

Memorandum

From: Karin Mac Donald, Director, Statewide Database (SWDB)

To: California Citizens Redistricting Commission; State of California Legislative Leadership

Date: 3/26/2021

Re: Census Bureau Legacy Data

On February 12, 2021, the Census Bureau published a revised timeline for the delivery of the P.L. 94-171 Redistricting Datafiles. These data had been anticipated for release by July 30, 2021 and that date has now been changed to a release by September 30, 2021. This announcement resulted in considerable push-back and legal action by states with redistricting and election deadlines that could not be met with that schedule. Subsequently, the Census Bureau announced that their continued evaluation of plans and processes had resulted in a determination that an interim version of the P.L. 94-171 dataset could be released at an earlier date: by mid to late August of 2021. This interim data product is referred to as the 'legacy' dataset.

Upon receipt of this information, Statewide Database staff immediately began to research four issues related to the 'legacy' data:

1. What exactly are these 'legacy' data and how do they differ from the P.L. 94-171 that will be released by census by September 30, 2021?
2. If the actual data in the legacy product are the same, can they be accurately converted into the format that will be released by census by September 30, 2021?
3. If the legacy data are the same and they can be accurately converted, how long will this process take?
4. What are the cost implications and what supplemental funding, if any, would need to be requested from the Legislature?

1. Our research found, and conversations with the Census Bureau's Redistricting and Voting Rights Data Office confirmed, that the 'legacy' data are the same data that will be released by September 30, 2021 in the 'traditionally' formatted P.L. 94-171 file. The 'legacy' datafiles consist of the final data product that is essentially not user friendly and necessitates more advanced database, analysis and manipulation skills to be usable for redistricting purposes than the later release of the P.L. 94-171.

Having confirmed in that these data are identical and will not be changed by the Bureau between August and September, we moved on to answer question 2.

2. The 'legacy' dataset will be released in the same format as a prototype dataset that is available on the census website. SWDB staff downloaded the prototype data (which consists of data for a small state) and evaluated the specific tasks that needed to be undertaken to manipulate and extract the necessary data, along with the complexity of implementing these tasks for a state as large as California, with it's ~530,000 census blocks. We determined that we are well equipped to undertake these tasks and to produce an accurate dataset, considering that SWDB has a 25year+ track record of producing much more complex datasets on even more units of analysis.

Because SWDB would, upon completion of the reformatting of the 'legacy' dataset, immediately move into the 30-day process of building the State's redistricting database, we determined that independent verification of accuracy prior to the Census Bureau's September release of the data, would be a well-advised best practice. SWDB subsequently reached out to California's State Demographer, Dr. Walter Schwarm, who serves as the Chief of the Demographic Research Unit (DRU) of the State's Department of Finance. The Demographic Research Unit represents the State in national programs relating to population statistics including the Federal-State Cooperative Program for Population Projections and Estimates, and the Census Data Center Programs. Collectively we determined that the best way to ensure accuracy of the dataset would be to set up parallel processing of the data at SWDB and the DRU, which will allow for data to be processed using different systems independent of each other, and be compared along the same parameters.

In sum, we are confident that we will be able to process and reformat these data accurately.

3. Once the collaboration and basic methodology had been established, we determined that a 2-week period of time from the release of the 'legacy' data would be a reasonable timeframe to complete the processing, run accuracy checks, compare the data and allow for any discrepancies to be analyzed and resolved. This 2-week period would be in addition to, and precede, the 30-day period of time that the SWDB will require to build the State's official redistricting database.

4. We are still awaiting technical documentation for the data and are preliminarily working on the code to set up the processing. At this point, we have not completed our budget estimate for this process.