Government Statistical Service

Best Practice and Impact

GSS Methodology Advisory Committee

Differential privacy: an introduction for statistical agencies

Dr. Hector Page, Privitar
Charlie Cabot, Privitar
Professor Kobbi Nissim, Georgetown University

A contributing article to the National Statistician's Quality Review into Privacy and Data Confidentiality Methods

PRIVITAR

December 2018

Contents

1. Introduction[1]

Differential privacy is a formal mathematical model of privacy [1]. Informally, differential privacy requires that the output of an analysis should reveal (almost) no information specific to any individual within that dataset, where by "specific" we mean information that could not be inferred about the individual from statistical facts about other individuals. Differential privacy was invented just twelve years ago, yet it has already seen significant adoption first in academia, then by leading government agencies and technology companies. Its appeal comes from two primary characteristics: a formally defined strong privacy protection and a measure of privacy leakage across multiple statistical releases. These characteristics are especially important in today's world, where more sensitive data is created than ever, more data releases are made than ever, new privacy attacks are being discovered, and existing attacks are steadily strengthening.

One specific motivating factor for the adoption of differential privacy has been the discovery of a new, serious type of attack called a reconstruction attack. In a reconstruction attack, an attacker is able to use statistics released about a sensitive dataset to infer with high accuracy a significant portion of the dataset itself. Reconstruction attacks hence cast serious doubts on the ability of traditional statistical disclosure control (SDC) methods[2] to adequately protect systems that release aggregate statistics[3]. The US Census recently reported "*serious vulnerabilities*" to reconstruction attacks in its 2000 and 2010 Census data releases, and consequently has designed a differentially private system for the 2020 Census [4]. This demonstrates not only that these attacks, formerly considered primarily a theoretical risk, are of real concern in practice but also that differential privacy might prove to be a suitable defence.

Differential privacy is still a relatively young field of research and users are still learning how to apply it effectively in practice. Difficulties in transitioning from research to practice include limited experience in managing differential privacy's parameters and achieving a suitable level of privacy without affecting the quality of the analysis (a problem referred to as the "privacy-utility trade-off"). However, data privacy researchers continue to find improvements to differentially private algorithms, with state of the art approaches being developed for a wide variety of analyses.

Organisations that release sensitive data should assess the impact of traditional and new privacy attacks, and evaluate whether differential privacy is a suitable and beneficial new defence. Differential privacy is particularly well-suited to the use case of releasing statistics about national populations to the public, because differentially

[1] The authors would like to thank the Privitar team, especially Guy Cohen and Dr Jason McFall, for their many helpful comments and edits.
[2] For an introduction to SDC methods, see [2].
[3] Reconstruction attacks (also referred to as blatant non-privacy) were introduced in [3].

private algorithms perform best in use cases with pre-determined statistics, large sample sizes and requirements for measurable, transparent, and comprehensive privacy controls. This article aims to inform statisticians, data analysts, and relevant policy makers about the motivation behind differential privacy and its privacy guarantees, and when, why, and how to incorporate differential privacy into projects that involve releasing statistics about sensitive data.

What is differential privacy? Differential privacy is a formal privacy model with associated methods for limiting statistical disclosure and controlling privacy risk. Differential privacy is not a specific data release method, but rather a definition, or a standard, that specifies a particular requirement that data release methods may or may not satisfy. If a data release method satisfies the requirement, then it would protect an individual's information essentially as if his or her information were not used in the analysis at all [5][4]. This is a strong assurance to individuals that information that is specific to them will not be learned from their inclusion in the dataset, where "specific" refers here to information that cannot be learned about the individual by examining other individuals' information. There are many differentially private data release techniques, many of which involve releasing aggregate statistics perturbed with random noise. During the computation of the release, noise is added in such a way to provide privacy while maximising accuracy of results. Differential privacy is a characteristic of all these techniques, not a specific technique itself.

2. What motivated the development of differential privacy?

Traditional SDC methods were devised for a data ecosystem very different from that of today. Computing power is increasing, and more, richer, data is being collected and made publicly available. Simultaneously, new attack methodologies are being developed, many of which exploit weaknesses in the protection provided by traditional methods. These trends, which are likely to continue, mean that not only are new privacy attacks becoming possible, but existing attacks are becoming stronger. They highlight a need for a fresh, rigorous, look at privacy.

2.1 The world is becoming richer in data

In modern society, every individual leaves an extensive trail of potentially sensitive data. Whilst government agencies still collect traditional data about individuals, like tax and census data, on top of this there are several new sources of sensitive data. For example, mobile phones, inseparable from their users, log "*a detailed chronicle of a person's physical presence compiled every day, every moment, over several years*" [6]. Mobile phone location data is already being captured and used by Transport for

[4] "Essentially" here means that the outcome distribution (i.e. the probability of each given outcome) of the differentially private analysis would be almost the same whether the individual's information would or would not be used. That is, no outcome would change in probability by very much. The exact measure of difference is part of the formal definition of differential privacy and is presented in Section 4.

London and usage of this type of data will likely be adopted more broadly in coming years [7]. Relatedly, wearable devices such as fitness trackers are increasingly being recognised as posing privacy risks [8]. The devices we carry with us do not reveal only our locations; social media providers collect detailed usage information such as scrolls, swipes, and clicks in apps and websites, and large technology and advertising companies are building detailed pictures of our browsing and purchasing habits via the third-party web tracking ecosystem. The scale of this data collection is enormous: as of Q1 2018 there were 2.13 billion people on Facebook alone [9]. As another example, Internet of Things (IoT) data from new technologies, such as smart homes and smart meters, contains highly detailed energy usage information that can reveal lifestyle habits. Any of these data sources may one day be used as or combined with government data.

2.2 Demand for open data is increasing

As well as being collected more extensively, data is being disseminated more widely. Demand for open data in both the private and public sectors is increasing with the growing adoption of analytics and data science. Given the demand for data-driven decision making, it is no longer appropriate to ensure privacy simply by locking data away behind access controls. There is therefore an escalating tension between extracting useful insights from sensitive data and protecting individuals' privacy, putting pressure on data privacy techniques to enable privacy and utility simultaneously. Even non-sensitive open datasets have a privacy implication: they raise the level of background knowledge available to someone conducting privacy attacks on other datasets. For example, such datasets may make it possible to identify individuals in otherwise anonymised datasets[5]. This growing risk has been recognised by the UK Government Statistical Service (GSS), in their online guidance on SDC methods, where it is stated that "*Recent technological advances, along with the requirement for open data, have dramatically increased the volume of data available to companies and individuals. These advances have given the issue of disclosure control greater importance, with increasing accessibility translating into higher risk of successful identifications from published statistics*" [10]. This trend towards more, and more open, data necessitates privacy protection that is meaningful in light of expanding auxiliary knowledge.

There is widespread recognition of the potential of data-driven societal benefits, motivating the move towards open data and greater sharing of data. This desire for open data is reflected by the "open by default" approach of the G8's 2013 Open Data Charter (2013), while the UK Digital Economy Act (2017) states that barriers to data sharing cause inefficiency and limit the use of data as a valuable asset. A number of notable policy papers have echoed this sentiment, but warn too of potential privacy concerns. The official 2017 government transformation strategy is focused on better

[5] How this might happen is discussed below in Section 3: Attacks motivating differential privacy research.

use of data for the public good, yet acknowledges the need to address privacy concerns when doing so [11]. Other examples include a) Professor Dame Wendy Hall and Jérôme Pesenti's AI report [12], which notes that opening up data access is vital to improving AI for healthcare and should form a key part of the UK's AI strategy, but that privacy concerns must be adequately addressed in order to do so; and b) the 2017 life sciences industrial strategy report from Professor Sir John Bell [13], which mentions that whilst data can improve research and patient care, privacy is an important aspect of enabling data sharing within the NHS.

It is not only official data releases that increase privacy risks. Data breaches have also increased the amount of data available [14]. Bad actors may be able to use these breached datasets in the same way as official data releases.

2.3 Privacy attacks are becoming more powerful

Privacy attacks are becoming more powerful as they are backed not just by richer background knowledge from new data sources, but by greater computing power and increasingly sophisticated techniques. Below are some examples of innovative privacy attacks, of several types, that have been performed in recent years, showcasing the overall rise in sophistication:

- Example 1: De-identified NYC taxi data was re-identified by reversing the hashing of direct identifiers with a brute force attack known as a rainbow table attack. The taxi rides were then linked to a variety of publicly available data to further re-identify the dataset [15].
- Example 2: Researchers used neural networks to identify faces in images that had been pixelated or blurred in an attempt to de-identify them [16].
- Example 3: Robust matching algorithms were used to match sparse de-identified Netflix user data with IMDb user names, resulting in re-identification [17].
- Example 4: Re-identification of individuals' de-identified Twitter accounts based on metadata was performed using multinomial logistic regression, random forests, and k-nearest neighbor models [18].
- Example 5: Researchers demonstrated a reliable method to determine the presence of trace amounts (<1%) of an individual's DNA within a mixture of DNA [19]. This is an example of a membership inference attack, discussed in Section 3.4, which can lead to sensitive attribute disclosure (e.g. if a DNA mixture is of people with a given medical condition, membership in this mixture discloses that an individual has the condition).

2.4 Public concern over threats to privacy is growing

The general public is becoming more aware of privacy concerns and data misuse, meaning societal demands and expectations of data holders or processors are higher than ever before. New data privacy laws reflect these sentiments and make stricter regulatory demands. The EU's General Data Protection Regulation (GDPR) has led

the way, inspiring similar legislation worldwide, such as the California Consumer Privacy Act (CCPA), and moves towards tougher privacy regulations throughout the Asia-Pacific region [20].

Meanwhile, in the private sector, the Cambridge Analytica app permissions scandal was a heavily covered news story, resulting in Mark Zuckerberg being questioned by the US Congress and the EU parliament, as well as being formally summoned to testify before MPs at the UK parliament. News stories such as this indicate growing public awareness of privacy and security related issues[6].

Differential privacy itself is beginning to see adoption in industry, including its well-publicised use by major technology companies such as Google [21] and Apple [22].

2.5 Future-proof privacy protections are desired

Many traditional SDC methodologies were developed and evaluated in terms of threat models detailing what were considered realistic adversaries at the time of their development. Such methods risk losing effectiveness over time, either as new unforeseen attacks arise, or as the definition of a reasonable adversary becomes outdated. Assumptions about current attacker computing power and background information mean that defences are likely to be broken in the future, because both attacker power and background knowledge will increase.

Data releases can persist indefinitely, and whilst some sensitive information contained in such releases becomes less sensitive over time, other information will remain sensitive indefinitely. For example, affiliation with religious or activist groups, psychiatric treatments, chronic conditions, or personal relationships can remain sensitive for many years. Hence, it is not satisfactory to know that sensitive data is protected only in the present time of a release. Even assuming what is sensitive and what is not can be dangerous: some information that currently seems innocuous may become sensitive, either as new attacks emerge or in conjunction with other pieces of information. For instance, smart meter data may once have been thought safe, but it has now been shown to reveal household occupants' daily habits, whether specific medical devices are in use in the home, and even what TV channel is being displayed [23, 24].

There is a comparison to be drawn between future-proof data privacy and future-proof cryptography. In cryptography, security parameters such as key length are set with the intention of keeping encryption schemes secure not just for the present, but for many

[6] Public conceptions of privacy and security often conflate these two closely related concepts. We use "security" to refer to restricting access to sensitive, or potentially sensitive, data. Relatedly, we use "privacy" to refer to allowing selective or specific access to sensitive data (e.g. for analytics), but making sure this access (e.g. aggregate statistics returned from analyses run on the dataset) does not reveal individuals' sensitive data.

years into the future. For example, NIST recommends considering the expected lifetime of data when considering an encryption key size, and makes different recommendations if the lifetime extends before or after 2030 [25].

3. Attacks motivating differential privacy research

The previous section demonstrated that today's world is rich with data and computing resources which, as well as providing many benefits, create privacy risk. This section outlines some specific classes of attacks against which modern privacy techniques should defend.

3.1 Linkage attacks

Masking of direct identifiers in row-level data (also known as microdata or record-level data) with the aim of de-identification has proven to be vulnerable to re-identification through the use of auxiliary data in an attack known as a linkage attack. In a linkage attack, quasi-identifiers such as age, gender, and post code are used in combination to determine the identity of a de-identified record, by linking to another dataset containing the same fields. This type of attack is particularly serious given the wealth of rich auxiliary data easily available to adversaries today. In the US, a President's Council of Advisors on Science and Technology (PCAST) 2014 report states: "*Anonymization is increasingly easily defeated…as the size and diversity of available data grows, the likelihood of being able to re-identify individuals … grows substantially*" [26]. This assertion is backed up by multiple examples of row-level data, deemed to be safely 'anonymised', later being found vulnerable to linkage attack.

- Example 1: Latanya Sweeney demonstrated that medical data stripped of direct identifiers could be re-identified via linkage with publicly available voter registration data. The author suggested *k*-anonymity as a model of row-level privacy [27][7].
- Example 2: The Netflix prize dataset revealed to be vulnerable to linkage attacks based on background knowledge from IMDb, with the authors asserting that anonymisation methods such as *k*-anonymity will fail for sparse, high-dimensional datasets [17].
- Example 3: Australian health records were re-identified via linkage, leading the authors to suggest that it is impossible to make a release of record-level data that is private without affecting accuracy of results so severely as to render the release useless [29]. They suggested releasing aggregate statistics with

[7] *k*-anonymity requires that the identifying (or quasi-identifying) information for each person in a release should be identical to the respective information of at least *k*-1 other persons in the release. While *k*-anonymity may protect against some re-identification attacks, data that has been *k*-anonymised may remain vulnerable to a host of other attacks, including composition attacks [28]. Additionally, k-anonymity can have an unacceptable effect on data utility. This is especially true if datasets are high-dimensional.

differential privacy as an alternative, because it balances reductions in data accuracy with privacy gain.

3.2 Differencing attacks and other composition attacks

Contemporary attempts at data privacy have shifted away from releasing de-identified microdata (individual record level data) to instead releasing aggregate statistics or permitting aggregate queries of datasets [30]. However, aggregate statistics can still pose privacy risks and still require privacy techniques to avoid being disclosive.

