What's Driving Conflicts Around Differential Privacy for the U.S. Census

Priyanka Nanayakkara and Jessica Hullman | Northwestern University

The U.S. Census Bureau's use of differential privacy has been fiercely debated among interested parties. Accuracy loss has been at the forefront, but conflicting confidentiality notions help explain why common ground is lacking. We propose three ways of understanding confidentiality conflicts and offer suggestions for researchers and organizations adopting formal privacy.

Every 10 years, the U.S. Census Bureau seeks to enumerate the nation's population through the decennial census. These data are used for reapportionment, redistricting, funding allocation, social science research, and many other purposes. The bureau is legally required to both provide states with population tabulations at different geographical levels as well as maintain the confidentiality of individual-level records (under 13 U.S. Code § 9 —https://www.census.gov/about/policies/privacy/data_stewardship/title_13_-_protection_of_confidential_information.html).

To prevent outside parties from gaining unauthorized access to census records, the bureau has historically relied on several methods of "disclosure avoidance." In 1850, this meant that, for the first time, it would refrain from publicly posting responses. In recent decades, disclosure avoidance methods have included techniques like swapping data from carefully selected households and suppressing tables on small-area data.

The bureau concluded that previous methods were inadequate after an internal investigation on published 2010 census tables. This analysis involved simulating an attacker obtaining the published statistics and commercially available data with names and addresses, then applying standard algorithms for inferring and solving a system of equations from the data to identify a proportion of the population. Alarmed by results like these, the bureau turned to differential privacy (DP),[1] the state of the art in data privacy research, in releasing results of the 2020 Census.

DP[1] is a definition of privacy that limits the extent to which results of an analysis differ with the inclusion of any given individual's information, thus providing guarantees around individual-level privacy. Algorithms that achieve DP often work by adding a calibrated amount of noise to results. In general, the higher the amount of noise added, the stronger the privacy protections. Thus, inherent in DP is a tradeoff between privacy and accuracy (how closely the published results match the collected data).

Because of the wide range of census data use cases, many interested parties are impacted by the shift in the disclosure avoidance system (DAS). These include members of the public, nongovernmental organizations, demographers, social science researchers, and legislative bodies. For example, demographers and social science researchers, who rely heavily on census data—particularly through Integrated Public Use Microdata Series (IPUMS), a database that provides access to census and survey data—have been vocal about how DP could impact published census figures.

Multiple nongovernmental organizations have provided input about the new DAS, including identity-based

Digital Object Identifier 10.1109/MSEC.2022.3202793
Date of current version: 27 September 2022

organizations, such as the National Congress of American Indians (NCAI), the National Association of Latino Elected and Appointed Officials Educational Fund, and Asian Americans Advancing Justice. Redistrictors and legislative bodies have also weighed in on the implementation of the new DAS as census data are necessary for political processes like reapportionment, redistricting, and upholding the Voting Rights Act.

Public discussions on the new DP-based DAS among some groups have skewed heavily toward implications to accuracy. These discussions have been bolstered by various analyses by interested parties of "demonstration data products"—versions of 2010 Census data products that the bureau released under the new DAS to encourage feedback—for fitness of use. Fitness of use has been framed as hinging on accuracy but without also considering privacy. At times, discussions have grown contentious. For example, in March 2021, Alabama sued the bureau for intending to "provide the States purposefully flawed population tabulations," claiming the tabulations would be unfit for redistricting purposes (*Alabama v. U.S. Department of Commerce*, 2021—https://www.brennancenter.org/our-work/court-cases/alabama-v-us-dept-commerce).

It is easy to understand how different tolerances for accuracy loss drive debate about DP at the Census. However, less attention has been paid to other important contributors to the lack of common ground between the U.S. Census Bureau and many members of interested parties. Acknowledging the underlying conflicts in how people conceive of confidentiality requirements and risks is also critical to understanding debates. We point to three dimensions along which the understanding of confidentiality loss implications varies: 1) *the bureau's role in confidentiality threats*, 2) *approaches for validating confidentiality risks*, and 3) *the bureau's responsibility to protect sensitive data*. These dimensions offer the following respective lessons for privacy researchers:

- *Account for improper data sharing*: Researchers should work toward expanded models of confidentiality threats that take into account the potential for improper data sharing from a data collection agency.
- *Employ strategic communication around new techniques*: Researchers should use strategic communication to preemptively deter confusion about motivations for new privacy techniques.
- *Explicitly discuss harms of confidentiality loss*: Researchers should more explicitly discuss harms that could come from confidentiality loss.

Background

In 1954, the U.S. Census Bureau's mandate to maintain the confidentiality of census records was codified in 13 U.S. Code § 9. At a high level, 13 U.S. Code § 9 states that the bureau or its employees may not use census data for purposes other than those for which they were collected, publish individually identifiable data, or allow unauthorized parties to view individual records.

