

# A Refreshment Stirred, Not Shaken (III): Can Swapping Be Differentially Private?

James Bailie, Ruobin Gong and Xiao-Li Meng\*

April 16, 2025

## Abstract

The quest for a precise and contextually grounded answer to the question in the present paper’s title resulted in this *stirred-not-shaken* triptych, a phrase that reflects our desire to deepen the theoretical basis, broaden the practical applicability, and reduce the misperception of *differential privacy* (DP)—all without shaking its core foundations. Indeed, given the existence of more than 200 formulations of DP (and counting), before even attempting to answer the titular question one must first precisely specify what it actually means to be DP. Motivated by this observation, a theoretical investigation into DP’s fundamental essence resulted in Part I of this trio, which introduces a five-building-block system explicating the *who*, *where*, *what*, *how* and *how much* aspects of DP. Instantiating this system in the context of the United States Decennial Census, Part II then demonstrates the broader applicability and relevance of DP by comparing a swapping strategy like that used in 2010 with the TopDown Algorithm—a DP method adopted in the 2020 Census. This paper provides nontechnical summaries of the preceding two parts as well as new discussion—for example, on how greater awareness of the five building blocks can thwart privacy theatrics; how our results bridging traditional SDC and DP allow a data custodian to reap the benefits of both these fields; how invariants impact disclosure risk; and how removing the implicit reliance on aleatoric uncertainty could lead to new generalizations of DP.

**Keywords:** Statistical Data Privacy, Confidentiality, Statistical Disclosure Control, Data Swapping, US Decennial Census, TopDown Algorithm.

## 1 WHAT MOTIVATED THIS STIRRED-NOT-SHAKEN TRIO?

Since its invention nearly two decades ago by Dwork et al. (2006b), differential privacy (DP) has received a tremendous amount of attention in research and practice. As a field, it aims to provide a tractable, mathematical framework to quantify and operationalize the evasive concept of privacy within the context of sharing (i.e. releasing) statistical data. Yet, driven by a myriad of constraints (e.g., challenges in establishing theoretical guarantees or in practical implementation), attempts to alter, enhance and relax the original, so-called ‘pure’  $\epsilon$ -DP definition have led to a plethora of ‘impure’ DP formulations. As such, the term DP today encompasses a broad class of technical standards conceptualizing what it means for a data release algorithm to be ‘private.’ (See Desfontaines and Pejó (2020) for a dizzying but still partial enumeration of this class.)

From a practical perspective, the interest in DP, and the subsequent explosion of different formulations, is easy to understand. General concerns of privacy breaches have increased substantially as our society adventures deeper into the digital age. Organizations in all sectors and at all levels are compelled to expend effort addressing the issue of data privacy, whether for noble reasons or for fear of liability; yet each entity

---

\*jamesbailie@g.harvard.edu, ruobin.gong@rutgers.edu, meng@stat.harvard.edu

is faced with its own unique set of concerns, necessitating its own custom solution (Schneider et al., 2025). The adoption of a form of DP by the United States Census Bureau (USCB) for its 2020 Decennial Census of Population and Housing is a shining example of organizational effort—one that has generated much theoretical contemplation and methodological advances, as well as controversies and emotions ranging from excitement to frustration; see the special issue of the *Harvard Data Science Review*, “Differential Privacy for the 2020 US Census: Can We Make Data Both Private and Useful?” (Gong et al., 2022).

Indeed, this trio of articles owes its existence to the bureau’s adoption of DP. Because data privacy has been a central concern of the census for well over a century, seeing the USCB’s recent development of the DP-inspired TopDown Algorithm (TDA) (Abowd et al., 2022) for its 2020 Census, one naturally may wonder in what ways it improves upon their past methods for statistical disclosure control (SDC). In particular, the data swapping strategy used in 2010, just like the TDA, involves injecting artificial randomness into the published census tables. So could it be a form of DP as well, which would then make it easier to compare both methods within a single, unified framework?

Those who adhere to the definition of pure  $\epsilon$ -DP may immediately declare that any form of swapping cannot satisfy DP because swapping leaves some aggregations of the data unaltered, and hence some statistics—termed its *invariants*—will be released without any artificial noise injected. By the same logic, the TDA also cannot satisfy DP because it is designed to maintain various total counts (e.g., state population counts) in order to comply with mandates set by the US Constitution and other external policy considerations. Yet such a strict adherence to a narrow perspective not only greatly limits the applicability of DP, but also fundamentally misperceives its essence. After all, DP is not concerned with protecting absolute privacy—no data release method can (Kifer & Machanavajjhala, 2011). Rather, it concerns the protection of information that can be revealed by an individual’s confidential data but is otherwise unavailable. The constitutional and policy mandates reveal information that cannot be protected, and hence any adoption of DP—or any other data protection standard for that matter—must take that into account. (Besides which, the narrow perspective that DP does not allow any invariants is misplaced. DP has in fact accommodated some invariants from its very inception: the number of records in the confidential dataset can be released exactly under many DP formulations, including pure  $\epsilon$ -DP.)

For the case of swapping, the matter becomes muddier as its invariants are not externally mandated but instead are inherent to its design. One can see immediately the potential complications with, and debates about, different designs and their resulting invariants. Confusions may easily result from comparing SDC methods that are based on different postulates of what is already known or considered not to need protection, akin to the trouble of comparing two distributions when they’re conditioned on different variables.

Such a problem is only one of many nuanced issues we have had to deal with as we seek precise and contextual answers to the question in this article’s title, as an impetus for a deep study of DP to reveal its statistical essence and to contemplate its practical complications. Consequently, it should come as little surprise that our investigation took considerably more time, and pages, than initially expected. In fact, it resulted in two ‘prequels’ which lay the groundwork necessary to answer the titular question—“Five Building Blocks of Differential Privacy” (referred to herein as Part I) and “Invariant-Preserving Deployments of Differential Privacy for the US Decennial Census” (Part II) (Bailie et al., 2025a, 2025b). This third and final part therefore will first provide an intuitive summary and explanation of both preceding parts before elaborating on their broader and deeper implications, and discussing some of the subtle issues that tend to invite misunderstanding, misuse, and misplaced expectations of DP.

More specifically, through five ‘questions of protection,’ Section 2 provides a nontechnical introduction to the system of DP specifications developed in Part I. To demonstrate the necessity of all five of these questions, Section 3 gives examples of how one might ‘game’ DP by trading off the strength of one’s answers to some of the questions with weaker answers to the other questions. Of course, we are not endorsing these examples; rather we use them to highlight possible ‘privacy theatrics’ that can be exploited in a mathematical and

seemingly objective framework such as DP. Section 4 is an accessible overview of Part II’s main results—it describes the evolution of the US Census’s SDC protection. In particular, with Part II’s DP specification for a particular type of data swapping similar to that used in 2010, Section 4 makes comparisons between the censuses in 2010 and 2020 using DP-based descriptions of their SDC protection.

This DP specification also serves as a bridge between traditional SDC and DP, allowing a data custodian to reap the benefits of both these fields—for example, mathematical guarantees and composition from DP; facial validity and logical consistency from traditional SDC (Section 5). Yet despite swapping’s advantages, potential disclosure risks arise from the fact that it leaves much of the data untouched (because of its numerous invariants). Section 6.1 grapples with the question of how these potential disclosure risks are compatible with the fact that swapping can satisfy a DP specification; and to mitigate these risks, Section 6.2 describes three extensions to swapping that induce fewer invariants, including one which is already deployed at some national statistical offices. Returning to our overarching aim to reap the benefits of both DP and traditional SDC, Section 6.3 explores the apparent tension between DP’s principle of ‘transparency for free’ and the notion of ‘privacy through obscurity’ found in traditional SDC. This tension hinges on the two fields’ differing views of the role of epistemic uncertainty in privacy protection, an observation that opens up new research directions—namely, extensions of existing DP definitions to account for epistemic uncertainty via imprecise probabilities, so as to retain rigorous guarantees of protection while preserving a data user’s ability to conduct valid statistical analysis.

Integrating the three parts together, this triptych reflects our triple ambition: firstly, to leverage the merits of DP, including its mathematical assurances of protection and algorithmic transparency, without sidelining the advantages of classical SDC; secondly, to unveil the nuances and potential pitfalls in employing DP as a theoretical yardstick for SDC procedures; and thirdly, to build connections between social and mathematical conceptualizations of data privacy by outlining real-world considerations behind the five-building-blocks system developed in Part I.

## 2 HIGHLIGHTS OF PART I: FIVE BUILDING BLOCKS OF DP

Every formulation of DP is a way to measure the ‘privacy’ of a *data release mechanism*, a function which transforms a confidential dataset into publishable statistics (also referred to as an SDC method, a data sharing algorithm, or similar). As we will review shortly, the core idea behind all of these formulation is to quantify privacy—or rather loss of privacy—as the amount of change in a mechanism’s likelihood function, relative to a corresponding change in the individual data points. Indeed, this narrow, mathematical conceptualization of data privacy in terms of a rate of change highlights DP’s central tenet that privacy is controlled by bounding the ‘derivative’ of a mechanism, as is alluded to by its nomenclature ‘differential.’

A technically oriented reader may immediately ask a host of questions. On what space is the likelihood function defined? How is relative change metricized? What constitutes an “individual data point”? What does “control” mean? And so on. Indeed, these questions are key to formalizing the above intuitive idea of DP into a concrete, mathematical definition and their many possible answers are reflected in the numerous formulations of DP found in the literature. However, while it may appear there are many questions to formalize, it turns out five are sufficient (and necessary) to fully characterize most existing definitions of DP—these are the five *questions of protection*, each of which correspond to what we call a *building block* of DP: the domain  $\mathcal{X}$  (who?); the multiverse  $\mathcal{D}$  (where?); the input premetric  $d_{\mathcal{X}}$  (what?); the output premetric  $D_{\mathcal{P}_r}$  (how?); and the protection loss budget  $\varepsilon_{\mathcal{D}}$  (how much?). A different choice for any one of these building blocks—i.e. a different answer to any one of the questions of protection—results in a different formulation of DP. As defining DP requires answering all five questions, instantiations for each of the building blocks collectively form a *DP specification*—a complete, self-contained formalization of the intuitive idea of DP as a bound on a mechanism’s rate of change.

## 2.1. Questions of Protection

To explain how these five building blocks specify DP, we start with the most basic question concerning a mechanism, “*who is eligible for protection?*” which is addressed by specifying all the actual, potential or counterfactual datasets that could be inputted into the mechanism. The collection of all such datasets is called the data space, or the *domain*, and is denoted by  $\mathcal{X}$ . Because setting down all these datasets requires situating the actual dataset  $\mathbf{x}$  in the context of its lifecycle (how else would one know what other, counterfactual datasets are possible?)—not just specifying the mathematical schema of  $\mathbf{x}$ —the domain  $\mathcal{X}$  provides the meaning of who  $\mathbf{x}$ ’s data subjects are and how  $\mathbf{x}$  represents them.

As briefly mentioned at the start of this section, an SDC method computes some output statistics based on the actual, confidential dataset  $\mathbf{x}$ . (It is worthwhile to emphasize that in the context of DP, while SDC methods are typically random functions of  $\mathbf{x}$ , the randomness is not ‘in’  $\mathbf{x}$ , but rather it is artificially injected into the output statistics to reduce their information content.) From an attacker’s perspective, the confidential dataset  $\mathbf{x}$ —which always belongs to  $\mathcal{X}$  by design—is the unknown ‘parameter’ to be inferred from the output statistics, hence the choice of  $\mathcal{X}$  is conceptually analogous to the choice of a parameter space in standard statistical inference.

Explicating  $\mathcal{X}$ , however, is only the first step. The next question is “*to where does the protection extend?*” which can be answered by specifying a *multiverse*  $\mathcal{D}$ , which is a collection of the possible *universes*  $\mathcal{D}$ , be they actual or hypothetical. The need for—and the distinction between—the data multiverse and the actual data universe is well illustrated by the application of DP to the US Decennial Census. Suppose the enumerated US population size  $N_{US}$  is 330 million. Then, in order to comply with the constitutional mandate that the state population totals must be released exactly as enumerated, any hypothetical census dataset that does not yield  $N_{US} = 330,000,000$  will not be within the scope of protection, since any attacker can easily rule out such datasets. In this instance, if  $N_{US}$  is the only aggregation that must be disclosed as is, then the corresponding data universe is simply all datasets in  $\mathcal{X}$  such that the corresponding  $N_{US} = 330,000,000$ .

However, it would be rather unwise to design a mechanism that works only when  $N_{US}$  is exactly 330 million. The enumerated US total population count varies from census to census, and indeed it varies within each census due to corrective adjustments (e.g., for reducing the impact of under-counting). While it is essential for any data release mechanism to respect constitutional mandates—i.e., to disclose the total state enumerations—the actual value of the disclosed  $N_{US}$  is rather accidental. A sensible design should work regardless of the value of  $N_{US}$ , at least for values within a reasonable range (e.g., one may argue that it is not sensible to require a mechanism to work for implausible values, such as  $N_{US} = 330$  billion). A data multiverse therefore collects all data universe within each of which the count  $N_{US}$  is a constant, so that all datasets in  $\mathcal{X}$  are protected—but only within the scope of their respective universes.

Next is the question, “*what is the granularity of protection?*” which can be a source of confusion, as well as an opportunity for manipulation in capable but malicious hands. Recalling that a DP specification is a bound on a mechanism’s rate of change, the granularity of protection informally corresponds to the entities whose data is counterfactually altered when measuring this rate of change. These entities are termed the *protection units* (or, elsewhere in the literature, the privacy units). Individual persons or business entities are common choices for the protection units, but they are not the only ones. For example, for a dataset of electronic communications, the protection unit may be defined as a single message sent by a person, rather than the sender herself. As an individual may send many messages a day, such a fine-grained protection unit allows a social media platform to declare a nominal level of DP protection which appears impressively high, even though the actual risks to the sender remain exponentially large.

