



NATIONAL CONGRESS OF AMERICAN INDIANS

POLICY RESEARCH CENTER

May 2021



Research Policy Update

Differential Privacy and the 2020 Census: A Guide to the Data and Impacts on American Indian/Alaska Native Tribal Nations

The purpose of this research brief is to update and supplement the NCAI Policy Research Center's September 2019 Research Policy Update on *Differential Privacy and the 2020 U.S. Decennial Census*. This update describes more recent information on the U.S. Census Bureau's new adoption of Differential Privacy to protect the confidentiality of individuals who responded to the 2020 Decennial Census and how these privacy measures may negatively impact the accuracy and usability of American Indian and Alaska Native (AI/AN) census data. This update concludes with a guide to additional resources to learn more about the impacts of the Disclosure Avoidance System on AI/AN 2020 Census data.

Introduction to 2020 Census Data

The Decennial Census is a complete count of the U.S. population that occurs every ten years that the federal government is required to conduct by the U.S. Constitution.¹ In 2020, the U.S. Census Bureau aimed to count every individual living within the United States as of April 1, 2020 for the Decennial Census.

The 2020 Census included nine questions for households with one individual and seven questions for each additional individual in the household. The questions were self-response and up to the individual on how to answer. To learn about the questions and the information gathered from the 2020 Census, view **Figure 1** on the next page or visit <http://bit.ly/2vxRCRR> to view a sample of the 2020 Census Questionnaire.

In previously updates, we have described priority uses for Decennial Census data that are important to Tribal Nations, including apportionment to determine how many representatives each state receives in the U.S. House of Representatives, redistricting at federal, state, and local levels, federal funding formulas, local tribal governance, and research for the next ten years. While keeping the information submitted by individuals confidential is important, accurate and usable data is critical for these priority uses of Decennial Census data. To learn more about apportionment and other key uses of the Decennial Census data, read our [Decennial Census: Key Uses of the Data](#) Research Policy Update.

Figure 1. 2020 Decennial Census Questions

The U.S. Decennial Census included the following questions in the 2020 questionnaire:

1. How many people were living or staying in the residence on April 1, 2020?
2. Were there any additional people staying in the residence on April 1, 2020 that were not included in the previous question?
3. Is the house, apartment, or mobile home owned with a mortgage/loan, owned without a mortgage/loan, rented, or occupied without payment or rent?
4. What is your telephone number?

The following Decennial Census questions are repeated for every person counted in the household:

5. What is your first and last name?
6. Are you male or female?
7. What is your age and date of birth?
8. Are you of Hispanic, Latino, or Spanish origin?
9. What is your race?

Questions for each additional person in the household:

10. Does this person usually live or stay somewhere else?
11. How is this person related to Person 1?

To view a sample copy of the 2020 Census Questionnaire visit: <http://bit.ly/2vxRCRR>

U.S. Census Bureau Data Privacy – *Why the Privacy Methods Changed*

In 2019, the U.S. Census Bureau reported that they conducted an experiment to determine to what extent public 2010 decennial census data could be re-constructed by external parties.² Database re-construction is when two or more datasets are compared to determine likely characteristics (e.g. location, race, gender, age) of an individual or individuals within a privacy protected dataset of interest. Database re-construction is not the same as re-identification. Re-identification is a step further and is the process of figuring out a specific person's identity within the data, and re-construction only identifies likely corresponding characteristics.

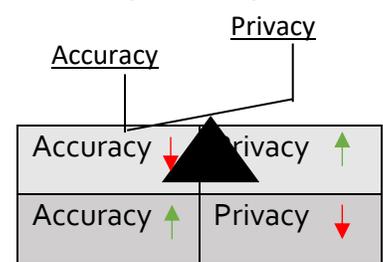
The U.S. Census Bureau's experiment succeeded in reconstructing parts of the 2010 Census public data.³ The reconstruction was able to find likely identities for some of the data with comparisons to commercial datasets.⁴ The likely identities found through the comparison between external commercial databases were then compared to secure private Census data not available outside of the U.S. Census Bureau to determine the accuracy of those likely

identities. Some of the data was re-identified correctly and some was not.⁵ The only way to know which identities were correct would be to have access to the private Census records.