By definition, DP defines confidentiality protection at the level of individual-level records. If a mechanism satisfies DP, then its output should be similar to the output it produces when any given individual's information is not included in the computation.

Formally, the definition for the simplest form of DP, ε-DP, is as follows:[1]

Suppose that D and D' are databases that differ by one record/individual. A randomized mechanism M satisfies DP if the following holds, where o is an output of M:

$$\Pr[M(D)=o] \le e^{\varepsilon} \Pr[M(D')=o]$$

Note that the new DAS is based on a different variation of DP, zero-concentrated DP.

Mechanisms satisfying DP often inject into results noise drawn from a specified probability distribution parameterized in part by a "privacy budget" parameter, ε. The privacy budget determines the strength of privacy protections: smaller values of ε correspond to higher levels of added noise, which further obscures each individual's contribution to the data and affords stronger privacy guarantees. Furthermore, DP represents a framework for accounting privacy loss: ε composes across queries such that the total privacy loss over multiple runs of differentially private mechanisms can be characterized by the sum of ε values for each run. Additionally, all parameters can be made public. Doing so does not hinder the stated privacy guarantees. Actually setting ε, however, is challenging as it requires determining how to prioritize accuracy versus privacy, a question that naturally requires input from a wide range of parties.

Methods

Our analysis is based on our close following of, and participation in, public discussions on DP and the 2020 Census over the past year. After becoming aware of the bureau's use of DP from colleagues doing privacy research, we embarked on a collaborative sense-making process aimed at better understanding why the new DAS was contentious. As part of this process, both authors attended multiple Census Quality Reinforcement Task Force meetings that brought together several interested parties. This familiarized us with a range of viewpoints and relevant materials, including public-facing or scholarly reports on or directly related to the new DAS by the interested parties previously described (see Table 1 for a selection of documents on which our analysis is based).

Table 1. Resources on DP and the 2020 U.S. Census. Full references available at https://priyakalot.github.io/DP-census/.

Category	Resources
Analyses and responses related to new DAS	boyd and Sarathy, "Differential Perspectives." Bun et al., "Statistical Inference Is Not a Privacy Violation." Christ et al., "DP and Swapping." Kenny et al., "Use of DP for census data." Petti and Flaxman, "DP in the 2020 US census." Ruggles, "Census Bureau has reluctantly acknowledged." Ruggles and Van Riper, "Role of Chance." Steed et al., "Policy impacts of statistical uncertainty and privacy."
External summaries of new DAS	IPUMS, "Changes to Census Bureau Data Products." National Conference of State Legislatures, "DP for Census Data Explained." Roubideaux and Evans-Lomayesva, "Price of Privacy?"
Materials by U.S. Census Bureau leadership	Abowd, "How will statistical agencies operate?" Abowd, "Protecting Confidentiality of America's Statistics." Abowd, "US Census Bureau Adopts DP." Jarmin, "Census Bureau Adopts Cutting Edge Privacy Protections."
Public opinion on census participation	Center for Survey Measurement, "Memorandum for Associate Directorate." Cohn et al., "Most Adults Aware of 2020 Census." McGeeney et al., "2020 Census Barriers, Attitudes, and Motivators Study."
Reidentification or reconstruction studies and DAS history	Hansen, "To Reduce Privacy Risks." McKenna, "U.S. Census Bureau Reidentification Studies." U.S. Census Bureau, "Privacy and Confidentiality."
U.S. Census Bureau handbook and FAQs	U.S. Census Bureau, "Disclosure Avoidance for 2020 Census." U.S. Census Bureau, "Disclosure Avoidance: Latest FAQs."
U.S. Census Bureau presentations and products	Hawes, "DP and the 2020 Decennial Census." Hawes, "DP 101." Hawes, "Simulated Reconstruction-Abetted Re-identification Attack." Hawes and Ratcliffe, "DP 201 and TopDown Algorithm." Rodríguez, "Disclosure Avoidance and the American Community Survey." U.S. Census Bureau, "2010 Demonstration Data Products."

The second author wrote multiple blog posts synthesizing themes in the debates over DP over this period, which were published to a broad, quantitative audience. Together, this exposed us to comments and e-mails expressing views both in favor of and opposed to the bureau's use of DP. Over the course of this process, we frequently discussed with one another different potential sources of conflict, seeking to arrive at a relatively concise account of where people did not see eye to eye. This led us to observe that differing views on confidentiality requirements and validation underlay many points of debate but were not as widely acknowledged as disagreements over acceptable loss to accuracy.