The simplest form of attack on aggregate statistics is a differencing attack. A very simple differencing attack involves finding two groups that differ by one individual and using difference in group values to infer an individual's value. For example, the difference in total salary between two groups, where those groups differ by one individual, will reveal that individual's salary. More complex variants exist, using not just the difference between two groups, but the differences of multiple groups and differences of differences. In practice, attackers able to scan through large numbers of statistics can construct attacks based on complex sets of differences.

Even when applying suppression of small-count statistics (e.g. statistics about 1-5 people), it has been shown via Dorothy Denning's work on tracker attacks in 1978 that singling individuals out from aggregate statistics is still possible [31]. Background information/auxiliary information, which yields linkage attacks on row-level data, remains an issue for aggregate statistics as it can facilitate the construction of tracker attacks.

There are multiple real-world examples of Web-based query systems releasing only aggregate statistics being broken by differencing attacks, such as:

- Example 1: Matthews et al. demonstrated the prevalence of differencing attack vulnerabilities in practice [32]. They reviewed multiple US state-level web-based data query systems allowing interactive query of public health data in the US to give flexible tabular outputs. Despite being designed with SDC methods in place (e.g. small count suppression), nine systems were found vulnerable to differencing attacks. Notably, the authors report that all systems were tested by them for a maximum of 30 minutes, demonstrating that sophisticated techniques were not required.
- Example 2: There have been multiple attacks on Facebook users using microtargeted advertising, despite Facebook's internal SDC methodologies, which can involve differencing as part of their methodology [33, 34, 35].

Output perturbation via zero-mean random noise is one method that has been used to mitigate differencing attacks. Such noise is randomly drawn each time an aggregate statistic is released, giving a different noisy answer for the same statistic. However,

this approach is vulnerable to averaging attacks, whereby noise can be averaged out across multiple statistics.

A possible means to defend against averaging attacks is to add noise which is deterministically calculated from the individuals of a given query set (e.g. which individuals are counted in a counting query). This approach is exemplified by the cell-key perturbation method [36]. This yields the same noisy answer whenever the same set of individuals are used in a query, thereby avoiding multiple noisy answers to the same query being averaged out. However, deterministic noise addition methods such as the cell-key perturbation method are vulnerable to subtle differencing attacks such as fishing attacks [37]. Adding in a second layer of noise deterministically based on the syntax of queries[8], rather than the identities of the individuals contributing to the query, might prevent fishing attacks, but this too may be vulnerable to subtler differencing attacks or other new classes of attack [37]. Indeed, there is no firm proof that subtler differencing attacks won't or can't be found. The cycle of new defences followed by new workarounds, which started with Denning's tracker attacks as a workaround to small count suppression, is still continuing.

The weakness of any variety of random noise addition is that, regardless of the noise added, if too many statistics are released, the true values of the underlying dataset can be accurately recovered via reconstruction attacks, which are discussed below. Therefore, any variety of noise addition can only avoid the risk of averaging or reconstruction attacks provided there is some limit on the number of statistics that are released about a given dataset.

3.3 Reconstruction attacks (aka blatant non-privacy)

When released summary statistics contain enough information about the underlying dataset, an adversary may be able to reconstruct, either exactly or with very high accuracy, the entire dataset from these summaries. This is known as a *reconstruction attack*. At the information-theoretic level, even if information is only released in the form of aggregate statistics, the more information that is released, the more determined the underlying data is. In other words, the set of possible datasets consistent with the released information is reduced with each release to the point that it is eventually fully determined. For statistics which can be expressed as a linear combination of the underlying data, like sums, counts, and averages, there is a general purpose reconstruction attack strategy which is to express the statistics as linear equations (or linear inequalities) about the individual sensitive variables, and solve the resulting system of linear constraints.

[8] For example, counting the number of individuals of each gender in each department of an organisation could be expressed in the SQL query language as "*SELECT COUNT(*) FROM dataset GROUPBY gender, department*". This statement could be edited to return the same set of people with different syntax by adding the (irrelevant) condition "*WHERE age<1000*".

Even perturbed statistics may be vulnerable to reconstruction attacks. For example, Dinur and Nissim showed in 2003 that releasing too many randomly selected statistics with high accuracy will allow reconstruction with extremely high probability [3] . This result was further strengthened and generalised in follow up work, including Dwork, McSherry and Talwar (2007) [38], Dwork and Yekhanin (2008) [39], Choromanski and Malkin (2012) [40], and Kantor and Nissim (2013) [41]. The result is informally referred to as the Fundamental Law of Information Recovery [42].

There are examples of reconstruction attacks on more than just sum queries. Reconstruction is possible, in practice, based on *any release of statistics which are linear functions of the underlying data or can be expressed as linear*. This encompasses more complex statistics such as M-estimators and decision tree/classifier error rates [43].

As with differencing attacks, background knowledge can aid reconstruction attacks, as it may result in some private information already being known to the adversary, or otherwise restrict the set of datasets consistent with the release. Reconstruction attack theorems usually make no assumptions on the type of noise which is introduced in the computation of the statistics (e.g. whether it is biased or not, or how this noise is distributed), only on the magnitude of this noise. This means that, regardless of what type of noise has been added, as long as the statistics are accurate enough, reconstruction follows[9].

The US Census Bureau's internal experiments have identified reconstruction attacks to be a high enough risk to incorporate differential privacy into the SDC methodology of the 2020 Census. Specifically, they state that "*reconstruction is a serious disclosure threat that all statistical tabulation systems from confidential data must acknowledge*" [44]. They further state that the SDC methodology employed for the 2010 Census is not suited to defend against reconstruction attacks, which are the "*death knell*" for traditional data publication systems from confidential sources [45].

3.4 Membership inference attacks

Other attacks on aggregate statistics focus on merely determining whether or not someone is in a dataset, which can itself be sensitive. These are known as membership inference attacks. Several instances have been shown in practice: a membership inference attack on the results of Genome-Wide Association Studies (GWAS) has been demonstrated [19], and further extensions have demonstrated that this type of membership inference works even when the statistics are noisy [46]. Other work demonstrated membership inference on aggregate location data. The effectiveness of this particular attack is shown to be increased with background

[9] An important consequence of a reconstruction attack not making assumptions on the type of noise added is their generalit, see [3].

knowledge [47]. Membership inference attacks have also been performed on machine learning models, whereby an adversary determines whether a given individual's data was used to train a model [48].

3.5 Why traditional SDC falls short in defending against these attacks

Privacy can be hard, or even impossible, to assess empirically: the fact that a system designer or tester did not find an attack does not mean that an attack does not exist. It is difficult to detect and prevent complex differencing attacks. This is particularly true given that there are currently a variety of types of differencing attacks, and it is not clear whether a future attack type might be conceived that creates unforeseen privacy risk. Such attacks may use independent, uncoordinated releases, which are hard to anticipate. Attacks that use external information are hard to design specific defences against, as the system designer cannot anticipate the various information sources that would be available to a future attacker.

The disclosure risk of multiple uncoordinated releases is hard to monitor, and traditional SDC does not provide a calculus for measuring accumulated privacy risks across multiple analyses. The existence of reconstruction attacks suggests that the output of any algorithm run on a database, even if the algorithm includes noise perturbation, will leak some information about the database itself. As more data is released, there is an increasing risk that the accumulated information implicates the privacy of one or more individuals. Without tracking this accumulated risk, it is unclear whether a release is safe or not. These issues can be aggravated when constructing more complex analyses or algorithms, such as machine learning models, or when making many different types of releases.

Approaches that focus on defending against currently-identified threats are inherently poorly suited to defence against unidentified threats or availability of new auxiliary knowledge in the future. Auxiliary knowledge will often make attacks on aggregate statistics easier, just as it strengthens attacks on row-level data. Any methodology that relies on a specific attacker profile (in terms of current availability of auxiliary knowledge and current attacks) to assess and mitigate disclosure risk is in danger of not being sufficiently future-proof.

Some traditional SDC systems rely on keeping parts of their inner workings secret. For instance, the 2010 US Census relied on record swapping, and kept the swap rate secret [44]. However, usage of secret parameters or processes introduces a point of failure if privacy depends on it. Additionally, a lack of fully public processes can make it hard for those using data output to correct for errors or reason about data utility.

To defend comprehensively against reconstruction attacks, one must limit information release in a quantified and systematic way that evaluates cumulative risk. No statistical disclosure control method can avoid the reality that too many statistics released with

too high accuracy may result in a privacy breach. What can be done is to measure and evaluate the cumulative risk of outputting the results of statistical analyses on private data. This is the approach taken by differential privacy. Treating privacy as a consumable resource, this measurement of privacy risk can then be used to guide choices about the number and type of statistics to release and the accuracy of these statistics.

3.6 Summary

The volume of available data is increasing inexorably. Robust and forward-looking technology is needed to defend against adversaries who make use of this data in combination with sophisticated attack methodologies and increasing computing power. Differencing and reconstruction attacks pose a real threat. We therefore need a way to reason about privacy in a manner that:

- Defends against future, unforeseen, attacks.
- Provides meaningful privacy protection in presence of arbitrary auxiliary information.
- Is transparent, interpretable, and auditable.
- Cannot be rendered ineffective at a future date via further processing.
- Allows a formal means by which the increasing risk of reconstruction attacks can be bounded when running an analysis or set of analyses on the same data.

Differential privacy was conceived of as a means by which to help address these needs.

4. A technical description of differential privacy

This section defines differential privacy (with several interpretations, including in both everyday and mathematical terms), describes mechanisms which achieve differential privacy, and tasks for which differentially private algorithms have been developed. To help delineate which use cases are a good fit for differential privacy, the section also details the sorts of problems differential privacy is *not* intended to solve.

4.1 Background: Previous privacy models and differential privacy

The search for a good model of privacy has a long history, and the SDC literature includes several such models presented in varying levels of accuracy and rigor, the most notable being notions of de-identification and *k*-anonymity[10] and Dalenius' definition[11] of disclosure [52]. Differential privacy, presented in 2006 by Dwork, McSherry, Nissim, and Smith, is the latest of several attempts, and has emerged, in

[10] These approaches equate privacy protection with preventing an individual's information from being linked with the individual. Variants of the approach include *k*-anonymity, l-diversity [49] and t-closeness [50].

[11] Dalenius defines disclosure as follows: "*If the release of the statistics S makes it possible to determine the value D_k more accurately than is possible without access to S, a disclosure has taken place*" (D_k refers to the personal information of subject k). See [51].

part, from understanding deficiencies in these models [1, 53]. Differential privacy was designed to be a robust, yet achievable, privacy definition. Leading to this definition of privacy were a series of works with the goal of putting privacy research on solid mathematical grounds [3, 54, 55, 56].

4.2 Definition of differential privacy

Differential privacy is a privacy model that bounds privacy risk. It is not itself a method, but rather a property that an analysis (or algorithm) may have. Furthermore, differential privacy is not a property merely of a particular output of an analysis, but of the information relationship the analysis creates between its input and its output. This is in contrast to privacy approaches such as *k*-anonymity that treat privacy as a syntactic property of the output[12]. Rather than a syntactic restriction on the output of an analysis, differential privacy is a semantic property of the relationship between analysis inputs (e.g. the dataset) and the output distribution, restricting what an observer of the output of an analysis may learn about the contribution of any single individual. This restriction can be shown, in turn, to relate to a well-defined notion of privacy risk. Furthermore, the differential privacy guarantee is a worst-case guarantee as it does not depend on the specific inputs or outputs, nor does it depend on the specifics of an attempted attack. Differential privacy has a parameter epsilon (ε) which represents the strength of the privacy guarantee. The parameter ε is a non-negative numerical value, and can be used as a means of quantifying privacy risks. The differential privacy framework provides tools for reasoning about how ε changes as a result of multiple uses of differentially private analyses.

4.2.1 An informal presentation of the definition

Differential privacy formalises the following intuitive view of privacy: the output of an analysis should reveal (almost) no information specific to any individual within that dataset. By “specific” we mean information that could not be inferred about the individual from statistical facts about other individuals. The formalisation requires that the outcome distribution of the analysis should remain almost identical whether an individual’s information is provided to the analysis or omitted from its input (or alternatively, whether an individual's true information is provided to the analysis or is replaced by arbitrary values). From the point of view of an individual within the dataset, differential privacy guarantees that an analysis will reveal (almost) no information that could not be learned about the individual if their information was not used as part of the input to the analysis.

[12] By syntactic we mean a property of the output’s format (e.g., that every combination of quasi identifiers should appear at least *k* times). A syntactic property does not have direct implications on the informational content of the output.

To illustrate, consider two scenarios. In the first, an analysis is run on a dataset with Bob's data included (data present), and in the second, the analysis is run on a hypothetical dataset which differs from the real one only in that Bob's data is not included (data absent)[13]. Differential privacy states that the outcome distribution of the analysis will be approximately the same in each scenario, and therefore very little information specific to Bob can be learned from the outcome of the analysis. This guarantee that the analysis' outcome should be (almost) independent of Bob's data applies not only to Bob, but to each individual in the dataset.

Ideally for privacy, the outcome distribution of the analysis would be identical in the data present and data absent scenarios, as this would mean that the outcome of the analysis is independent of any individual's input. However, such a strict requirement would rule out producing any utility from the data[14]. This is why differential privacy allows for some discrepancy between the output distributions in the data present and data absent scenarios, and the differential privacy guarantee is that the two scenarios are *almost* indistinguishable. Any observer considering the data present and data absent scenarios cannot, given a sample of the outcome of the analysis, distinguish with too high certainty which scenario generated this outcome. This allowed discrepancy is quantified by a parameter usually denoted by ε. To see why having identical output distributions destroys utility, note that if the output distributions were identical with and without Bob's data, then Bob's data must be irrelevant to the analysis and can be deleted without consequence. By similar argument, all other persons' data could be deleted without consequence. And so clearly, if all the data can be deleted from a data analysis without consequence, the data analysis cannot be meaningful [5].

The differential privacy guarantee is achieved by means of adding uncertainty into the outcome of an analysis, often in the form of noise perturbation. The perturbation magnitude affects the level of privacy protection, in terms of how similar are the outcome distributions in the presence and absence of an individual's data. It also affects the accuracy of the computed analysis. As a rule of thumb, achieving a higher level of privacy (i.e. more similar outcomes in the presence or absence of any single individual's data) requires adding a higher level of noise and hence results in a lesser accuracy. Although differential privacy requires adding uncertainty, note that there are usually other sources of uncertainty in an analysis already, such as sampling error and collection errors. In many use cases, the uncertainty needed for differential privacy is less than these other types of uncertainty.

