Attachment B: Description of Approved Research Projects

1. **Research, Testing and Operations for the 2020 Census**

   The Census Bureau intends to test and use data from many sources, including public and private sources, to design and conduct the 2020 Census. The Census Bureau is using state program information, along with data from federal, state, and private sources to pursue the following four research initiatives leading to the integration of these data to augment or replace Census operations:

   A. validating and enhancing the Master Address File (MAF);[1]
   B. designing and assigning resources to carry out the 2020 Census;[2]
   C. unduplicating public, private, and census lists; and
   D. imputing missing data.[3]

   To determine the feasibility of incorporating the administrative records data into Census operations, the data will be explored using 2010 decennial census data and intercensal test data to assess the quality and utility of household roster, location, and demographic characteristics data. Disagreement observed in content across sources will be analyzed to develop approaches such as modeling or business rules to resolve conflicting information. Administrative records will be assessed to tabulate data at various levels of geography including the housing unit, block, tract, and state levels. These tabulations will be compared to the 2010 Census and intercensal test data to assess accuracy based on the level of aggregation. Research and testing includes simulations with 2010 Census data as well as using these data in in operational site tests with the end goal of eventually incorporating the data into decennial census operations.

2. **Evaluation and Improvement of Demographic Surveys Using Tribal, State and Federal Administrative Records**

   The U.S. Census Bureau administers multiple household and business surveys that are designed to characterize the program participation characteristics of the U.S. population. This project will evaluate the accuracy of such survey data, explore the determinants of misreporting and the impact of misreporting on estimates of population characteristics. This project will also evaluate and improve current procedures for editing and/or imputing survey data related to program participation. The project will also use the tribal and state data to investigate survey frame coverage, survey effects and survey non-response.

---

1 Examples of possible Census Bureau uses include updating and evaluating the Master Address File.

2 Examples of possible data uses include simulating a targeted address canvassing, modeling to determine data collection modes, modeling to predict self-response probabilities, simulating non-response follow-up designs, directed advertising campaigns, and special enumeration operations planning.

3 Examples of possible Census Bureau uses include item imputation and count imputation using information from a composite of public and private data sources.

This project will focus on evaluating and improving the measurement of tribal and state programs in the American Community Survey (ACS), the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP). In the case of SIPP, this evaluation will also inform the redesign of the SIPP survey instrument that is currently underway.

This project will include additional studies that use administrative records from tribal, state and federal programs (including the state/tribal data) and/or Census Bureau survey data, and that can inform the Census Bureau about the quality of its data products and the feasibility of using administrative records to improve Census Bureau data products. These additional studies could include linkages to surveys conducted by the National Center for Health Statistics (NCHS), and linkages to datasets created from joint statistical projects between the Census Bureau and NCHS, the USDA Economic Research Service, the Bureau of Justice Statistics, the Department of Health and Human Services and the Department of Veterans Affairs.

3. **Characterizing WIC participants and likely eligible non-participants of WIC**

   WIC data linked to ACS data can provide characteristic information for women, infants, and children participating in the WIC program. Additionally, infants and children with a high probability of being eligible for WIC can be estimated using ACS information on family structure, reported income, and public assistance.[4] The estimated eligible infants and children can be matched to the administrative WIC data to determine who is estimated to be eligible but not participating in the program, helping agencies determine unmet need. The project results can help tribal and state agencies improve outreach efforts. This research helps the Census Bureau better understand their survey and census data.

4. **Improve Census Bureau Record Linkage Methods**

   The Census Bureau's Person Identification Validation System (PVS) uses administrative records from a variety of sources to assign, where possible, unique person identifiers to each record in its surveys, Censuses and administrative records. However, coverage for some populations may be incomplete in the PVS. Incomplete coverage in the PVS would have important implications for the Census Bureau's ability to evaluate and improve its data products using administrative records. The use of the state/tribal program administrative records will permit an assessment of the coverage completeness in PVS while improving upon coverage in the PVS of the populations participating in tribal or state programs.

5. **Research, Testing and Operations for the Decennial Census Programs and Current Surveys**

   The Census Bureau intends to test the use of administrative records data from many sources, including public and private sources, for use in decennial programs to include the Decennial Censuses, American Community Survey, and current economic and demographic surveys. The Census Bureau plans to use federal, state, tribal, and third party source data for cost reduction and quality improvement initiatives. These data will

   ---
   4 Eligible pregnant and breastfeeding women cannot be identified in ACS data.

be considered for use to augment or replace data collection and processing operations. Further, development of effective linkages between the Census Bureau's foundational datasets (including, but not limited to LEHD's job frame, the Business register, MAF/TIGER, and the to-be-developed demographic frame) will facilitate reuse of administrative data across the enterprise.