While we did not formally limit the time frame of documents we analyzed, we tended to focus on documents from the last few years following the bureau's announcement of the use of DP for the Census. Our analysis is informed by several types of documents, including U.S. Census Bureau materials on the new DAS, scholarly computer science articles describing how notions of confidentiality have evolved, scholarly analyses of the new DAS, historical works describing previous U.S. censuses, and reports or commentaries on the new DAS published by various interested parties. When reviewing documents, we paid close attention to statements or sections about DP or confidentiality more broadly. We analyzed these sections allowing themes in arguments to emerge. We then iteratively developed dimensions along which arguments appeared to conflict, discussing these items among ourselves after each iteration.

Based on this close analysis, we propose the three previously mentioned dimensions along which confidentiality concerns conflict: 1) the bureau's role in confidentiality threats, 2) approaches for validating confidentiality risks, and 3) the bureau's responsibility to protect sensitive data.

The Bureau's Role in Confidentiality Threats

Discussions among the U.S. Census Bureau and interested parties imply different views on how threats to confidentiality of census data are likely to be mediated. While U.S. Census Bureau communications imply that outsider-mediated threats (where a party outside the bureau reconstructs records using published census tables and then reidentifies people) are of highest concern, insider-mediated threats (where the bureau directly provides confidential census records to an unauthorized party) appear to be of more concern among some parties.

Outsider-Mediated Confidentiality Threats

In conversations about the new DAS, the U.S. Census Bureau has focused on preventing outsider-mediated confidentiality threats. That is, the new DAS is intended to make it difficult for an outside party to reconstruct records and reidentify respondents. The U.S. Census Bureau's handbook on the new DAS emphasizes the vulnerability of census data to outside attacks:

> Census data present an enticing target for re-identification attacks. As the federal government's largest statistical agency, the Census Bureau publishes a very large number of statistics. The 2010 Census data products included over 150 billion statistics based on 309 million people and 1.9 billion confidential data points. This wealth of published statistics suggests that highly accurate reconstruction of census records may be possible, and, if it is possible, that many re-identifications not attributable purely to statistical information may also be possible, especially in small blocks and subpopulations.

The handbook and the bureau's website's FAQ page also note the specific types of harms that could result from disclosures:

> The disclosure of [race, ethnicity, and household relationships] could not only make it easier to target individuals—particularly in vulnerable populations such as communities of color, same-sex couples, older adults, or parents of very young children—for fraud, enforcement actions, disinformation, or physical or virtual abuse, but it could also undermine the public's trust in the confidentiality of its census response, which could cause people to be less likely to respond to future censuses …

Hence, the bureau is concerned not only with the harms directly resulting from the misuse of inappropriately disclosed census data, but also with the reduced trust in census operations that could result from an outsider-mediated attack.

Some other interested parties also focus attention on outsider-mediated threats. For instance, a group of data privacy experts, some of whom are computer science researchers, provided an amicus brief for Alabama's lawsuit (*Alabama v. U.S. Department of Commerce,* 2021) where they describe outsider-mediated threats that the new DAS seeks to address:

> Reconstruction-abetted reidentification attacks could create risks to national security. Entities who possess substantial troves of nonpublic personal data about the U.S. population are particularly well positioned to perform re-identification attacks on reconstructed datasets … For example, a foreign power could undermine confidence in the Census Bureau and depress

> future participation in the census by using Facebook or another social media platform to reveal to 50 million Americans that their data can be reconstructed and re-identified from census responses.

Like the U.S. Census Bureau, these data privacy experts also draw a connection between outsider-mediated attacks, implied to pose a direct threat to the privacy of individuals represented in the attacked data, and a more indirect threat in the form of reduced public trust in the U.S. Census Bureau impacting future censuses.

Insider-Mediated Confidentiality Threats

We see concerns about insider-mediated confidentiality threats in conversations around the census more generally (not necessarily those specific to the new DAS), including recent surveys on members of the broader public's concerns. For example, a Pew study found that a major reason for hesitancy in participating in the census could be mistrust in the government and a belief that the government may misuse the information, indicating that insider-mediated threats may be of concern. In particular, the study found that 21% of adults surveyed expressed hesitancy about whether they would participate in the census. Of these respondents, 60% said that mistrust in how the government would use the collected information was either a major or minor reason for reluctance in participating in the census.

The 2020 Census Barriers, Attitudes, and Motivators Study (CBAMS) Survey, conducted by the bureau, found that 28% of respondents were very or extremely concerned that the bureau would not keep census responses confidential. Furthermore, 24% of respondents were very or extremely concerned that the bureau would share census responses with other government agencies. The survey's results also showed that all racial and ethnic groups were more concerned than non-Hispanic Whites about the bureau not keeping responses confidential and sharing responses with other government agencies. During cognitive interviews of the CBAMS, some Spanish-speaking respondents were concerned about whether their answers could be shared with other government agencies. For example, one respondent referred to fears of getting arrested because of his undocumented status, possibly by an authority like the U.S.'s Immigration and Customs Enforcement. In focus groups, many Chinese-speaking respondents were concerned about how data about immigration status would be used, and some Arabic speakers also expressed concerns about deportation. (This work was likely conducted during the time that the Trump administration proposed adding a citizenship question to the census.)