Mathematically, the granularity of protection is formally conceptualized via the input premetric  $d_{\mathcal{X}}(\mathbf{x}, \mathbf{x}')$ , which is a measure of the difference (or ‘distance’) between two datasets  $\mathbf{x}$  and  $\mathbf{x}'$  in the domain  $\mathcal{X}$ . Furthermore, the protection units correspond conceptually to unit differences in  $d_{\mathcal{X}}$ —i.e. the differences between *neighboring* (or adjacent) datasets, which are pairs of counterfactual datasets  $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$  with  $d_{\mathcal{X}}(\mathbf{x}, \mathbf{x}') = 1$ .

(More exactly, a protection unit is what differs during the generating processes of two neighboring datasets.) While neighboring datasets could differ by the deletion of one record, or the alteration of a single attribute, to properly define the protection units requires an appreciation of what a unit difference in  $d_{\mathcal{X}}$  actually represents in the real world. This in turn necessitates placing  $\mathbf{x}$  within the context of its data pipeline, highlighting once again the importance of the domain  $\mathcal{X}$ ’s contextual definition over and above a purely mathematical description of it.

As we have repeatedly described, DP quantifies loss of ‘privacy’ in terms of the rate of change in the variations of a data release mechanism’s output. For each specification of DP, this rate of change is calculated with respect to the specification’s input premetric  $d_{\mathcal{X}}$ , but *how is the change in output variations measured?* A premetric is also used to address this question, specially the specification’s output premetric  $D_{\text{Pr}}$ . While  $d_{\mathcal{X}}(\mathbf{x}, \mathbf{x}')$  measures differences between datasets  $\mathbf{x}$  and  $\mathbf{x}'$ , the output premetric  $D_{\text{Pr}}(\mathbb{P}_{\mathbf{x}}, \mathbb{P}_{\mathbf{x}'})$  is a measure of the difference between probability distributions  $\mathbb{P}_{\mathbf{x}}$  and  $\mathbb{P}_{\mathbf{x}'}$ . Here,  $\mathbb{P}_{\mathbf{x}}$  denotes the likelihood function—i.e. the probability distribution of the released statistics, as a function of the confidential dataset  $\mathbf{x}$ , where the randomness in  $\mathbb{P}_{\mathbf{x}}$  is introduced solely by the data release mechanism. This is a sensible approach to SDC: as all statistical information is created by variations, by limiting the difference between the output distributions  $\mathbb{P}_{\mathbf{x}}$  and  $\mathbb{P}_{\mathbf{x}'}$ , we limit an attacker’s ability to distinguish between  $\mathbf{x}$  and  $\mathbf{x}'$ . Furthermore, by controlling the rate of change, rather than the absolute change, we allow large differences between  $\mathbb{P}_{\mathbf{x}}$  and  $\mathbb{P}_{\mathbf{x}'}$  when  $d_{\mathcal{X}}(\mathbf{x}, \mathbf{x}')$  is also large (intuitively meaning that the difference between  $\mathbf{x}$  and  $\mathbf{x}'$  is not considered confidential information); while still restricting  $D_{\text{Pr}}(\mathbb{P}_{\mathbf{x}}, \mathbb{P}_{\mathbf{x}'})$  when  $d_{\mathcal{X}}(\mathbf{x}, \mathbf{x}')$  is small.

This brings us to our fifth and final question, “*how much protection is afforded?*” whose answer is given by the bound the DP specification imposes on a mechanism’s rate of change. This bound is given by the building block  $\varepsilon(\cdot)$ , a function mapping each universe  $\mathcal{D} \in \mathcal{D}$  to a non-negative (potentially infinite) value  $\varepsilon(\mathcal{D})$ . For notational convenience, we denote both the value  $\varepsilon(\mathcal{D})$  and the function  $\varepsilon(\cdot)$  by  $\varepsilon_{\mathcal{D}}$  except when we need to explicitly differentiate between them.

For each universe  $\mathcal{D}$ , the quantity  $\varepsilon_{\mathcal{D}}$  is the specification’s bound on the rate of change within the universe  $\mathcal{D}$ ; we say that a mechanism satisfies a DP specification if its rate of change in each universe  $\mathcal{D}$  is bounded by the value  $\varepsilon_{\mathcal{D}}$ . Smaller values of  $\varepsilon_{\mathcal{D}}$  are more restrictive and hence supply higher levels of SDC protection to  $\mathcal{D}$  than larger values of  $\varepsilon_{\mathcal{D}}$ . We allow the bound  $\varepsilon_{\mathcal{D}}$  to vary between universes so that different universes can be afforded different degrees of protection. This is desirable for example when, due to data utility concerns, the data custodian wants some entities to receive less protection than others; or when they would like to measure protection “per-attribute” (Ashmead et al., 2019; Seeman et al., 2023).

We propose the term *protection loss budget* for the function  $\varepsilon_{\mathcal{D}}$  instead of the commonly used phrase “privacy loss budget” (although we maintain the same abbreviation, PLB). This reflects our desire to avoid running the risk of misrepresenting DP as a panacea for the diversity of data privacy issues found in modern society (Nissenbaum, 2010). In fact, DP only addresses a very narrow, but legitimate, concern (Seeman & Susser, 2024)—an observation which also explains our use of quotation marks when discussing DP as a conceptualization of ‘privacy’: While we strive to describe DP in a way that stays faithful to its literature, we also want to remain cognizant of the fact that, even in the specific context of statistical data sharing, there are dimensions to privacy which are outside the gamut of DP (Baillie & Gong, 2023).

## 2.2. What Is a Specification of DP, Formally?

Now, having described each of the five building blocks of a DP specification in turn—the domain  $\mathcal{X}$ , the multiverse  $\mathcal{D}$ , the input premetric  $d_{\mathcal{X}}$ , the output premetric  $D_{\text{Pr}}$  and the PLB  $\varepsilon_{\mathcal{D}}$ —we can now formally tie them together with the following definition: A *DP specification* is the condition on a data release mechanism that

$$\frac{D_{\text{Pr}}(\mathbb{P}_{\mathbf{x}}, \mathbb{P}_{\mathbf{x}'})}{d_{\mathcal{X}}(\mathbf{x}, \mathbf{x}')} \leq \varepsilon_{\mathcal{D}}, \quad \text{or more generally, } D_{\text{Pr}}(\mathbb{P}_{\mathbf{x}}, \mathbb{P}_{\mathbf{x}'}) \leq \varepsilon_{\mathcal{D}} d_{\mathcal{X}}(\mathbf{x}, \mathbf{x}'), \quad (1)$$

for all  $\mathbf{x}, \mathbf{x}' \in \mathcal{D}$  and all  $\mathcal{D} \in \mathcal{D}$ . We provide two expressions here because the first one captures the intuition of DP as a bound on the mechanism’s rate of change (i.e. the change in  $D_{\text{Pr}}$  per corresponding change in  $d_{\mathcal{X}}$ ); while the second shows that mathematically a DP specification is simply a Lipschitz continuity condition on  $\mathbb{P}_{\mathbf{x}}$  as a function of the input data  $\mathbf{x}$ . (A Lipschitz condition on a function is a generalization of the condition that the function’s derivative is bounded.)

A DP specification is characterized by its choices for the five building blocks, with different choices generating different DP specifications. While similar ideas already exist (some of which directly inspired this work), the usefulness of our system of DP specifications comes from the fact that: 1) it is a sufficiently general theory to encompass most published formulations of DP, thereby providing a unifying framework for a vast swathe of literature; and yet 2) each DP specification is a complete, mathematically precise definition of DP. In contrast, many common notions of DP are concerned with only one building block, leaving the others unspecified, and hence are not well-defined formulations of DP. For example, pure  $\varepsilon$ -DP, approximate  $(\varepsilon, \delta)$ -DP, and  $\rho$ -zero concentrated DP ( $\rho$ -zCDP) specify choices for the output premetric  $D_{\text{Pr}}$  only. This is evidenced by the fact that there are both bounded and unbounded versions of these three notions—and yet the bounded-vs-unbounded distinction itself only partly specifies the choice of  $d_{\mathcal{X}}$  (as either a Hamming or symmetric difference distance) since it still leaves unspecified the protection unit of  $d_{\mathcal{X}}$  (i.e. at what granularity the Hamming or symmetric difference distance is defined). Even after a data custodian’s choices for all of the above have been clarified, an observer still does not have a complete understanding of the custodian’s formulation of DP (since they do not know what form of the data is even being protected)—a confusing situation which warrants systematization. Our aim with Part I’s system of DP specifications is to resolve this confusion.

In summary, a DP specification is a technical standard against which a data release mechanism can be assessed (either a mechanism satisfies the condition given in Equation (1), in which case it meets the DP specification, or it does not), and it is defined by its five components (or building blocks), which are:

1. The *domain*  $\mathcal{X}$ : the set of all actual, potential or counterfactual datasets to be protected under the DP specification. Intuitively speaking,  $\mathcal{X}$  describes the ‘domain of protection’ and answers the question, “*who* is eligible for protection?”
2. The *multiverse*  $\mathcal{D}$ : the collection of the DP specification’s universes, which are subsets of  $\mathcal{X}$  and collectively instantiate the ‘scope of protection’ provided by the specification.  $\mathcal{D}$  answers the question, “*to where* does the protection extend?”
3. The *input premetric*  $d_{\mathcal{X}}$ : the DP specification’s measure of difference (or ‘distance’, loosely speaking) between any two datasets  $\mathbf{x}$  and  $\mathbf{x}'$  in the domain  $\mathcal{X}$ .  $d_{\mathcal{X}}$  conceptualizes the ‘protection unit’ and answers the question, “*what* is the granularity of protection?”
4. The *output premetric*  $D_{\text{Pr}}$ : the DP specification’s measure of difference between any two of the released data’s possible probability distributions (e.g.  $\mathbb{P}_{\mathbf{x}}$  and  $\mathbb{P}_{\mathbf{x}'}$ ).  $D_{\text{Pr}}$  captures the specification’s ‘standard of protection,’ answering the question, “*how* to measure change in output variations?”
5. The *protection loss budget*  $\varepsilon_{\mathcal{D}}$ : the function  $\varepsilon_{\mathcal{D}}$  which supplies a value  $\varepsilon_{\mathcal{D}}$  for each universe  $\mathcal{D} \in \mathcal{D}$ , quantifying the DP specification’s ‘intensity of protection’ and answering the question, “*how much* protection is afforded?”

Much of the literature on DP treat the PLB as the main measure of the strength of a DP formulation. For example, Abowd and Schmutte (2015) uses it as the sole quantity for balancing the tradeoff between privacy and utility. However, the five-building-blocks system reveals that the PLB can only be meaningfully defined after the *DP flavor*—that is, the collection of the first four building blocks—has been fully spelled out. That is to say, only with answers to ‘who,’ ‘where,’ ‘what’ and ‘how,’ can we render the answer to ‘how much’