The experiment to reconstruct the data and then attempt to re-identify the data with commercial datasets led to a concern for census data privacy.⁶ The U.S. Census Bureau has a requirement under Title 13 to keep census information private.⁷ The U.S. Census Bureau’s interpretation of Title 13 has shifted with the 2020 Census from protecting the privacy of individual responses to protecting the data from being compromised by external actors reconstructing and creating a potentially re-identified dataset.⁸

The U.S. Census Bureau plans to use an approach called Differential Privacy as a part of its Disclosure Avoidance System to protect the confidentiality of data in the 2020 Decennial Census dataset and its public tabulations. To learn more about the basics of Differential Privacy, read [Differential Privacy and the 2020 U.S. Decennial Census: Impact on American Indian and Alaska Native Data](#). The use of Differential Privacy to manage the privacy of 2020 Census Data is meant to prevent the re-construction and re-identification of the public census data from external sources. However, this approach creates a tradeoff (**Figure 2**). The more privacy protections added by the application of Differential Privacy, the less accurate and usable the data becomes. If the data is made to be more accurate, there will be less privacy protection and a greater potential for the use of external commercial datasets to find likely re-identifications of individual characteristics.

Figure 2. Differential Privacy: Accuracy v. Privacy



Census Data Privacy – Implementation and Assessment

The plans of the U.S. Census Bureau to apply Differential Privacy to 2020 Census data has generated significant concerns from many stakeholders on the likely impacts on census data quality and usability. In response to stakeholder concerns, the U.S. Census Bureau decided in 2019 to release demonstration data products using 2010 public census data with their proposed privacy methods applied that external users could analyze. The demonstration data products illustrated the impacts of the planned implementation strategies for application of Differential Privacy on the potential accuracy and usability of census data.

Figure 3. Census Disclosure Avoidance System Demonstration Product Releases

Demonstration Product	Date Released
Demonstration 1	October 29, 2019
Demonstration 2	May 27, 2020
Demonstration 3	September 17, 2020
Demonstration 4	November 16, 2020
Demonstration 5 – PLB 12.2	April 28, 2021
Demonstration 6 – PLB 4.5	April 30, 2021

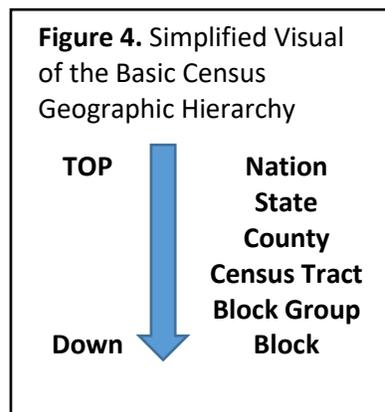
As of May 2021, the U.S Census Bureau has released six demonstration products for data users to analyze with differing applications of Differential Privacy and modifications to how the data

will be processed using census geographies (**Figure 3**). All of the demonstration products were compared to the published 2010 Decennial Census Summary File 1 data to understand the impact of the privacy methods on different population groups, communities, and census geographies compared to the public 2010 census data. The U.S. Census Bureau also created error metrics to compare the various demonstration products in terms of their impact on the accuracy of the data.

To date, the demonstration products reveal that the series of proposed strategies and adjustments to the privacy measures result in significant negative impacts on data accuracy and usability, especially for small, rural, and remote populations. The impacts are random given the statistical nature of these strategies, and some Tribal Nations gain population counts in each strategy and other Tribal Nations lose significant counts from their actual numbers. Some of the gains and losses are small numbers, but some Tribal Nations end up losing half or more of their census data counts, and some even lose their entire population and end up with population counts of zero. Tribal Nations with populations below 500 are particularly impacted. The results from the demonstration products are not exactly what will happen with the 2020 Census data, since each demonstration product uses 2010 Census data as an example, but the results so far demonstrate the potential impacts of the U.S. Census Bureau plans to implement these privacy protections on the 2020 Census data.