We might further understand concerns about insider-mediated threats by considering historical instances of harm related to census data. These examples may not only help inform why some interested parties are concerned about improper data sharing by the bureau but also help make insider-mediated harms more concrete in illustrating how they have transpired in the past. In these cases, regardless of whether the bureau was technically in violation of the law, data were shared by the bureau with other parties, resulting in harms.

In 1910, President Taft issued a Census Proclamation formally assuring the American public—in an attempt to increase census participation—that responses would not be used against respondents.[2] However, in 1918, after the passage of a war powers act, the bureau provided individual-level data on names and ages to the U.S. Department of Justice.[2] Following Japan's attack on Pearl Harbor in 1941, the U.S. Census Bureau produced tabulations of Japanese Americans that were used to identify them for forced incarceration in internment camps.[2] Most recently, in 2002, the bureau provided tabulations of Arab Americans by zip code to the U.S. Department of Homeland Security. As reported in *The New York Times*,[3] the bureau did not violate the law in providing these tabulations, but the incident represented a breach of public trust.

Account for Improper Data Sharing

The difference between outsider- versus insider-mediated threats lies primarily in how exactly the data will be acquired by the wrong party—through the sole efforts of an outside party or through data sharing by the bureau.

DP, as implemented by the bureau, aligns with preventing outsider-mediated threats as it assumes the bureau to be trustworthy. The new DAS is based on the central model of DP,[1] where unnoised data are collected by a central agency (in this case, the U.S. Census Bureau), and noise is applied to the data before aggregate statistics are made publicly available. Thus, the central agency is assumed to be trustworthy and assumed to not inappropriately share data (which the bureau is legally prohibited from doing under 13 U.S. Code § 9). In surfacing and acknowledging this assumption, privacy researchers may be better positioned to implement DP or design privacy techniques in ways that protect against risks of a central governing body directly sharing information.

This is not a faraway possibility, considering other existing DP models. In the local model,[1] responses are perturbed by individuals before being collected by the central agency responsible for analyses. Thus, even if raw data are handed over to law enforcement or the central agency's databases are hacked, individual responses have limited usability. While there may be practical limitations involved with perturbing each individual's responses before they are sent to the bureau, we might consider other ways of implementing DP that limit the amount of raw data directly collected by the bureau. Data collection may occur at the local level, such that unperturbed responses are collected by local government agencies or community organizations, which then noise responses and send them to the bureau. We note that this plan may receive pushback, however, as there may not be trust in local governments to not inflate population counts.

The Validation of Confidentiality Risks

While the bureau has used both theoretical and empirical methods of justifying the need for the new DAS, some interested parties have focused almost exclusively on empirical methods and results. Differences in modes of validating threats appear to be a significant source of miscommunication between the bureau and these interested parties.

The Bureau's Establishment of Confidentiality Loss

The bureau's validation of confidentiality loss has drawn on advances from computer science research and results from its internal investigation on the 2010 Census.

Theoretical approach. The bureau draws on theoretical advances in computer science research that are informed by a definition of disclosure put forth in 1977 by statistician Tore Dalenius, who was motivated in part by the census use case. Dalenius[4] reasoned that a disclosure has taken place if it is "possible to determine the value D_k more accurately than is possible without access to S," where D_k is an attribute of a person, for instance, and S is a set of released statistics. In other words, released statistics should not help with learning individual attributes any more than is possible without the released statistics. Computer scientists Dwork and Naor[5] showed that avoiding a disclosure under this definition is impossible to achieve while maintaining data utility. That is, to avoid disclosure completely would require releasing "perfectly useless" statistics, in the words of the bureau's chief scientist John Abowd. While we do not find that official materials from the bureau refer to Dalenius's definition or Dwork and Naor's[5] impossibility result, Abowd recounted these findings in a 2016 talk at the bureau which outlined motivations for the adoption of DP.

Closely related to the impossibility of complete disclosure avoidance is the Database Reconstruction Theorem. In 2003, Dinur and Nissim[6] proved that a dataset can be correctly reconstructed if enough aggregate statistics computed over the dataset are released. In this way, each statistic "leaks" information about the dataset, and after enough releases, the entire dataset has effectively been released. Thus, if it is taken as premise that any release from a dataset will result in some confidentiality loss, the benefits of DP naturally follow because it provides a formal way of accounting for this loss (in terms of the privacy budget) and for precisely quantifying the loss and placing limits on it. Previous disclosure avoidance methods, on the other hand, did not allow for such a precise accounting of confidentiality loss. There is significant evidence that the bureau subscribes to the view of statistics resulting in some amount of disclosure; for example, Abowd has cited the Database Reconstruction Theorem in explaining the need for a new DAS, and bureau webinars have also cited the theorem in explaining the necessity of DP.