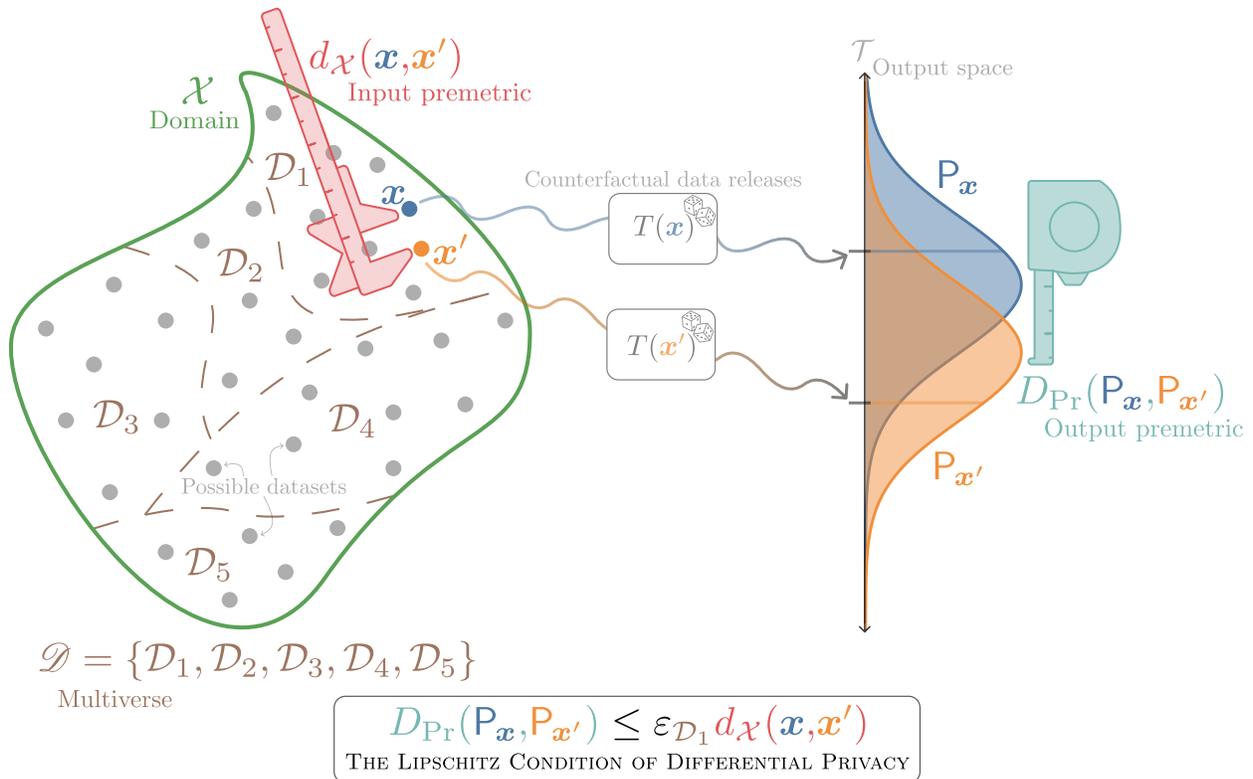


Figure 1: *Schematic of a differential privacy specification  $\epsilon_{\mathcal{D}}\text{-DP}(\mathcal{X}, \mathcal{D}, d_{\mathcal{X}}, D_{Pr})$ .* The *domain*  $\mathcal{X}$  is the set of all possible datasets (be they actual, potential or counterfactual). We denote two arbitrary datasets by  $x$  and  $x'$ ; other possible datasets are depicted by gray circles. The *multiverse*  $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5\}$  is a collection of sets of datasets—these sets are called *universes*. (In this schematic,  $\mathcal{D}$  partitions the domain  $\mathcal{X}$ , as would happen when  $\mathcal{D}$  encodes invariants. In general, this need not be the case. In fact, often the universes may be overlapping.) A data release mechanism  $T$  transforms a dataset  $x$  to a random output  $T(x)$ , which is a draw from the probability distribution  $P_x$ . Intuitively, differential privacy requires that similar datasets  $x$  and  $x'$  have similar *output distributions*  $P_x$  and  $P_{x'}$ . This is formalized by the Lipschitz condition  $D_{Pr}(P_x, P_{x'}) \leq \epsilon_{\mathcal{D}_1} d_{\mathcal{X}}(x, x')$ , which states that the ‘distance’  $D_{Pr}(P_x, P_{x'})$  between the output distributions is at most a constant multiple  $\epsilon_{\mathcal{D}_1}$  of the ‘distance’  $d_{\mathcal{X}}(x, x')$  between the corresponding input datasets. Here, similarity (or ‘distance’) between datasets is measured by the DP specification’s *input premetric*  $d_{\mathcal{X}}$ , visualized above as a caliper, and similarity between probability distributions of the output under different inputs is measured by the DP specification’s *output premetric*  $D_{Pr}$  (the tape measure). For simplicity, we depict the output space  $\mathcal{T}$  as one dimensional, although in practice it is frequently a high-dimensional space, or even a union of many different probability spaces (as is the case for local DP). (The PLB above,  $\epsilon_{\mathcal{D}_1}$ , has the subscript  $\mathcal{D}_1$  because the Lipschitz condition is applied to the datasets  $x$  and  $x'$  which are members of the universe  $\mathcal{D}_1$ , and because the PLB is allowed to vary between universes, potentially taking up to five different values,  $\epsilon_{\mathcal{D}_1}, \epsilon_{\mathcal{D}_2}, \dots, \epsilon_{\mathcal{D}_5}$ .)

a concrete meaning. In fact, a DP flavor can be thought of as a yardstick measuring the protection given by a data release mechanism, while the PLB is a specific measurement on that yardstick; and choices for both the yardstick and its measurement can be used when finding the optimal privacy-utility tradeoff. To make another analogy, PLBs are to their DP flavors as banknotes are to their currencies. At the end of the day, it is not the number on the banknote (or the value of the PLB) that matters, but its purchasing power—which is determined by the strength of the note’s currency (or the PLB’s DP flavor). For example, if Japan announces that tomorrow their currency will be denominated by “centiyen” (which is equal to one hundred yen), the number in a Japanese person’s bank account would change, but the number of Macbooks they could afford would not. In the same way, a ‘redenomination’ of the data custodian’s DP flavor will change their PLB without affecting the actual level of SDC protection, as we explore in the next section. Moreover, this analogy reminds us that just like currencies are proxies to purchasing power, determined by their complex economic and political environment, DP flavors are proxies to the contextual, fluid concept of privacy.

### 3 HOW TO REDUCE ‘PRIVACY LOSS’ WITHOUT ADDING MORE NOISE: A PERVERSE GUIDE

One lesson of the previous section is that the PLB is deficient as an unqualified measurement of SDC protection. Any description that omits the other building blocks opens the door to a variety of ways to ‘cheat’—i.e. ways to reduce the PLB without adding more noise. While these ways are not merely mathematical possibilities or pathological cases, but rather consequences of taking the value of  $\varepsilon_{\mathcal{D}}$  nominally and out of context, their mathematical feasibility can be explained rather simply: In order to alter one component of a DP specification—in this case the PLB—one can maneuver the other components ( $\mathcal{X}$ ,  $\mathcal{D}$ ,  $d_{\mathcal{X}}$  or  $D_{Pr}$ ), or the parameters of the data release mechanism, to that end. These maneuvers may be valid, particularly when the data custodian is transparent about the resulting DP specification, but they can also be used to misleadingly promote a high nominal level of protection (small PLB) especially if shortcomings in the other building blocks are deliberately hidden. Needless to say, our intention is not to encourage unscrupulous behavior, but rather to expose the inherent weaknesses that are open for exploitation in what might appear to be an objective and mathematically absolute framework for ‘privacy’ protection. These warnings may be particularly relevant to commercial implementations of DP, where conflicts of interest are commonplace (see e.g. Waldman, 2021). For example, when the data custodian and data user are the same entity (such as a tech company collecting data about their customers), they may be tempted to cut some differentially-private corners—or engage in some *privacy theatrics* (Smart et al., 2022)—so as to simultaneously improve data utility while maintaining prima facie privacy protection to assuage their data contributors.

#### 3.1. Cheating the System of DP Specifications

In cataloging some of these ways to ‘cheat,’ we start with the first ingredient of a DP flavor: the protection domain,  $\mathcal{X}$ . The less to protect, the less protection budget needs to be spent. Therefore, choosing or interpreting  $\mathcal{X}$  more restrictively may lead to a smaller PLB, even without actually adding more noise to the data. This potentially perverse incentive should receive more scrutiny, because often the choice of  $\mathcal{X}$  is an unconscious one on the data custodian’s part—they are simply given the data they are given—but is indispensable to the interpretation of the DP specification. The domain  $\mathcal{X}$  frames the variables that are contained in the confidential dataset  $\mathbf{x}$  and thus determines the socio-cultural sensitivity of what is being protected—an integral part of any understanding of privacy (Nissenbaum, 2010). More generally, every choice of  $\mathcal{X}$  encodes a data conceptualization (Leonelli, 2019)—a representation *by* the data *of* the individuals who contributed their information. Viewing the data release mechanism as a constituent phase in a data life cycle,  $\mathcal{X}$  specifies the starting point of that cycle. The impact of this choice permeates through other phases in the cycle, notably before data privatization including conceptualization, coding, cleaning, imputation, (sub-)sampling, and so on (Meng, 2021). Restrictions imposed by each of these steps may impact  $\mathcal{X}$  before

the privatization step, thus affecting the PLB (Hu et al., 2024). Take the concrete example of training a large machine learning model, which often involves an initial “pretraining” step, followed by “finetuning.” In this setting, some DP implementations take  $\mathcal{x}$  to be the data used for finetuning; by considering the pretrained model as given, any data used in pretraining is not included in the domain and hence not subject to DP protection (Tramèr et al., 2024). For another example, consider implementing DP in the context of a statistical survey, where the data release mechanism could plausibly start before or after the sampling step—a choice that would change the interpretation of what is being protected, as well as what assumptions about the attacker are required to ensure the nominal level of DP protection (Bailey & Drechsler, 2024).

The second way to apparently spend less PLB without adding more noise is to change the multiverse  $\mathcal{D}$ , the second ingredient of a DP flavor. Piling on more invariants, i.e., summaries of data that will be published exactly by the data release mechanism, is one of such examples, because it creates a more shattered data multiverse  $\mathcal{D}$ . (A multiverse  $\mathcal{D}$  is a shattering (or refinement) of another multiverse  $\mathcal{D}'$  if every  $\mathcal{D} \in \mathcal{D}$  is a subset of some  $\mathcal{D}' \in \mathcal{D}'$ .) For those who appreciate DP as a framework for protecting the *relative* privacy or information, this possibility is rather obvious. The more one discloses via the invariants, the less information left in the data that require protection, and hence a smaller PLB is incurred. Take swapping as a concrete case: as shown in Part II and briefly recapped in Section 4.1, the PLB  $\varepsilon$  of the Permutation Swapping Algorithm is determined by the swap rate  $p$  and the largest stratum size  $b$ . To decrease the nominal value of  $\varepsilon$ , one can either increase  $p$  (up to a point) or decrease  $b$ . When the dataset has a fixed size, the simplest way to decrease  $b$  is to define the stratifying variables at a finer resolution, resulting in smaller strata within which swapping is confined. As illustrated in Part II, the various choices of the stratifying variables at different levels of geography, with or without crossing with the household size variable, result in  $b$  ranging from as small as 11.7 thousand to as large as 13.7 million, and a nominal  $\varepsilon$  from 12.31 to 19.38 (respectively) at  $p = 5\%$ .

A third way to achieve a nominal reduction of the PLB is to redefine the protection units—as captured by the third ingredient of a DP flavor,  $d_{\mathcal{X}}$ —to have a finer granularity. With all else being equal, a DP specification with coarser protection units packs more weight in its PLB; it offers a stronger protection guarantee at the same nominal budget than a specification with on finer protection units. In the opposite direction, when a more expansive protection unit is supplanted by a narrower one, the input premetric  $d_{\mathcal{X}}$  becomes inflated in that a unit change in the former sense may amount to multiple units of change in the latter sense, “watering down” the PLB by the same amount.

This maneuver has been recognized, and to some extent utilized, in the literature on the design of data release mechanism for complex data structures. For example, the choice of neighbors is particularly important for network data: Are neighbors defined by removing a node or an edge from the network, that is, are protection units edges or nodes (Raskhodnikova & Smith, 2016)? For business databases, does a company constitute a unit, or should units be employees, or both (Haney et al., 2017; He et al., 2014; Schmutte, 2016)? Or should they be the company’s transactions? Similarly for large personal databases in commercial settings, should an individual constitute a unit, or should each of their interactions with the platform (such as a post or a ‘like’) be protection units, or should units be the set of a user’s interactions within a given time period (e.g. a single day) (Desfontaines, 2023; Kenthapadi & Tran, 2018; Messing et al., 2020)? Finally, when publishing social statistics, do households deserve SDC protection above and beyond the protection afforded to their individual members (Machanavajjhala, 2022)? Evidently, a data custodian often has a choice regarding which types of entities should be the protection units—a choice that can lead to meaningful protection, or to a superficial nominal guarantee that masks substantive vulnerabilities.

A fourth way to gain nominal PLB out of thin air is to artificially introduce an output premetric  $D_{\text{Pr}}$  that systematically assesses two distributions to be closer than would otherwise be the case. Technically speaking, the relaxation from  $\varepsilon$ -DP to  $(\varepsilon, \delta)$ -approximate DP (ADP) (Dwork et al., 2006a) can be understood as a maneuver of this type.<sup>1</sup> This can be seen by writing out the choice of output premetric corresponding to

<sup>1</sup>This is not to say that  $(\varepsilon, \delta)$ -ADP is never a valid choice—in certain situations the gains to data utility may legitimately

$(\varepsilon, \delta)$ -ADP (see Part I):

$$D_{\text{MULT}}^\delta(\mathbf{P}, \mathbf{Q}) = \sup_{S \in \mathcal{F}} \left\{ \ln \frac{[\mathbf{P}(S) - \delta]^+}{\mathbf{Q}(S)}, \ln \frac{[\mathbf{Q}(S) - \delta]^+}{\mathbf{P}(S)}, 0 \right\}, \quad (2)$$

where  $[x]^+ = \max\{x, 0\}$ ;  $\mathbf{P}$  and  $\mathbf{Q}$  are two probability measures on the same output space  $\mathcal{T}$ ; and  $\mathcal{F}$  is the  $\sigma$ -algebra on  $\mathcal{T}$  (i.e. the collection of events on this output space that are of interest and that permit logically coherent probabilistic assignment). Clearly, for any  $\delta > 0$  and  $\mathbf{P} \neq \mathbf{Q}$ , we have that  $D_{\text{MULT}}^\delta(\mathbf{P}, \mathbf{Q}) < \sup_{S \in \mathcal{F}} |\ln \mathbf{P}(S) - \ln \mathbf{Q}(S)|$ . Since the right hand side of this equation is the output premetric corresponding to pure  $\varepsilon$ -DP, we have reduced the PLB without changing the actual data release mechanism.

### 3.2. Other $\varepsilon$ Evasion Tactics

Because the PLB is used to bound the worst-case rate of change, as seen in Equation (1), any strategy which reduces the extremeness of the worst case will permit a smaller PLB without injecting any additional noise. For example, we can reduce the  $\sigma$ -algebra  $\mathcal{F}$  by pretending that we are interested in fewer possible output events. This can be seen clearly in Equation (2), where the right-hand side is the largest possible value over any possible event  $S$  in  $\mathcal{F}$ . If we reduce  $\mathcal{F}$  to a sub- $\sigma$ -algebra, then the largest possible value may decrease, affording us with a smaller PLB without actually changing the behavior of the mechanism.

The concept of subspace differential privacy (Gao et al., 2022) reflects a maneuver of this type, as it requires control over the output only within sets in the Borel  $\sigma$ -algebra generated by a linear subspace of the ambient output space. Coarsening the  $\sigma$ -algebra associated with the output space signifies a weaker standard against which the data release mechanism is assessed, yet it does not compel the mechanism to be non-measurable with respect to another richer  $\sigma$ -algebra. Therefore, for subspace differential privacy, the mechanism could still take values outside the linear subspace, but no guarantees of SDC protection would be ensured by DP in such cases. (This is different from the requirement of invariants, which operates on the input space rather than the output space of the data release mechanism.)