[13] Note that these "data present" and "data absent" scenarios are hypothetical: differential privacy is unrelated to actually giving data participants opt-in/opt-out controls.

[14] As discussed below, if an analysis outcome is independent of any individual's input then the analysis must ignore its entire input and hence no utility is produced regardless of how utility may be defined in specific settings. The issue of ensuring acceptable utility is discussed in Sections 5.3.2 and 6.3.3.

The differential privacy guarantee of near indistinguishability between the data present and data absent scenarios is preserved even with arbitrary background knowledge [57]. In particular, even if every other individual's information is revealed to the adversary, the outcome of a differentially private analysis would still not reveal whether Bob's information was included. This means that the only information that can be learned about Bob is information that can be revealed about him from other individuals' information. Information that is specific to him would not be revealed. Consider, for example, the case where the number of people with a particular medical condition is released. An adversary is unsure only of whether Bob has the condition, but she knows that there are 42 people other than Bob with the condition. If the released count under differential privacy is 43, she cannot be sure whether this increase is due to Bob having the condition or due to the noise perturbation added, as the outcome 43 is almost as likely whether he has the condition or not. She might be able to work out that Bob is likely to have the condition based on who else does (e.g. if there emerges a correlation between certain traits and the condition), but then this knowledge is not due to Bob's inclusion in the release.

Furthermore, the differential privacy guarantee continues to hold even if Bob participated in several analyses satisfying the differential privacy guarantee. This important property of differential privacy, *composition*, is discussed in more detail below.

4.2.2 The mathematical formulation

Let X be an arbitrary data domain, and let T be an arbitrary output domain. A randomised algorithm mapping a dataset to output $A : X^n \to T$ satisfies ε-differential privacy if for all dataset pairs $x, y \in X^n$ where x and y differ by a single entry, and for all subsets S of the output domain T:

$$\Pr[A(x) \in S] \leq e^{\epsilon} \Pr[A(y) \in S],$$

where the probability is over the randomness of the algorithm A [1].

In this sense, ε>0 measures privacy risk as the information leakage of an analysis, as it sets a bound on the probability change across all possible outcomes of an analysis. At most, the probability of any one outcome can change by a factor of e^{ϵ}.

From the point of view of an individual in the dataset, one of the two datasets (x and y) can represent the data as supplied to an analysis (the individual's data present scenario) and the other can represent the data in a thought experiment where the individual's information is omitted or arbitrarily changed (the individual's data absent scenario). The parameter ε therefore bounds how much the individual's information may affect the outcome of the analysis. In turn, this is a proxy for how much their individual privacy might be impacted by this analysis. This view neatly mirrors the

perspective an individual might take if they were deciding whether or not to contribute their data to a particular analysis, or whether or not to contribute their true information. The differential privacy guarantee is a worst-case guarantee in the sense that it holds for every possible dataset and holds no matter what background information the attacker may have.

From the point of view of the entity running the algorithm and potentially responsible for protecting the privacy of the contributing individuals, ε bounds the maximum amount by which the probability of an outcome could change if any single individual's data were not included[15]. This "worst-case" guarantee is taken across all individuals and possible datasets and ensures that everyone is protected.

A larger value for ε means less privacy, whilst smaller ε means more privacy. In particular, setting ε=0 we get e^{ϵ}=1 and hence such setting corresponds to requiring that no individual's input should have any effect on the outcome, in which case we get perfect privacy but (as discussed above) no utility. With increasing values of ε, the bounds on the probability distributions, obtained by applying the analysis A on the neighboring databases x and y, become more relaxed. This corresponds to allowing specific information about individuals to be leaked with less uncertainty. This relaxation also allows for extracting more utility from the underlying data. At ε=infinity, the differential privacy requirement is void - it is satisfied by any analysis, privacy preserving or not. The question of choosing a value for ε is hence a question of trading off utility and privacy - while a small value for ε such as 0.01 or 0.1 is desirable, setting ε to be this low can sometimes result in loss of utility, especially with smaller dataset sizes. Some real world implementations have reported using ε to be in the range of 1-10. The construction of differentially private algorithms for a variety of tasks is subject to significant research effort. The principal aim of such research is improve the privacy-utility trade-off of differentially private algorithms so as to allow use of lower values of ε while maintaining an acceptable accuracy of the analysis.

4.3 Fundamental properties: post-processing, composition, group privacy

Differential privacy has several appealing properties not found in older conceptions of privacy. These properties facilitate both reasoning about privacy and the construction of sophisticated differentially private algorithms from simpler building blocks.

Consider a transformation consisting of first applying an ε-differentially private analysis A on the input and then applying an arbitrary transformation B on the resulting outcome. The resulting transformation which we will denote $B \circ A$ is also ε-

[15] Note that this refers to a change in probability distributions of the output. The algorithms are randomised, so they may output a different result each time.

differentially private. This property is resilience to post-processing[16]. It implies that the output of a differentially private analysis can be reused for a variety of purposes without creating additional privacy risk[17]. We emphasise that these guarantees hold even when the details of the analysis A, the transformation B, and the value of ε are made public[18]. We also note that while the application of the transformation B on the outcome of A cannot affect privacy adversely, certain transformations can sometimes result in *improved* privacy (i.e. a lower ε).

As a general rule, privacy risk will accumulate across multiple data releases, and differential privacy is equipped with tools for reasoning about this accumulated risk [58]. For example, the composition of two differentially private analyses with parameters ε1 and ε2 respectively can be viewed as a single differentially private analysis, and the resulting privacy parameter is bounded by ε1 + ε2. This robustness under composition means that privacy risks of multiple analyses can be bounded.

A consequence of composition is that the total desired ε can be treated as a "privacy budget" and consumed in parts by each individual data release. This budget can be thought of as analogous to any other budget. If the analyst has £500 to spend on all analyses overall, then they could perform 5 analyses at £100 each. Similarly, if the analyst requires an overall ε = 0.5, they could perform 5 analyses at ε=0.1 each. This ε budget can be split unevenly in the same manner as any budget, allowing fine tuning of noise addition across analyses. Most other privacy frameworks do not allow for such compositional reasoning. Differential privacy's compositional bounding of risk based on ε means that going over budget (i.e. returning a set number of statistics in a differentially private manner to give an overall ε, then returning more statistics) does not mean that the dataset is open to immediate attack. Instead, it means that the overall level of ε has increased, and the privacy guarantee has weakened in a measurable way.

Sometimes, differentially private releases compose better than this simple additive rule dictates. For example, if a histogram is released where bin counts correspond to disjoint subsets of individuals, each bin count can be assigned the full privacy budget. Put another way, a disjoint histogram where each bin count has budget ε has an overall budget of ε too. This improved composition rule is known as parallel composition.

[16] Resilience to post-processing is analogous to the data processing inequality from information theory. The data processing inequality states that the information content of a signal cannot be increased via its processing.

[17] One particular application is when an outcome of the differentially private analysis may contain inconsistencies or negative counts due to noise addition. While experts can take the noise into account in their analyses, agencies may choose to modify the public facing releases by making them consistent, or by replacing negative counts with zeros.

[18] In particular, the guarantees of differential privacy described in this and the following section would continue to hold when the transformation B is a privacy attack, and even if B implements computational methods unknown or unavailable at the time when the differentially private analysis A was designed and implemented.

It should be noted that some of the variants of differential privacy (discussed in the section below, “Variants of differential privacy”) allow for improved (sub-additive) composition (i.e. for N analyses with budget ε their overall budget is less than Nε and closer to sqrt(N)*ε) by virtue of more efficient composition theorems existing for these variants.

Differential privacy can allow reasoning about the privacy of groups as well as individuals: for all datasets x, y that differ on at most N entries, and for all subsets S of the outcome space, it holds that

$$\Pr[A(x) \in S] \leq e^{\epsilon N} \Pr[A(y) \in S]$$

This implies that the privacy guarantees that apply to an individual with the privacy parameter ε apply to a group of size N with the privacy parameter becoming Nε. The property of group privacy allows one to define a level of protection for families, for instance.

This combination of properties allows for complex analyses to be built up from well-understood differentially private primitives, and the risk to privacy can still be formally reasoned about even in the face of multiple complex analyses being performed on the same dataset.

4.4 Interpretations of the definition

Differential privacy can be interpreted as providing plausible deniability to individuals whose personal information is used in an analysis. This is in the sense that the output of the analysis does not allow an observer to distinguish whether a particular individual provided truthful or false inputs, or whether they participated in the analysis or not. This is because differential privacy requires that the observed outcome (i.e. one point on the outcome probability distribution) would be almost as probable (within a factor of e^{ϵ}) regardless of whether a given individual gives their real input or replaces it with an arbitrary value, or whether they do or do not participate in the analysis.

The parameter ε can be related to bounding the privacy risk due to an individual’s participation in an analysis in the sense that the probability of any event can change by a factor of at most e^{ϵ} due to the inclusion of the individual’s information in the analysis. Consider for example an individual who is worried that the inclusion of her medical information in an analysis may adversely affect her chances of obtaining a job position. If the analysis is differentially private, then the chances of not obtaining the position can grow by a factor of at most e^{ϵ} when compared to when the same analysis

is performed with the individual's information excluded[19]. Similarly, the chances of obtaining the position can diminish by a factor of at most e^{ϵ}.

In this way, the parameter ε can be thought of as bounding privacy loss. Setting of ε (i.e. the upper bound on privacy loss) is a policy decision that should be guided by a technical understanding of what such a decision would mean in terms of changes to privacy risk and utility of the analysis. The parameter ε can be shown to bound the vulnerability to reconstruction attacks, as a dataset can never be more accurately reconstructed than allowed by ε, a property that extends into the future regardless of auxiliary knowledge and attacker capabilities[20].

Comparing the privacy guarantee offered by differential privacy with that traditionally sought by SDC methods, we note that traditional disclosure risk assessment is often based on specific assumptions made about the data itself (e.g. that it is sampled from an i.i.d. distribution) and, furthermore, assumptions on the attacker's knowledge and methods. In contrast, the differential privacy guarantee is made without needing to rely on any such assumptions. The differential privacy guarantee is made in terms of comparing an individual's risk with and without her information being included in the analysis. This differential privacy guarantee does not directly relate to ensuring that an individual's secrets would not be disclosed, and indeed information about the individual could be disclosed if this information can be learned by observing the information of other individuals[21]. The guarantee, however, ensures that information that is specific to the individual, i.e., information that cannot be learned about her unless her information is included, would not be revealed.

Abowd and Schmutte have suggested an economic interpretation of differential privacy, with privacy protection and statistical accuracy as social goods [59]. In their interpretation, the fact that increasing privacy (i.e. lowering ε) means reduced utility is phrased as a resource allocation problem where an optimal choice of ε depends on weighing demand for statistical accuracy (i.e. utility) against privacy. This is an excellent example of how differential privacy opens up the possibility of reasoned balancing of privacy and utility at a policy level, and of how differential privacy can be connected to other disciplines.

[19] For a more detailed example of how the parameter ε can be related to privacy risk see [5]

[20] To see how choice of ε is related to protection against reconstruction attacks consider an ε-differentially private mechanism that is executed on dataset x where every entry is chosen at random from the data domain X. Note that if we replace an entry x_i in x with a x'_i which is randomly selected from X then the mechanism cannot reveal any information about x_i and hence a reconstruction attempt would guess x_i correctly with probability 1/|X|. It follows from the definition of differential privacy that when the mechanism is applied on the unmodified x the reconstruction of the i-th entry can succeed with probability at most $e^{\epsilon}/|X|$.

[21] For example, if a study shows that smokers are more prone to lung cancer, then this would be learned about individuals whether they participate in the study or not.

4.5 Variants of differential privacy

The definition presented above is often referred to as *pure* differential privacy. Several newer variants of ε-differential privacy have been proposed since 2006, each of which conveys a slight relaxation of the original formal requirement (that no outcome should change in probability by an amount more than e^{ϵ}). These variants have been devised principally to improve the privacy-utility trade-off of differential privacy. That is, a stronger privacy guarantee with the same level of utility, or a higher level of utility for the same privacy guarantee. They can be particularly important for use cases in which the noise added by traditional ε-differential privacy for the desired ε level affects utility too much. Variants typically consider a slightly weaker guarantee than ε-differential privacy, and thereby require less noise. These variants include approximate differential privacy [60], concentrated differential privacy [61, 62], and truncated concentrated differential privacy [63].

Approximate (ϵ, δ)-differential privacy is a variant that has received significant research attention. It includes a second parameter, delta (δ). It states that, for some δ>0:
$\Pr[A(x) \in S] \leq e^{\epsilon} \Pr[A(y) \in S] + \delta$

This additional δ allows for a probability of approximately δ that the likelihood ratio between the data present and data absent scenarios will be higher than e^{ϵ}. That is, with a probability of approximately δ, a given outcome can become more or less probable, between the data present and absent scenarios, by a factor greater than e^{ϵ}. Where this is the case, approximate differential privacy does not place a bound on how much greater than e^{ϵ} this factor can be. The parameter δ is typically chosen to be very small: significantly smaller than 1/N where N is the number of people whose data is in a dataset [58]. Approximate differential privacy can be interpreted as allowing for privacy breaches to happen with extremely small probability δ, akin to the accepted practice in cryptography of allowing for security breaches to happen with negligible probability.

Approximate differential privacy allows for added utility, leveraging noise addition techniques that do not satisfy pure differential privacy, as well as more efficient composition theorems than are possible for pure differential privacy.

One early example of approximate differential privacy being used in practice is the work of McSherry and Mironov, who created a (ϵ, δ) differentially private recommender system based on Netflix data [64]. Their assessment of the algorithm showed good utility, above that of the Cinematch baseline offered by Netflix. Additionally, Rinott et al. have discussed how approximate differential privacy might be applicable to releasing perturbed frequency tables with truncated cell perturbations of the sort suggested by the Australian Bureau of Statistics' Table Builder project, in order to increase utility [65].

4.6 Differentially private mechanisms and algorithms

This section details some mechanisms for achieving differential privacy, and some example tasks to which they have been applied.