Empirical approach. The bureau has historically conducted reidentification studies to assess previous confidentiality protections. After the 2010 Census, a team at the bureau conducted an internal reconstruction attack on 2010 Census data (see Figure 1), which also

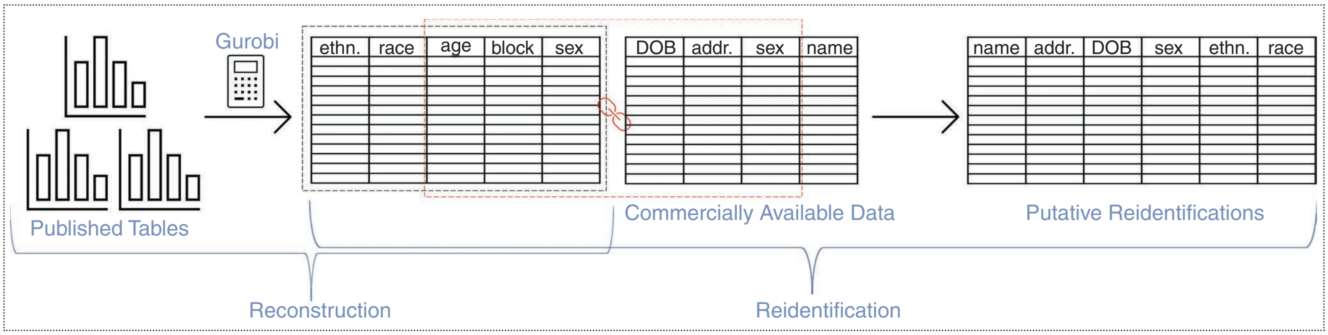

Figure 1. An internal team at the bureau first reconstructed individual-level 2010 Census records by solving a system of linear equations consistent with several published 2010 Census tables. They then linked reconstructed records to commercial data to attempt to reidentify individuals. ethn: ethnicity; DOB: date of birth; addr: address.

seemed to serve as a demonstration of the Database Reconstruction Theorem and the need for a change in the DAS.

In this internal attack, a team at the bureau began by setting up a system of equations consistent with several published tables on race, ethnicity, sex, and age from the 2010 Census. Using Gurobi, a mixed-integer linear programming software, they solved this system of equations and reconstructed individual-level records with block, race, ethnicity, sex, and age attributes. Of the reconstructed records, 46.48% had exact matches on all five attributes when linked to confidential census responses [the Census Edited File (CEF)]. It was found that 70.98% of the reconstructed records were "fuzzy matches," meaning that all attributes matched the attributes of a CEF record exactly except age, which matched plus or minus one year. For the reidentification portion of its attack, the team then linked reconstructed records to person-level commercial data from the 2010 time period, including name, address, sex, and birthdate attributes. The bureau calls linked records *putative reidentifications* because they are not yet confirmed to be actual reidentifications. The team found that at least 17% of 2010 records were confirmed reidentifications using proprietary census files.

Apart from its own internal investigation, the bureau has also relied on examples of attacks on noncensus data to further establish the practical need for DP and evidence that its previous methods needed reexamination. For example, a March 2020 presentation by Michael Hawes, Senior Advisor for Data Access and Privacy at the bureau, cites examples of reidentification attacks from the computer science literature and media, and notes that "[r]econstruction and [r]e-identification are not just theoretical possibilities … they are happening!" The bureau's handbook on the new DAS additionally asserts that "a dramatic increase in the availability of both large-scale computing resources and commercial-strength optimizers that can solve systems of billions of simultaneous equations … have changed the threat of database reconstruction from a theoretical risk to an issue that [they] must address." The handbook also notes that, while there may not be documented reidentification attacks by "bad actors," they have "documented reidentifications that users have brought to [their] attention through Reidentification Studies." Further details on these studies are not provided, though it is plausible that the bureau is referring to examples such as a reconstruction attack described in *The New York Times* on Manhattan's census data conducted by academic statisticians for illustration.

Responses to the U.S. Census Bureau's Establishment of Confidentiality Loss

In questioning whether there is sufficient evidence to establish confidentiality loss, and by extension the necessity of the new DAS, some interested parties have focused on questioning the practical significance of the bureau's empirical results, specifically of the internal investigative attack.

Demographers Ruggles and Van Riper,[7] for instance, argue that the results of the reconstruction portion of the bureau's internal attack are similar to the rate of reconstruction they achieved through a randomized method based on aggregate census counts. The authors' reasoning appears to be that since the reconstruction results obtained by the bureau do not outperform what they consider to be a reasonable baseline, the results are not particularly alarming or cause for changes to the DAS.

