CHAPTER 6

Designing Access with Differential Privacy

Alexandra Wood (Harvard University)
Micah Altman (Massachusetts Institute of Technology)
Kobbi Nissim (Georgetown University)
Salil Vadhan (Harvard University)

6.1 Introduction and Overview

This chapter explains how administrative data containing personal information can be collected, analyzed, and published in a way that ensures the individuals in the data will be afforded the strong protections of *differential privacy*.

It is intended as a practical resource for government agencies and research organizations interested in exploring the possibility of implementing tools for differentially private data sharing and analysis. Using intuitive examples rather than the mathematical formalism used in other guides, this chapter introduces the differential privacy definition and the risks it was developed to address. The text employs modern privacy frameworks to explain how to determine whether the use of

Cite as: Wood, Alexandra, Micah Altman, Kobbi Nissim, and Salil Vadhan. “Designing Access with Differential Privacy.” In: Cole, Shawn, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber (eds.), *Handbook on Using Administrative Data for Research and Evidence-based Policy*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab. 2020.

differential privacy is an appropriate solution in a given setting. It also discusses the design considerations one should take into account when implementing differential privacy. This discussion incorporates a review of real-world implementations, including tools designed for tiered access systems combining differential privacy with other disclosure controls presented in this Handbook, such as consent mechanisms, data use agreements, and secure environments.

Differential privacy technology has passed a preliminary transition from being the subject of academic work to initial implementations by large organizations and high-tech companies that have the expertise to develop and implement customized differentially private methods. With a growing collection of software packages for generating differentially private releases from summary statistics to machine learning models, differential privacy is now transitioning to being usable more widely and by smaller organizations.

6.1.1 Organization of this Chapter

We place differential privacy in a general framework—introduced by Altman et al. (2015) and an alternative to the Five Safes framework (Desai, Ritchie and Welpton, 2016) used throughout this Handbook—that involves selecting combinations of statistical, technical, and administrative controls to mitigate risks of harm to individuals resulting from access to data. The framework discusses differential privacy as an approach to employ together with other tools, including consent mechanisms, data use agreements, and secure environments. Some of the content in this chapter (Sections 6.1–6.3) is excerpted from, adapted from, or otherwise based, in part, on Wood et al. (2018) and Altman et al. (2015).

The chapter is organized as follows: Section 6.2 explains the differential privacy guarantee in more detail using stories to illustrate what differential privacy does and does not protect. Section 6.3 places differential privacy in a general framework of complementary privacy controls and characterizes principles for selecting differential privacy in conjunction with other controls. These principles include calibrating privacy and security controls to the intended uses and privacy risks associated with the data, and anticipating, regulating, monitoring, and reviewing interactions with data across all stages of the lifecycle (including the post-access stages), as risks and methods will evolve over time. Section 6.4 presents succinct summaries of several deployment cases. These provide selected concrete examples of data dissemination that illustrate some key design choices and their implications.

More technical discussions of several topics are included in an extensive online appendix. A discussion of different technical approaches to disseminating data with differential privacy can be found in Appendix A, which also characterizes the key design choices and trade-offs across them. Appendix B elaborates on the implications of differential privacy for data collection, use, and dissemination with a special emphasis on how differential privacy affects data collection and data repository practice and policy. Appendix C provides a list of selected tools and resources for implementing differential privacy protections.

Section 6.2 is recommended for policymakers as well as for analysts and communications professionals seeking to explain differential privacy to policymakers, data users, and data subjects. Sections 6.3 and 6.4, in combination with Appendix B, are recommended for organizational directors and principal investigators responsible for identifying where differential privacy is appropriate as part of a project or organization-level data-protection strategy. Appendices A, B, and C are recommended for those with a technical background aiming to design and deploy differential privacy addressing specific data dissemination requirements.

6.1.2 Motivation: Formal Guarantees are Needed to Protect Data against Growing Privacy Risks

Government agencies and research organizations are utilizing increasingly greater quantities of personal information about individuals over progressively longer periods of time. Powerful analytical capabilities, including emerging machine learning techniques, are enabling the mining of large-scale data sets to infer new insights about human characteristics and behaviors and driving demand for large-scale data sets for scientific inquiry, public policy, and innovation. These factors are also creating heightened risks to individual privacy.