To determine the feasibility and value of incorporating administrative data into census and survey operations, the data will be linked to decennial census data, survey data, and intercensal test data to assess household roster, location, and demographic and economic characteristics data. Disagreement observed in content across sources will be analyzed to develop approaches such as modeling or business rules to resolve conflicting information. Administrative data will be assessed to tabulate data at various levels of geography, including the housing unit, block, tract, municipality, county, state and tribal levels. These tabulations will be compared to previous Census data, survey data, and intercensal test data to assess accuracy based on the level of aggregation. Research and testing includes simulations with decennial Census data, as well as using these data in operational test sites to incorporate the data into administrative data repositories for use in decennial census and current survey operations.

This work will be done in coordination with the other projects listed.

6. **Leveraging Administrative Data to Support Evidence-Building and Evidence Based Decision Making**

The acquisition and use of administrative records data is becoming increasingly central to the Census Bureau's mission. Like many agencies in the Federal Statistical System, the Census Bureau uses administrative records data to reduce inquiries on the public and to promote the efficient use of federal funds. In addition, the 2018 Foundations for Evidence-Based Policymaking Act has resulted in the Census Bureau focusing on ways to leverage research to support evidence-based decision-making. Central to these goals is the Census Bureau's Data Linkage Infrastructure as well as internal and external research conducted through the Federal Statistical Research Data Center (FSRDC) network.

This project supports the use of administrative records data to foster program and policy research and the efficient use of public funds through evidence-building activities. Evidence-building is broadly defined as the production of estimates that can support policy and program evaluation. This activity does not necessarily involve a direct evaluation of a program or policy, but rather produces statistics that can be used to support evidence-based decision-making.

This project includes research that will:

- Explore the use of administrative records to promote the efficient use of federal funds, and to produce and improve statistics on the American population and economy.

- Develop population estimates on demographic characteristics such as race, ethnicity, income, socioeconomic status, migration, food security, inequality, housing, health status and outcomes, and other characteristics that can be used to evaluate federal programs and policies and be used by policy makers to make

evidence-based decisions.

- Directly evaluate federal, state, local, and tribal programs and policies.

As Census and other Federal Statistical Agencies integrate administrative records data for research purposes, they acquire and use the data within the context of comprehensive data stewardship programs that ensure strong confidentiality and privacy protections. Data access is limited to Census Bureau staff and Special Sworn Status (SSS) researchers.

Both internal and external proposals requesting access to some (*linked, PII removed*) state level administrative records data, prior to Census Bureau approval, may be forwarded for determination to the federal agency which administers the state level program, for a determination whether that project furthers administration of the federal/state level program and to ensure it meets the requirements of authorizing statutes.

7. **Characterizing likely eligible non-participants of SNAP**

Persons with a high probability of being eligible for SNAP can be modeled using ACS data on household composition and reported income. The modeled eligible persons can be matched to the administrative SNAP data to determine who participated in the program. This research is important for the Census Bureau as it permits a greater understanding of whether SNAP take-up is uniformly distributed across demographic and geographic groups. Disparities among persons with different languages spoken at home, race or ethnic identification, or urbanicity could affect interpretations of Supplemental Poverty Measure estimates. This research also benefits Food and Nutrition Service (FNS), and tribal and state agencies by assessing the characteristics of persons with unmet needs. The project results in aggregate information that tribal and state agencies can use to improve outreach efforts.

8. **Longitudinal Data for Research**

The Food and Nutrition Service (FNS) and the Census Bureau intend to establish a longitudinal database that contains information under a unique and anonymized identifier about households and members of households that receive Supplemental Nutrition Assistance Program (SNAP) benefits in each State, per the Agricultural Improvement Act of 2018 (2018 Farm Bill).

The SNAP longitudinal database will include monthly information about households, and members of households, that receive benefits under SNAP in the participating State taken from existing information collected by the State agency, such as –

1. Demographic Characteristics;
2. Income and financial resources (as described in section 5(g) of the 2018 Farm Bill);
3. Employment status;
4. Household circumstances, such as deductible expenses; and
5. Amount of monthly allotment received under SNAP; and

6. MAY include information from other State data sources such as –

    a) Earnings and employment data from the State department of labor;
    b) Health insurance program data; or
    c) Data from participation in other programs administered by the State.

Data contained within the SNAP longitudinal database will be used solely to conduct research on participation in and the operation of SNAP, including duration of participation in SNAP and also for Census Bureau research related to data linkage improvement, survey improvement, decennial research and operations. The term "solely" refers to the requirement that the SNAP longitudinal database must be used for research to inform program administration of SNAP.

Benefit to FNS: This data will provide an opportunity to conduct longitudinal research on SNAP participation, across States and over time, to better understand the dynamics of participation patterns to inform SNAP policy decisions.

Benefits to Census Bureau: These data will provide an opportunity to conduct research related to data linkage improvement, survey improvement, and decennial research and operations.