Multiple groups have also argued against the practical significance of the internal attack's reidentification results. For example, Ruggles has noted on social media that confirming the correctness of purported reidentifications requires access to confidential census records. That is, if a third party were to make supposed reidentifications but were unable to confirm whether any of them are true reidentifications, then the usefulness of the supposed reidentifications would be severely limited. In an NCAI webinar, Yvette Roubideaux, Vice President for Research and Director of the Policy Research Center at NCAI, made a similar argument and further noted that "the concerns over privacy … at this point are all theoretical," thus indicating that the bureau's theory-based reasoning around confidentiality loss may not have been compelling for some interested parties. While not necessarily arguing against the new DAS, the National Conference of State Legislatures implies in an explainer webpage that theoretical reasoning around confidentiality loss does not constitute "evidence" of confidentiality loss ("[t]here is no evidence that confidentiality has been compromised so far, but that doesn't change the theoretical possibility that it could happen").

Employ Strategic Communication Around New Techniques

In the emphasis around empirical reasoning—both by the bureau in terms of its communication strategy and in terms of some interested parties' preference in reasoning style—the full value of DP, particularly in how it allows for precise accounting of confidentiality loss, is not conveyed. DP follows first from a theoretical understanding of confidentiality loss. As Bowen and Garfinkel[8] have written, "[t]he math of [DP] tells us there is a real cost to every data release. There is a running bill, even if we do not choose to acknowledge it." DP provides a formal way of keeping track of the "bill" and allows for choosing exactly how much to "spend."

While the bureau's theoretical reasoning on confidentiality loss is grounded in well-known computer science research findings, we posit that results like the Database Reconstruction Theorem[6] are difficult to make compelling to a wider audience to justify a change in the DAS. Additionally, while the more empirical internal investigation may have been conducted in line with previous practices for identifying areas for improvement in disclosure avoidance methods, the bureau may have chosen to emphasize these results over more theoretical reasoning for rhetorical purposes. In short, empirical results appear to more immediately convey a sense of alarm.

The pushback to the bureau's empirical reconstruction attack results relates to the inherent ambiguity around the relationship between empirical and theoretical findings in establishing confidentiality loss. Outside of any formal model of the cost of implementing privacy protection relative to its benefits, whether theoretical evidence is enough to warrant the change becomes a matter of taste. However, interested parties who are concerned about the possibility of less accurate data are motivated to question any arguments for the new DAS that they perceive as lacking evidence.

While the precise tradeoff between theoretical and empirical motivations for using DP cannot be analyzed in the abstract, organizations that adopt state-of-the-art privacy protection techniques, which often come with theoretical motivations, might learn from the bureau's encounters. In particular, organizations adapting new techniques may benefit from thinking about a communication strategy hierarchically. In other words, if multiple approaches (for example, empirical and theoretical) are taken to establish the need for a new method, communication around findings from these approaches should be conveyed in a way that makes clear the order in which these approaches were taken or considered and their relative importance in establishing the need for a new technique. This may at least more quickly focus conversation on how confidentiality loss is assessed by various interested parties.

The Bureau's Responsibility to Protect Sensitive Data

Arguments around DP can arise from disagreements about the scope of harms that disclosure avoidance methods should prevent. The most expanded scope of harms includes any potential harm—at the individual or group level—that could arise from the use of census statistics, by themselves or in conjunction with other information. A limited scope of harms that can be addressed by DP precludes harms that arise from the use of published statistics to learn statistical patterns at a population level, even if they are used to malicious ends. The extent of harm scoping is not new to census confidentiality discussions: Cox and Nelson,[9] for example, in a 1986 article described how political and marketing consultants have used census data on income by zip code to categorize geographic areas for targeted messaging. Such messaging could be harmful, but under a limited scope would not be considered a confidentiality violation as it relies on population patterns.

In discussions around the new DAS, one type of argument assumes an expanded scope of harms and shows that the new DAS does not prevent against harms falling under this broader conceptualization of confidentiality violations. For example, Kenny et al.,[10] scholars working on redistricting, argue that individual-level race information can still be predicted with high accuracy using census data processed through the new DAS in conjunction with other data. They specifically rely on the Bayesian improved surname geocoding (BISG) methodology, which uses individuals' names, addresses, and census block information to predict race. They argue that if race and ethnicity are indeed sensitive attributes, DP would not maintain confidentiality over these attributes as BISG performs well with 2010 Census data (as published) and under DP at different levels of ε (that is, data demonstration products the bureau released with the rollout of the new DAS). BISG works well in part because data that associate surnames with races are available and are presumably of high quality. Kenny et al.'s[10] argument is essentially that other available data used in combination with DP-noised census data can still reveal individual-level attributes, from which they imply that the new DAS is not sufficiently protecting confidentiality.

