The Operational Design of the 2020 Census: Overview of the Current Status

Robin A Pennington
U.S. Census Bureau

The United States decennial census has a long history of being a monumental effort of considerable complexity. Questionnaires and field data collection procedures have to account for a wide variety of situations. At the same time, cultural norms related to relationship status, race and ethnicity, and even sex of the household members are always evolving. The need to understand family and household configurations is increasing, even as these concepts increase in complexity. This means questionnaire design, response options, and field procedures must continue to be updated. In addition, these updated options need to be accommodated within online self-response and automated field data collection instruments.

At the same time, there is an expectation that the 2020 Census will cost no more per household than the 2010 Census did, in 2010 dollars. This creates pressure to find ways to cut costs but also provides incentive to find major innovations in how census data will be collected (Blumerman, et al). For example, collection of the census data by Internet is now an expectation. There are considerable costs associated with this option, as well as security risks that must be addressed. On the other hand, there are cascading savings to be derived from using less paper. The potential savings start with printing and mailing significantly fewer paper questionnaires and from fewer respondents mailing back their completed questionnaires to data processing centers. More savings would accrue from having less paper processing, due to the costs of the building footprint, scalability of systems to handle huge quantities of paper questionnaires through various processes, and staffing of the data processing centers.

Another cost-saving path is related to creation of the frame for taking the census. From creating the address list at the time of enumeration up until the 1970 Census through maintaining and updating the address list in a database since 1990 to collecting Global Positioning System (GPS) coordinates on housing structure records during the 2010 Census, there have been major innovations in census frame development in recent decades. In this decade there has been continued effort to use the United States Postal Service (USPS) addresses from the Delivery Sequence File (DSF) and inputs from local, tribal, and state governmental units to keep the address list current. These efforts in tandem make it possible to consider eliminating the necessity to perform a nationwide in-field canvass to update the address list prior to the 2020 Census. Additional in-office work at the Census Bureau, described below, makes this option viable.

The prime target for cost savings in the decennial census is the Nonresponse Followup (NRFU) operation. This operation, in which the population that does not respond in the self-response phase is enumerated, is the most expensive census operation. Efforts to minimize the NRFU workload are one key factor to achieving cost savings in the

decennial census. This is planned to be achieved through multiple strategies, the first of which is maximizing self-response, especially through the Internet. Another is automating field operations, with one key cost-avoidance feature being that the NRFU universe can be regularly updated throughout the NRFU operation. Any household can submit a self-response throughout the course of the census and be removed from the universe of addresses requiring follow-up. With continued updating of the NRFU universe, the fieldwork requires planning strategies such as routing and work assignments that are redefined as the universe is updated. Less field management is needed as a result of the automation. Additionally, continued efforts to contact household members to complete the enumeration often prove to be quite expensive, and sometimes do not result in high-quality data. Administrative record or other third-party data of sufficiently high quality may be available such that expensive follow-up efforts to obtain the enumeration can be curtailed. Automation of all aspects of NRFU can also result in cost savings stemming from not having to process and store paper questionnaires and payroll transactions. The automation and redesign of NRFU thus entails multiple opportunities to re-envision processes and potentially find cost savings.

These opportunities have been categorized into four major areas of innovation planned for implementation in the 2020 Census. These are: Reengineering Address Canvassing, Optimizing Self-Response, Utilizing Administrative Records and Third-Party Data, and Reengineering Field Operations. This paper provides the context for how these programs fit together to create a comprehensive plan for completing a census of the same quality as the 2010 Census without increasing per household cost. Some key features of these programs will be discussed in the papers related to this overview document.

The motivation for Reengineering Address Canvassing lies in understanding what the operation has been up until this census. The fieldwork to create the address list of the enumeration universe for the entire country is responsible for much of the cost of the address frame creation. In addition, while address lists built by the Census Bureau using a variety of sources were used in 1970, 1980, and 1990 in areas where questionnaires were mailed and expected to be returned by mail, much of the country still required hand delivery of questionnaires or enumeration at the door (U.S. Census Bureau, 1990; U.S. Census Bureau, 2000). An early innovation applied to this process was that the Address Control File of 1990 Census addresses was input into a database. The Census Bureau's Geography Division maintains this database, called the Master Address File (MAF). The Geography Division also maintains the Census Bureau digital map, the Topologically Integrated Geographic Encoding and Referencing System (TIGER). An extract of addresses from the MAF served as the control address file for both Census 2000 and the 2010 Census, and TIGER was used for the maps for field operations. Numerous additional updates have been made to these databases. For the MAF, this includes changes resulting from census operations and periodic updates from the DSF. Prior to the 2010 Census the TIGER's street map network was updated with spatially accurate data, including GPS coordinates, and this database was integrated with the MAF, forming the MAF/TIGER System. The field updating of the address list during the Address

