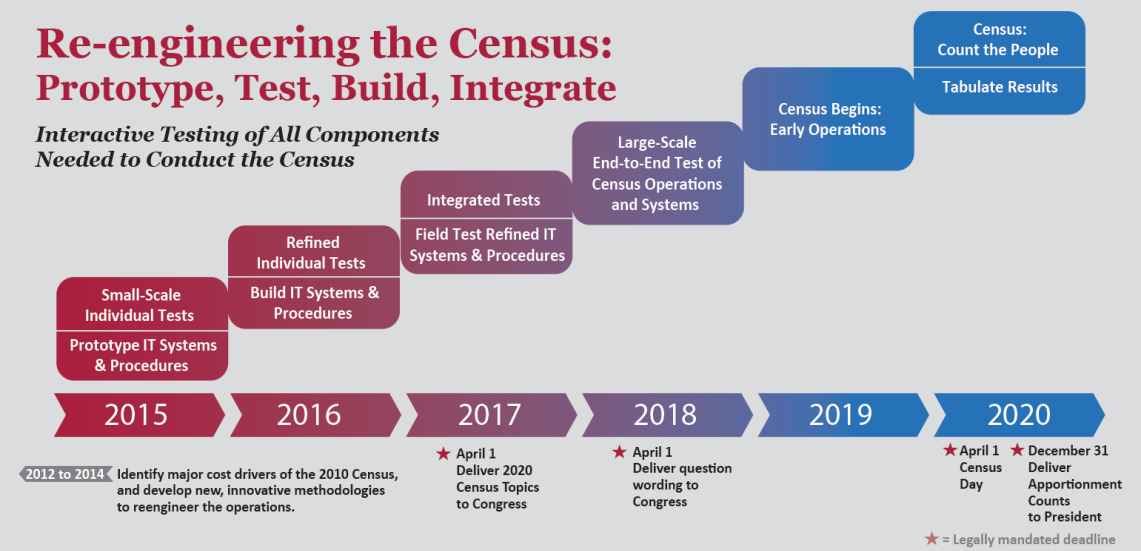
7. Unless significant changes are made to the goals and the way we conduct the census, the costs will continue to rise at a similar rate by 2020. If we were to conduct the 2020 census like the 2010 census, we estimate cost projections to be $17.7 billion over the life cycle. Congress has indicated, and the Census Bureau agrees, that these escalating costs are unsustainable.

8. Subsequent analysis of the 2010 Census revealed that five factors were driving costs. These factors are as follows:

- Increased diversity of the population
- Demand for the Census Bureau to strive for improving accuracy over previous censuses
- Lack of full public participation in the self-response phase of the census, requiring hiring of a large field staff for nonresponse follow-up
- Challenges managing major acquisitions, schedule, and budget
- Substantial investments in major, national updating of the address list prior to enumeration

9. Further, analysis into the operations and activities associated with these cost drivers resulted in identifying key innovation areas for 2020. As shown in the figure below, over the course of the next six years, from 2015 to 2020, the Census Bureau has charted a path for prototyping, testing, building and integrating all components needed to conduct the census to achieve cost savings.

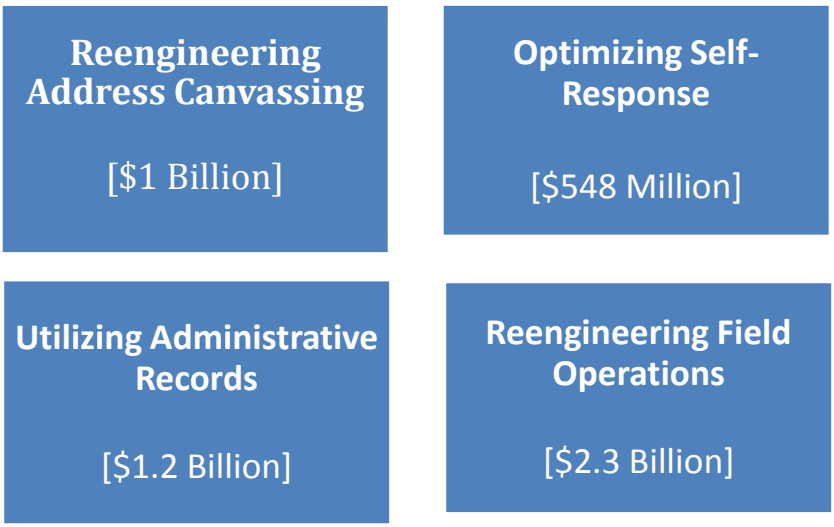
Figure 1
Path for prototyping, testing, building and integrating components