Arguments in support of DP that respond to arguments like that of Kenny et al.[10] defend a more limited scope of harms, as is commonly assumed in the DP literature in computer science. In response to Kenny et al.,[10] a group of 10 computer scientists (including three of the four creators of DP) explained that using BISG to accurately predict race would not represent a confidentiality violation since the method relies only on aggregate census statistics. They explain that DP is intended to provide group-level insights while preserving individual-level privacy and that "[t]he BISG prediction is not about the individual" as a person's BISG prediction "is a statistical relationship between name, geography, and ethnicity" and one that changes as a person moves locations. Echoing this statement, the U.S. Census Bureau has also emphasized that the "[DAS] permits accurate inferences based on aggregate statistical information about groups" in the handbook on the new DAS. In fact, that the new DAS preserves aggregate information well enough to make accurate inferences is considered an important feature among computer science researchers and the bureau.

Explicitly Discuss Harms of Confidentiality Loss

We might draw parallels between a limited and expanded scope of harms and concepts of "safety" and "security," as proposed by Tawana Petty.[11] Petty argues that security usually involves "securing items, property, or even their identity," but that "[v]ery often, this mindset does not have a human factor involved." In other words, security does not necessarily imply that actual harm is prevented (that is, that people are safe).

Considering census data through a safety frame, or through an expanded scope of harms, can also help take into account harms that align with historical instances of inappropriate data sharing (for example, when the bureau provided data that were used to facilitate forced incarceration of Japanese Americans). Clearly outlining potential harms of confidentiality loss and delineating limitations of a new privacy technique can also help interested parties assess its appropriateness for a given context and determine the extent to which it will address relevant concerns.

Furthermore, arguments that point to limitations of DP in protecting against using aggregate statistics to make accurate individual-level inferences align with recent works published in law discussing the ways in which individual data privacy rights may be limited. Solow-Niederman[12] argues for an updated legal framework for data privacy that takes into account the ways aggregate information can be used to make highly accurate predictions about individuals, forming an "inference economy." Similarly, Viljoen's[13] relational theory of data governance argues for the need to consider horizontal relationships among data subjects, not only vertical relationships among data subjects and data collectors. This theory also acknowledges the ways in which one person's information may be used to infer information about another. Bridging these privacy concepts with technical practices is an important area of future privacy research.

While we have primarily discussed harms resulting from disclosures, as suggested previously, preventing less direct "discredit harm,"[14] or harms that result from the perception of census data not being confidential, also plays an important role in census confidentiality measures. As the bureau and other interested parties have noted, one discredit harm could be decreased trust in census operations, which could lead to reduced participation in future censuses.

The impacts of discredit harms are difficult to assess; in particular, there is a lack of understanding of how the accuracy implications of fewer census responses due to a lack of trust in the bureau compare to accuracy loss stemming from the DP-based DAS. Future research in this area could help organizations like the U.S. Census Bureau communicate more formally around important tradeoffs that motivate their privacy strategies. In general, more research into laying out specific disclosure and discredit harms and their costs, specifically in the census case, could provide empirical findings and conceptual frameworks to ongoing conversations about the confidentiality versus accuracy tradeoffs. Whether all interested parties will be receptive to more explicit evidence on risks like discredit harm remains to be seen.

Generalization to Other Legal Privacy Contexts

Conflicts about confidentiality that we identify may also arise in other contexts where technical solutions are applied to meet legal privacy requirements, such as the American Community Survey (ACS) and enforcement of the European Union's General Data Protection Regulation (GDPR).

Like the decennial census, the ACS is also subject to confidentiality requirements under 13 U.S. Code § 9. To protect confidentiality, the U.S. Census Bureau plans to release the ACS as fully synthetic data in upcoming years. Before generating synthetic data, the bureau plans to collect survey responses that, if inappropriately shared with unauthorized parties, could compromise people's privacy and potentially cause greater harm than the decennial census, given the more detailed nature of ACS questions. There has also already been pushback to the shift to synthetic data from researchers who use the ACS in their work. Thus, the bureau's motivation for the change to synthetic data, and in turn how they validate confidentiality loss, may again become an area of disagreement. We might also expect disagreements about whether synthetic data protect against the right scope of harms. Synthetic data mimic the original data, meaning that aggregate statistics from the ACS should still reflect actual patterns and can thus be used to make predictions about people's individual-level attributes.

The GDPR aims to protect personal or identifiable information. Computer science researchers have already begun to show how DP could exempt some data from GDPR requirements.[15] If an organization first collects unanonymized data records before applying DP, insider-mediated threats may still need to be protected against. In addition, much like tensions around differences in how confidentiality loss is validated, how data are validated as anonymous may be a source of miscommunication or disagreement. Finally, there may again be variations in the scope of harms protected against by using DP to exempt data from restrictions, particularly if aggregate statistics produced under DP can still be used to predict people's individual information with high accuracy.