Randomised response is a technique presented by Warner in 1965 as a way for respondents to answer sensitive or stigmatising survey questions while maintaining their privacy [66]. It is mostly useful for answering questions with binary Yes/No answers, and confidentiality is achieved by the respondent flipping her answer with a fixed probability. It is straightforward to show that if the flipping probability is set to $1/(1+e^{\varepsilon})$ then ε-differential privacy is achieved. For example, setting the flipping probability to 33% corresponds to having $\varepsilon = \ln 2$. Over a large population, this randomisation is averaged away and trends emerge. Randomised response underlies current implementations of differentially private analysis by Google and Apple. Google has published a system for telemetry based on randomised response called RAPPOR, which it uses for privacy-preserving telemetry in Google Chrome [21]. More broadly, the scenario where individuals add noise before submitting their data, so as to ensure they obtain the differential privacy guarantee for themselves, is referred to as the "local model". Local differential privacy mechanisms have the benefit of not requiring the analyst to be trusted, hence their appeal in industry. However, they have worse privacy-utility trade-offs than differential privacy mechanisms computed on data held centrally by a trusted curator (a setting referred to as the "curator model"). Local differential privacy mechanisms are effective mostly when there are very many data subjects [67].

Protecting summary statistics, such as sums, counts, and histograms, can be done with the Laplace mechanism [1]. This mechanism perturbs output by adding random noise that is calibrated to a property of the approximated statistic called "global sensitivity". Global sensitivity of a function f mapping databases into a real number is the amount that any single individual can influence the output of the function f:

$$\Delta_f = max|f(x) - f(y)|$$

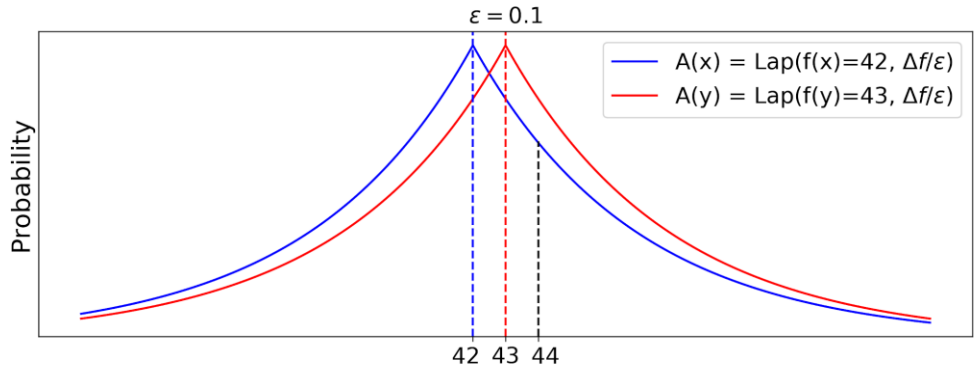
where the maximum is taken over all datasets x, y that differ on a single entry. For example, the sensitivity of count queries is 1, because a single individual can affect a count by at most 1 by being present or absent in the query set[22].

The Laplace mechanism adds noise, drawn from a Laplace distribution, to the output of a function f of a dataset[23]. These functions could be, for example, count queries or

[22] This notion can be generalised to functions that map databases into points in the d-dimensional Euclidean space: $\Delta_f = \max|f(x) - f(y)|_1$ where $|.|_1$ denotes the 1-norm.

[23] The Laplace distribution, denoted Lap(λ), is a continuous probability distribution with probability density function $h(z)=1/(2\lambda)\, e^{(-|z| / \lambda)}$.

machine learning training algorithms. The Laplace distribution is centred at 0 and has scale $\Delta f/\epsilon$ where Δf is the global sensitivity of the function. This means that with increasing values of ε the variance of the distribution is smaller (i.e. less noise is added), which corresponds to larger values of ε meaning less privacy in theory. Smaller ε yields higher variance and therefore more noise, which corresponds to more privacy. Similarly, smaller ε means less utility, whilst larger ε means the opposite.



Differential privacy via the Laplace mechanism is illustrated in the above figure. In this example, consider a function f that outputs the number of people who have a given medical condition in Bob's data present scenario (dataset y, f(y) = 43) and data absent scenario (dataset x, f(x) = 42). The Laplace mechanism A releases these counts with noise drawn from the Laplace distribution parameterised by $\Delta f/\epsilon$. Visualised in blue and red, respectively, is the probability that A(x) and A(y) will output a particular value for the number of people with the medical condition. For example, the probability that the value output in Bob's data present scenario P[A(y) = 44] is found where the line x = 44 intersects the red distribution, and likewise for P[A(x) = 44] and the blue distribution. The Laplace mechanism satisfies the requirements of differential privacy because, for all possible values v that could be output by A, the ratio of P[A(y) = v] to P[A(x) = v] is between $e^{-\varepsilon}$ and e^{ε} [24].

Intuitively, the noise added in the Laplace mechanism masks the additive contribution of any single individual to the function f, hence the noise is calibrated in proportion to the largest possible individual additive contribution Δf. A smaller ε (better privacy) results from a higher noise magnitude, hence the inverse proportion of privacy to ε [25]. Tuning the noise added to an algorithm's sensitivity in this manner can intuitively be understood as masking the contribution of any one individual. It has been formally proved that adding Laplace noise calibrated to sensitivity in this manner satisfies ε-differential privacy.

[24] More accurately, the probability of any event can change by a factor of at most e^{ε} between Bob's data present and data absent scenarios. Here, an event occurs when the output of A falls within some (measurable) set such as an interval or a union of intervals. For example, the ratio of the probabilities $P[A(y) \in [44, 45]]$ and $P[A(x) \in [44, 45]]$ is bounded between $e^{-\varepsilon}$ and e^{ε}.

[25] For the simple proof formalising this intuition see [1].

The Gaussian mechanism [58] is similar to the Laplace mechanism, with the following differences: it uses Gaussian noise instead of Laplace noise and achieves approximate differential privacy [60] instead of pure differential privacy.

In addition to these broadly applicable standard techniques (Randomised Response, Laplace mechanism, and Gaussian mechanism), there are a number of state of the art differentially private algorithms that attempt to achieve an optimal privacy-utility trade-off for specific tasks. For example, there are differentially private algorithms that produce categorical output [68], synthetic datasets [69], and algorithms that train machine learning models [70] (e.g. clustering [71], linear regression [72], and classification [73]).

Differentially private synthetic data sets, also referred to as private data releases, are row-level datasets produced from sensitive data, such that they maintain much of the statistical structure of the original data while preserving differential privacy. Some synthetic data generation procedures involve creating a statistical model of the dataset and sampling rows from that model. A critical difference between traditional and differentially private ways of doing so is that traditional techniques provide a heuristic notion of privacy and hence carry a risk that the resulting dataset may allow an attacker to learn information which is specific to rows in the original dataset (e.g. by first learning the model from the synthetic rows, then the original rows from the model) [74]. If the model is learned with differential privacy then the released synthetic data provably protects against an attacker learning specific information about rows in the original dataset[26]. It may therefore be beneficial to consider releasing differentially private synthetic data where organisations currently use traditional synthetic data techniques.

4.7 Scope of differential privacy usage

Differential privacy is not intended for de-identifying row-level datasets, but works instead on releasing aggregate statistics, like summary statistics, or publishing synthetic datasets[27]. Essentially, differential privacy hides the presence or absence of any individual row, where rows can correspond to an individual. This rules out specifying any information about single rows.

Differentially private algorithms work best in circumstances when data exists in abundance, and in particular when the noise added for achieving differential privacy is comparable to, or dominated by, other statistical errors (e.g. due to sampling and measurement). In such cases, the use of differentially private algorithms may benefit utility as they would allow use of data that would otherwise need to be redacted.

[26] In this situation, generation of the model itself is differentially private, whilst sampling from the model and releasing synthetic data is post-processing (which does not adversely affect the differential privacy guarantee).

[27] From the perspective of differential privacy, machine learning models are another type of aggregate statistic because, much like summary statistics, they represent data about groups rather than about individuals.

Examples of tasks that can be performed with large datasets include summary statistics, distribution learning, generation of synthetic databases, learning of classifiers, and clustering. On the other hand, it is immediate from the definition of differential privacy that when applied to very small databases (the extreme example being a dataset of only one person), differentially private algorithms would not yield a good privacy-utility trade-off.

Differential privacy inherently limits how much information can be extracted from datasets. In particular, differential privacy may not be achievable in use cases where synthetic datasets cannot be used and a dataset needs to be queried repeatedly and without limit, because such usage of the dataset may lead to the possibility of a reconstruction attack. It is worth noting that the framework of differential privacy may still be useful in providing weaker notions of privacy in some of these use cases. For example, if the data consists of events generated by individuals over time, then it may be possible to use differentially private computations to protect each of the events generated by each individual (called *event privacy*), but the overall nature of the activity of individuals may be revealed.

5. How differential privacy fits in with traditional SDC approaches

The previous sections outlined the contemporary context which has motivated the development of differential privacy, and described differential privacy itself. We now compare and contrast differential privacy with pre-existing SDC methods.

5.1 Similarities of traditional SDC methods and differential privacy methods

Differential privacy can in fact be viewed as a logical continuation of traditional SDC methods, of which noise perturbation is already a variety. Differential privacy adds to the progression of SDC methods a formal manner of reasoning about privacy loss, which in turn yields a rich variety of perturbation techniques. Furthermore, some specific SDC techniques can be analyzed through the lens of differential privacy and hence find their way into differentially private algorithms. Most prominently, randomised response, which was introduced over 30 years ago, is used as a building block in current large scale implementations of differentially private analyses including Google’s RAPPOR differential privacy system and Apple’s implementation of differential privacy.

Crucially, both traditional SDC and differential privacy methods may be subject to reconstruction attacks if too many statistics are released. For both traditional SDC and differential privacy methods, releasing the output of any useful analysis, even if perturbed, inherently involves an increase of disclosure risk, which grows with an increasing number of releases. An important advantage of differential privacy is that it provides a calculus for reasoning rigorously about this accumulated risk, whereas most or all traditional SDC methods do not exhibit such a calculus.

5.2 Benefits of differential privacy over traditional SDC

Differential privacy has certain benefits that distinguish it from traditional SDC. These are listed below:

Better quantification of privacy loss: Using analyses that satisfy differential privacy allows access to a rigorous framework in which privacy loss can be reasoned about and quantified. For example, knowing that any analysis leaks information, differential privacy quantifies this amount, even taking into account multiple analyses on the same dataset. This ability to formally quantify and track privacy risks from an analysis, or even across a set of analyses, is increasingly important for any privacy-aware organisation working with sensitive data. As there is a measurable and interpretable quantity, ε, to work with, agencies are able to make a quantifiable statement about individuals' expectations of privacy. Additionally, ε is associated with a specific level of noise perturbation added to the analysis. This means that the distortion level applied to the data can be mapped to utility within the specific definition of utility according to the use case. This allows informed balancing of privacy against data utility considerations.

Privacy risk metric reliant on few assumptions: Unlike differential privacy, traditional SDC methodologies are not founded on an easily quantifiable definition of privacy that is tied to individuals' expectations of privacy. Measures of disclosure risk do of course exist for traditional SDC, but are typically tied to assumptions about the adversary in terms of prior knowledge and attack methodologies [65]. Differential privacy's guarantee is agnostic to these attacker profile considerations. Realistically, then, the alternative with traditional SDC is a labour-intensive process of manually assessing risk that is less quantifiable.

Additionally, because differential privacy ensures that an analysis output is indistinguishable from the case where any single person's *entire* record had been absent from the input, differential privacy renders it unnecessary to decide which part of a data record is sensitive and which part is identifying. This is safer because judging sensitivity and identifiability is fallible, given that sensitivity and background knowledge levels change over time.

Better future-proofing: Many traditional SDC methods were conceived of to defend against attacks and background knowledge available at the time. That is, their ability to mitigate disclosure risk depends on assumptions about the attacker profile. This inherently limits the period for which they're effective, and they can be broken by newly discovered attacks or new auxiliary knowledge on the part of an adversary. In contrast, as discussed in Section 4.3, differential privacy is maintained under post-processing and hence an attack applied on the outcome of a differentially private analysis cannot break the differential privacy guarantee.

Public parameters: Traditional SDC methods often keep their parameters secret for security purposes, and this is common practice in statistical agencies. For example, the US Census kept their rate of record swapping secret in the last Census [44]. In fact, the US Census described their statistical disclosure avoidance process before differential privacy as having "*aspects of the black arts*," in that "*knowledge of the actual disclosure avoidance techniques and parameters was restricted to a small group of specialists, and the remainder of the agency treated disclosure avoidance as a black box that input dangerous data and output clean, safe data*" [75]. It is better to have parameters public because:

- Understanding of how risk composes across multiple analyses on the same underlying data is facilitated.
- Security-through-obscurity means that the public cannot scrutinise the protection method, and a point of failure is introduced.
- Incorrect conclusions from the released data are less likely to be caused by perturbation, as this perturbation can be known and accounted for. Indeed, if noise of a known distribution is added, it is possible for those making use of outputs from differentially private analyses to correct for the noise [76].

Synergy with protections against overfitting: The emphasis of differential privacy on preserving group insights can encourage good statistical practices. Use of noise to obscure individual contributions avoids overfitting and spurious results, but group trends and findings are preserved. In machine learning terms, this can be restated as "*differential privacy helps avoid overfitting*" [77], but more generally can be thought of as: if a finding is robust and genuine, then it should be robust to a small amount of random perturbation [78].

5.3 Current challenges to differential privacy and SDC

5.3.1 Determining acceptable privacy risk and setting parameters

It is not agreed upon, or fully understood, what ε should be set at to provide an acceptable level of privacy. The parameter ε bounds worst-case privacy loss, but how that relates to practical risk in real-world use cases is still an area of research. Differential privacy practitioners are building up experience which will provide insight on this issue in years to come. Academic papers typically use example ε's in the 0.01 - 1 range. However, industry implementations so far have set ε in the range of 1 - 10 (Apple set ε to 2, 4, and 8 in different applications [79]; Google uses $\ln(3) \approx 1.1$ [80]; Census OnTheMap used 8.99 [81]). Using high ε's weakens the worst-case guarantee of differential privacy, but leaves in place other desirable properties of differential privacy such as composability, transparency and quantifiability of privacy risk. Some work has done using auction theory to set ε, but this is based on the assumption that individuals must be compensated for their privacy loss, which is often not true in practice [82].