10. Thus far, we have concentrated primarily on technology-based and methodology innovations with the greatest opportunity to address the major cost drivers, as well as those that further enhance our understanding of the response process and therefore the quality of the census itself. We then developed a research and testing plan that carries us to September 2015 for the key design decisions. From there we need to refine field operations, build and test IT systems, and then integrate all parts of the plan in time to conduct an end-to-end test in 2018.

III. Innovation and Research Areas

11. We identified four major innovation areas where we believe can save more than $5 billion over the life cycle of the 2020 census, conducing it for an estimated $12.2 billion instead of the projected $17.7 billion cost. These innovation areas cover four main areas as follows:



12. These innovation areas show the estimated cost savings based on analysis conducted in 2015. In September 2015, we will re-evaluate these estimates based on tests conducted to-date, with the new estimates released in conjunction with the operational plan mentioned earlier. Each of these innovation and research areas are discussed in the sections that follow.

IV. Reengineering Address Canvassing

Goal: To reduce the need for a nationwide in-field address canvassing operation. Instead, the Census Bureau plans to develop innovative alternative methodologies for updating the Master Address File (MAF)/Topologically Integrated Geographic Encoding and Referencing (TIGER) System throughout the decade, using a combination of in-field and in-office techniques.

These design changes have the potential to save the Census Bureau $1 billion.

13. For the 2020 Census, we have redefined address canvassing to be a combination of in-office and in-field canvassing where we will continue to canvass every block, but we will only conduct in-field, on the ground, canvassing where it is necessary.

14. The goal of reengineering address canvassing is to eliminate a nationwide in-field address canvassing in 2019, as was done for the prior 2010 census. Instead, the Census Bureau proposes to utilize the use of aerial imagery and change detection techniques, supported by statistical models, to identify the areas where in-field canvassing is required. Over the past two years, we have worked to acquire data to inform the new process.

15. We now have terabytes of data from federal, state and local government sources, as well as the private sector to update our Master Address File (MAF) and determine where in-field work is required. Specific federal files we are working with include the following: extracts from the United States Postal Service, Indian Health Service Registration File, U.S. Department of Housing and Urban Development Public and Indian Housing Information File, and the Selective Service Registration File. In addition to federal files, we maintain an extensive partnership program with state and local governments. For years, we have received, processed and ingested "partner" files into our address frame. We are also using private sector data where we have found gaps that exist in the governmental data.

16. We just concluded our first test of the use of statistical models to detect change. Tested in 2014, and referred to as the Address Validation Test, we are currently conducting data analysis of the test. In this test, we developed two independent statistical models that predict the stability of a census block, and then we conducted a dependent listing operation of 10,100 blocks in the United States. Currently, we are comparing the result of the dependent listing to the output from the statistical models to assess the veracity of the models. A second component of this test identified 29 counties in the United States for which we had aerial imagery in-house that overlapped with the 10,100 blocks. We used both automated and manual change detection techniques in those 29 counties to identify blocks with change. Of those, we identified approximately 700 blocks as suitable for an in-field canvass. We are comparing the results of this in-field canvass with the results of the change detection work, as well as the results of the dependent listing operation.

V. Optimizing Self-Response

Goal: To communicate the importance of the 2020 Census to the United States population in order to generate the largest possible self-response. This in turn will reduce the need to follow up with non-responding households in person, which has always been the most expensive operation of the census. Changes under consideration for 2020 include providing an Internet response option, using targeted digital advertising via social media, encouraging the use of the Internet as the primary response mode through advertising and partnership efforts, and allowing respondents to submit a questionnaire without a unique identification code. We are studying some fundamental questions on optimizing the content and methods of contacting respondents through email.

These design changes have the potential to save the Census Bureau $548 million.