Debates around the U.S. Census Bureau's use of DP for the 2020 Census provide insight into disagreements likely to arise more generally when technical approaches are applied to satisfy legal requirements. We see conflicting notions of confidentiality around the bureau's role in confidentiality threats, the validation of confidentiality risks, and the bureau's responsibility to protect sensitive data. These dimensions, along which notions of confidentiality conflict, shed light on multiple ways in which privacy researchers and those implementing new techniques can work to improve the development and rollout of such techniques.

We suggest that researchers should work toward expanded models of confidentiality threats that account for the possibility of the data collector improperly sharing data, employ strategic communication around motivation for these techniques, and more explicitly discuss harms of confidentiality loss to inform future research in preventing a broader space of harms. As privacy regulation expands and evolves, it is critical for researchers developing or implementing technical solutions to help anticipate and resolve potential conflicts, in turn providing more appropriate and comprehensive privacy protections. ■

Acknowledgment

We would like to thank the organizers and attendees of the CS+Law Monthly Workshop for their thoughtful comments and feedback.

References

1. C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, nos. 3–4, pp. 211–407, 2014, doi: 10.1561/0400000042.
2. M. J. Anderson, *The American Census: A Social History*. New Haven, CT, USA: Yale Univ. Press, 2015.
3. L. Clemetson. "Homeland security given data on Arab-Americans." NY Times. Accessed: Sep. 14, 2022. [Online]. Available: https://www.nytimes.com/2004/07/30/us/homeland-security-given-data-on-arab-americans.html
4. T. Dalenius, "Towards a methodology for statistical disclosure control," *Statistisk tidskrift*, [Online]. vol. 15, pp. 429–444, 1977. Available: https://ecommons.cornell.edu/bitstream/handle/1813/111303/dalenius-1977.pdf
5. C. Dwork and M. Naor, "On the difficulties of disclosure prevention in statistical databases or the case for differential privacy," *J. Privacy Confidentiality*, vol. 2, no. 1, pp. 1–12, 2010, doi: 10.29012/jpc.v2i1.585.
6. I. Dinur and K. Nissim, "Revealing information while preserving privacy," in *Proc. 22nd ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst.*, 2003, pp. 202–210, doi: 10.1145/773153.773173.
7. S. Ruggles and D. Van Riper, "The role of chance in the census bureau database reconstruction experiment," *Population Res. Policy Rev.*, vol. 41, no. 3, pp. 1–8, 2021, doi: 10.1007/s11113-021-09674-3.
8. C. Bowen and S. Garfinkel, "The philosophy of differential privacy," *Notices Amer. Math. Soc.*, vol. 68, no. 10, pp. 1–13, 2021, doi: 10.1090/noti2363.
9. L. H. Cox and D. Nelson, "Confidentiality issues at the United States Bureau of the Census," *J. Official Statist.*, vol. 2, no. 2, pp. 135–160, 1986.
10. C. T. Kenny, S. Kuriwaki, C. McCartan, E. T. Rosenman, T. Simko, and K. Imai, "The use of differential privacy for census data and its impact on redistricting: The case of the 2020 US Census," *Sci. Adv.*, vol. 7, no. 41, p. eabk3283, 2021, doi: 10.1126/sciadv.abk3283.
11. T. Petty, "Safety vs. security: Are you safe or are you secure?" in *Digital Defense Playbook: Community Power Tools for Reclaiming Data*, S. P. Gangadharan, T. Petty, T. Lewis, and M. Saba, Eds. Detroit: Our Data Bodies, 2018, pp. 1–49.
12. A. Solow-Niederman, "Information privacy and the inference economy," *Forthcoming Northwestern Univ. Law Rev.*, pp. 1–67, Sep. 2021, doi: 10.2139/ssrn.3921003. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3921003
13. S. Viljoen, "A relational theory of data governance," *Yale Law J.*, vol. 131, no. 2, p. 573, 2021.
14. M. Anderson and S. E. Fienberg, "US census confidentiality: Perception and reality," presented at the International Statistical Institute Biennial Meeting (Seoul), 2001.
15. A. Cohen and K. Nissim, "Towards formalizing the GDPR's notion of singling out," *Proc. Nat. Acad. Sci.*, vol. 117, no. 15, pp. 8344–8352, 2020, doi: 10.1073/pnas.1914598117.

Priyanka Nanayakkara is a Ph.D. candidate in computer science and communication studies at Northwestern University in Evanston, IL 60208 USA. Her research interests include the societal implications of algorithms and supporting reasoning around their tradeoffs. Nanayakkara received a master's degree in computer science and communication studies from Northwestern University. Contact her at priyankan@u.northwestern.edu.

Jessica Hullman is the Ginni Rometty Associate Professor of Computer Science and Engineering at Northwestern University in Evanston, IL 60208 USA. Her research interests include interactive visualization, the representation of uncertainty, and limitations of statistical inference. Hullman received a Ph.D. in information science from the University of Michigan. Contact her at jhullman@northwestern.edu.