In contrast, because traditional SDC methods have been around for longer, there is more experience built up about how to translate between acceptable privacy loss at a practical level and at the level of method parameters. However, as discussed above, traditional SDC methods have been shown vulnerable to attack in spite of this accumulated experience in setting parameters.

5.3.2 Ensuring sufficient utility

Differential privacy has been criticised as having overly punitive effects on utility[28] compared to traditional SDC methods [83, 84]. These criticisms may stem from the impression that in many real world applications getting a low ε involves adding a lot of noise. In truth, differential privacy is a tunable model and there is a privacy-utility trade-off which can swing fully in both directions (i.e. maximising utility vs. maximising privacy). This re-expresses the problem of understanding whether higher values of ε (i.e. more utility) offer some degree of safety. Additionally, differential privacy's concept of a privacy budget may be seen as limiting use of data in ways traditional SDC does not. This view is, however, incorrect: all SDC techniques must limit use of data or else they will be vulnerable to reconstruction attacks. Differential privacy just makes this limit explicit.

Furthermore, differential privacy as a theory allows the privacy-utility trade-off to be formally understood in terms of ε, but the balance of these two factors is not dependent solely on the differential privacy approach itself, but rather the particular algorithm or mechanism employed and the particular use case. Past research has optimised the privacy-utility trade-off for specific use cases, and future research may lead to lower values of ε for the same utility in a broader variety of use cases, as well as improved algorithm-selection criteria or guidance.

All traditional SDC methods also have a privacy-utility trade-off. For example, when using small count suppression, raising suppression threshold improves privacy but harms utility as more statistics get suppressed. The key contrast with differential privacy is that for many traditional SDC methods the levels of privacy and utility provided are not as easily measurable as with differential privacy.

5.3.4 Ensuring consistency among statistics and avoiding negative counts

Because differential privacy adds noise to statistics, it can lead to situations where statistics are inconsistent. One example is simply regenerating precisely the same statistic, which may result in a different answer due to differentially private algorithms'

[28] Utility can be considered as informational content. Attempting to maximise utility by this definition is likely synergistic with any higher-level concept of utility, and this discussion applies to any definition of utility. However, it is worth being aware that the definition of utility as a practical concept, embodied by the simple question "*is this noisy data still useful?*", will depend on the intended use of the data. The possible need to assess utility via an appropriate metric is discussed in Section 6.3.3.

randomised nature[29]. Another example is the accuracy of table marginal counts. For instance, a differentially private release may say that there are 47 men with diabetes, 53 women with diabetes, and 99 people total with diabetes. Some efforts have been made to create differentially private mechanisms that can enforce consistency [85, 86]. It should be noted that creating differentially private synthetic data, and subsequently using that synthetic data to generate statistics, is always consistent. The US Census is taking this approach [75].

Similarly, the noise addition required to achieve differential privacy can sometimes lead to unexpected statistical estimates, such as when counts become negative because of noise addition. Negative counts can, however, be post-processed and replaced with zeros without adversely affecting privacy (though this may cause bias in the overall data release). Alternatively, certain differentially private algorithms, like synthetic data generation algorithms, avoid the issue of negative counts.

Consistency and negative counts are not differential privacy-specific problems as some traditional SDC methods suffer from the same problems (for instance, cell key perturbation). Teaching data analysts to better manage noisy statistics would benefit them whether using differential privacy or traditional SDC methods. In particular, analysts should be aware that some existing statistics software may not produce correct answers if fed inconsistent data or data with negative counts.

5.3.5 Impact on statistical tests

Problems can arise when taking statistics generated from a differentially private mechanism and performing a statistical test directly on them. For instance, running a chi-squared independence test on noisy counts will not give accurate p-values. There is currently active research on methods to perform statistical tests under differential privacy [87, 88].

For SDC methods, as with consistency, this problem is method dependent. Any method (traditional or differentially private) that involves perturbation of data can alter some correlations between variables and can affect statistical tests. Given these possible effects on statistical tests, the fact that differentially private methods are public and have public parameters is especially advantageous, because the effects on the statistical tests can be fully understood and accounted for.

5.3.6 Time-series data and releases over time

Most differentially private mechanisms are intended to work on rectangular data where there is one row per person. Much data nowadays is time-series data which is of higher dimensionality and where an individual row could correspond to an event, such as a

[29] It should be noted that this example is illustrative and is a poor use of privacy budget in practice. Identically re-querying an unchanging dataset still comes at a cost the overall privacy budget as it makes averaging attacks possible.

transaction, and an individual may appear in multiple rows. This data necessitates different approaches from the majority of those considered in differential privacy research. For instance, one must decide whether to protect each event or each individual. Additionally, multiple releases over time - where there are correlations between releases - pose similar problems. For instance, what happens if a data point hardly varies at all over time and is included in many sequential releases? There has been some work in this area, but there are currently no mature solutions [89, 90].

Most traditional SDC methods (record swapping, small count suppression) are also intended to work only on rectangular data and single releases. They too, therefore, run into some of the same issues as differentially private methods when dealing with non-rectangular datasets which may involve time-series data.

5.3.7 User education

As differential privacy has only a few real-world deployments so far, analysts in general are less familiar and less comfortable with the techniques used to achieve it. This issue becomes more pronounced as differentially private algorithms become more complex. Users need to understand differential privacy in terms of what it means when they receive differentially private noisy statistics, how to select between different differentially private algorithms, and how to determine whether a use case is suitable for differential privacy. Security and privacy departments, whether in academia or in government, need to be educated on the privacy benefits of differential privacy, its limitations, and how to set and interpret parameters.

Since traditional SDC methods are older, analysts are more familiar with them and there is widespread experience in their use. However, the ability to understand the effectiveness of these techniques is limited because of their heuristic nature.

6. Applications of differential privacy

Having explained the context and details of differential privacy, including its relationship to other SDC methodologies, we now turn to a discussion of applying differential privacy in practice. In this section, we outline how differential privacy might be useful to statistical agencies, some different pathways towards differential privacy adoption, some practical strategies in coping with the challenges it faces, and some general principles of good practice when using differential privacy.

6.1 Differential privacy for statistical agencies

Government statistical agencies may find differential privacy a valuable tool for several reasons. First, national statistics (that is, statistics about the population of a country) usually have large sample sizes. This is beneficial, because statistical releases about larger populations can achieve a given value of ε with less impact on accuracy.

Second, the transparency of differential privacy (i.e. that all details of the differentially private mechanism can safely be made public) is a benefit to statistical agencies. Use of public parameters and noise addition mechanisms enables recipients of national statistics to take noise perturbation into account. The transparency allows the effects of noise perturbation to be managed properly: for instance, the confidence intervals of each statistic can be calculated. An additional benefit of the open nature of differential privacy is that the public, including the academic community, can fully audit the data privacy mechanisms in place. This allows for verification of privacy protection and engenders public trust. If a national statistical agency were to suffer a privacy attack, it might risk loss of public trust and reputational damage. While this risk is not entirely avoided with differential privacy (e.g. incorrect implementations, poor selection of ε, and allowing for unlimited queries could each still lead to disclosure), the risk is mitigated if ε and the differentially private implementation are chosen transparently and with input from the public.

Third, the differential privacy guarantee is especially suitable when data is released to the public (e.g. online). In such cases, the data is likely to be exposed to adversaries who can execute advanced and evolving privacy attacks. Being future proof, differentially private algorithms defend against vulnerabilities that could emerge with the use of heuristic SDC approaches [91].

Fourth, many of the statistics published by government agencies are summation and counting queries, which have received a lot of attention in the differential privacy research literature and have mature differentially private mechanisms with good privacy-utility trade-offs. In particular, advanced approaches for differentially private histograms have been developed, improving on the privacy-utility trade-offs of the simple approach of adding Laplace noise to each bin. Collaborative academic work undertaken by the US Census Bureau has yielded algorithms for releasing workloads of counting queries (i.e. contingency tables) over high-dimensional data with improved privacy-utility trade-off and scalability [92]. The same programme of collaboration has yielded efficient differentially private algorithms for counts of counts (e.g. for every output area, count how many households contain x people, for different values of x) [93].

Fifth, statistical agencies generally have a clearer idea than other organisations of a specific, finite set of statistics they want to release. This is beneficial because differential privacy cannot easily handle systems that answer infinite different queries: either the finite ε privacy budget will be overspent and the desired level of privacy will not be achieved, or the budget will have to be split, possibly unevenly, between so many queries that the resulting high level of noise destroys utility[30]. Indeed, no SDC method that releases useful results can release such results indefinitely without being open to reconstruction attack.

[30] Note: this would also require knowing how many queries were to be asked in advance.

The fit of differential privacy to national statistics use cases has been investigated by others. Rinott and et al. published work in 2018 on the applicability of differential privacy to table builder tools at the US Census Bureau and Australian Bureau of Statistics [65]. They state several key requirements of a system releasing frequency tables that can be achieved under differential privacy, including preservation of structural zeros, unbiased perturbations, and truncated output based on bounded absolute difference between original and perturbed values (at the cost of moving from pure to approximate differential privacy). Their work highlights the fact that differential privacy is a fitting privacy model for national population statistics.

The good fit between differential privacy and national statistics does not, of course, guarantee that differential privacy will always apply to national statistics use cases problem-free. Common challenges to adopting differential privacy, such as setting ε, educating users in how to account for perturbations, and ensuring consistency, will still require work to overcome.

6.2 US Census case study

The US Census Bureau has conducted internal experiments confirming that confidential microdata from the 2010 Census can be reconstructed quite accurately, and they now state publicly that reconstruction attacks are a recognised vulnerability for census releases [45, 94]. Motivated by this issue, they have adopted a new differentially private mechanism for statistical disclosure control in the 2020 Census [44]. This mechanism was purpose-built for the US Census in order to have a strong privacy-utility trade-off while satisfying certain census-specific requirements. At a high level, the mechanism starts by creating a differentially private histogram of demographic attributes and sampling the right number of rows to create a synthetic dataset. It then iteratively assigns the rows to a state, county, tract, and block. This results in hierarchical tiers of geographic attributes, which is desired. The US Census was actually an early user of differential privacy with the OnTheMap project in 2008 [95]. That project opted for approximate differential privacy, as it provided better utility. However, the 2020 Census will use pure differential privacy.

The US Census Bureau is actively engaged in addressing the challenges of applying differential privacy in practice. At the time of writing, it is unclear exactly how they will set ε, but they have said that a policy committee (the Data Stewardship Executive Policy Committee), not technical staff, will decide on the value of ε. It is likely that committee thinking will be informed by Abowd and Schmutte's recent work on privacy and utility as public goods [59]. The Census Bureau chose a differentially private mechanism that produces synthetic data, rather than one that produced statistics directly, because synthetic data files can work with legacy software and ensure consistency. This minimises the disruption of introducing a new SDC methodology. They have given several educational presentations, both internally and externally,

about differential privacy and the reasons why it was chosen. These reasons are principally that a) it's easy to understand, b) the privacy guarantee is both provable and tunable, c) the privacy guarantee does not depend on assumptions about external data, d) reconstruction attacks are defended against, and e) the guarantee is composable across multiple releases [44].

A key challenge the Census Bureau will face when applying differential privacy is managing the privacy budget, which equates to setting the noise level such that the Census has acceptable levels of both utility and privacy. To address concerns about utility, they've educated users that every mechanism has a privacy-utility trade-off, and that differential privacy simply exposes this trade-off explicitly, with public parameters [94]. They also plan to reserve some privacy budget for the publishing of accuracy information at the end of the process, helping with post-hoc interpretation of differentially private statistics [44]. To our knowledge, at the time of preparing this document they have not yet publicly addressed the question of statistical tests done on Census data releases.

The Census Bureau has tailored their mechanism to the specifics of their use case, in order to achieve better privacy and utility. For instance, it is a legal requirement that the Census report correct block-level populations, which are used to calculate things like the number of Representatives apportioned to each state. The Census 2020 mechanism therefore ensures that some statistics are preserved exactly, while others are perturbed with noise addition. On this front, they are actively engaged in collaborations with differential privacy academics from a range of institutions such as Colgate University, Penn State, Duke University, the University of Massachusetts, Georgetown University, Harvard University, Boston University, and Purdue University. Such collaborations allow the Census Bureau to expand their expertise and address challenges specific to their use case. Academic publications resulting from their joint work can be publicly scrutinised, improving the quality of and encouraging confidence in their methods.

The 2020 Census may be just the start of differential privacy usage at the Census Bureau. Garfinkel, Abowd, and Powazek write, "*Beyond the 2020 Census, the Census Bureau intends to use differential privacy or related formal privacy systems to protect all of its statistical publications.*" They write of differential privacy that "[t]he methods put in place for the 2018 and 2020 implementations will act as templates, greatly easing its adoption in future statistical projects," highlighting the long-term value of the first major differential privacy deployment. And, despite the challenges associated with the first deployment, they believe that the Census Bureau will be rewarded: "With skilled staff and effective methodology in place, differential privacy can make lasting improvements to privacy protection at the federal government's largest statistical agency" [75].

6.3 Where and how to start applying differential privacy

This section discusses some pathways to differential privacy application, detailing types of use case well suited to an initial deployment of differential privacy in any organisation. Also covered are practical strategies for addressing differential privacy's challenges and general best practices when applying differential privacy.

6.3.1 Path to deployment A: Differentially private synthetic data

One potential pathway to effectively deploying differential privacy is to use differentially private synthetic data. In many cases where microdata is used for analysis, synthetic data can be substituted for the microdata directly. Additionally, existing non-differentially private synthetic data release use cases can also be replaced by differentially private synthetic data releases, which may be beneficial because releasing synthetic data without any perturbation can still be disclosive. Synthetic data is derived from statistics about the original data, and poorly protected synthetic data can reveal these statistics which in turn can reveal original records [74].

There are several practical advantages of beginning the adoption of differential privacy with a synthetic data use case. Differentially private synthetic data can be very useful for organisations which already have pipelines and processes in place to manage row-level data, and can facilitate the uptake of differential privacy as a means of thinking. Importantly, query results are the same every time, as no noise is added past the stage of creating the data. This means there is no confusion caused by inconsistent results. A synthetic dataset can be queried indefinitely to give aggregate statistics and, because of differential privacy's post-processing property, no further privacy risk is incurred.