To further emphasize the importance of understanding the vulnerabilities and subtleties inherent to DP, consider the case where the data custodian uses a constant data release mechanism—that is, a mechanism of the form  $T(\mathbf{x}) = C$ , where  $C$  does not change with  $\mathbf{x}$ . Clearly such a data release mechanism has zero PLB regardless of the DP flavor under consideration, because it is completely insensitive to any manipulation of the input data  $\mathbf{x}$ . But what if the data curator chooses  $C$  to be the same value as the very data or query they are ostensibly trying to protect? Surely that means the PLB should be infinite, since the actual query or data is disclosed exactly. Whereas this may appear to be a pathological case, it carries a critical message: the design of a mechanism cannot be permitted to depend on the confidential dataset itself. This is the very reason that, when designing the 2020 DP mechanisms, the USCB used the 2010 Census data, rather the 2020 data (US Census Bureau, 2023c).

Yet 2010 and 2020 census data are highly correlated. Indeed, if a respondent’s data did not change in 2020 as compared to 2010 (except obviously adding ten years to their age), and the attacker knew this fact, can we really say their data was protected under the DP specification of the 2020 Census? Not necessarily: in the worst case, the DP specification of any data release mechanism is conditional on the attacker knowing all of the data used in the development of the mechanism. This might prompt a data custodian to take a meta-perspective, where the data-dependent design of the mechanism is itself considered as part of a ‘meta-mechanism.’ For the constant mechanism  $T(\mathbf{x}) = C$  described above, where  $C$  was designed to be the actual confidential dataset, the corresponding meta-mechanism is the identity function  $T_{\text{meta}}(\mathbf{x}) = \mathbf{x}$ , whose PLB, infinity, reflects the true level of protection given by  $T$ . The meta-perspective therefore resolves the paradox

---

outweigh the loss to SDC of adopting  $(\varepsilon, \delta)$ -DP. However, often there are other choices which allow for the same gains in utility while requiring that a data release mechanism ‘fail gracefully’ rather than allowing a non-zero probability of ‘catastrophic failure’ (see e.g. Near & Abuah, 2025, Chapter 7).

of this perverse toy example. But real world examples often have complex data-dependent pipelines which are difficult to fully ‘algorithmize’ into meta-mechanisms, making a true implementation of “end-to-end” DP challenging (see e.g. Drechsler & Bailie, 2024).

## 4 HIGHLIGHTS OF PART II: THE US CENSUS’S EVOLVING DATA PROTECTION

The Decennial Census of Population and Housing is a critical piece of US infrastructure. It determines the apportionment of seats in the House of Representatives; it is relied upon for allocating trillions of dollars in federal funding each year; and it informs the decision making of businesses, urban planners and hospitals, amongst many others (National Research Council, 1995; Reamer, 2019; Villa Ross, 2023). Safeguarding such an important data source requires robust SDC, a task that the USCB has long taken seriously. In fact, the bureau has been managing the risk of indirect disclosure in the Decennial Census from 1940 onward (US Census Bureau, 2019). In the 1990 Census, protections were strengthened and a new SDC method was introduced: *data swapping* (Dalenius & Reiss, 1982; Fienberg & McIntyre, 2004; McKenna, 2018).

Data swapping (or record swapping) is a general concept that encompasses a broad class of algorithms. These algorithms select a set of records and then shuffle the values of certain variables among these records. We call the variables whose values are shuffled the *swapping variables*; all other variables are called the *holding variables*. Usually, records are partitioned into groups (or strata) according to the values they take on a subset of the holding variables we call the *matching variables*; and records are only shuffled within their matching group.

The primary SDC methods used in the 1990, 2000 and 2010 Censuses were forms of data swapping, the full technical details of which have not been made public due to confidentiality concerns. We know however that the USCB swapped entire households, rather than shuffling person-level data between households; that the swapping variable was geographic (e.g. block group, tract, or county); and that the matching variables included broader levels of geographies (i.e. tract, county or state) as well as the household’s total counts of adults and children. Furthermore, unique or unusual households that the bureau believed had higher disclosure risk had a higher chance of being swapped.

In the 2010s, spurred by an increasing awareness of privacy risks in statistical products (Dinur & Nissim, 2003), the USCB conducted a *reconstruction attack* on the 2010 Census (Abowd et al., 2023). Using published tables, and publicly available information on the relationships between these tables, they were able to determine with a high degree of confidence much of the underlying post-swapped microdata which produced these tables (see Section 6.1). This led to a revolution at the bureau—the 2020 Census would not be protected by using data swapping, as was the case for the previous three decades, but rather by brand new SDC methods which were explicitly designed to satisfy DP (Abowd, 2018). Yet, does the USCB’s official adoption of DP, on its own, truly represent a sea change in how it protects the census? What if, in fact, the census was already protected in 2010 by a DP method—or at least by a method which is very similar to a DP one?

### 4.1. A DP Guarantee for the 2010 Census?

While data swapping was not originally invented with DP in mind, it may still be possible for it to satisfy some DP specification. However, it is effectively a method for adding noise *only* to the relationship between the swapping and holding variables within each matching group. As such, the marginal distributions of these three sets of variables are invariant under swapping, as is the joint distribution of the swapping and matching variables. Data swapping can therefore only be DP subject to the invariants it induces—i.e. it can satisfy a DP specification only if the specification’s data multiverse respects its invariants. This limiting of the scope of protection greatly reduces the SDC guarantee provided by DP, as we discuss extensively in Section 6.

Moreover, some of the technical implementation details of the 2010 swapping procedure preclude it from satisfying a pure  $\varepsilon$ -DP specification (i.e. a specification whose output premetric  $D_{Pr}$  is the same as the one used in the original DP specification of Dwork et al. (2006b)). Nevertheless, the *Permutation Swapping Algorithm* (PSA)—which keeps to the spirit of the 2010 procedure, if not the exact implementation—does satisfy pure  $\varepsilon$ -DP subject to the invariants it induces.

The PSA is very simple to describe: It selects records independently with probability  $p$  (the ‘swap rate’) and then permutes the values of those records’ swapping variables.<sup>2</sup> The following statement is an informal version of the main result of Part II, which proves that the PSA satisfies a DP specification.

**Theorem II.1** (informally). *Subject to the invariants induced by it, the PSA is  $\varepsilon$ -differentially private, with*

$$\varepsilon \leq \begin{cases} \ln(b+1) - \ln o & \text{if } 0 < p \leq 0.5, \\ \max\{\ln o, \ln(b+1) - \ln o\} & \text{if } 0.5 < p < 1, \end{cases}$$

where  $o = p/(1-p)$ , and  $b$  is the size of the largest matching group.

In nontechnical terms, the first four building blocks of the PSA’s DP specification are as follows. Firstly, the protection domain  $\mathcal{X}$  is any set of datasets which all share the same variables. Secondly, what “subject to the invariants” means is that among all possible datasets in  $\mathcal{X}$ , the only ones that carry the protection guarantee are the ones that share the same invariant values with the actualized, confidential dataset. This is because other datasets will be excluded from consideration by any competent attacker due to the fact that their invariant values are different to those published. Mathematically, the invariants partition the domain  $\mathcal{X}$  into universes, with all datasets in each universe sharing the same invariant values. By “subjecting DP to the PSA’s invariants,” we mean that DP’s Lipschitz condition (see Equation (1)) is restricted to datasets in the same universe. Hence, comparisons between datasets with different values of the invariants are excluded from consideration.

Thirdly, the granularity of the PSA’s DP specification is equal to the resolution of the PSA’s swaps. For example, if the PSA swaps person records, then the granularity of protection is person records. More technically, we mean that the PSA’s input premetric  $d_{\mathcal{X}}$  is the Hamming distance on person records. Thus, if a number  $n$  of person records are changed, the distribution of the PSA’s output will change by at most  $\varepsilon n$  units, as measured by the output premetric of the PSA’s DP specification (assuming that the changes to these records do not result in a change in the value of any of the invariants). Fourthly, this output premetric is given by the maximum likelihood ratio—that is, the maximum value, over all possible outputs  $t$ , of the relative likelihood of observing output  $t$ , under input dataset  $\mathbf{x}$  as compared to under the some alternative input  $\mathbf{x}'$ . Hence, the output premetric of the PSA’s specification corresponds to the notion of  $\varepsilon$ -DP.

Part II then instantiates the PSA to align as closely as possible with what we know about the 2010 swapping procedure. This gives us a concrete DP specification reflecting the SDC protection afforded to the 2010 Census—although with the caveat that this assumes this census was protected by the PSA. As such, we emphasize that we do not provide a DP specification for the actual 2010 Census since such a specification must reckon with the exact implementation details of the 2010 SDC methods, not the PSA. However, because we believe the PSA can closely parallel the 2010 swapping procedure by appropriately choosing its implementation parameters, the resulting DP specification is nevertheless a useful perspective on the protection provided to the 2010 Census.

To mimic the 2010 swapping procedure, the choices for the PSA’s implementation parameters are as follows. The 2010 DAS was integrated into the census data pipeline after all imputation and editing processes were completed; we place the PSA at the same place in the pipeline. This means the PSA’s protection domain is the set  $\mathcal{X}_{\text{CEF}}$  of all possible *Census Edited Files*—i.e. all hypothetical outputs resulting from the first stages

<sup>2</sup>Incidentally this idea, under the name  $n$ -Cycle swapping, was under active investigation by the USCB up until the bureau redirected its research efforts towards DP (McKenna & Haubach, 2019).

of the census data pipeline through to the imputation and editing processes. This has important implications on what data is actually being protected by the PSA: the edited and imputed records, not the ‘raw’ Census responses. Moreover, because  $\mathcal{X}_{\text{CEF}}$  determines what it means to counterfactually alter data, it also has implications on what the protection unit in 2010 was. Indeed, mirroring the 2010 swapping procedure, we set the PSA to swap household-level records, so that its input premetric  $d_{\mathcal{X}}$  is the Hamming distance on household records; yet this does not imply the 2010 protection units were households—because a change in a single record of the Census Edited File does not always correspond to a single household changing their Census responses. Instead, the 2010 protection units are ‘post-imputation households’—imaginary entities which can alter their own records in the Census Edited File freely without affecting other, imputed records. (See Part II for an explanation of how the PLB must be inflated in order to have households as the protection unit.)

We set the PSA’s matching variables to be the household’s state and size, and its swapping variables to be the household’s county. This results in a multiverse  $\mathcal{D}_{2010}$  where all statistics at the state and national levels are invariant, as well as the counts of households by size at the block level. Finally, we set the swap rate  $p$  to be 2-4%, as Boyd and Sarathy (2022) states was used in 2010. This results in a PLB  $\varepsilon$  between 18.29 and 19. However, since the 2010 swapping procedure also included the number of voting-aged people as a matching variable, 18.29-19 is an upper bound for 2010’s approximate PLB. (We cannot compute the PLB when the number of adults is invariant because the necessary statistic—the value of  $b$  in Theorem II.1—is not publicly available.) However, this also implies that  $\mathcal{D}_{2010}$  gives a lower bound on the invariants: in addition to the statistics reported above, the block-level breakdown of households by the number of adult occupants (and hence also by the number of children occupants) were invariant.

#### 4.2. The DP Guarantee of the 2020 Census

Having established that the 2010 Census may be analyzed from the perspective of DP, it is fruitful to compare it with the DP specification of the 2020 Census. We start this comparison by examining each of the five DP building blocks in turn. As in 2010, the 2020 disclosure avoidance system (DAS) also took the Census Edited File as input. Hence, the protection domain remains constant across the two census; in both cases it is the set of all possible Census Edited Files  $\mathcal{X}_{\text{CEF}}$ . Secondly, the granularity of protection in 2020 was person-level records. All other components being equal, this would imply weaker protection than in 2010, which protected household-level records; but, as we will see, the three remaining components are not equal.

Most importantly, the 2020 DAS has far fewer invariants than was the case in 2010. The 2020 invariants were carefully considered and minimized to those required by operational and constitutional mandates. These invariants are the state populations as well as the counts at the block level of housing units and of each type of occupied group quarters. As we will discuss in Section 6, initial analysis suggests that these invariants have minimal effect on SDC; the same cannot be said about 2010’s invariants (Abowd et al., 2023).

The output premetric (i.e. ‘standard of protection’) used in 2020 is what we call the *normalized Rényi metric*; this corresponds to the notion of zero-concentrated DP (zCDP) (Bun & Steinke, 2016). Under these settings for the first four components, the 2020 Census’s PLB is given by  $\rho^2 = 55.371$ . (We follow the standard convention of using  $\rho$  to denote the PLB in the context of zCDP.) Additionally, we may translate from zCDP to  $(\varepsilon, \delta)$ -DP and thereby also express the 2020 PLB by  $\varepsilon = 126.78$  with  $\delta = 10^{-10}$ . (To be clear, in this translation across DP specifications, the first three building blocks stay the same, while the output premetric changes from the normalized Rényi metric (defined in Part I) to the  $\delta$ -approximate multiplicative divergence, defined in Equation (2).)

It is worth noting that the above DP specification only assesses the disclosure risk associated with the primary 2020 Census products. As of the time of writing, there have already been other releases of the 2020 Census data, and more are planned for the future.<sup>3</sup> While we have not been able to obtain information on

<sup>3</sup>The primary 2020 Census data products are the P.L. 94-171 Redistricting Summary File (US Census Bureau, 2021a, 2021b),

the SDC of these releases, they will necessarily increase the 2020 PLB and may possibly weaken the other components of the 2020 DP specification. A degradation of the DP specification due to additional data releases is not a concern under data swapping; the DP specification from the previous subsection covers all the 2010 Census products because they were all generated from the post-swapped microdata (see the end of Section 5).

