

WRITTEN STATEMENT OF
JOHN THOMPSON
DIRECTOR, U.S. CENSUS BUREAU
U.S. DEPARTMENT OF COMMERCE

ON THE
“2020 CENSUS: CHALLENGES FACING THE BUREAU FOR
A MODERN, COST-EFFECTIVE SURVEY”

BEFORE THE
SENATE COMMITTEE ON HOMELAND SECURITY AND GOVERNMENTAL AFFAIRS

April 20, 2015

On behalf of the Department of Commerce and the U.S. Census Bureau, I thank Chairman Johnson and Ranking Member Carper for the opportunity to testify today regarding the preparation for the 2020 Census. I look forward to today’s hearing and to our ongoing discussions as we prepare for the 2020 Census.

Looking back, as we began this decade, the Census Bureau, with the guidance of Congress, established an important goal to design and conduct the 2020 Census in a manner that costs less per housing unit than the 2010 Census and to maintain quality. The Census Bureau then set out to identify the major cost drivers and to develop innovative enumeration methods aimed at reducing these costs. With Congress’ support, we focused our research and testing in these early years on major innovations oriented around technology and the strategic use of information to rein in those cost drivers. When I returned to the Census Bureau in 2013, it was because I believed the goal of an accurate, cost-effective census was possible. Today, I would like to describe our work in four key innovation areas.

We believe that investing now in these four key innovation areas can yield up to \$5 billion in savings, but I need to underscore that in order to deliver these savings, we must complete the current and future tests. The tests we conducted in 2013 and 2014, and the four tests we are conducting this year are helping us along the critical path toward the 2020 Census design decisions, which we will deliver later this year. I will discuss how these tests have informed our planning throughout the testimony, and how critical the research and testing we will be conducting in the next two years will position us for the end-to-end test in 2018. The alternative to implementing the new design is repeating the 2010 Census design in 2020, which would forfeit the potential savings of \$5 billion.

These are the four key areas of innovation and the potential savings they represent:

- **Better Address Validation:** by using the U.S. Postal Service and other information sources, including aerial imagery, we plan to avoid walking every street in the nation to validate and update the address list. We can potentially save approximately \$1 billion.
- **Better Response Options:** by making responding to the census more convenient whether through the Internet, phone, or by mail, we can potentially save approximately \$548 million.

- ***Better Use of Existing Information***: by using existing government and commercial information to reduce the need to follow up with non-responding housing units, we can potentially save \$1.2 billion.
- ***Better Field Operations***: by effectively using technology to manage and track cases, as well as route the census takers who will be using smartphones and tablets, rather than pencil and paper, we can potentially save \$2.3 billion.

While these innovation areas highlight potential savings, we are also mindful of the importance of quality. We believe that achieving important and meaningful cost-savings in these areas, we will be able to concentrate more effectively in reaching hard-to-reach communities and ensuring that everyone has the opportunity to participate in the 2020 Census.

The foundation of an accurate census is an accurate address frame, which includes both the address and the geospatial location of all the housing units in the United States. For the past several censuses, this has required walking every street in every community in the United States to verify the designation and location of every housing unit the year before the census. Our vision for Address Canvassing has changed, and we now see the possibility of “in-office” canvassing in addition to “in-field” canvassing. Recognizing the technological advances and the increasing availability of data over the past few years, we may no longer have to validate every address by personal visit. In-office address canvassing is now possible relying on a mix of automated and interactive techniques, and we can do this throughout the decade without having to wait for the year before the census. Moreover, even if we have to send people out to the field, we may only need them to validate part of a census block, rather than walking every street to verify every address within the block. Now, we have to determine the best way to detect change and effectively manage those changes within the Master Address File (MAF). We are considering not only how to refine our old procedures, but also how we can integrate data and services from the private sector, and have recently issued a Request for Proposals (RFP) related to updating our geospatial assets. Specifically, we want to purchase address, road, and satellite or aerial imagery data, as well as potentially collect new data for new areas, and we looking to the private sector to help identify areas where change is occurring.