One simple way to implement differentially private synthetic data generation for a counting query use case is by expressing the data as a differentially private histogram, and then sampling rows from this histogram. For example, consider an illustrative simple dataset about gender, where there are 100 people: 60 females and 40 males. This data can be expressed in a histogram. Then the Laplace mechanism can be used to make a differentially private, noisy histogram. Say the noisy histogram listed a 66 frequency for female and a 39 frequency for male. Finally, rows can be sampled from this histogram - the resulting dataset will have a female-to-male ratio of approximately 66:39. The resulting synthetic data retains the approximate pattern of the original dataset, but is differentially private. Several more sophisticated algorithms for synthetic data generation have been published (e.g. MWEM [69] and PrivBayes [96]), and the US Census Bureau will shortly publish their synthetic data generation algorithm developed for the 2020 Census.

As with releases of aggregate statistics, differential privacy allows, via setting ε, balancing of privacy and utility for synthetic data. Similarly, it also allows measurement and tracking of privacy risk across several releases of synthetic data in the same

manner as any other aggregate statistics. However, extra care must be taken when considering utility for synthetic data. Synthetic data embeds patterns from the original data, but not all patterns, and the recipient of the synthetic data may not realise which patterns are and are not embedded. If they query the synthetic data for a statistic that was not preserved in the synthetic data generation process, they risk getting a very inaccurate result [97]. Furthermore, they may not realise this is happening. Thus, those using the synthetic data must be careful not to query it in a way not intended by the data holders.

6.3.2 Path to deployment B: Noise addition on aggregate statistics

An alternative route to differential privacy adoption is to add differentially private noise to a release of aggregate statistics. This can be done, for example, by using the Laplace mechanism for sum and count queries. As before, it is easiest to choose a use case of count queries, as count queries are currently best understood and handled in differential privacy.

The best use cases to begin with consist of releases of fixed, known-in-advance sets of statistics. There are many of these use cases in government: for instance, the Department for Education releases the same (or very similar) tables about looked-after children every year [98]. These tables break down counts of looked-after children by several attributes, including category of need and ethnic origin. Having fixed statistics is important because, as stated previously, it is impossible have an unlimited number of queries, as incremental risk of reconstruction cannot be avoided.

Once a set of count or sum statistics is chosen, the Laplace mechanism can be straightforwardly applied to add noise to the statistics. In order to apply the Laplace mechanism, however, a value for ε must be chosen. Each ε value translates to a magnitude of noise, so the distortive effects of the noise on the statistics can be analysed before releasing the statistics. There are a number of more advanced algorithms than simple addition of Laplace noise independently to each aggregate statistic, such as the Matrix mechanism for count statistics [99]. Choice of algorithm is discussed further below (“Practical strategies for addressing the challenges of differential privacy”).

When releasing the statistics, the data holder can include information about the distortion, such as confidence intervals. This information allows the recipient of the noisy statistics to understand how accurate each statistic is.

Lastly, it is advisable to manage all such releases of differentially private statistics centrally, so that the composition effects of multiple statistical releases can be tracked. For instance, if ten total releases are going to be made, it may make sense to use one tenth of the total desired ε on each release. Whichever strategy for managing the

privacy budget is chosen, it is helpful to track all statistical releases through the same centrally managed system.

6.3.3 Practical strategies for addressing the challenges of differential privacy

The challenges of using differential privacy in practice are discussed above. This section covers some practical measures to manage those challenges.

One practical approach to checking that a set value of ε gives the required level of privacy in a practical sense is to attempt to attack a differentially private statistical release with differencing and reconstruction attacks. This is a heuristic approach to tackling a key question: will an adversary be able to learn anything they shouldn't from these noisy statistics? This allows translation of an ε-differential privacy guarantee into a practical understanding of the likelihood of disclosure at this ε level. This approach is heuristic in the sense that whilst such an approach will reveal when a chosen ε is too large to be an effective privacy protection, it will not reveal if this value of ε is smaller than required. Stated otherwise, this intruder testing can prove the presence of a vulnerability, but cannot prove the absence of all vulnerabilities. In spite of its imperfections, intruder testing has already been used for traditional SDC approaches [100]. One way to structure intruder testing is to run a public bug bounty-style contest where anyone can attempt to attack a sample dataset. Another is to perform this internally, although this does not allow for public verification. Any indication that a given ε is "safe" (i.e. prevents attacks in practice) will depend on the quality of attacks leveraged, and such a challenge should therefore use as wide an array of up to date attacks as possible.

Similarly, as differentially private analyses often come with metrics about the accuracy of each statistic they produce, it is beneficial to consider whether these accuracy levels are practically acceptable for the use case at hand. That is, is the distortion low enough to make the data release useful in practice? For example, if the data is being used for a city planning project, will the city planners draw the same conclusions from it? Considering these questions will lead to a better analysis of the privacy-utility trade-off that differential privacy offers, and inform a better choice of ε.

Education about the meaning of the differential privacy guarantee is essential. This should be tailored to the audience, and likely involves translating the ε parameter into something more intuitively understandable for a given audience. Those who are deciding whether to contribute to the dataset would likely prefer an explanation in terms of the minimal increase in risk that results from their data being included in the analysis. However, statisticians and analysts making use of the dataset would likely prefer explanations in terms of privacy budget in order to spend this budget wisely in their analysis, and in terms of distortion in order to understand effects on statistical tests. In some cases, it may well be worth holding back some privacy budget to publish information as to the accuracy of differentially private summary statistics, or to provide

some guidance about the accuracy of analyses performed on synthetic data [101, 102].

Large sample sizes improve the privacy-utility trade-off of differential privacy, and very small sample sizes may have unworkable privacy-utility trade-offs. This is because the risk of attacks such as reconstruction relates to the size of the dataset: as the number of statistics released, relative to dataset size, increases, more noise must be added to counter the growing risk of attack. For this reason, if there are many more statistics being published than people in the sample, the privacy-utility trade-off may be unworkable. As stated above, this is not just true for differentially private techniques, but for all SDC techniques. The precise relationship between number of statistics and privacy-utility trade-off is nuanced and dependent on factors such as query type. For instance, an especially high number of statistics can be published from a dataset if they are all counting queries. As a general rule, the larger the ratio of population to statistics, the better the privacy utility trade-off. Indeed, for most differentially private algorithms performing on a set number of statistics, having a larger dataset with the same ε is functionally equivalent to having a larger ε with the same dataset (i.e. less noise) [103]. Education about designing and performing differentially private analyses is therefore also important: analysts should be aware of these issues and exercise restraint in which statistics they choose to include in an analysis.

Applying differential privacy in practice means deciding which differentially private mechanism or algorithm is optimal for the task at hand. This involves discovering what is the best utility that can be achieved for a set value of ε. A broad selection of state of the art differentially private algorithms exist for a number of specific tasks. Selecting which is most appropriate in a given situation can be facilitated by starting with the simplest algorithms, and only moving towards more complex ones if a simpler approach does not give an acceptable privacy-utility trade-off for the specific situation. It should also be noted some differentially private algorithms (e.g. some synthetic data generation algorithms such as MWEM) involve free parameters, and that careful free parameter tuning can often improve utility. Research into the best method of algorithm selection is underway, with formal algorithm benchmarking frameworks being developed [103].

Currently, few broadly-applicable libraries or products for differential privacy are available. Applying differential privacy in practice often involves some software development, and this is more likely when moving beyond simpler algorithms. This could involve adapting existing algorithms to the specific use case, or developing new ones in-house. However, conducting bespoke development, possibly in collaboration with differential privacy researchers, can enable use-case specific improvements to the privacy-utility trade-off, as exemplified by the case of the US Census. Because differential privacy research is a young field, differential privacy expertise is currently concentrated primarily in academia, with smaller, but growing, communities in government and industry.

Lastly, it can be useful to post-process differentially private statistics with techniques like small count suppression for practical benefit. One useful pattern is applying the Laplace mechanism to a statistical release of count statistics, and then suppressing small counts: this retains full differential privacy protections, but improves the statistical release's ease-of-use, because it suppresses the statistics that are likely most distorted by differentially private noise addition and least useful, thus leaving only the most reliable statistics. Suppressing these least reliable statistics also reduces the size of the data release, leading to faster processing and less memory usage.

6.3.4 Best practices in differential privacy application

This section contains some suggested best practices when using differential privacy. These originate partially from Dwork's work with the Commission on Evidence-based Policymaking [104]. The first is to establish a set of differentially private mechanisms and get them approved externally. This may involve open sourcing code for core differential privacy primitives, similar to the OpenSSL library. Open-source differentially private mechanisms will speed differential privacy adoption, reduce duplicated work, and increase confidence that differential privacy is being implemented correctly. Where possible, bespoke developments in applying differential privacy to a given use case should also be made public, in order to build up a public library of useful differentially private algorithms and reduce future overhead of fitting differential privacy to a given application.

Another best practice is to make ε, and the ε selection process, public. This aids the conversation around how to take differential privacy from theory to practice, and engenders public trust. There are currently few examples of how ε values have worked in practice, and sharing knowledge about what works and what does not aids in the process of making differential privacy more effective in practice. Publicising chosen values of ε sets public benchmarks of privacy-utility trade-offs to be matched and, eventually, surpassed. In terms of publicising the selection process, one method of selecting ε (as described above, "Practical strategies for addressing the challenges of differential privacy") is to use intruder testing methodologies; if this method is used, the methodologies could be publicised for similar reasons.

6.4 Existing differential privacy tools

Here we present a (non-exhaustive) list of some of the major public differential privacy tools, which may be useful to differential privacy practitioners as references. It should be noted that there are also some commercial differential privacy products available.

- Google's RAPPOR [80].
- PINQ [105] (and its extension wPINQ [106]).
- Harvard Privacy Tools project: Private data Sharing Interface (PSI [87]).
- GUPT [107].

- FLEX [108].
- Diffpriv R toolbox [109].

7. Conclusion

Traditional SDC methods are being strained by increasing data volumes and powerful privacy attacks. In particular, reconstruction attacks call into serious question the abilities of traditional SDC methods to prevent unintended disclosure. This article presents differential privacy as a new alternative to replace or complement existing SDC methods, one that promises more robust privacy protection and novel capabilities for managing risk and increasing transparency. We recommend that organisations (i) identify the settings where differentially private methods may provide benefit over SDC techniques and (ii) conduct pilot deployments of differentially private methods, building expertise for wider deployments in the future.

7.1 Benefits of differential privacy

Differential privacy has several distinctive characteristics that benefit all stakeholders: data subjects, data holders (e.g. statistical agencies), and data users.

Differential privacy provides a mathematical guarantee about the privacy afforded to each data subject in a data analysis. The guarantee has a compelling intuitive interpretation: informally, each individual's information is protected (almost) to the extent it would be protected had this individual's data been retracted from the analysis. The additional risk that an individual may incur in a differentially private computation (as compared with retracting their data) is quantifiable. Furthermore, this risk can be quantified across multiple differentially private releases.

The privacy guarantee of differential privacy is unconditional as it does not depend on any assumptions about the attacker's strategy. In this sense, differential privacy counters not only known attack types such as reconstruction, but also future, not yet known attack types. This reduces the risk of reputational damage to statistical agencies that may arise from unintended disclosures, as well as the risk of harm to individuals.

In contrast to some traditional SDC methods, the privacy guarantees of differentially private analyses are maintained even if their algorithms are disclosed in full. This allows statistical agencies to benefit from increased transparency, as they can publicly release full details of algorithms and implementations. This transparency enables external scrutiny of the validity and appropriateness of the methods employed and their implementations. This engenders public trust and supports collaboration between statistical agencies, the public, and academia.

The transparency of differential privacy, and the fact that its parameters (like ε) can be safely published, means that those using the data have a clear understanding of how data has been treated to control disclosure, including the type and magnitude of noise introduced. This understanding can help prevent data users from drawing incorrect inferences or conclusions.

7.2 Challenges to adopting differential privacy

Differential privacy is a relatively new concept, and deploying differentially private technology in the real world requires addressing some practical challenges.

Organisations must develop a satisfactory processes for configuring differential privacy (e.g. setting its parameter ε which governs the strength of the privacy protection) such that their data releases are both useful and of acceptably low privacy risk. Setting ε properly involves understanding the real-world impact of noise addition in the context of each relevant use case. It also involves understanding how differential privacy conforms with legal and ethical privacy standards.

At this early stage of deploying differential privacy, organisations will also face the difficulty of finding or training qualified personnel to help reason about, design, build, and verify the validity of differentially private algorithms. Selecting or designing the best differentially private mechanism for each use case is crucial to attaining a good privacy-utility trade-off. Importantly, organisations will also need to invest in educating data holders, data analysts, and even data subjects about differential privacy. Collaboration among government, industry, and academia is key to addressing these challenges effectively.

7.3 Recommended next steps

This document is intended as an introductory step to understanding differential privacy, its strengths, and its limitations. For organisations who need to protect privacy in statistical releases, we recommend the following next steps:

- Gauge the risk of reconstruction attacks, and other state of the art privacy attacks, in existing statistical releases, taking into account the number of releases from the same source data and the existence of publicly available databases. Focus first on tools such as flexible table builders that release many statistics.

- Identify use cases suitable for piloting differential privacy. Examples include releases of descriptive statistics (e.g. frequency tables), where differentially private algorithms will fit with existing data analysis pipelines and cause minimal disruption. Consider lower-sensitivity datasets as good candidates for a first differentially private analysis in order to gain experience with differential privacy without undue disclosure risk.

- Engage with policymakers, legal scholars, differential privacy researchers, and other relevant stakeholders in order to start a wider discussion around what is an appropriate level of privacy protection.

- Strengthen relationships with differential privacy communities in academia and industry, as a way to influence research in directions that are relevant for the statistical agency (e.g. toward supporting weighted data or time-series data) and to more effectively bridge the gaps between the theory and practice of differential privacy.

With these steps, organisations can begin to use differential privacy for more robust and transparent privacy protection, to address existing vulnerabilities, and to prepare themselves for new privacy challenges ahead.

8. Recommended reading

Ordered from general to technical:

Garfinkel, S., Abowd, J. and Powazek, S. (2018). “Issues Encountered Deploying Differential Privacy.” *ArXiv*, September 6, 2018. https://doi.org/10.1145/3267323.3268949

- A recent account of issues in deploying differential privacy at the US Census and the reasons for pursuing it despite the challenges.