Beyond the core details of the 2020 DP specification explained above, several other aspects merit attention. Firstly, as in 2010, setting the protection domain to be  $\mathcal{X}_{\text{CEF}}$  in 2020 has important implications: The 2020 Census does not have “end-to-end” DP protection (c.f. Hu et al., 2024); its protection units are ‘post-imputation persons’ and as such does not provide protection to individuals’ census responses directly. Secondly, a more nuanced perspective on the 2020 DAS would examine its *per-attribute* PLBs (Ashmead et al., 2019). A per-attribute analysis considers a DP specification in which only one variable (i.e. attribute) is allowed to vary within each data universe. This allows for a more fine grained assessment of SDC, rather than assuming the worst-case possibility of complete dependence between variables when composing the per-attribute budgets into a single total budget. Apart from the following two observations, we leave this important discussion to future work: The per-attribute budgets are much smaller than the overall 2020 PLB; and a per-attribute analysis is not applicable to data swapping since its DP specification does not rely on the composition theorem of DP.

## 5 REAPING THE BENEFITS OF BOTH TRADITIONAL SDC AND DP

In the previous section, we analyzed data swapping, a quintessential example of traditional SDC, via the system of DP specifications, and we used this analysis to compare the SDC protections provided in the 2010 and 2020 Censuses, sketching an evolution of the USCB’s approach to disclosure avoidance. In this section, we demonstrate how this analysis also allows data swapping to enjoy some of the distinctive advantages of DP without forgoing the strengths of traditional SDC. While other work have examined other traditional SDC methods through the lens of DP to varying degrees of success (Bailie & Chien, 2019; Chien & Sadeghi, 2024; Neunhoeffer et al., 2024; Rinott et al., 2018), continuing to bridge the gap between these two paradigms bestows opportunities to reap the best of both worlds, since both approaches each have their own unique advantages (Slavković & Seeman, 2023).

On one hand, the PSA’s DP specification gives a precise, mathematical formulation of the SDC it provides. It delimits the information the PSA does not protect (its invariants) and the extent to which it protects the remaining information. It describes the granularity at which attackers are limited in learning about aspects of the confidential data that are not disclosed by the invariants alone, and the standard against which this learning is measured. In short, the PSA’s DP specification answers the ‘who’, ‘where’, ‘what’, ‘how’ and ‘how much’ questions of protection. These answers serve as a useful aid in assessing whether the PSA is appropriate for a given data release, and, if so, in choosing its implementation parameters.

While this paper has largely focused on its ability to describe SDC protection, DP is not just a descriptive framework. It also provides a calculus for reasoning about how protection loss accumulates across multiple data products, a tool which is becoming increasingly more valuable as national statistical offices diversify their offerings (Kitchin, 2015). Moreover, adopting a DP flavor as the yardstick for measuring SDC permits complete transparency of the data release mechanism, since DP guarantees do not degrade with the attacker’s *knowledge of the mechanism* (in contrast, degradation can occur with the attacker’s knowledge of the relationships among the data subjects). For example, even when armed with the complete knowledge of the PSA’s implementation details (including the values of all its parameters, such as the swap rate or

---

the Demographic and Housing Characteristics (DHC) File (US Census Bureau, 2023a), the Detailed DHC-A and -B Files (US Census Bureau, 2023b, 2024a), the Supplemental DHC (US Census Bureau, 2024b) and related auxiliary products. Examples of future releases that rely on the 2020 Census include the annual Population and Housing Unit Estimates and the National Population Projections (US Census Bureau, 2023d, 2023e). See Part II for more details.

swap key, but excluding, of course, the value of its random seed), it is still impossible for an attacker to thwart the SDC protections described by its DP specification. While transparency does not assist attackers in breaking DP’s protection guarantees, it is important to legitimate data users as an essential prerequisite for valid statistical analysis of privacy-protected data (Gong, 2022). Indeed, in Section 6.3 we explain how, by allowing quantitative analysts such as statisticians and social scientists to correct for the statistical errors induced by SDC protection, transparency increases data utility and supports robust research findings.

On the other hand, traditional SDC techniques also have their own value, with data swapping in particular enjoying advantages that most DP methods do not. For example, swapping maintains *facial validity*—the 2010 Census outputs all look ‘reasonable’ in the sense that there are no negative or fractional counts, nor are there implausibly large or small reported values. More generally, the 2010 publications pass the sanity checks an observant reader might make, which is useful for building trust in the census among the general public. From the opposite perspective, a lack of facial validity is an important concern for statistical agencies like the USCB. It presents an issue for data users who are confused and disinclined to use seemingly erroneous data; it erodes the public image of the agency; and it hampers efforts to improve differential response rates among disadvantaged communities (boyd & Sarathy, 2022; Drechsler, 2023; Oberski & Kreuter, 2020).<sup>4</sup>

*Logical consistency* is another advantageous property of data swapping, as it ensures all statistics produced from the 2010 Census align with one another. For example, in any contingency table released in 2010, the sum of cells in a single row or column always matches the marginal total. Likewise, reported values for the same count remain consistent across different publications. Additionally, the 2010 Census outputs respect the structural zeroes and edit constraints present in the underlying confidential microdata. In contrast, many of the 2020 Census outputs will not be consistent across publications and even within the same publication, row- and column-sums will not match the reported totals. (However, the outputs produced by the TDA—the PL and DHS files—are logically consistent, although as we will see, this comes at a cost.)

Like facial validity, logical consistency is important for users and advocates of census data (boyd & Sarathy, 2022; Hotz & Salvo, 2022; Ruggles et al., 2019). Yet, many DP methods do not maintain facial validity nor logical consistency, and others, such as the TDA, cannot satisfy these properties without partially destroying statistical transparency.<sup>5</sup> This is because the majority of DP methods rely on optimization-based post-processing to restore facial validity and logical consistency (e.g. Barak et al., 2007; Hay et al., 2010). Optimization based post-processing can be algorithmically transparent but in most cases it destroys the statistical transparency of the resulting two-step privacy mechanism—a crucial requirement for principled statistical analysis (Gong, 2022). The recent proposal by Dharangutte et al. (2023) does away the need for post-processing when the noise infusion is additive. However, it relies on MCMC sampling and hence is nontrivial to implement for large-scale data products. In contrast, swapping achieves facial validity and logical consistency automatically without the need for additional computation.

Furthermore, data swapping—like other traditional SDC techniques but unlike many DP methods (such as those used in 2020)—is easy to communicate and understand at a high level by a broad, nontechnical audience. This is important for building trust and maintaining the buy-in of data providers, custodians and other stakeholders. Swapping is also easily implementable and amenable to the types of data collected by government agencies, as evidenced by its use in the US, the UK and the EU (de Vries et al., 2023; McKenna, 2018; Office for National Statistics, 2023).

---

<sup>4</sup>Related, but distinct, to facial validity is the concept of *face privacy*, which is the requirement that the data release mechanism produce output which appears to the casual observer to offer privacy protection (Hod & Canetti, 2025). In contrast, facial validity requires that the output is a plausible representation of the real world, even to a data user who is unaware that artificial noise was added for SDC protection.

<sup>5</sup>A data release mechanism  $T$  is statistically (or probabilistically) transparent if the conditional probability distributions  $P_{\mathbf{x}}(T \in \cdot)$  are public knowledge (Gong, 2022). Statistical transparency is distinct to algorithmic transparency, which requires that the source code of the mechanism  $T$  is disseminated. For all practical purposes, statistical transparency is a stronger requirement than algorithmic transparency since  $T$ ’s source code may be so complex that it is practically impossible to derive the conditional distribution it induces.

Finally, as a pre-tabular perturbation method, swapping also has the advantage of producing a ‘synthetic’ dataset that serves as the source for all census publications. This simplifies the data release process as all outputs are derived from this ‘post-swapped’ data without requiring additional SDC treatment. This also explains why the 2010 publications maintain logical consistency; because no further noise is introduced, all outputs are consistent with the post-swapped data and hence also with each other. Moreover, this approach ensures that the release of additional data products in the future not only maintains *backwards compatibility* with previous products, but also does not degrade their DP specification. Indeed, as long as all products are based on the same post-swapped microdata, they are all covered under the DP specification of the swapping algorithm which produced this microdata (by DP’s post-processing theorem). In this way, data swapping allows the statistical agency to publish a single DP specification which encompasses all existing and future publications. As mentioned in Section 4.2, this stands in contrast to the 2020 Census. Each publication from the 2020 Census has its own DP specification, and to understand the SDC provided to the census data as a whole, one must aggregate these DP specifications into a single comprehensive one—a process which must be repeated with each new publication. And, because every data product introduces additional disclosure risk, the overall 2020 DP specification weakens after each new release, in comparison to the single, upfront DP specification associated with data swapping.

## 6 DISCLOSURE RISK, TRANSPARENCY, AND DATA UTILITY

### 6.1. Understanding the Impact of Invariants on Disclosure Risk

A major criticism of the swapping method implemented in the 2010 Census is that it induces too many invariants. One salient consequence of a plurality of invariants is that it severely constrains the permissible values for the confidential data. Indeed, the larger the number of invariants, the more datasets an attacker can rule out as impossible, and, consequently, the higher the risk of disclosure. As Abowd and Hawes (2023) discuss, the invariants in the 2010 Census elevate disclosure risk because, not only are they numerous, but they also include information at a very fine granularity, for example, the total and voting age populations at the block level.

The USCB’s *reconstruction-abetted re-identification attack* against their 2010 Census (Abowd et al., 2023) provides some understanding of the impact of its invariants on disclosure risk, although, as we will see, there were also other confounding factors at play. Generally speaking, a reconstruction-abetted re-identification attack consists of two steps: a reconstruction attack, which creates a permissible version of the confidential microdata; followed by a re-identification attack, which links this ‘reconstructed’ dataset to an external source in order to attach names or other personally identifying information to (some or all of) the reconstructed records. The reconstruction step works by collating many publicly available, aggregate statistics about the confidential (unknown) microdata and then constructing a dataset which agrees with these statistics, or—in the case where the published statistics are noisy functions of the confidential microdata —*inferring* a dataset based on a statistical model of the published data.<sup>6</sup> This reconstructed dataset is a plausible guess for the confidential microdata, since—in the case where the statistics are deterministic functions of the confidential microdata—it generates identical statistics to the ones generated by the microdata. When the statistics are noisy, the situation is more complex: the reconstructed dataset does not necessarily reproduce the published statistics exactly due to their stochasticity, but nevertheless it still could plausibly be the microdata which generated these statistics. In any case, the larger the number of published statistics and the more accurate they are, the more heavily they constrain the possible configurations of the reconstructed dataset, and hence the more likely it is that this reconstruction agrees with the true confidential microdata.

---

<sup>6</sup>This model has the unknown microdata as its parameters, the released statistics as its data, and the SDC method which generated the published statistics from the confidential microdata as its data generating process. For example, the reconstructed dataset could be the maximum likelihood estimate under this model, or it could be a draw from a Bayesian posterior which is compatible with this model.

For the reconstruction attack on the 2010 Census, the underlying microdata the USCB targeted was not the Census Edited File  $\mathbf{x}_{\text{CEF}}$ , but the post-swapped data—i.e. the resulting records after swapping had been applied to  $\mathbf{x}_{\text{CEF}}$ . This avoids the complication of designing a reconstruction attack which accounts for the noise introduced by swapping. Yet, because of the low swap rate and the large number of invariants in the 2010 DAS, there is a high degree of alignment between the reconstructed data and  $\mathbf{x}_{\text{CEF}}$ . As a result, linking the reconstructed data to an external dataset containing personally identifiable information allows for the possibility of learning potentially sensitive data: the race and ethnicity of census respondents. The USCB’s reconstruction experiments further suggest that the rate of swapping must be significantly increased to achieve what it deemed as an acceptable level of protection (Abowd & Hawes, 2023). It is from these observations the bureau concluded an urgent need to revamp their swapping-based SDC. However, the question remains open: in what ways does imposing a specific set of invariants impact the disclosure risk of the resulting data product?

To understand the relationship between swapping’s invariants and disclosure risk, two caveats are worth noting at the outset. First, the degree of vulnerability to a reconstruction attack is a measure of *absolute disclosure risk* (Duncan & Lambert, 1986; Reiter, 2005), defined as the degree of certainty with which an attacker can make inferences about confidential information from the published data. Unless strong assumptions about the attacker’s prior knowledge are made, DP does not directly translate into any quantifiable degree of control over the absolute disclosure risk; see e.g. Dwork and Naor (2010), Hotz and Salvo (2020), Kifer and Machanavajjhala (2011), and McClure and Reiter (2012). Similarly, absolute measures of reconstruction attack success (such as percentages of successfully reconstructed records, as reported by the USCB) are not controlled by DP (Francis, 2022; Kenny et al., 2021). Second, invariants are not unique to swapping, nor should they be viewed as a static byproduct. The final choice of invariants used in the 2020 TDA was arrived at by the USCB through an iterative process. For example, block-level populations were once considered as an invariant, but were ultimately not included (Abowd et al., 2022; Ashmead et al., 2019; Kifer, 2019).