This year’s Address Validation Test encompasses two components that will also bring insights into this innovation strategy. The purpose of the MAF Validation Test component is to assess the ability of statistical models to predict change, whether that is new roads, new housing units, or other changes, as well as to validate the quality of the MAF. This is a nationwide test of more than 10,000 blocks. The Partial Block Canvassing Test component focuses on the use of in-office methodologies such as imagery to detect change and update our records, as well as our ability to implement an in-field canvassing of partial census blocks. We expect to have data from both of these components by the end of May.

After better address validation, the next area of innovation focus is developing better response options. We are moving away from the traditional method of relying solely on the mailout questionnaire and the enumerator to count every household. We are expanding options for people to self-respond by Internet, either at home or on a mobile device, and by phone. Those individuals and households without access to or the desire to use the Internet could still respond by paper.

Our test in Savannah, Georgia, which began earlier this year, allows us to conduct early research on the use of advertising and partnerships to engage and motivate respondents and to validate Internet responses in a way to organize our field workload and ensure accuracy. So far, we are seeing a self-response rate of about 35% for the Maricopa site, and about 31% for the Savannah site. At this same point in time during the 2010 Census (17 days into the self-response phase), the Maricopa site had a self-response rate of about 47%, and the Savannah site had a self-response rate of about 41%. We expect to see lower response rates in tests compared to the actual decennial census, but we are encouraged to see a differential between the two sites that is less than what it was in 2010. I think that is almost certainly due to the effect of the more extensive advertising, social media, and partnership efforts in the Savannah site.

We want to make the census as mobile as possible, and this means allowing respondents to answer the census without entering a “census-ID,” which is a unique number used to identify each housing unit on our Master Address File. In fact, we anticipate that by promoting the Internet option there will be an increased number of these responses in 2020. In the past, we allowed a very limited number of non-ID responses, and therefore the process of validating, or “non-ID” processing, was limited, and meant having to verify the response at the back-end. For 2020, we need to be able to “validate” these responses quickly, in real-time. This will permit us to remove the housing unit from the workload and not send an enumerator to the door, which is also a cost-savings. We also need to be able to validate the response to ensure that it is accurate, and not a duplicate or fraudulent. We are researching and testing ways to compare responses to data we have in-house, whether from other government agencies or commercial sources, as well as the potential of having a third party perform the validation.

The use of these types of data highlights the third area of innovation: better use of existing information. The use of administrative records from other federal and state government agencies and the use of third-party commercial data has tremendous potential to reduce costs. The use of administrative records is not entirely new to the decennial census. We have used records in certain types of enumeration situations, such as prisons, colleges, and nursing homes. We have relied on the records from the Departments of Defense and State, as well as other agencies, to produce the tallies of the military and federal families serving overseas.

Under the Census Bureau’s authorizing legislation, we currently use a variety of administrative records and third party data across many programs to accomplish our mission more efficiently, to save taxpayer money, and to produce innovative products, all while protecting privacy and confidentiality. We receive data from many federal agencies, such as the Internal Revenue Service, the Social Security Administration, the Postal Service, and the Centers for Medicare & Medicaid Services, as well as state governments and commercial vendors. Using administrative records is highly cost-effective and reduces the reporting burden on the American people by utilizing information they have already provided. All of this information is protected and is only used to create statistical products. We protect these data in the very same ways we protect the information we collect from individuals and businesses, through rigorous physical, procedural, and information security protocols that meet the Census Bureau’s legal obligations, as well as the obligations imposed by the tax code and other federal information security requirements, such as the Federal Information Security Management Act (FISMA).

Based on our knowledge and experiences with administrative records and commercial data, we want to expand our use of these data for the 2020 Census. We are exploring several options and two of the most promising are using these data to help manage and even reduce the field workload, which is historically the most expensive operation of the decennial census. The potential cost savings represented by this area of innovation is \$1.2 billion. We already tested some of our strategies for using these data in the 2013 and 2014 tests, and we are testing key enumeration approaches in Maricopa County, Arizona. We are specifically looking at administrative records to remove vacant housing units and to use the information on persons to enumerate occupied housing units, thereby removing them from the non-response follow-up (NRFU) operations.