Nissim, K., Steinke, T., Wood, A., Altman, M., Bembenek, A., Bun, M. et al. (2018). “Differential Privacy: A Primer for a Non-Technical Audience,” 2018. *Harvard Privacy Tools Group.* Accessed 10 October 2018. https://privacytools.seas.harvard.edu/publications/differential-privacy-primer-non-technical-audience-preliminary-version.

- A primer describing differential privacy with minimal technical language but with accuracy.

Heffetz, O. and Ligett, K. (2014) “Privacy and Data-Based Research.” *Journal of Economic Perspectives* 28, no. 2, May 2014, 75–98. https://doi.org/10.1257/jep.28.2.75.

- An introduction of differential privacy with focus on empirical economic research.

Dwork, C., McSherry, F., Nissim, K, and Smith, A. (2017) “Calibrating Noise to Sensitivity in Private Data Analysis”. Journal of Privacy and Confidentiality 7 (3), 17-51, 2017. https://doi.org/10.29012/jpc.v7i3.405.

- An updated version of the paper that originally introduced differential privacy. Provides a general background for the concept.

Dwork, C., and Roth, A. (2014). “The Algorithmic Foundations of Differential Privacy.” *Foundations and Trends in Theoretical Computer Science* 9, no. 3–4 (August 11, 2014): 211–407. https://doi.org/10.1561/0400000042.

- A detailed mathematical introduction to the theory of differential privacy.

Vadhan, S. (2017). “The Complexity of Differential Privacy.” In *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, edited by Yehuda Lindell, 347–450. Information Security and Cryptography. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-57048-8_7.

- A concise review of theoretical results in differential privacy focusing on computational and sample complexity of various fundamental tasks.

9. References

[1] Cynthia, D., McSherry, F., Nissim, K. and Smith, A. (2017). “Calibrating Noise to Sensitivity in Private Data Analysis.” In *Theory of Cryptography*, edited by Shai Halevi and Tal Rabin, 265–84. Lecture Notes in Computer Science. Springer Berlin Heidelberg. See also: Cynthia, D.,McSherry, F., Nissim, K. and Smith, A. “Calibrating Noise to Sensitivity in Private Data Analysis.” *Journal of Privacy and Confidentiality* 7, No. 3 (May 30, 2017): 17–51. https://doi.org/10.29012/jpc.v7i3.405.

[2] Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Longhurst, J. et al. (2006).“Handbook on Statistical Disclosure Control”.

[3] Dinur, I. and Nissim, K. (2003). “Revealing Information While Preserving Privacy.” In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 202–210. PODS ’03. New York, NY, USA: ACM, 2003. https://doi.org/10.1145/773153.773173.

[4] Abowd, J.M. (2018) “The U.S. Census Bureau Adopts Differential Privacy.” In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2867–2867. KDD ’18. New York, NY, USA: ACM, 2018. https://doi.org/10.1145/3219819.3226070.

[5] Nissim, K., Steinke, T., Wood, A., Altman, M., Bembenek, A., Bun, M. et al. (2018). “Differential Privacy: A Primer for a Non-Technical Audience,” 2018. *Harvard Privacy Tools Group.* Accessed 10 October 2018. https://privacytools.seas.harvard.edu/publications/differential-privacy-primer-non-technical-audience-preliminary-version.

[6] Carpenter v. United States. No. 16-402. Supreme Court of the United States. June 22 2018.

[7] “Wifi Data Trial – Understanding London Underground Customer Journeys.” *TfL Digital Blog*, November 23, 2016. Accessed 10 October 2018. https://blog.tfl.gov.uk/2016/11/23/wifi-data-trial-understanding-london-underground-customer-journeys/

[8] Aktypi, A., Nurse, J. and Goldsmith, M. (2017). “Unwinding Ariadne’s Identity Thread: Privacy Risks with Fitness Trackers and Online Social Networks.” In *Proceedings of the 2017 on Multimedia Privacy and Security*, 1–11. MPS ’17. New York, NY, USA: ACM, 2017. https://doi.org/10.1145/3137616.3137617.

[9] “Number of monthly active Facebook users worldwide as of 2nd quarter 2018 (in millions).” *Statista.* 2018. https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/.

[10] “Statistical Disclosure Control – GSS.” Accessed October 10, 2018. https://gss.civilservice.gov.uk/guidances/methodology/statistical-disclosure-control/.

[11] “Government Transformation Strategy.” Accessed October 10, 2018. https://www.gov.uk/government/publications/government-transformation-strategy-2017-to-2020/government-transformation-strategy-better-use-of-data

[12] Hall, W.and Pesenti, J. (2018). “Growing The Artificial Intelligence Industry In The UK.” Accessed October 10, 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

[13] Bell, J. (2018). “Life Sciences Industrial Strategy.” Accessed October 10, 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/650447/LifeSciencesIndustrialStrategy_acc2.pdf

[14] ”World’s Biggest Data Breaches.” *Informationisbeautiful.net*. Accessed October 10, 2018. http://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/

[15] Pandurangan, V. (2018). “On Taxis and Rainbows.” Accessed October 10, 2018. https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1

[16] McPherson, R., Shokri, R. and Shmatikov, V. (2016). “Defeating Image Obfuscation with Deep Learning.” *ArXiv*, September 1, 2016. http://arxiv.org/abs/1609.00408.

[17] Narayanan, A. and Shmatikov, V. (2008). “Robust De-Anonymization of Large Sparse Datasets.” In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, 111–125. SP ’08. Washington, DC, USA: IEEE Computer Society, 2008. https://doi.org/10.1109/SP.2008.33.

[18] Perez, B., Musolesi, M. and Stringhini, G. (2018). “You Are Your Metadata: Identification and Obfuscation of Social Media Users Using Metadata Information.” Proceedings paper. AAAI Conference on Web and Social Media (ICWSM), June 25, 2018. https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17838.

[19] Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J. et al. (2008). “Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays.” *PLoS Genetics* 4, no. 8 (August 29, 2008). https://doi.org/10.1371/journal.pgen.1000167.

[20] “Asia Data Protection And Cyber Security Guide” *hldataprotection.com.* Accessed October 10, 2018. https://www.hldataprotection.com/files/2018/06/Hogan_Lovells_Asia_Data_Protection_and_Cyber_Security_Guide_2018.pdf

[21] “Learning Statistics with Privacy, Aided by the Flip of a Coin.” *Google AI Blog*. Accessed October 10, 2018. http://ai.googleblog.com/2014/10/learning-statistics-with-privacy-aided.html.

[22] “This Is How We Protect Your Privacy”. *apple.com.* Accessed October 10, 2018. https://www.apple.com/uk/privacy/approach-to-privacy/

[23] Murrill, B.J. and Liu, E. (2012). “Smart Meter Data: Privacy and Cybersecurity.” (February 3, 2012) *marylandsmartmeterawareness.org.* Accessed October 10, 2018. http://marylandsmartmeterawareness.org/wp-content/uploads/2012/07/Congress-Research-Service-SM-privacy-and-cybersecurity.pdf

[24] Greveler, U., Justus, B. and Loehr, D. (2012). “Multimedia Content Identification through Smart Meter Power Usage Profiles.” *Proceedings of the International Conference on Information and Knowledge Engineering (IKE).* 2012.

[25] Barker, E. (2018). “Recommendation for Key Management”. *Nist.gov.* Accessed October 10 2018. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-57pt1r4.pdf

[26] “Big Data and Privacy: A Technological Perspective.” (May 2014) *Nist.gov.* Accessed October 10, 2018 https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf

[27] Sweeney, L. (2002). “K-Anonymity: A Model for Protecting Privacy.” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, no. 05 (October 2002): 557–70. https://doi.org/10.1142/S0218488502001648.

[28] Ganta, K. and Smith, A. (2008).“Composition Attacks and Auxiliary Information in Data Privacy.”*ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2008): 265-273.

[29] Culnane, C., Rubinstein, B. and Teague, V. (2017). “Health Data in an Open World,” December 15, 2017. https://arxiv.org/abs/1712.05627.

[30] Gomatam, S., A. F. Karr, J. P. Reiter, and A. P. Sanil. (2005). “Data Dissemination and Disclosure Limitation in a World Without Microdata: A Risk–Utility Framework for Remote Access Analysis Servers.” *Statistical Science* 20, no. 2 (May 2005): 163–77. https://doi.org/10.1214/088342305000000043.

[31] Denning, D., Denning, P. and Schwartz, M. (1979). “The Tracker: A Threat to Statistical Database Security.” *ACM Trans. Database Syst*, (March 1979): 76–96.

[32] Matthews, G., Harel, O. and Aseltine, R. (2017). “A Review of Statistical Disclosure Control Techniques Employed by Web-Based Data Query Systems.” *Journal of Public Health Management and Practice* 23, no. 4 (2017): e1–4. https://doi.org/10.1097/PHH.0000000000000473.

[33] Korolova, A. (2010). “Privacy Violations Using Microtargeted Ads: A Case Study.” In *2010 IEEE International Conference on Data Mining Workshops*, 474–82, 2010. https://doi.org/10.1109/ICDMW.2010.137.

[34] Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K., Loiseau, P. and Goga, O. (2018). “Privacy Risks with Facebook’s PII-Based Targeting: Auditing a Data Broker’s Advertising Interface.” In *2018 IEEE Symposium on Security and Privacy (SP)*, 89–107, 2018. https://doi.org/10.1109/SP.2018.00014.

[35] Faizullabhoy, I, and Korolova, A. (2018). “Facebook’s Advertising Platform: New Attack Vectors and the Need for Interventions.” *ArXiv*, March 27, 2018. http://arxiv.org/abs/1803.10099.

[36] Fraser, B. and Wooton, J. (2005). “A Proposed Method for Confidentialising Tabular Output to Protect against Differencing,” (November 2005).

[37] Chipperfield, J., Gow, D. and Loong, B. (2016). “The Australian Bureau of Statistics and releasing frequency tables via a remote server.” *Stat. J. IAOS* 32 53–64, 2016.

[38] Dwork, C., McSherry, F. and Talwar, K. (2007). “The Price of Privacy and the Limits of LP Decoding.” In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, 85–94. STOC ’07. New York, NY, USA: ACM, 2007. https://doi.org/10.1145/1250790.1250804.

[39] Dwork, C. and Yekhanin, S. (2008). “New Efficient Attacks on Statistical Disclosure Control Mechanisms.” In *Advances in Cryptology – CRYPTO 2008*, edited by David Wagner, 5157:469–80. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. https://doi.org/10.1007/978-3-540-85174-5_26.

[40] Choromanski, K, and Malkin, T. (2012). “The Power of the Dinur-Nissim Algorithm: Breaking Privacy of Statistical and Graph Databases.” In *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, 65–76. PODS ’12. New York, NY, USA: ACM, 2012. https://doi.org/10.1145/2213556.2213570.

[41] Kantor, A. and Nissim, K. (2013). “Attacks on Statistical Databases: The Highly Noisy Case.” *Information Processing Letters* 113, no. 12 (June 30, 2013): 409–13. https://doi.org/10.1016/j.ipl.2013.03.005.

[42] Abowd, J., Alvisi, L., Dwork, C., Kannan, S., Machanavajjhala, A. and Reiter, J. (2017). “Privacy-Preserving Data Analysis for the Federal Statistical Agencies.” *ArXiv E-Prints* 1701 (January 1, 2017): arXiv:1701.00752.

[43] Kasiviswanathan, R. and Smith, A. (2013). “The Power of Linear Reconstruction Attacks.” In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, 1415–33. Proceedings. Society for Industrial and Applied Mathematics, 2013. https://doi.org/10.1137/1.9781611973105.102.

[44] Garfinkel, S. (2018). “Modernizing the Disclosure Avoidance System for the 2020 Census”. Accessed October 10 2018 http://simson.net/ref/2018/2018-02-14%20Garfinkel%20Gerogetown%20Modernizing%20the%20DAS%20for%20the%202020%20Census.pdf

[45] Abowd, J. (2018). “Staring Down the Database Reconstruction Theorem.” Accessed October 20, 2018. https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/jsm/jsm-presentation-database-reconstruction.pdf

[46] Dwork, C., Smith, A., Steinke, T., Ullman, J. and Vadhan, S. (2015). “Robust Traceability from Trace Amounts.” In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, 650–69, 2015. https://doi.org/10.1109/FOCS.2015.46.

[47] Pyrgelis, A., Troncoso, C. and Cristofaro, E. (2018). “Knock Knock, Who’s There? Membership Inference on Aggregate Location Data.” In *Proceedings 2018 Network and Distributed System Security Symposium*. San Diego, CA: Internet Society, 2018. https://doi.org/10.14722/ndss.2018.23183.

[48] Shokri, R., M. Stronati, C. Song, and V. Shmatikov. (2017).“Membership Inference Attacks Against Machine Learning Models.” In *2017 IEEE Symposium on Security and Privacy (SP)*, 3–18, 2017. https://doi.org/10.1109/SP.2017.41.

[49] Machanavajjhala, A., Kifer, D., Gehrke, J, and Venkitasubramaniam, M. (2007). “L-Diversity: Privacy Beyond K-Anonymity.” *ACM Trans. Knowl. Discov. Data* 1, no. 1 (March 2007). https://doi.org/10.1145/1217299.1217302

[50] Li, N., Li, T. and Venkatasubramanian, S. (2007). “T-Closeness: Privacy Beyond k-Anonymity and l-Diversity.” In *2007 IEEE 23rd International Conference on Data Engineering*, 106–15, 2007. https://doi.org/10.1109/ICDE.2007.367856).

[51] Dalenius, T. (1977). “Towards a methodology for statistical disclosure control.” *Statistik Tidskrift* 15, pp. 429–444, 1977.

[52] Nissim, K. and Wood, A. (2018). “Is Privacy *Privacy* ?” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no. 2128 (September 13, 2018): 20170358. https://doi.org/10.1098/rsta.2017.0358

[53] Dwork, C. and Naor, M. (2010). “On the Difficulties of Disclosure Prevention in Statistical Databases or The Case for Differential Privacy.” *Journal of Privacy and Confidentiality* 2, no. 1 (September 1, 2010). https://doi.org/10.29012/jpc.v2i1.585.

[54] Evfimievski, A., Gehrke, J. and Srikant, R. (2003). “Limiting Privacy Breaches in Privacy Preserving Data Mining.” In *Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 211–222. PODS ’03. New York, NY, USA: ACM, 2003. https://doi.org/10.1145/773153.773174.