Notwithstanding these caveats, it is worthwhile to inquire, to the extent possible, about the impact of invariants on disclosure risk through the lens of DP. Such an inquiry can be challenging within the standard DP paradigm, because the impact of invariants cannot be adequately captured by the PLB. (The PLB will be infinite whenever there are invariants which are not permitted by the DP flavor in question, regardless of the number of such invariants.) By contrast, the system of DP specifications laid out in Part I is more dexterous as invariants can be explicitly included through the multiverse  $\mathcal{S}$ . An analysis of the impact of invariants on disclosure risk therefore amounts to a five-dimensional comparison between alternative DP specifications that differ on  $\mathcal{S}$  and potentially on other building blocks as well. A comprehensive description of the five-way dynamics remains open for future research, although investigations with a restricted scope (e.g. only varying two or three dimensions at a time, rather than all five) can already be informative. For example, it can be shown that reconstruction attacks can be increasingly successful if applied to DP protected data when more invariants are imposed on them (Protivash et al., 2022). Analysis in Part II also indicates that the granularity of swapping’s invariants has a large numerical impact on the PLB  $\varepsilon$  (through the largest stratum size  $b$ ), while the swap rate has comparatively little influence. This suggests a reduction of invariants may have a larger impact on SDC protection compared to an increase in the swap rate—at least when protection is measured by a DP flavor. Finally, crude comparisons can be made between swapping and the TDA based on their vulnerability to reconstruction attacks. From experiments in Abowd et al. (2023), swapping with a high swap rate of 50% is roughly comparable in its protection against reconstruction attacks as the TDA (using the 2020 production settings). This suggests that the reduction in invariants from the 2010 data swapping method’s invariants (which are numerous) to the TDA’s invariants (which are few) is equivalent to an increase in PLB from  $\varepsilon = 15.19$  at the household level to  $\varepsilon = 52.83$  (with  $\delta = 10^{-10}$ ) at the person level.

## 6.2. Mitigating the Impact of Invariants on Disclosure Risk

Because some small number of invariants is frequently mandated, while a large number may have an adverse impact on disclosure risk, the modern statistical agency needs methodologies that allow for the specification of invariants in a flexible and precise manner. To this end, swapping—as instantiated either in 2010 or in our work—does not suffice because its invariants are largely hardwired into its mechanics. However, several extensions to data swapping enable more customization in the choice of invariants, thereby allowing for a better balance between SDC protection and accuracy targets. (On this topic, it is also worth noting that the TDA allows for a range of invariant choices.)

One such extension is *probabilistic unit matching*, which was considered by the USCB as part of its comparative analysis between data swapping and the TDA. Instead of using the swapping key to form hard strata within which swapping is confined, this method permits units across different strata to be swapped with a small probability, which could be inversely proportional to some distance metric on the strata. For example, consider using county as the swapping variable, with state and household size as the matching variables. Suppose that, for some  $\alpha > 0$ , a household chosen for swapping would have a  $(1 - \alpha)\%$  chance of being swapped with another household of the same size, but an  $\alpha\%$  chance of being swapped with a differently sized household. Doing so retains the county-wide household counts as invariant, but the county-wide total populations would no longer be invariant.

Two other approaches to remove some of swapping’s invariants are *pre-* and *post-swap perturbation*. As their names suggest, the former infuses noise into the confidential records prior to applying swapping (Hawes & Rodríguez, 2021, p. 23), whereas the latter perturbs an intermediate data product after applying swapping. Notably, data swapping followed by tabular perturbation is a common SDC strategy; for example, it is the approach taken by the Office of National Statistics (ONS) for the protection of the 2021 UK Census (Office for National Statistics, 2023). In this case, the cell key method (CKM) is employed to perturb the cells of contingency tables after targeted swapping has been applied to the underlying microdata (Fraser & Wooton, 2005; Marley & Leaver, 2011; Thompson et al., 2013). Notably, the CKM procedure has been analyzed through the lens of DP (Bailie & Chien, 2019; Chien & Sadeghi, 2024; Rinott et al., 2018). In addition to its use at the ONS, applying swapping and then CKM perturbation is also recommended by Eurostat’s *Centre of Excellence on Statistical Disclosure Control* for EU censuses (Glessing & Schulte Nordholt, 2017).

We leave to future work a full investigation of the protections supplied by probabilistic matching and by swapping combined with pre- or post-swap perturbation. Note that compared to standard swapping algorithms such as the PSA, all three of the above procedures introduce strictly more auxiliary randomness into the data product. It would therefore be reasonable to expect the resulting algorithms to enjoy DP guarantees while supplying fewer and more flexible choices of invariants. One salient question for this line of research is to determine the DP specification for the ‘chaining’ (i.e. sequential composition) of two mechanisms (e.g. swapping followed by tabular perturbation), when these mechanisms satisfy different DP specifications.

## 6.3. Transparent Privacy: Epistemic Uncertainty and Data Utility

A common principle in traditional SDC is that privacy requires some degree of obscurity. A data custodian should not fully reveal the implementation details of their SDC method, because these details could be used by an attacker to unpick the method’s SDC protection (see Slavković and Seeman, 2023 and references therein). As such, the privacy protection offered by any randomized SDC algorithm is not solely due to the aleatoric uncertainty associated with the noise it injects, but also due to the epistemic uncertainty arising from the attacker’s deficient knowledge of the algorithm. The inherent randomness of the SDC method provides direct protection, and the plausible deniability surrounding how exactly the method was implemented provides indirect protection. However, epistemic uncertainty is unfortunately harder than its aleatoric counterpart to model and reason about; see for example the extensive literature on imprecise probabilities (IP) (Augustin et al., 2014; Shafer, 1976; Walley, 1991). Consequently, the privacy protection

provided by the epistemic uncertainty of an SDC method is difficult to describe with mathematical guarantees and, as far as we are aware, has not been systemically studied. Nevertheless, the principle of ‘privacy through obscurity’ is frequently invoked by national statistical offices in reference to their SDC methods (Chipperfield et al., 2016; McKenna, 2018; UK Statistics Authority, 2021).

Establishing a DP guarantee for a traditional SDC method provides a rationale for reducing obscurity: as the DP guarantee does not degrade with knowledge of the mechanism (as explained in Section 5), the data custodian is justified in being publicly transparent about the SDC method. This would provide crucial insights into the quality of existing data products in ways that were not previously possible. For example, data swapping is not traditionally a transparent SDC technique; in fact, the currently available public documentation on the 2010 DAS is deliberately deficient, stymieing researchers’ ability to appropriately account for the noise it injects into census data (Kenny et al., 2024). The explicit statement of the DP specification of the PSA provides a theoretical justification for releasing the implementation details of the 2010 swapping procedure. Indeed, the publication of a complete description of the 2010 DAS, as the USCB has done for the 2020 DAS, would greatly facilitate any comparative analysis at the heart of the on-going debates about the tradeoff between SDC and data utility in the census.

There are more reasons to publish the algorithmic specification of SDC procedures like swapping. Despite being the main SDC method for the census in 1990-2010, it remains difficult to fully quantify swapping’s adverse impact on data quality. A peek into the technical specification of swapping can add tremendous utility to the data products it protects. Data users who conduct complex statistical modeling with official data products may particularly benefit from the transparent knowledge of the SDC methods used in creating those products. For example, it is well understood in the literature that data swapping negatively affects the quality of downstream data analyses. Specifically, because swapping adds noise to the relationships between the swapping and holding variables, it reduces researchers’ ability to study these relationships. Mitra and Reiter (2006) and Drechsler and Reiter (2010) demonstrate that even low swap rates (e.g. 5%) can substantially reduce the effective coverage of confidence intervals for the regression coefficient between swapping and holding variables. Such deterioration in coverage is in no small part due to performing naïve statistical analyses which does not account for the SDC mechanism. As Gong (2022) demonstrates, performing standard regression analyses on data protected via DP noise infusion results in similar types of coverage deterioration. Furthermore, this deterioration can be restored once the DP method is statistically modeled—at the expense of wider, though valid, intervals. With that said, an analyst cannot be blamed for performing naïve analyses on swapped data when details of the swapping procedure are not public knowledge.

Publishing the details of the USCB’s swapping methods will also help to dispel the “statistical imaginaries” associated with the censuses of 1990-2010 (boyd & Sarathy, 2022; Sarathy & boyd, 2024). It is easier to argue that census data contain errors due to SDC protection when one can point precisely to how these errors were introduced and quantify their distribution exactly. Transparency enables concrete statements like, ‘the expected error due to SDC protection is  $x\%$ ,’ in place of vague expressions such as, ‘these counts may have some error as their contributing households could have been swapped.’ In general, raising awareness about an official data product’s SDC errors by publicly documenting the methods used to protect them is a step towards “shifting the statistical imaginary to account for uncertainty” (boyd & Sarathy, 2022), thereby improving the legitimacy of the data product and the perception of accountability of the data custodian. Such a paradigm shift occurred in 2020 due to the USCB’s transparency about their new SDC methods; a similar shift in the imaginary of earlier census data could be sparked by a comparable level of transparency about their 2010 SDC method.

Before closing, we caution that the idea a DP method can be transparently disseminated at no cost to privacy rests on two premises. The first is that the epistemic uncertainty of an SDC method is not a legitimate form of privacy protection. DP follows an intellectual tradition originating in cryptography which dismisses epistemic uncertainty as a brittle form of protection—brittle because it depends on a watertight security system to safeguard an SDC method’s implementation details. Indeed, any measure of SDC protection which relies

on epistemic uncertainty must depend on assumptions about the attacker’s knowledge of the data release mechanism; therefore, epistemic-based SDC can degrade as an attacker learns new information. As such, existing measures of protection found in the DP literature—i.e. DP specifications—are solely based on the aleatoric uncertainty arising from an SDC method’s stochasticity, justifying DP’s tenet of ‘transparency for free.’ Nevertheless, our system of DP specifications could in principle be extended to also incorporate protection based on epistemic uncertainty.

Towards this end, one open research direction is to leverage machinery from IP to quantify the epistemic uncertainty due to partial knowledge of the data release mechanism. If, instead of full statistical transparency of the data release mechanism (i.e. public dissemination of the likelihood  $P_{\mathbf{x}}$ ), the data custodian could opt to publish an IP version of  $P_{\mathbf{x}}$ . This would provide a level of transparency which is sufficient for principled statistical analysis because an analyst can correct for the noise introduced by the mechanism using IP tools. At the same time, an attacker’s ability to infer confidential information would be reduced because they no longer have access to the precise  $P_{\mathbf{x}}$ , but only to its IP version. This increase in protection could be incorporated into a DP specification through the use of an output premetric which quantifies the difference between two IP objects, rather than between two precise probabilities. Research along these lines will explore new Pareto frontiers in the tradeoff between SDC and utility by introducing—and rigorously quantifying—a controlled amount of ‘privacy through obscurity’ without losing the transparency required for statistical analysis.

The second premise is that the dissemination of a DP method’s specification does not itself incur a cognizable privacy cost. This premise becomes questionable when the specification includes invariants. In the case of the PSA (or any other invariant-preserving DP mechanism), publishing its specification necessitates revealing the complete list of all its invariants. However, as we have repeatedly emphasized, there may be a disclosure risk associated with the knowledge of what swapping’s invariants are, in and of itself—especially when its invariants are at a fine level of granularity. For example, the 2010 invariants are privacy-revealing in an intuitive sense, even if they are not technically privacy revealing in the sense that their publication does not increase the PLB under the PSA’s DP flavor. Indeed, too many invariants supply the attacker with confidence in their efforts to reconstruct the microdata and reidentify individual records. With many invariants present, the DP specification itself could carry a non-trivial amount of information, and some could consider that a breach of privacy, even though the sense of privacy here extends beyond the scope of the DP specification in question. We are sympathetic to this view and therefore advocate for careful deliberation, weighing the cost of making public the invariants (through releasing the DP specification) against the benefits of algorithmic transparency this allows.

## 7 CONCLUSION: So... Is SWAPPING DP?

In this trio of papers, we have taken a “stirred, not shaken” approach to the two central subjects under study: DP and data swapping. Our goal is neither to revamp these notions nor to rob them of their essence. On the contrary, our overarching goal is to *reveal* their essence, which is achieved by laying down a rigorous explication of DP through our five-building-block system, and by reexamining data swapping through the lens of this system. We want to be clear that we are not advocating to reverse the progress that the USCB and the data privacy research community have made to advance SDC. In fact, we seek to continue this progress by building stronger connections between DP and traditional SDC so as to reap the benefits of both fields. We believe that this approach ultimately facilitates the modernization of SDC in a way that is helpful to data custodians, responsible to data contributors, and respectful to data users.

Returning to the titular question of this work, we observe that the term DP has been used to refer to multiple distinct notions, potentially leading to some understandable confusion on the part of the reader. Indeed, while summarizing an entire field of work is difficult, we identify four broad concept categories associated with

DP. Firstly, we have primarily used the term DP to refer to the class of technical standards encapsulated by the Lipschitz condition in Equation (1), which we have variously called DP definitions, formulations, or specifications. Secondly, DP as a field also consists of its data release mechanisms, which can come in three different representations: as a mathematical abstraction (for proving the mechanism satisfies a DP specification); as an algorithmic implementation (for using the mechanism); and as a sociotechnical deployment (for embedding the mechanism into its real world context). (See Seeman and Susser (2024) for a discussion on the distinctions between these three representations of a data release mechanism, and why these distinctions matter.) Thirdly, DP can be considered as a framework by augmenting the first two concepts—DP specifications and mechanisms—with a set of tools for reasoning about these concepts (e.g. composition theorems or programming frameworks (Gaboardi et al., 2024; Steinke, 2022)).