Canvassing operation incorporated the collection of GPS coordinates for residential structures and the addition of these data points to the MAF/TIGER System. As a result of these innovations, the activity of walking the ground to update the address list prior to the 2020 Census can be greatly reduced from walking every street in the country as was performed in prior censuses. In much of the country the DSF reflects address and housing changes. Similarly, local governments can also provide updates on the housing stock under their jurisdiction through programs offered through the Census Bureau's Geography Division. Therefore, this Address Canvassing operation will be needed only in those areas where change in the housing stock is not captured through other means. The challenge is identifying these areas without incurring the cost of going into the field.

In the fall of 2015, the Census Bureau began the In-Office Address Canvassing operation, using more scientific means of identifying housing change that has not been captured in other sources. For example imagery review of two different vintages can show areas where change has occurred, and a review of the current MAF will show if the updates have been incorporated. Various tests are planned to validate the in-office procedures or provide insight into how procedures could be improved. The In-Office review of the address frame is planned to continue through early 2019. As of the date of this publication, over 50 percent of blocks have been completed through the initial In-Office review to determine if the MAF matches the housing inventory on the ground.

A procedure that complements the In-Office Address Canvassing is known as the MAF Coverage Study. This planned ongoing field work provides a measure of the coverage of the address frame on the MAF in total and for various types of areas. The MAF Coverage Study will be used to evaluate the In-Office Address Canvassing results and potentially to improve In-Field Address Canvassing data collection methodologies. Boyd, *et al*'s paper provides details on this operation and how it works to complete our understanding of coverage and quality in advance of the census.

The second major innovation area for the 2020 Census is Optimizing Self-Response. There are several innovations related to this area, which work simultaneously. The biggest innovation is incorporation of an Internet response option, available for all households. The self-response modes that were available in previous censuses — paper questionnaires as the primary response mode and enumeration by a telephone call to the Census Bureau used to a much lesser extent — will still be available in the 2020 Census, as well. Respondents are also allowed to answer through the Internet or telephone modes even without the Census ID that is used for tracking responses back to the census universe. This is known as Non-ID Processing. The option to provide a questionnaire without the Census ID was available in previous censuses in a more limited sense. It was implemented primarily as a means of including people in the Census who did not believe they had been counted, either because their housing unit did not receive a questionnaire or because they believed they were left off the questionnaire submitted for the household. Electronic and real-time means of linking the response for the reported address to the

address record on census files even without the Census ID allows this program to be expanded to all households for the 2020 Census.

The Non-ID Processing operation is just one means of optimizing self-response to counteract the decrease in response rates observed across the spectrum of survey and census data collections. The hope is that the availability of Internet for any type of response will help make response rates as high as possible. In addition, there has been a great deal of testing related to the delivery of materials and the timing of the delivery in order to enhance response rates and minimize follow-up operations. Bentley's paper provides background and test results that have guided decisions about communications, mailings, and response options.

The highest cost operation is the follow-up on nonresponding units, or those units that do not respond after the Optimizing Self-Response strategies are implemented. Traditionally this Nonresponse Followup (NRFU) operation has been entirely a field data collection operation, with a large field office management infrastructure. A number of major innovations are related to reducing the cost to get responses for the nonresponding units. A primary area of research is the use of administrative records and third-party data to reduce the data collection in the field.