17. The design changes include testing different strategies for contacting respondents and motivating them to self-respond. We are testing various self-response modes such as Internet, telephone response, and paper. The scope of the tests included using digital micro-targeted advertising methods to increase awareness and engagement of the public and to allow them the option to respond without a unique Census-provided identification code (also known as the "Non-ID" option). With Non-ID processing, we tested the ability to do real-time processing, which matched response to information on the Census Bureau's master address and geospatial files. Additionally we are studying the use of the Internet data collection application for mobile devices.

18. Next year, we will continue tests for optimizing self-response by conducting a 2016 Field Operations Test. For the first time, the Census Bureau will test the use of a cloud-based infrastructure to host its 2016 Field Operations Test questionnaire. The test includes the best use of targeted communications and partnerships to promote language support options and to reach historically hard-to-count populations. Additionally, this test will continue refining methods to process responses in real time submitted without a unique Census-provided identification code. The test will also determine system performance parameters by measuring the systems' abilities to manage a significant number of concurrent users during self-response.

VI. Utilizing Administrative Records

Goal: To use administrative record data (e.g., information from federal and state governments) and third-party data (e.g., information from commercial sources) to reduce the Nonresponse Follow-up (NRFU) workload. The data sources may also be used to enumerate the population in cases of nonresponse.

These design changes have the potential to save the Census Bureau $1.2 billion.

19. The use of these data is not new for the United States. We have used administrative data, such as tax data, for decades to improve data collections. Today, there is a new generation of big data – as the electronic environment flourishes – as a means to improve statistics. We are researching ways to utilize these new data sources in new ways to increase efficiencies and reduce costs. At the same time, we must also continue to maintain the quality of the official statistics.

20. One application of the use of administrative data in the 2020 Census design is in our planning and operationalization of our nonresponse follow-up operation. During the 2010 Census, the nonresponse follow-up universe included 50 million housing units. Each of

those units received at least one personal visit, resulting in the identification of 31 million occupied and 14 million vacant housing units. We deleted another five million units that did not meet the Census Bureau's definition of a housing unit. Vacant and deleted units accounted for about 38 percent of the non-responding universe. The use of administrative data to avoid the expense of conducting a personal visit to an address to discover vacant units for which no questionnaire was, or could be, returned is one of the key cost drivers of the Census. By using administrative data to identify these vacant units, we can eliminate and reduce the field effort.

21. In terms of our initial research and testing in this area, our primary focus has been on the use of government provided files (both federal and state level files). Where necessary, we used commercial files to supplement existing federal data in our research and testing. We have conducted three field tests of these new methodologies, and to date, the research has been promising and has informed changes in the methodologies employed in each round of testing.

VII. Reengineering Field Operations

Goal: To use technology to more efficiently and effectively conduct and manage the 2020 Census fieldwork. The Census Bureau plans to develop an operational control system that manages tasks and makes some of the decisions typically made by humans (e.g., case assignments, timing and number of contact attempts). The work also will develop more effective and efficient ways of implementing and managing field operations through a new field structure, revised and new field staff roles, new approaches to scheduling work, and changes to staffing ratios.

These design changes have the potential to save the Census Bureau $2.3 billion.

22. The three components of field reengineering are the operational control system, the COMPASS application, and the field organizational structure.

A. Operational Control System

23. The Census Bureau plans to develop an operational control system that manages tasks and makes decisions typically made by humans (e.g., case assignments, contact attempts). Additional modernization includes a streamlined approach to implementing and managing field operations through a new field management structure, including the infrastructure, field staff roles, work schedules, and staffing ratios.

24. Specifically, we are utilizing big data to help us in the efficient assignment of work. Our operational control system (case management system) will consider outstanding work, enumerator home location, and enumerator attributes to tailor a work assignment that is appropriate for each enumerator on a daily basis. Enumerator assistance from remote operations centers also reduces the enumerator-supervisor ratio.

25. The operational control system will use real-time paradata to manage the work and the enumerators by providing alerts. For example, in real-time the system will alert supervisory level employees about specific performance or conduct issues. An alert could be as simple as the employee completed cases for the day but did not send in their corresponding payroll and expense information. It can be as complex as an employee's device location being far from where we believe the work to be.

26. Behind the scenes, a significant component of this effort, and the use of big data, is in the development of the Constructive Simulation Model. This model is core to the

operational control system. It is a three-dimensional model that utilizes 2010 Census data, data from the American Community Survey, and real-time paradata to determine the best time for our enumerators to contact households. This information is combined with household data, enumerator data, and business stopping rules to determine an enumerator's case assignments and the route they should follow each day.