The U.S. Census Bureau's geographic spine plays an important role in how they apply their Disclosure Avoidance System. The U.S. Census Bureau publishes census data in various ways including for various geographies such as counties, census tracts, and blocks in a specific geographic hierarchy or "geographic spine" that is used in the processing of census data (**Figure 4**). The U.S. Census Bureau created American

Indian/Alaska Native (AI/AN) geographies to represent census data in defined "tribal lands." The AI/AN tribal lands are not part of and do not fit neatly in the main Census geography hierarchy. Because the AI/AN geographies were never on the main "geographic spine" and the Census Bureau has used other methods to protect and anonymize the data, there have been accuracy concerns before.⁹ Accurate data in tribal geographies or tribal lands is important for many reasons, including for accurate local tribal decision-making and for use in federal funding formulas. There was some inaccuracy in smaller tribal geographies in the 2010 Census due to privacy protections, but the new application of Differential Privacy for the 2020 Census likely means more inaccuracy; especially for small, rural, and remote Tribal Nations but potentially for any Tribal Nation.



The U.S. Census Bureau adjusted their planned privacy methods in demonstration products three and four to include AI/AN tribal geographies on the privacy system's geographic spine.¹⁰ This means that before the data is released to the public, the privacy algorithm processes the data within AI/AN geographies more directly than previously. This was done in response to

tribal consultation input identifying AI/AN geographies as a priority and requesting more accurate data for Tribal Nations.

Before the data is compiled into the tables and data files we are used to, the data collected from the AI/AN geographies gets processed through the geography hierarchy (**Figure 4**) to add the privacy protections. While the privacy protections are being added to the data, the AI/AN tribal geography data is kept within the states they belong to but the individual tribal areas are not separated in the data.

Once the AI/AN data is processed down the Census geographic hierarchy with the privacy protections, census block and block group data for are added together to rebuild and create the public data for each individual AI/AN tribal land. This system for adding privacy to and reforming the AI/AN geographies is being called the Optimized Spine and was used for the two April 2021 demonstration products.

In demonstration products three and four, the sum of all AI/AN tribal geography total populations in were held invariant at the state level, which meant that the total population of all the tribal areas within a state added together would equal 100 percent of the correct total number of people counted in the Decennial Census on those areas. The privacy protections applied to the data would then cause individual counts for some Tribal Nations to see a loss or gain of people in their tribal area from other tribal areas in the state but the sum of the AI/AN tribal area state population totals added together would be accurate.

However, the U.S. Census Bureau's Data Stewardship Executive Policy Committee (DSEP) decided in November 2020 to not hold AI/AN tribal geography populations invariant at the state level for unclear reasons.¹¹ They indicated that they would make the data more accurate using other methods such as applying more Privacy Loss Budget (PLB) in the application of Differential Privacy to the AI/AN geography data to improve its accuracy. However, the current demonstration data reveal that there is little or no increase in accuracy for Tribal Nations. The U.S. Census Bureau received input to make all individual AI/AN tribal geographies invariant i.e. the Tribal Nation totals would be the actual counts, but they have not implement that recommendation. Tribal Nations have made it clear that they want accurate census data to be reported from the 2020 Census.

To read about the impacts of the latest application of Differential Privacy and the Optimized Spine on AI/AN tribal areas in demonstration products five and six, read our Research Policy Update brief [Impacts of the April 2021 Census Disclosure Avoidance System on Tribal Nations](#). To view our ArcGIS maps that visualize the impacts of the planned privacy methods for each Tribal Nation, visit <https://arcg.is/1fWG4uo>. Of note, this demonstration product is meant to illustrate possible impact of the plans for privacy protection. The DSEP will make final decisions in June 2020 that will be used in the final processing step for the 2020 Census data, and the impacts may be different.