In particular, there are two stages within the NRFU operation where these data have been applied with success. The first is to remove vacant units from the workload. In previous censuses where we used an initial address frame, we have not allowed or provided the means for a respondent to inform us that a unit is vacant, but rather have required a personal visit during the NRFU operation to determine the status of the housing unit. We have also required this status to be confirmed by two separate enumerators or processes, to minimize the risk of excluding people from the census. One of the allowed processes was a status code from the USPS indicating that the questionnaire could not be delivered to the housing unit because it was vacant. The plan for the 2020 Census could be seen as a major expansion of this effort. We will continue to use the status code from the USPS and other third-party data that we believe to be of high quality to derive a status of vacant for units on our census address frame and avoid or minimize field work for that case. At the same time, we will send these units a final mailing to encourage response in case this derived status is in error.

A second effort, with more complicated constraints, is related to deriving the criteria under which we would remove occupied units from the NRFU workload because we believe we have data of sufficiently high quality from administrative records and third-party data sources. Any such data used for this purpose must be able to provide the number of residents of the household along with their demographic and relationship data. Mule, *et al*'s paper describes the research on both of these efforts and the current status of our plans for implementation in the 2020 Census.

Another of the NRFU innovations falls under the category of Adaptive Design. This is an effort to model what approach to getting a response for cases in this universe is likely to be most effective. That is — when is somebody most likely to be at home, and how many attempts should be made to collect the data at a given household — using paradata and other sources of information as input. In previous censuses an enumerator received a paper listing showing all the units in a geographic area that required follow-up for enumeration and worked the listing until it was completed. With the implementation of an adaptable contact strategy and the automation of all processes, an enumerator's field work assignments are now recalculated every day based on responses received and other data. Potentially a householder could become aware of NRFU field work happening in the area through communications about the Census, a notice of visit left at the door by a NRFU enumerator, or such things as conversations with neighbors, and could provide a response by Internet, by telephone, or by mailing back a paper questionnaire. The case can be removed from the NRFU workload once an enumeration is received and processed through any of the available modes. Daily field assignments will be made through automation using inputs such as enumerator schedule availability, enumerator distance to assignment area, contact strategy using Adaptive Design, and an updated universe file. Even the routing for the assignment will be provided to the enumerators. The management for the NRFU operation can now be more centralized and automated. In addition, electronic means of training the field staff and the automation of the collection of their time and mileage data creates more efficiencies and cost savings. Adams presents results of testing these new procedures in the 2015 Census Test.

The driving force behind the design of the 2020 Census has been to minimize cost while maintaining quality. The Census Bureau has parsed these Innovation Area projects into individual efforts in order to describe the innovations and facilitate progress on the individual components. In reality, each innovation has impacts on other efforts, and the software systems responsible for control of the individual operations need to account for all the connections, as well. Architectural diagrams, which are products of our systems development design, identify all the connections. Diagrams for the particular operations identified in the 2020 Census Operational Plan (U.S. Census Bureau, 2015) will be presented in the Detailed Operational Plans that are currently in the process of development for these operations. The Detailed Operational Plans will provide the level of detail and specificity that is required for an understanding of the output and contributions of each individual operation. Altogether the complete documentation will provide a thorough understanding of how the census will be performed with full automation of all field operations, multiple response options available for all, including the Internet, continually updated universes through a variety of inputs, and more efficient and adaptive processes, allowing for cost savings as compared to repeating the 2010 Census design.

References:

Adams, T., Field Re-engineering for the 2020 Census, 2016, in *JSM Proceedings*, American Statistical Association

Bentley, M., Optimizing Self-Response for the 2020 Census, 2016, in *JSM Proceedings*, American Statistical Association

Blumerman, L., Bishop, D., Dinwiddie, J., June 2016, Plans and Innovations for the 2020 Decennial Census of the United States, *Statistical Journal of the International Association for Official Statistics*

Boyd, E., Letourneau, E., Williams, A., 2016, An Overview of the Master Address File Coverage Study, in *JSM Proceedings*, American Statistical Association

Mule, V., Keller, A., 2016, Using Administrative Records to Identify Occupied and Vacant Units, in *JSM Proceedings*, American Statistical Association

U.S. Census Bureau. History of the 1990 Census [Internet]. Available from: http://www.census.gov/history/pdf/1990proceduralhistory.pdf

U.S. Census Bureau. History: 2000 Census of Population and Housing [Internet]. Available from: http://www.census.gov/history/pdf/Census2000v2.pdf.

U.S. Census Bureau. November 2015, 2020 Census Operational Plan -- a new design for the 21st century [Internet]. Version 1.1. Available from: http://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan.pdf.