27. Inputs into the model include GIS inputs, 2010 Census data (Census response data, payroll data, and employee location data), information from the Master Address File, and paradata from the American Community Survey. The resulting outputs of this model are an assigned daily case list for enumerators, predicted travel time for all legs within the optimized route, estimated time to find location for all case contacts, estimated interview time, and contact outcome for all case contacts (in order to set a standard number of cases to provide enumerators per shift).

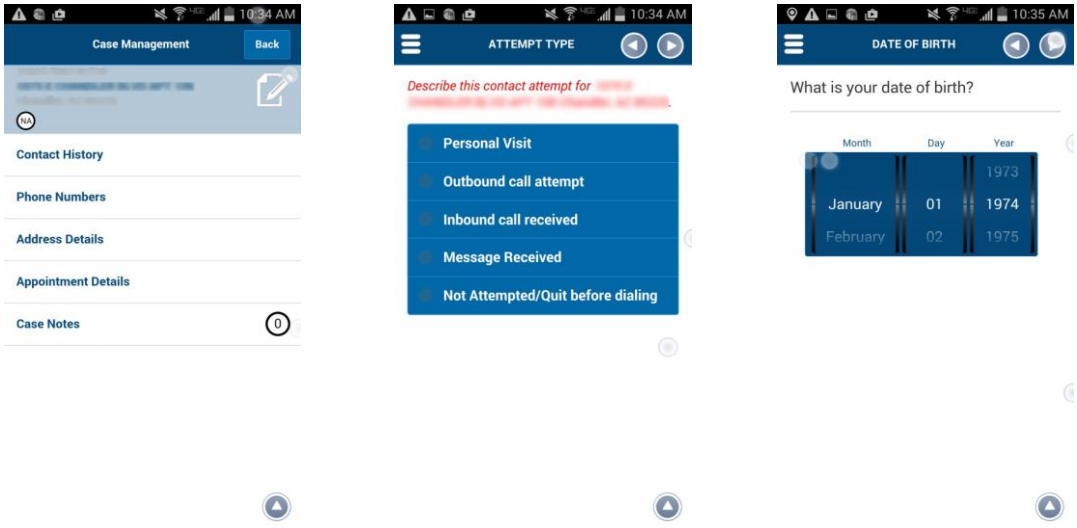
28. We conducted one successful large-scale simulation of this prototype. We put it in the field for the first time in the 2015 Census Test. We are very excited about the promise our new control system holds for the Census and the surveys conducted by the Census Bureau.

B. COMPASS Mobile Application

29. COMPASS is a mobile application that will help enumerators conduct the non-response follow-up survey. The device will allow enumerators to manage the field telephone workload, collect CAPI responses, and collect paradata. Additionally, enumerators will be able to submit their time and attendance sheets from their phone, further eliminating the paper processing of time sheets associated with the last decennial operation.