The U.S. Census Bureau also published metrics that track changes in error rates for the data in each of the demonstration products. The metrics showed how the changes to the privacy protections improved or worsened with the changes. Error rates have improved over time, but Tribal Nations must determine if errors of five percent or even ten percent of their total population are an acceptable outcome of these enhanced privacy protections. There is a price to accuracy due to privacy protections, and each Tribal Nation needs to determine the relative importance of both privacy and accuracy. However, if Tribal Nations experience undercounts or even zero counts in the 2020 Census data due to privacy protections, the price of privacy is likely too great. The U.S. Census Bureau needs to hear from Tribal Nations as soon as possible on this topic with information on how accurate the 2020 Census data needs to be for tribal priority uses.

The U.S. Census Bureau plans to hold a tribal consultation on its 2020 Census Disclosure Avoidance System on May 19, 2021 to provide a final opportunity for Tribal Nations to give input on the proposed privacy protections as shared in the latest demonstration product data and their priorities for accurate census data. The 2020 Census data cannot be released publicly with the actual counts due to privacy considerations, but the application of Differential Privacy and the Disclosure Avoidance System methods should respect the political status of Tribal Nations and adjust the data to the accuracy levels they need. Tribal consultation input for priority uses of census data needs to be incorporated into the final 2020 Census Disclosure Avoidance System on a government-to-government basis with respect to the political status of Tribal Nations, and not merely as another racial/ethnic group.

Additional Resources – *Understanding Impacts on AI/AN Census Data*

The table below provides several resources on the 2020 Census Disclosure Avoidance System and the potential impacts to AI/AN data accuracy and usability. These resources can help Tribal Nations and their technical staff prepare for the May 19, 2021 tribal consultation. This is the last consultation before the U.S. Census Bureau makes the final decisions on how they plan to implement Differential Privacy in their 2020 Census Disclosure Avoidance System. There is still time for the U.S. Census Bureau to make adjustments to improve the accuracy and usability of census data for Tribal Nations based on tribal priorities.

What information do you need?	Where to find the resources?
<p>Why does Census Accuracy Matter?</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Decennial Census: Key Uses of the Data (NCAI Policy Research Center) <input type="checkbox"/> U.S. Census Bureau Tribal Consultation on Differential Privacy and Access to AI/AN Data – Webinar (Randall Akee)

<p>Introduction to Census Privacy Measures</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Differential Privacy and the 2020 U.S. Decennial Census: Impact on American Indian and Alaska Native Data (NCAI Policy Research Center) <input type="checkbox"/> A Guide to 2020 Census Data Privacy – Webinar (NCAI Policy Research Center) <input type="checkbox"/> Implications to Montana of Differential Privacy for Census Bureau 2020 Data Dissemination – PowerPoint (Mary Craigle) <input type="checkbox"/> Changes to Census Bureau Data Products – Resource Page (IPUMS) <input type="checkbox"/> ADVANCED: Workshop on 2020 Census Data Products: Data Needs and Privacy Considerations – Video Presentations (National Academies CNSTAT)
<p>2020 Census Privacy Impacts on Tribal Nations</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Changes to the Census Could Make Small Towns Disappear – Article (Gus Wezerek and David Van Riper) <input type="checkbox"/> Population Counts on American Indian Reservations and Alaska Native Villages with and without the Application of Differential Privacy – Video Presentation (Randall Akee) <input type="checkbox"/> Impact of DP on AI/AN Tribes – Video Presentation (Norm DeWeaver)
<p>April 2021 Demonstration Data</p>	<ul style="list-style-type: none"> <input type="checkbox"/> Impact of Privacy Measures on American Indian/Alaska Native Census Data – ArcGIS Interactive Maps (Beau Spencer and NCAI Policy Research Center) <input type="checkbox"/> How to Use the Impact of Privacy Measures on AI/AN Census Data ArcGIS Interactive Maps – Video Tutorial (NCAI Policy Research Center) <input type="checkbox"/> Impacts of the April 2021 Census Disclosure Avoidance System on Tribal Nations (NCAI Policy Research Center) <input type="checkbox"/> Fact Sheet: 2010 Demonstration Privacy-Protected Microdata Files 2021-04-28 (U.S. Census Bureau) <input type="checkbox"/> Understanding the April 2021 Demonstration Data – Webinar Series (U.S. Census Bureau)