[55] Dwork, C. and Nissim, K. (2004) “Privacy-Preserving Datamining on Vertically Partitioned Databases.” *CRYPTO*, 528–544. Springer, 2004.

[56] Blum, A., Dwork, C., McSherry, F. and Nissim, K. (2005). “Practical Privacy: The SuLQ Framework.” In *Proceedings of the Twenty-Fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 128–138. PODS ’05. New York, NY, USA: ACM, 2005. https://doi.org/10.1145/1065167.1065184.

[57] Kasiviswanathan, S. and Smith, A. (2014). “On the ‘Semantics’ of Differential Privacy: A Bayesian Formulation.” *Journal of Privacy and Confidentiality* 6, no. 1 (June 1, 2014). https://doi.org/10.29012/jpc.v6i1.634.

[58] Dwork, C, and Roth, A. (2014). “The Algorithmic Foundations of Differential Privacy.” *Foundations and Trends in Theoretical Computer Science* 9, no. 3–4 (August 11, 2014): 211–407. https://doi.org/10.1561/0400000042.

[59] Abowd, J. and Schmutte, I. (2018). “An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices,” August 20, 2018. https://arxiv.org/abs/1808.06303.

[60] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I. and Naor, M. (2006). “Our Data, Ourselves: Privacy Via Distributed Noise Generation.” In *Advances in Cryptology - EUROCRYPT 2006*, edited by Serge Vaudenay, 486–503. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2006.

[61] Dwork, C. and Rothblum, G. (2016). “Concentrated Differential Privacy,” March 6, 2016. https://arxiv.org/abs/1603.01887.

[62] Bun, m. and Steinke, T. (2016). “Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds.” In *Theory of Cryptography*, edited by Martin Hirt and Adam Smith, 635–58. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2016.

[63] Bun, M, Dwork, C. Rothblum, G. and Steinke, T. (2018). “Composable and Versatile Privacy via Truncated CDP.” In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, 74–86. STOC 2018. New York, NY, USA: ACM, 2018. https://doi.org/10.1145/3188745.3188946.

[64] McSherry, F, and Mironov, I. (2009). “Differentially Private Recommender Systems: Building Privacy into the Netflix Prize Contenders.” In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 627–636. KDD ’09. New York, NY, USA: ACM, 2009. https://doi.org/10.1145/1557019.1557090

[65] Rinott, Y., O’Keefe, C., Shlomo, N. and Skinner, C. (2018). “Confidentiality and Differential Privacy in the Dissemination of Frequency Tables.” *Statistical Science* 33, no. 3 (August 2018): 358–85. https://doi.org/10.1214/17-STS64

[66] Warner, S. (1965). “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias.” *Journal of the American Statistical Association* 60, no. 309 (March 1, 1965): 63–69. https://doi.org/10.1080/01621459.1965.10480775.

[67] Kasiviswanathan, S., Lee, H., Nissim, K., Raskhodnikova, S. and Smith, A. (2011). “What Can We Learn Privately?” *SIAM Journal on Computing* 40, no. 3 (January 1, 2011): 793–826. https://doi.org/10.1137/090756090.

[68] McSherry, F. and Talwar, K. (2007). “Mechanism Design via Differential Privacy.” In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS’07)*, 94–103, 2007. https://doi.org/10.1109/FOCS.2007.66

[69] Hardt, M., Ligett, K. and Mcsherry, F. (2012). “A Simple and Practical Algorithm for Differentially Private Data Release.” In *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 2339–2347. Curran Associates, Inc., 2012. http://papers.nips.cc/paper/4548-a-simple-and-practical-algorithm-for-differentially-private-data-release.pdf

[70] Abadi, M., Erlingsson, U., Goodfellow, I., McMahan, B., Papernot, N., Mironov, I. et al. (2017). “On the Protection of Private Information in Machine Learning Systems: Two Recent Approaches.” In *Proceedings of 30th IEEE Computer Security Foundations Symposium (CSF)*, 1–6, 2017. https://arxiv.org/abs/1708.08022

[71] Su, D., Cao, J., Li, N., Bertino, E. and Jin, H. (2016). “Differentially Private K-Means Clustering.” In *Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy*, 26–37. CODASPY ’16. New York, NY, USA: ACM, 2016. https://doi.org/10.1145/2857705.2857708

[72] Wang, Y. (2018). “Revisiting Differentially Private Linear Regression: Optimal and Adaptive Prediction & Estimation in Unbounded Domain.” *ArXiv*, March 7, 2018. http://arxiv.org/abs/1803.02596

[73] Liu, X., Li, Q., Li, T. and Chen, D. (2018). “Differentially Private Classification with Decision Tree Ensemble.” *Applied Soft Computing* 62 (January 1, 2018): 807–16. https://doi.org/10.1016/j.asoc.2017.09.010

[74] Hu, J., Reiter, J. and Wang, Q. (2014). “Disclosure Risk Evaluation for Fully Synthetic Categorical Data.” In *Privacy in Statistical Databases*, edited by Josep Domingo-Ferrer, 185–99. Lecture Notes in Computer Science. Springer International Publishing, 2014.

[75] Garfinkel, S., Abowd, J. and Powazek, S. (2018). “Issues Encountered Deploying Differential Privacy.” *ArXiv*, September 6, 2018. https://doi.org/10.1145/3267323.3268949

[76] Balle, B. and Wang, Y. (2018). “Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising.” In *International Conference on Machine Learning*, 394–403, 2018. http://proceedings.mlr.press/v80/balle18a.html

[77] Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O. and Roth, A. (2015). “Generalization in Adaptive Data Analysis and Holdout Reuse.” In *Advances in Neural Information Processing Systems 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, 2350–2358. Curran Associates, Inc., 2015. http://papers.nips.cc/paper/5993-generalization-in-adaptive-data-analysis-and-holdout-reuse.pdf.

[78] “Privacy and Machine Learning: Two Unexpected Allies?” *cleverhans-blog*, April 29, 2018. http://cleverhans.io/privacy/2018/04/29/privacy-and-machine-learning.html.

[79] “Learning with Privacy at Scale.” *Apple Machine Learning Journal*. Accessed October 10, 2018. https://machinelearning.apple.com/2017/12/06/learning-with-privacy-at-scale.html

[80] Erlingsson, U., Pihur, V. and Korolova, A. (2014). “RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response.” In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 1054–1067. CCS ’14. New York, NY, USA: ACM, 2014. https://doi.org/10.1145/2660267.2660348

[81] “Formal Privacy Guarantees and Analytical Validity of *OnTheMap* Public-Use Data”. Presented at 3rd IAB Workshop on Confidentiality and Disclosure, (November 2008). Accessed October 10, 2018. http://doku.iab.de/fdz/events/2008/SDC-Workshop_Presentation_Andersson.pdf

[82] Ghosh, A. and Roth, A. (2015). “Selling Privacy at Auction.” *Games and Economic Behavior* 91 (May 1, 2015): 334–46. https://doi.org/10.1016/j.geb.2013.06.013.

[83] Clifton, C. and Tassa, T. (2013). “On Syntactic Anonymity and Differential Privacy.” In *2013 IEEE 29th International Conference on Data Engineering Workshops (ICDEW)*, 88–93, 2013. https://doi.org/10.1109/ICDEW.2013.6547433.

[84] Bambauer, J., Muralidhar, K. and Sarathy, R. (2013). “Fool’s Gold: An Illustrated Critique of Differential Privacy.” SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, September 15, 2013. https://papers.ssrn.com/abstract=2326746.

[85] Barak, B., Chaudhuri, K., Dwork, C., Kale, S., McSherry, F. and Talwar, K. (2007). “Privacy, Accuracy, and Consistency Too: A Holistic Solution to Contingency Table Release.” In *Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 273–282. PODS ’07. New York, NY, USA: ACM, 2007. https://doi.org/10.1145/1265530.1265569

[86] Hay, M., Rastogi, V., Miklau, G. and Suciu, D. (2010). “Boosting the Accuracy of Differentially Private Histograms Through Consistency.” *Proc. VLDB Endow.* 3, no. 1–2 (September 2010): 1021–1032. https://doi.org/10.14778/1920841.1920970

[87] Gaboardi, M., Lim, H., Rogers, R. and Vadhan, S. (2016). “Differentially Private Chi-Squared Hypothesis Testing: Goodness of Fit and Independence Testing.” In *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48*, 2111–2120. ICML’16. New York, NY, USA: JMLR.org, 2016. http://dl.acm.org/citation.cfm?id=3045390.3045613

[88] Wang, Y., Lee, J. and Kifer, D. (2015). “Revisiting Differentially Private Hypothesis Tests for Categorical Data,” November 11, 2015. https://arxiv.org/abs/1511.03376.

[89] Rastogi, V. and Nath, S. (2010). “Differentially Private Aggregation of Distributed Time-Series with Transformation and Encryption.” In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, 735–746. SIGMOD ’10. New York, NY, USA: ACM, 2010. https://doi.org/10.1145/1807167.1807247

[90] Papadimitriou, S., Li, F., Kollios, G. and Yu, P. (2007). “Time Series Compressibility and Privacy.” In *Proceedings of the 33rd International Conference on Very Large Data Bases*, 459–470. VLDB ’07. Vienna, Austria: VLDB Endowment, 2007. http://dl.acm.org/citation.cfm?id=1325851.1325905

[91] Gaboardi, M., Honaker, J., King, G., Murtagh, J., Nissim, K., Ullman, J. and Vadhan, S. (2016).“PSI ({\Psi}): A Private Data Sharing Interface,” September 14, 2016. https://arxiv.org/abs/1609.04340.

[92] McKenna, R., Miklau, G., Hay, M. and Machanavajjhala, A. (2018). “Optimizing Error of High-Dimensional Statistical Queries under Differential Privacy.” *Proceedings of the VLDB Endowment* 11, no. 10 (June 1, 2018): 1206–19. https://doi.org/10.14778/3231751.3231769

[93] Kuo, Y., Chiu, C., Kifer, D., Hay, M. and Machanavajjhala, A. (2018). “Differentially Private Hierarchical Count-of-Counts Histograms.” *Proceedings of the VLDB Endowment* 11, no. 11 (July 1, 2018): 1509–21. https://doi.org/10.14778/3236187.3236202

[94] Abowd, J. (2018). “Why the Census Bureau Adopted Differential Privacy for the 2020 Census of Population.” 6 June 2018. Webinar. https://www.nsf.gov/events/event_summ.jsp?cntn_id=245477&org=NSF

[95] Machanavajjhala, A., D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber. (2008). “Privacy: Theory Meets Practice on the Map.” In *2008 IEEE 24th International Conference on Data Engineering*, 277–86, 2008. https://doi.org/10.1109/ICDE.2008.4497436

[96] Zhang, J., Cormode, G., Procopiuc, C., Srivastava, D. and Xiao, X. (2017). “PrivBayes: Private Data Release via Bayesian Networks.” *ACM Trans. Database Syst.* 42, no. 4 (October 2017): 25:1–25:41. https://doi.org/10.1145/3134428

[97] Bowen, C. and Liu, F. (2016). “Comparative Study of Differentially Private Data Synthesis Methods.” *ArXiv*, February 2, 2016. http://arxiv.org/abs/1602.01063

[98] “Children Looked after in England Including Adoption: 2016 to 2017.” *gov.uk*. Accessed October 10, 2018. https://www.gov.uk/government/statistics/children-looked-after-in-england-including-adoption-2016-to-2017

[99] Li, C., Miklau, G., Hay, M., McGregor, A. and Rastogi, V. (2015). “The Matrix Mechanism: Optimizing Linear Counting Queries under Differential Privacy.” *The VLDB Journal* 24, no. 6 (December 1, 2015): 757–81. https://doi.org/10.1007/s00778-015-0398-x

[100] Tudor, C, Cornish, G. and Spicer, K. (2011). “Intruder Testing on the 2011 UK Census: Providing Practical Evidence for Disclosure Protection.” *Journal of Privacy and Confidentiality* 5, no. 2 (February 1, 2014). https://doi.org/10.29012/jpc.v5i2.632

[101] Barrientos, A., Reiter, J., Machanavajjhala, A. and Chen, Y. (2017). “Differentially Private Significance Tests for Regression Coefficients.” *ArXiv:1705.09561 [Stat]*, May 26, 2017. http://arxiv.org/abs/1705.09561

[102] Yu, H, and Reiter, J. (2018). “Differentially Private Verification of Regression Predictions from Synthetic Data,” 2018, 19.

[103] Hay, M., Machanavajjhala, A., Miklau, G., Chen, Y. and Zhang, D. (2016). “Principled Evaluation of Differentially Private Algorithms Using DPBench.” In *Proceedings of the 2016 International Conference on Management of Data*, 139–154. SIGMOD ’16. New York, NY, USA: ACM, 2016. https://doi.org/10.1145/2882903.2882931

[104] Dwork, C. (2018). “Differential Privacy and its Properties”. *Cep.gov*. Accessed October 2, 2018. https://www.cep.gov/content/dam/cep/events/2016-09-09/2016-09-dwork.pdf

[105] McSherry, F. (2010). “Privacy Integrated Queries: An Extensible Platform for Privacy-Preserving Data Analysis.” *Communications of the ACM* 53, no. 9 (September 1, 2010): 89. https://doi.org/10.1145/1810891.1810916

[106] Proserpio, D, Goldberg, S. and McSherry, F. (2014). “Calibrating Data to Sensitivity in Private Data Analysis: A Platform for Differentially-Private Analysis of Weighted Datasets.” *Proc. VLDB Endow.* 7, no. 8 (April 2014): 637–648. https://doi.org/10.14778/2732296.2732300

[107] Mohan, P., Thakurta, A., Shi, E., Song, D. and Culler, D. (2012). “GUPT: Privacy Preserving Data Analysis Made Easy.” In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, 349–360. SIGMOD ’12. New York, NY, USA: ACM, 2012. https://doi.org/10.1145/2213836.2213876

[108] Johnson, N., Near, J. and Song, D. (2017). “Towards Practical Differential Privacy for SQL Queries.” *ArXiv*, June 28, 2017. https://doi.org/10.1145/3177732.3177733

[109] Rubinstein, B. and Alda, F. (2017). “Diffpriv: An R Package for Easy Differential Privacy,” *Journal of Machine Learning Research*, 18, 1-5, (July 2017). https://cran.r-project.org/web/packages/diffpriv/vignettes/diffpriv.pdf