30. These three pictures, with data blurred out, are what an enumerator sees on their handheld device.

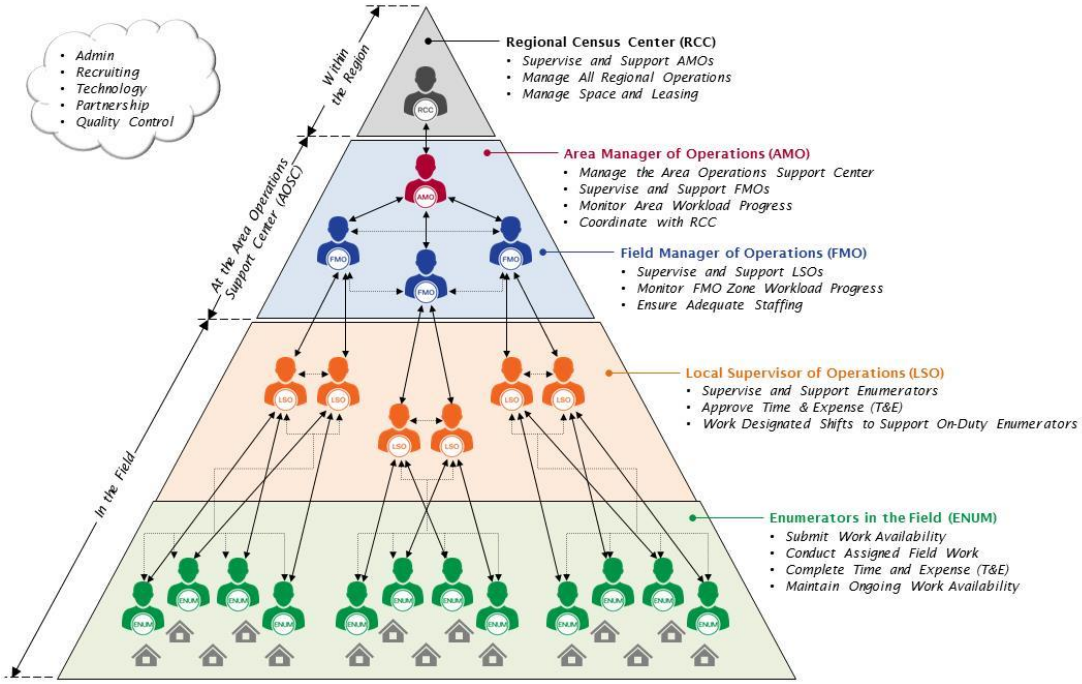
Figure 2
COMPASS Mobile Application



C. Field Organizational Structure

31. The elimination of paper and the introduction of automated enumerator assignments change the traditional field organizational structure. The following diagram illustrates that with a new organizational structure we are looking to separate the management of staff from the management of workload.

Figure 3
Field Organizational Structure



32. Looking at the pyramid, at the bottom level we have the enumerators who will perform data collection. At the next level, we have the Local Supervisor of Operations, or LSOs, who will supervise and manage enumerators and staff resources in the field. Typically, the ratio would be eight enumerators to one LSO under the old structure, but with the automation, the ratio will increase significantly.

33. At the next level up in the pyramid are the Field Managers of Operations, or FMOs, and the Area Manager of Operations, or AMO, who will work out of the Area Operations Support Center or AOSC. The FMOs will manage operational performance and will supervise the LSOs. Each FMO has a primary geographic area of responsibility, but they will work as a team to ensure the completion of work for the entire AOSC.

34. A professional level Area Manager of Operations, AMO, will manage an Area Office Support Center (AOSC), supervise FMOs, and ensure completion of attempt workload in the time expected for their assigned area.

35. The new structure reflects less management levels than the last decennial. Additionally, we have yet to determine the number of temporary structures to house the management for administration, recruiting, quality assurance, and technology for the fieldwork. Largely, this will depend upon automation and centralization of certain functions related to these operations.

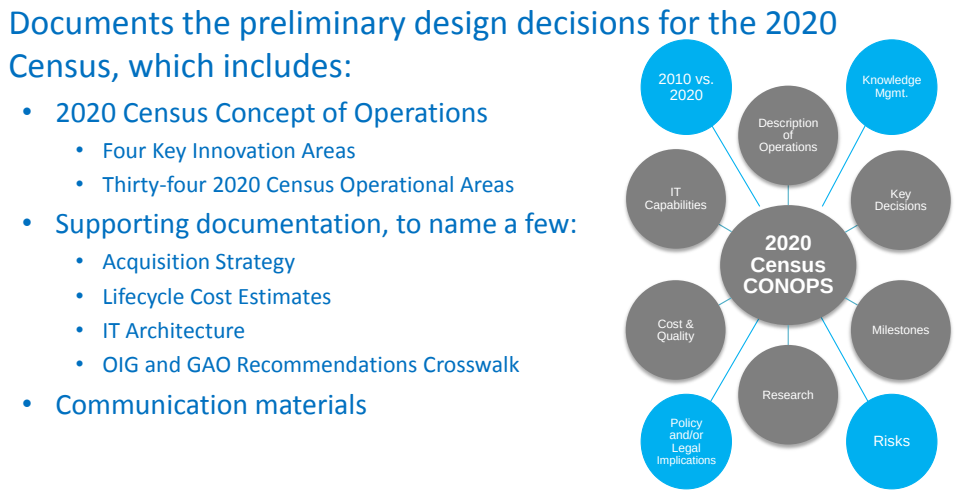
1. 2020 Census Operational Plan

36. We are at the mid-point in the decade where we must make the choices that set the major operations in motion. The last two years have focused on extensive research, testing, and design. In 2015, we spent the better part of the year narrowing our efforts to set major design decisions.