For example, during the 2010 Census, the field workload included 50 million housing units. Each of these housing units received at least one in-person visit. Of these 50 million housing units, 14 million were vacant and another 5 million were deleted because they were either non-existent or were not in fact housing units. We sent an enumerator to each one of these more than 19 million housing units just to discover that there was no one living there to answer the questionnaire. By using data from the Postal Service and other agencies, such as the Internal Revenue Service (IRS), we believe we can determine whether housing units are vacant and remove these units from the field workload achieving substantial cost-savings.

As we proceed in this area of innovation, I want to assure you that we are moving methodically and purposefully, and in full consideration of our obligations to protect privacy and confidentiality. We asked our National Advisory Committee to form a working group and to advise on the privacy and quality implications of using these data, and we are also seeking stakeholder input through various forums.

To effectively implement these strategies, we not only have to effectively utilize the information we are already receiving from other government agencies, but we also need to seek new sources of information. In some instances, such as tax data, we need to work with the IRS to negotiate the most thorough, effective and direct use of the data to enumerate non-responding housing units. The Census Bureau is also currently evaluating several federal records sources with the intention of potentially negotiating agreements with agencies such as the Departments of Defense and Homeland Security.

However, one of the most important sources, the National Directory of New Hires (NDNH), is currently unavailable to the Census Bureau because we do not have access to these data under the Social Security Act. This database supplies information on workers, including the newly hired, which we could use to corroborate and supplement other information on workers received in the tax data. NDNH also includes information on persons receiving unemployment benefits, which we could use to improve our coverage of this hard-to-reach population. Last year, we developed draft language at the request of this Committee in consultation with the Department of Health and Human Services. The language provided limited access for the Census Bureau to use these data for statistical purposes and ensured protection of those data under the Census Law. In addition, the President's FY 2016 Budget also highlights our need for these data. Without

NDNH, we lose out on an information asset that could be critically valuable in achieving our cost-savings goal.

Ultimately though, we will have to send census takers into the field to enumerate the remaining non-responding households, which underscores the importance of the fourth area of innovation: better field operations. As part of the research and testing, including a particular focus of the 2015 Census Test in Maricopa, we are taking a deep look at our field staffing structure and we are testing out several technological innovations. The goal of reengineering our field operations is to use technology more efficiently and effectively to conduct and manage the 2020 Census field workload.

In previous censuses, we did not equip the enumerators with any type of automated device, relying instead on paper and pencil to collect all of the information out of the field, including census responses and the enumerator time and attendance. We had limited ability to communicate with the staff in the field, instead relying on in-person meetings or telephone calls between the enumerators and their supervisors. The supervisors had to meet with the census takers at least on a daily basis, if not more frequently, to collect the census responses and all the administrative paperwork, and then the supervisors had to take the paperwork to the local census offices. Ultimately, all of this time spent collecting paperwork was a significant contributor to the overall cost of the previous census field operations. Moreover, because we relied on paper, there was very little analytical capacity that could be used to better manage the staff or redirect resources.

We plan to develop a sophisticated operational control system that utilizes information effectively to manage tasks and assignments, which we had in the past left to humans using their best judgment. Our goals are to incorporate operational best practices including the optimization of daily assignments, intelligent routing, and real-time, responsive management of issues. We are working with the private sector again, including companies such as the United Parcel Service and Esri, to learn more about these types of best practices. We also intend to take full advantage of technology to send our enumerators out with mobile devices, rather than pencil and paper. They will use devices to collect responses and report their time and attendance, instead of the paper forms we used in 2010.

The four key innovation areas represent significant cost savings that can be achieved only if we get the opportunity to fully test these alternative methods in the next two years, which will allow us the opportunity to conduct an end-to-end test in 2018. To get there, we will release a design document, called a concept of operations, at the end of this fiscal year. This document will describe: the proposed operations; key decisions and the basis and working assumptions for current and future decisions; milestones for future decisions; research efforts; cost and quality trade-offs; and the IT capabilities necessary to support the operational implementation of the 2020 Census.