NCAI and its Policy Research Center created several educational resources and wrote several letters to the U.S. Census Bureau with concerns about the planned 2020 Census privacy methods and the potential harms on Tribal Nations from the resulting accuracy and usability issues. To view more resources developed by NCAI Policy Research Center on this topic, go to the Research Recommendations website at <https://bit.ly/2N1lCND>.

Citation: NCAI Policy Research Center (2021). *Differential Privacy and the 2020 Census: A Guide to the Data Analyses and Impacts on AI/AN Data*. Washington DC: National Congress of American Indians, May 2021.

Questions: NCAI Policy Research Center – email: research@ncai.org; website: <http://www.ncai.org/prc>

Endnotes

- ¹ “Questions Asked About the 2020 Decennial Census.” United States Census 2020. U.S. Census Bureau, <https://2020census.gov/en/about-questions.html>.
- ² John M Abowd. Staring Down the Database Reconstruction Theorem. Presentation to AAAS Annual Meeting Feb 16, 2019, 2019. Accessed on May 16, 2021 at: <https://www2.census.gov/programs-surveys/decennial/2020/resources/presentations-publications/2019-02-16-abowd-db-reconstruction.pdf?>
- ³ John M Abowd. Staring Down the Database Reconstruction Theorem. Presentation to AAAS Annual Meeting Feb 16, 2019, 2019. Accessed on May 16, 2021 at: <https://www2.census.gov/programs-surveys/decennial/2020/resources/presentations-publications/2019-02-16-abowd-db-reconstruction.pdf?>
- ⁴ John M Abowd. Staring Down the Database Reconstruction Theorem. Presentation to AAAS Annual Meeting Feb 16, 2019, 2019. Accessed on May 16, 2021 at: <https://www2.census.gov/programs-surveys/decennial/2020/resources/presentations-publications/2019-02-16-abowd-db-reconstruction.pdf?>
- ⁵ Long, Gordon. Formal Privacy Methods for the 2020 Census. JASON The MITRE Corporation, Apr. 2020, <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/privacy-methods-2020-census.pdf>.
- ⁶ John M Abowd. Staring Down the Database Reconstruction Theorem. Presentation to AAAS Annual Meeting Feb 16, 2019, 2019. Accessed on May 16, 2021 at: <https://www2.census.gov/programs-surveys/decennial/2020/resources/presentations-publications/2019-02-16-abowd-db-reconstruction.pdf?>
- ⁷ Jason Gauthier, History Staff. Title 13, U.S. Code - History - U.S. Census Bureau, www.census.gov/history/www/reference/privacy_confidentiality/title_13_us_code.html.
- ⁸ Craige, Mary. “Implications to Montana of Differential Privacy for Census Bureau 2020 Data Dissemination.” MT Department of Commerce. May 2021. https://ftpaspen.msl.mt.gov/EventResources/20210413112706_19041.pdf.
- ⁹ Bureau, US Census. “A History of Census Privacy Protections.” The United States Census Bureau, 10 Oct. 2019, www.census.gov/library/visualizations/2019/comm/history-privacy-protection.html.
- ¹⁰ US Census Bureau. “2020 Disclosure Avoidance System Updates: 9/17/20 New Privacy Protected Microdata Files (PPMFs).” The United States Census Bureau, 17 Sept. 2020, www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates.html.
- ¹¹ US Census Bureau. “2020 Disclosure Avoidance System Updates: 11/25/20 Invariants Set for 2020 Census Data Products” The United States Census Bureau, 25 Nov. 2020, www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-updates.html.