37. The 2020 Concept of Operations (CONOPS) describes how the 2020 program will operate to execute the 2020 Strategic Plan. It describes the operational intent and context, and describes how the Census Bureau will execute the 2020 census.

38. As shown in the figure below, the CONOPS will have many components.

Figure 4
CONOPS components



39. Within the 2020 Census Operational Plan, the CONOPS will include heavy focus on the four key design areas (Reengineering Address Canvassing, Optimizing Self Response, Utilizing Administrative Records, and Reengineering Field Operations} as they relate to the survey life-cycle operations. The plan covers 34 major operations associated with the census. The narrative will have varying levels of maturity for those operations based on the research and testing completed to date. This is important to note because the CONOPS we release in October 2015 will document what we know to date about each operation – as well as what we still need to research and determine.

2. Cost/Quality Analyses

40. The Census Bureau has been researching and testing several new methods that we expect will produce significant costs savings, but that may affect the overall quality of the Census. We are evaluating these new methods, and testing that they can meet our high standard of quality while also meeting the goal to reduce costs.

41. There are many aspects of quality when applied to the census count. The key is accuracy, but accuracy can be measured in many ways. For example, under an alternative census design,

- how close would the census count be to "the truth," that is, some benchmark?
- how accurate are the counts for the 50 states and the District of Columbia?
- do we miss people in the inner city, or the rural parts of the country?

- do we miss people with specific demographic characteristics, such as Hispanics?
- do we count some people twice, or where they do not live most of the time?

42. We also measure and monitor characteristics of our operations that do not guarantee high quality by themselves, but have been shown to be highly correlated with the quality of the numbers we release. For example, evaluations have demonstrated that a higher response to our census questionnaire mail-out (the return rate) is an indicator of higher quality and fewer errors in our census count. Similarly, we monitor the completion rate for census enumerations with regard to the status of the housing unit (e.g., occupied, vacant, nonexistent) and completed characteristics (e.g., race, owner, renter). A higher completion rate by itself does not guarantee higher quality. However, higher accuracy generally follows when we obtain the response directly from a household resident, rather than from one of our methods of "imputation"—that is, statistically assigning a characteristic value.

VIII. Stakeholder and Oversight Activities

43. Another aspect of our efforts this decade is maintaining an open and transparent process with our stakeholders, including oversight groups.

44. For example, we produce a Monthly Status Report (MSR) for Census Bureau leadership and key staff, as well as key stakeholders. The stakeholders cover representatives (including program auditors from the Office of Inspector General) from the U.S. Department of Commerce (our parent agency), the Office of Management and Budget, staff from our authorizing and appropriations committees in Congress, and program auditors from the Government Accountability Office.

45. Once each quarter we conduct a Program Management Review (PMR) for these same audiences. The purpose of the PMRs is to provide information and status (accomplishments, risks, challenges, near term focus, etc.) for the 2020 Census program. Beginning this year, the PMRs are broadcast live (and later archived) on our website so that more stakeholders and interested parties can obtain current information on our plans and progress. This information is found at http://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/program-briefings.html.

46. We also make presentations and engage with many other groups, including our formal advisory committees, and a panel of experts at the National Academy of Sciences. In addition, we attend and make presentations at meetings of many other groups, such as the American Statistical Association, the Population Association of America, the U.S. Conference of Mayors, and at various international meetings.

IX. Conclusion

47. We are now half way through the decade, and in fiscal year 2016 we will move beyond the research and testing phase. We will have made our major design decisions at the end of September 2015; after, we begin the work of executing this design. The Census Bureau will continue to refine the design of the reengineered 2020 Census, as well as implement some early 2020 operations and legally mandated programs that must begin next year. See the companion paper for this conference: “Changing Times, Changing Methods, Changing Technologies: Innovation and Testing for the U. S. 2020 Population and Housing Census,” by Arona L. Pistiner, Senior Advisor for International Collaboration and Policy, Decennial Communication and Budget Office, United States Census Bureau.

48. The 2020 Decennial Census will look and operate quite differently than any previous census in the United States. Many of the innovations we plan for 2020 are due to

new or maturing technological capabilities. Our goal is to find ways to use these innovations effectively to control costs without sacrificing quality.