In reaching this point, we have received strong support from the Department of Commerce and the Office of Management and Budget. We have also received strong support from our overseers, including this Committee and the Government Accountability Office, as well as the Office of the Inspector General. We appreciate the opportunity to consult and receive guidance, especially as we plan, conduct, and evaluate the critical research and testing associated with

these areas of innovation, which we believe will generate even more refined and improved cost-estimates. We are also intent on improving our capabilities and skill sets in the area of cost estimation to further ensure the integrity of and support for the 2020 Census design moving forward. Finally, we also are committed to ensuring a transparent process, which incorporates comprehensive documentation and schedules, as a further measure to ensure the integrity of and support for the census.

I will close by reiterating the importance of the current testing and the continued research and testing, especially next year in FY 2016. We have very little time left to ensure the effective implementation of these innovations before we reach 2018, when the costs of the census begin to rise steeply. We need to conduct real-world tests now, not only to be ready, but also to ensure that we can maintain quality. These tests are leading us toward having complete methodological approaches and systems development for the end-to-end test in 2018. FY 2016 is also critically important to the 2020 Census timeline, because we will award several large contracts, such as the census questionnaire assistance and the advertising and partnership contracts.

Finally, FY 2016 is also important for the development of the key systems that we will use to support the infrastructure to handle data collection and processing. Known as the Census Enterprise Data Collection and Processing (CEDCaP) program, we are moving to a smarter, more cost-efficient enterprise level strategy to manage core information technology aspects for all of our censuses and surveys. In the past we have often built single one-off systems for the current surveys and the censuses. In the case of the decennial census, we built systems that were retired at the end of the each census. Now, we intend to create an integrated shared-service, which we can use continuously across all of our programs, that is adaptive and scalable. The CEDCaP effort will streamline the collection and processing for the Census Bureau as an enterprise, which will save taxpayer money. It is crucial to the 2020 Census that resources are available to develop these systems in FY 2016; otherwise, we may not be ready for the end-to-end test in 2018.

As a final point, I must mention our commitment to protecting the privacy and confidentiality of individuals' information. The Census Bureau takes cyber security very seriously and we stand by our record. To protect our information systems and the information we collect, the Census Bureau has implemented a robust, comprehensive, and layered cyber security program. Some of the key points of the program are:

- The Census Bureau employs a Managed Trusted Internet Protocol Service on all external Internet links and it utilizes the Department of Homeland Security's resources that further protect Internet traffic, by looking for malicious code and suspicious activity on these links.
- Other key safeguards include 2-factor authentication to access the Census Bureau network remotely; use of encryption in transmissions of data and data at rest; use of a Data Loss Prevention solution, which looks at outbound email traffic to ensure no sensitive information is transmitted in an unencrypted state; use of firewalls, Intrusion Detection Systems and Intrusion Prevention Systems and a dedicated cyber security staff that monitors these systems and investigates unusual activity; and the use of a Virtual Desktop Infrastructure that provides further security.
- All Census Bureau systems are compliant per National Institute of Standards and Technology (NIST) Federal Standards and the Census Bureau. The Census Bureau security program

incorporates continuous monitoring of all IT systems, and we have implemented a Risk Management Framework (RMF) program that is fully compliant with NIST standards and guidelines documentation.

- Finally, all employees receive annual training on data stewardship and security.

From CEDCaP and cyber security to each of the tests, each activity plays a significant role in helping the Census Bureau to design an accurate and cost-effective 2020 Census. While the census is simple in concept, it is in reality a complex program, ranging over many years and many operations, which must be synchronized to meet our ultimate mandate. The savings we predict cannot be effectively implemented without the rigorous testing that is designed to inform development of the systems and operations and without your support.

Thank you and I hope this update has been informative, and I look forward to your questions.