Finally, DP has a philosophy which formulates a particular way of thinking about data privacy (Bun et al., 2021; Dwork, 2006; McKay Bowen & Garfinkel, 2021). For our purposes, we focus on one aspect of this philosophy: By default, every attribute in the data should be treated as sensitive, since we do not know how it may be used by an attacker in the future. As such, every statistic—even if noisy and only partially informative of the dataset’s attributes—should be considered as privacy-revealing.<sup>7</sup> In contrast, the philosophy behind data swapping, as explained in Part II, is based on thwarting re-identification attacks by adding noise to the relationship between sensitive attributes and quasi-identifiers. Thus, swapping’s philosophy does not conform with DP’s because it only adds noise to one attribute of the data, not all of them.

This misalignment between their philosophies is a major reason to conclude that swapping is not DP. However, as we argue in Section 1, a strict adherence to this aspect of DP’s philosophy greatly limits its applicability, since it does not allow for the possibility that some information cannot be protected—either because it is already public, or because it is a mandated invariant. It is our view that a good theory of SDC should not delineate what does and does not need protection. Rather, it should provide a flexible framework that enables practitioners and policymakers—who possess the necessary contextual information—to decide what level of protection should be afforded to each data attribute. Indeed, this perspective is supported by the broader literature (He et al., 2014; Kifer & Machanavajjhala, 2014), even if it is misaligned with the default approach to DP.

Nonetheless, there is considerable alignment between the ideas behind DP and swapping. For one, they both agree on the tenet that SDC protection arises from epistemic uncertainty (see Section 6.3)—i.e. that sensitive attributes should be protected using randomization. This alignment was ultimately what allowed us to prove that the PSA satisfies a DP specification, placing it squarely within the framework of DP and answering the titular question.

## Acknowledgments

We thank Daniel Susser for his helpful and stimulating feedback on early drafts of this work, and Priyanka Nanayakkara for her proofreading of later drafts and valuable suggestions, particularly with regard to Figure 1. We are grateful to the participants of the National Bureau of Economic Research’s conference *Data Privacy Protection and the Conduct of Applied Research: Methods, Approaches and Their Consequences* (May 4-5, 2023); the 36th New England Statistical Symposium’s invited session *A Private Refreshment on Statistical Principles and Senses* (June 6, 2023); the 2023 Joint Statistical Meetings session *Methodological Approaches to Privacy Concerns Across Multiple Domains* (August 10, 2023); the CA Census retreat at the Boston University Center for Computing and Data Science (September 26, 2023); and the Statistics Canada Methodology Seminar (October 31, 2023) for their thoughtful comments and questions. We appreciate

---

<sup>7</sup>Although there are exceptions to this rule. For example, the dataset size in bounded DP is not considered privacy-revealing.

enormously the detailed feedback provided by Daniel Kifer, John Abowd, Philip Leclerc, Ryan Cummings, Rolando Rodriguez, Robert Ashmead, Sallie Keller and Michael Hawes at the US Census Bureau, which greatly improved and corrected all three parts of this trio of papers. All remaining errors are purely our own. JB gratefully acknowledges partial financial support from the Australian-American Fulbright Commission and the Kinghorn Foundation; RG and XLM acknowledge partial financial support from the NSF; and XLM acknowledges partial financial support from Harvard University’s Office of the Vice Provost for Research.

## REFERENCES

- Abowd, J., Ashmead, R., Cumings-Menon, R., Garfinkel, S., Heineck, M., Heiss, C., Johns, R., Kifer, D., Leclerc, P., Machanavajjhala, A., Moran, B., Sexton, W., Spence, M., & Zhuravlev, P. (2022). The 2020 Census disclosure avoidance system TopDown Algorithm. *Harvard Data Science Review*, (Special Issue 2). <https://doi.org/10.1162/99608f92.529e3cb9>
- Abowd, J. M. (2018). The U.S. Census Bureau adopts differential privacy. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*, 2867–2867. <https://doi.org/10.1145/3219819.3226070>
- Abowd, J. M., Adams, T., Ashmead, R., Darais, D., Dey, S., Garfinkel, S. L., Goldschlag, N., Kifer, D., Leclerc, P., Lew, E., Moore, S., Rodríguez, R. A., Tadros, R. N., & Vilhuber, L. (2023). *The 2010 Census confidentiality protections failed, here’s how and why* (Working Paper No. 31995). National Bureau of Economic Research. <https://doi.org/10.3386/w31995>
- Abowd, J. M., & Hawes, M. B. (2023). Confidentiality protection in the 2020 US Census of Population and Housing. *Annual Review of Statistics and Its Application*, 10(1), 119–144. <https://doi.org/10.1146/annurev-statistics-010422-034226>
- Abowd, J. M., & Schmutte, I. M. (2015). Economic analysis and statistical disclosure limitation. *Brookings Papers on Economic Activity*, 2015(Spring), 221–267. <https://doi.org/10.1353/eca.2016.0004>
- Ashmead, R., Kifer, D., Leclerc, P., Machanavajjhala, A., & Sexton, W. (2019). *Effective privacy after adjusting for invariants with applications to the 2020 Census* (tech. rep.). [https://github.com/uscensusbureau/census2020-das-e2e/blob/master/doc/20190711\\_0941\\_Effective\\_Privacy\\_after\\_Adjusting\\_for\\_Constraints\\_With\\_applications\\_to\\_the\\_2020\\_Census.pdf](https://github.com/uscensusbureau/census2020-das-e2e/blob/master/doc/20190711_0941_Effective_Privacy_after_Adjusting_for_Constraints_With_applications_to_the_2020_Census.pdf).
- Augustin, T., Coolen, F. P. A., de Cooman, G., & Troffaes, M. C. M. (Eds.). (2014). *Introduction to imprecise probabilities*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118763117>
- Bailie, J., & Chien, C.-H. (2019). ABS perturbation methodology through the lens of differential privacy. *Work Session on Statistical Data Confidentiality, UN Economic Commission for Europe*, 13.
- Bailie, J., & Drechsler, J. (2024). Whose data is it anyway? Towards a formal treatment of differential privacy for surveys. *Data Privacy Protection and the Conduct of Applied Research: Methods, Approaches and Their Consequences*, 33. [https://conference.nber.org/conf\\_papers/f194306.pdf](https://conference.nber.org/conf_papers/f194306.pdf)
- Bailie, J., & Gong, R. (2023). The Five Safes as a privacy context. *The 5th Annual Symposium on Applications of Contextual Integrity*.
- Bailie, J., Gong, R., & Meng, X.-L. (2025a). A refreshment stirred, not shaken (I): Five building blocks of differential privacy. *Submitted to the Journal of the Royal Statistical Society, Series B*.
- Bailie, J., Gong, R., & Meng, X.-L. (2025b). A refreshment stirred, not shaken (II): Invariant-preserving deployments of differential privacy for the US Decennial Census. *Submitted to the Harvard Data Science Review*. <https://arxiv.org/abs/2501.08449>
- Barak, B., Chaudhuri, K., Dwork, C., Kale, S., McSherry, F., & Talwar, K. (2007). Privacy, accuracy, and consistency too: A holistic solution to contingency table release. *Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems – PODS '07*, 273. <https://doi.org/10.1145/1265530.1265569>

- boyd, d., & Sarathy, J. (2022). Differential perspectives: Epistemic disconnects surrounding the U.S. Census Bureau’s use of differential privacy. *Harvard Data Science Review*, (Special Issue 2). <https://doi.org/10.1162/99608f92.66882f0e>
- Bun, M., Desfontaines, D., Dwork, C., Naor, M., Nissim, K., Roth, A., Smith, A., Steinke, T., Ullman, J., & Vadhan, S. (2021). Statistical inference is not a privacy violation. <https://differentialprivacy.org/inference-is-not-a-privacy-violation/>.
- Bun, M., & Steinke, T. (2016). Concentrated differential privacy: Simplifications, extensions, and lower bounds. In M. Hirt & A. Smith (Eds.), *Theory of cryptography* (pp. 635–658). Springer. [https://doi.org/10.1007/978-3-662-53641-4\\_24](https://doi.org/10.1007/978-3-662-53641-4_24)
- Chien, C.-H., & Sadeghi, P. (2024). On the connection between the ABS perturbation methodology and differential privacy. *Journal of Privacy and Confidentiality*, 14(2). <https://doi.org/10.29012/jpc.859>
- Chipperfield, J., Gow, D., & Loong, B. (2016). The Australian Bureau of Statistics and releasing frequency tables via a remote server. *Statistical Journal of the IAOS*, 32(1), 53–64. <https://doi.org/10.3233/SJI-160969>
- Dalenius, T., & Reiss, S. P. (1982). Data-swapping: A technique for disclosure control. *Journal of Statistical Planning and Inference*, 6(1), 73–85. [https://doi.org/10.1016/0378-3758\(82\)90058-1](https://doi.org/10.1016/0378-3758(82)90058-1)
- de Vries, M., Golmajer, M., Tent, R., Giessing, S., & de Wolf, P.-P. (2023). An overview of used methods to protect the European Census 2021 tables. *UNECE Conference of European Statisticians Expert Meeting on Statistical Data Confidentiality*, 16. <https://unece.org/statistics/documents/2023/08/working-documents/overview-used-methods-protect-european-census-2021>
- Desfontaines, D. (2023). A list of real-world uses of differential privacy. Blog post on *Ted Is Writing Things*. <https://desfontain.es/privacy/real-world-differential-privacy.html>.
- Desfontaines, D., & Pejó, B. (2020). SoK: Differential privacies. *Proceedings on Privacy Enhancing Technologies*, 2020(2), 288–313. <https://doi.org/10.2478/popets-2020-0028>
- Dharangutte, P., Gao, J., Gong, R., & Yu, F.-Y. (2023). Integer subspace differential privacy. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-23)*.
- Dinur, I., & Nissim, K. (2003). Revealing information while preserving privacy. *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 202–210. <https://doi.org/10.1145/773153.773173>
- Drechsler, J. (2023). Differential privacy for government agencies—Are we there yet? *Journal of the American Statistical Association*, 118(541), 761–773. <https://doi.org/10.1080/01621459.2022.2161385>
- Drechsler, J., & Bailie, J. (2024). The complexities of differential privacy for survey data. <https://doi.org/10.48550/arXiv.2408.07006>
- Drechsler, J., & Reiter, J. P. (2010). Sampling with synthesis: A new approach for releasing public use census microdata. *Journal of the American Statistical Association*, 105(492), 1347–1357. <https://doi.org/10.1198/jasa.2010.ap09480>
- Duncan, G. T., & Lambert, D. (1986). Disclosure-limited data dissemination. *Journal of the American Statistical Association*, 81(393), 10–18.
- Dwork, C. (2006). Differential privacy. *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Part II, 4052*, 1–12. [https://doi.org/10.1007/11787006\\_1](https://doi.org/10.1007/11787006_1)
- Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., & Naor, M. (2006a). Our data, ourselves: Privacy via distributed noise generation. In S. Vaudenay (Ed.), *Advances in cryptology - EUROCRYPT 2006* (pp. 486–503). Springer. [https://doi.org/10.1007/11761679\\_29](https://doi.org/10.1007/11761679_29)
- Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006b). Calibrating noise to sensitivity in private data analysis. In S. Halevi & T. Rabin (Eds.), *Proceedings of the Third Theory of Cryptography Conference, TCC 2006* (pp. 265–284, Vol. 3876). Springer. [https://doi.org/10.1007/11681878\\_14](https://doi.org/10.1007/11681878_14)
- Dwork, C., & Naor, M. (2010). On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. *Journal of Privacy and Confidentiality*, 2(1). <https://doi.org/10.29012/jpc.v2i1.585>

- Fienberg, S., & McIntyre, J. (2004). Data swapping: Variations on a theme by Dalenius and Reiss. *Privacy in Statistical Databases: PSD 2004 Proceedings*. [https://doi.org/10.1007/978-3-540-25955-8\\_2](https://doi.org/10.1007/978-3-540-25955-8_2)
- Francis, P. (2022). A note on the misinterpretation of the US Census re-identification attack. <https://doi.org/10.48550/arXiv.2202.04872>
- Fraser, B., & Wooton, J. (2005). A proposed method for confidentialising tabular output to protect against differencing. *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*.
- Gaboardi, M., Hay, M., & Vadhan, S. (2024). Programming frameworks for differential privacy. <https://doi.org/10.48550/arXiv.2403.11088>
- Gao, J., Gong, R., & Yu, F.-Y. (2022). Subspace differential privacy. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4), 3986–3995. <https://doi.org/10.1609/aaai.v36i4.20315>
- Glessing, S., & Schulte Nordholt, E. (2017). *Recommendations for best practices to protect the Census 2021 hypercubes*. Centre of Excellence on Statistical Disclosure Control, Eurostat. [https://cros-legacy.ec.europa.eu/content/recommendations-protection-hypercubes\\_en](https://cros-legacy.ec.europa.eu/content/recommendations-protection-hypercubes_en).
- Gong, R. (2022). Transparent privacy is principled privacy. *Harvard Data Science Review*, (Special Issue 2). <https://doi.org/10.1162/99608f92.b5d3faaa>
- Gong, R., Groshen, E. L., & Vadhan, S. (Eds.). (2022). Differential privacy for the 2020 US Census: Can we make data both private and useful? *Harvard Data Science Review*, (Special Issue 2). <https://hdr.mitpress.mit.edu/specialissue2>
- Haney, S., Machanavajjhala, A., Abowd, J. M., Graham, M., Kutzbach, M., & Vilhuber, L. (2017). Utility cost of formal privacy for releasing national employer-employee statistics. *Proceedings of the 2017 ACM International Conference on Management of Data – SIGMOD ’17*, 1339–1354. <https://doi.org/10.1145/3035918.3035940>
- Hawes, M., & Rodríguez, R. (2021). Determining the Privacy-loss Budget: Research into Alternatives to Differential Privacy. Presentation to the Census Scientific Advisory Committee. <https://www2.census.gov/about/partners/cac/sac/meetings/2021-05/presentation-research-on-alternatives-to-differential-privacy.pdf>.
- Hay, M., Rastogi, V., Miklau, G., & Suci, D. (2010). Boosting the accuracy of differentially private histograms through consistency. *Proceedings of the VLDB Endowment*, 3(1).
- He, X., Machanavajjhala, A., & Ding, B. (2014). Blowfish privacy: Tuning privacy-utility trade-offs using policies. *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data – SIGMOD ’14*, 1447–1458. <https://doi.org/10.1145/2588555.2588581>
- Hod, S., & Canetti, R. (2025). Differentially private release of Israel’s National Registry of Live Births. *Proceedings of the 2025 IEEE Symposium on Security and Privacy (SP)*. <https://doi.org/10.1109/SP61157.2025.00101>
- Hotz, V. J., & Salvo, J. (2020). *Assessing the use of differential privacy for the 2020 Census: Summary of what we learned from the CNSTAT workshop*. National Academies Committee on National Statistics. [https://www.amstat.org/asa/files/pdfs/POL-CNSTAT\\_CensusDP\\_WorkshopLessonsLearnedSummary.pdf](https://www.amstat.org/asa/files/pdfs/POL-CNSTAT_CensusDP_WorkshopLessonsLearnedSummary.pdf)
- Hotz, V. J., & Salvo, J. (2022). A chronicle of the application of differential privacy to the 2020 Census. *Harvard Data Science Review*, (Special Issue 2). <https://doi.org/10.1162/99608f92.ff891fe5>
- Hu, Y., Sanyal, A., & Schölkopf, B. (2024). Provable privacy with non-private pre-processing. *Proceedings of the 41st International Conference on Machine Learning*, 235, 19402–19437. <https://doi.org/10.5555/3692070.3692852>
- Kenny, C. T., Kuriwaki, S., McCartan, C., Rosenman, E. T., Simko, T., & Imai, K. (2021). The use of differential privacy for census data and its impact on redistricting: The case of the 2020 U.S. Census. *Science Advances*, 7(41), eabk3283. <https://doi.org/10.1126/sciadv.abk3283>
- Kenny, C. T., McCartan, C., Kuriwaki, S., Simko, T., & Imai, K. (2024). Evaluating bias and noise induced by the U.S. Census Bureau’s privacy protection methods. *Science Advances*, 10(18), ead12524. <https://doi.org/10.1126/sciadv.ad12524>

- Kenthapadi, K., & Tran, T. T. L. (2018). PriPeARL: A framework for privacy-preserving analytics and reporting at LinkedIn. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2183–2191. <https://doi.org/10.1145/3269206.3272031>
- Kifer, D. (2019). Design principles of the TopDown algorithm. Presentation at JASON, La Jolla, CA.
- Kifer, D., & Machanavajjhala, A. (2011). No free lunch in data privacy. *Proceedings of the 2011 International Conference on Management of Data – SIGMOD '11*, 193–204. <https://doi.org/10.1145/1989323.1989345>
- Kifer, D., & Machanavajjhala, A. (2014). Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems*, 39(1), 1–36. <https://doi.org/10.1145/2514689>
- Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, 31(3), 471–481. <https://doi.org/10.3233/SJI-150906>
- Leonelli, S. (2019). Data governance is key to interpretation: Reconceptualizing data in data science. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.17405bb6>
- Machanavajjhala, A. (2022). Candidate differential privacy algorithms for 2020 Decennial Census Group II products. Workshop on the Analysis of Census Noisy Measurement Files and Differential Privacy. <http://dimacs.rutgers.edu/events/details?eID=2038>.
- Marley, J. K., & Leaver, V. L. (2011). A method for confidentialising user-defined tables: Statistical properties and a risk-utility analysis. *Proceedings of the 58th World Statistical Congress*, 1072–1081.
- McClure, D., & Reiter, J. P. (2012). Differential privacy and statistical disclosure risk measures: An investigation with binary synthetic data. *Transactions on Data Privacy*, 5(3), 535–552.
- McKay Bowen, C., & Garfinkel, S. (2021). The philosophy of differential privacy. *Notices of the American Mathematical Society*, 68(10), 1. <https://doi.org/10.1090/noti2363>
- McKenna, L. (2018). *Disclosure avoidance techniques used for the 1970 through 2010 Decennial Censuses of Population and Housing* (working paper). The Research and Methodology Directorate – US Census Bureau. <https://www.census.gov/content/dam/Census/library/working-papers/2018/adrm/Disclosure%20Avoidance%20for%20the%201970-2010%20Censuses.pdf>
- McKenna, L., & Haubach, M. (2019). *Legacy techniques and current research in disclosure avoidance at the U.S. Census Bureau* (working paper). Research and Methodology Directorate, United States Census Bureau. [https://www.census.gov/content/dam/Census/library/working-papers/2019/adrm/5%20Legacy%20Techniques\(tagged\)%20CED-DA%20Report%20Series.pdf](https://www.census.gov/content/dam/Census/library/working-papers/2019/adrm/5%20Legacy%20Techniques(tagged)%20CED-DA%20Report%20Series.pdf)
- Meng, X.-L. (2021). Enhancing (publications on) data quality: Deeper data minding and fuller data confession. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(4), 1161–1175. <https://doi.org/10.1111/rssa.12762>
- Messing, S., DeGregorio, C., Hillenbrand, B., King, G., Mahanti, S., Mukerjee, Z., Nayak, C., Persily, N., State, B., & Wilkins, A. (2020). *Urls-v3.pdf*. In *Facebook privacy-protected full urls data set*. Harvard Dataverse. <https://doi.org/10.7910/DVN/TDOAPG/DGSAMS>
- Mitra, R., & Reiter, J. P. (2006). Adjusting survey weights when altering identifying design variables via synthetic data. *Privacy in Statistical Databases: PSD 2006 Proceedings*, 177–188.
- National Research Council. (1995). *Modernizing the U.S. Census* (B. Edmonston & C. Schultze, Eds.). National Academies Press. <https://doi.org/10.17226/4805>
- Near, J. P., & Abuah, C. (2025). *Programming differential privacy*. <https://programming-dp.com/>
- Neunhoeffer, M., Lattner, J., & Drechsler, J. (2024). On the formal privacy guarantees of synthetic data (generated without formal privacy guarantees). *National Bureau of Economic Research (2024): Data Privacy Protection and the Conduct of Applied Research: Methods, Approaches and Their Consequences*. [https://conference.nber.org/conf\\_papers/f194303.pdf](https://conference.nber.org/conf_papers/f194303.pdf)
- Nissenbaum, H. F. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law Books.
- Oberski, D. L., & Kreuter, F. (2020). Differential privacy and social science: An urgent puzzle. *Harvard Data Science Review*, 2(1). <https://doi.org/10.1162/99608f92.63a22079>

- Office for National Statistics. (2023). Protecting personal data in Census 2021 results. ONS Website, Methodology. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/protectingpersonaldataincensus2021results>.
- Protivash, P., Durrell, J., Ding, Z., Zhang, D., & Kifer, D. (2022). Reconstruction attacks on aggressive relaxations of differential privacy. <https://doi.org/10.48550/arXiv.2209.03905>
- Raskhodnikova, S., & Smith, A. (2016). Differentially private analysis of graphs. In M.-Y. Kao (Ed.), *Encyclopedia of Algorithms* (pp. 543–547). Springer. [https://doi.org/10.1007/978-1-4939-2864-4\\_549](https://doi.org/10.1007/978-1-4939-2864-4_549)
- Reamer, A. (2019). *Brief 7: Comprehensive accounting of census-guided federal spending (FY2017)*. George Washington Institute of Public Policy. <https://gwipp.gwu.edu/sites/g/files/zaxdzs6111/files/downloads/Counting%20for%20Dollars%202020%20Brief%20A%20-%20Comprehensive%20Accounting.pdf>
- Reiter, J. P. (2005). Estimating risks of identification disclosure in microdata. *Journal of the American Statistical Association*, 100(472), 1103–1112.
- Rinott, Y., O’Keefe, C. M., Shlomo, N., & Skinner, C. (2018). Confidentiality and differential privacy in the dissemination of frequency tables. *Statistical Science*, 33(3), 358–385. <https://doi.org/10.1214/17-STS641>
- Ruggles, S., Fitch, C., Magnuson, D., & Schroeder, J. (2019). Differential privacy and census data: Implications for social and economic research. *AEA Papers and Proceedings*, 109, 403–408. <https://doi.org/10.1257/pandp.20191107>
- Sarathy, J., & boyd, d. (2024). Statistical imaginaries, state legitimacy: Grappling with the arrangements underpinning quantification in the U.S. Census. *Critical Sociology*, 08969205241270898. <https://doi.org/10.1177/08969205241270898>
- Schmutte, I. M. (2016). Differentially private publication of data on wages and job mobility. *Statistical Journal of the IAOS*, 32(1), 81–92. <https://doi.org/10.3233/SJI-160962>
- Schneider, M. J., Bailie, J., & Iacobucci, D. (2025). Why data anonymization hasn’t taken off. *Submitted to Customer Needs and Solutions*.
- Seeman, J., Sexton, W., Pujol, D., & Machanavajjhala, A. (2023). Privately answering queries on skewed data via per record differential privacy. <https://doi.org/10.48550/arXiv.2310.12827>
- Seeman, J., & Susser, D. (2024). Between privacy and utility: On differential privacy in theory and practice. *ACM Journal on Responsible Computing*, 1(1), 3:1–18. <https://doi.org/10.1145/3626494>
- Shafer, G. (1976). *A mathematical theory of evidence*. Princeton University Press, Princeton, NJ.
- Slavković, A., & Seeman, J. (2023). Statistical data privacy: A song of privacy and utility. *Annual Review of Statistics and Its Application*, 10(1), 189–218. <https://doi.org/10.1146/annurev-statistics-033121-112921>
- Smart, M. A., Sood, D., & Vaccaro, K. (2022). Understanding risks of privacy theater with differential privacy. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 342:1–24. <https://doi.org/10.1145/3555762>
- Steinke, T. (2022). Composition of differential privacy & privacy amplification by subsampling. <https://doi.org/10.48550/arXiv.2210.00597>
- Thompson, G., Broadfoot, S., & Elazar, D. (2013). Methodology for the automatic confidentialisation of statistical outputs from remote servers at the Australian Bureau of Statistics. *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality*, 38. <https://www.unece.org/stats/documents/2013.10.confidentiality.html>
- Tramèr, F., Kamath, G., & Carlini, N. (2024). Position: Considerations for differentially private learning with large-scale public pretraining. *Proceedings of the 41st International Conference on Machine Learning*, 48453–48467. <https://proceedings.mlr.press/v235/tramer24a.html>
- UK Statistics Authority. (2021). *Transparency of SDC methods and parameters (post-meeting EAP paper for SDC for Census August 2021)* (tech. rep. No. EAP168). <https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2022/02/EAP168-Statistical-Disclosure-Control-for-Census.pdf>.

- US Census Bureau. (2019). A history of census privacy protections. <https://www.census.gov/library/visualizations/2019/comm/history-privacy-protection.html>
- US Census Bureau. (2021a). *2020 Census National Redistricting Data Summary File* (tech. rep.). [https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census\\_PL94\\_171Redistricting\\_NationalTechDoc.pdf](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_NationalTechDoc.pdf).
- US Census Bureau. (2021b). *2020 Census State Redistricting Data Summary File* (tech. rep.). [https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census\\_PL94\\_171Redistricting\\_StatesTechDoc\\_English.pdf](https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/summary-file/2020Census_PL94_171Redistricting_StatesTechDoc_English.pdf).
- US Census Bureau. (2023a). *2020 Census Demographic and Housing Characteristics File (DHC)* (tech. rep.). <https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/demographic-and-housing-characteristics-file-and-demographic-profile/2020census-demographic-and-housing-characteristics-file-and-demographic-profile-techdoc.pdf>.
- US Census Bureau. (2023b). *2020 Census Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) technical documentation* (tech. rep.). <https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/detailed-demographic-and-housing-characteristics-file-a/2020census-detailed-dhc-a-techdoc.pdf>.
- US Census Bureau. (2023c). Developing the DAS: Demonstration data and progress metrics. <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-development.html>.
- US Census Bureau. (2023d). *Methodology, assumptions and inputs for the 2023 National Population Projections* (tech. rep.). <https://www2.census.gov/programs-surveys/popproj/technical-documentation/methodology/methodstatement23.pdf>.
- US Census Bureau. (2023e). *Methodology for the United States Population Estimates: Vintage 2023* (tech. rep.). <https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020-2023/methods-statement-v2023.pdf>.
- US Census Bureau. (2024a). *2020 Census Detailed Demographic and Housing Characteristics File B (Detailed DHC-B) technical documentation* (tech. rep.). <https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/detailed-demographic-and-housing-characteristics-file-b/2020census-detailed-dhc-b-techdoc.pdf>.
- US Census Bureau. (2024b). *2020 Supplemental Demographic and Housing Characteristics File (S-DHC) technical documentation* (tech. rep.). <https://www2.census.gov/programs-surveys/decennial/2020/technical-documentation/complete-tech-docs/supplemental-demographic-and-housing-characteristics-file/2020census-supplemental-dhc-techdoc.pdf>.
- Villa Ross, C. (2023). *Uses of Decennial Census programs data in federal funds distribution: Fiscal year 2021*. US Census Bureau. <https://www.census.gov/library/working-papers/2023/dec/census-data-federal-funds.html>
- Waldman, A. E. (2021). *Industry unbound: The inside story of privacy, data, and corporate power*. Cambridge University Press.
- Walley, P. (1991). *Statistical reasoning with imprecise probabilities* (1st ed). Chapman and Hall.