A number of measures have been developed for sharing sensitive data while protecting the privacy of individuals. These interventions encompass a wide range of legal, procedural, and technical controls, from providing access to only trusted researchers, using data enclaves, and imposing restrictions as part of data use agreements, among others. One category of controls is a collection of *statistical disclosure limitation (SDL)* techniques, which are widely adopted by statistical agencies, research organizations, and data analysts to analyze and share data containing privacy-sensitive information with the aim of preventing users of the data from learning personal information pertaining to an individual. Statistical disclosure limitation encompasses a wide range of methods for suppressing, aggregating, perturbing, swapping, and generalizing attributes of individuals in the data.[1] SDL techniques are often applied with the explicit goal of *de-identification* (i.e., redacting or coarsening data with the goal of increasing the difficulty of linking an identified person to a record in a data release).[2]

Differential privacy is motivated by an ever-growing number of real-world examples of data releases that were thought to be sufficiently protective of privacy but were later shown to carry significant privacy risks. Over time, changes in the way information is collected and analyzed, including advances in analytical capabilities, increases in computational power, and the expanding availability of personal data from a wide range of sources, are eroding the effectiveness of traditional SDL techniques.

For over a century,[3] statistical agencies have recognized the need to protect against uses of data that would threaten privacy, and, for most of this time, the primary focus of formal protections has been to prevent re-identification (for an overview, see Willenborg and

[1] For an overview of traditional SDL techniques, see Harris-Kojetin et al. (2005) and chapter 5 in this handbook.

[2] For an introduction to de-identification techniques, see Garfinkel (2016).

[3] See, e.g., Chapter 2 Section 25 of the Thirteenth Census Act (The Statutes at Large of the United States of America, 1909).

De Waal, 1996). Re-identification attacks gained renewed attention in the privacy research literature in the late 1990s (Sweeney, 1997) and have become increasingly sophisticated over time, along with other emerging types of attacks that seek to infer characteristics of individuals based on information about them in the data (Narayanan and Shmatikov, 2008; de Montjoye et al., 2013; Calandrino et al., 2011). In particular, successful attacks on de-identified data have shown that traditional technical measures for privacy protection may be vulnerable to attacks devised after a technique's deployment and use. Some de-identification techniques, for example, categorize attributes in the data as (quasi-)identifying (e.g., names, dates of birth, or addresses) or non-identifying (e.g., movie ratings or hospital admission dates). Data providers may later discover that attributes initially believed to be non-identifying can in fact be used to re-identify individuals. De-identification hence requires a careful analysis—not only of present data sources that could be linked with the de-identified data toward enabling re-identification but also of future data sources and other hard-to-anticipate future sources of auxiliary information that can be used for re-identification.

Moreover, there are privacy attacks beyond record linkage attacks on de-identified records. A recent example illustrating the evolving nature of privacy attacks is the reconstruction and re-identification of the 2010 Decennial Census database. This example demonstrates that even publications of statistical tables transformed using traditional statistical disclosure limitation techniques may be vulnerable to privacy attacks.[4]

> In a paper published in 2018, researchers revealed that the underlying confidential data from the 2010 US Decennial Census could be reconstructed using only the statistical tables published by the US Census Bureau (Garfinkel, Abowd and Martindale, 2019). Researchers demonstrated a type of attack, called a *database reconstruction attack*, that leveraged the large volumes of data from the published statistical tables in order

[4] This example is reproduced from Fluitt et al. (2019).

to narrow down the possible values of individual-level records. The researchers were able to reconstruct with perfect accuracy the sex, age, race, ethnicity, and fine-grained geographic location (to the block-level) reported by Census respondents for 46 percent of the US population (Abowd, 2019). Researchers also showed that, if they slightly relaxed their conditions and allowed age to vary by up to only one year, these five pieces of information could be reconstructed for 71 percent of the population (Abowd, 2019).

Further, the researchers showed that the reconstructed records could be completely *re-identified*. They were able to assign personally identifiable information to individual records using commercial databases that were available in 2010 (Abowd, 2019). They concluded that, with this attack, they could putatively re-identify 138 million people, and they confirmed that these re-identifications were accurate for 52 million people, or 17 percent of the US population (Abowd, 2019).

These findings are startling. In 2012, the last time the Census Bureau performed a simulated re-identification attack on census data sets, the re-identification rate was only 0.0038 percent (Ramachandran et al., 2012). The test attack using the data published for the 2010 Decennial Census demonstrates that previous risk assessments underestimated the re-identification risk by a factor of at least 4,500 (Ramachandran et al., 2012).

The demonstration of a database reconstruction attack on the statistical tables published by the Census Bureau is just the latest in a long line of attacks illustrating the privacy risks associated with releasing and analyzing large volumes of data about individuals. In particular, it is a real-world manifestation of the growing risks from combining and analyzing multiple statistical releases—broadly referred to as risks from *composition* (Ganta, Kasiviswanathan and Smith, 2008; Fluitt et al., 2019). The modern mathematical understanding recognizes that any research output increases disclosure risk.[5] Although some increases in disclosure risk may be small, they accumulate, potentially to the point of a severe privacy breach. Taken together, the outputs may enable an accurate reconstruction of large portions of the data set, as seen in the reconstruction and re-identification of the 2010 Decennial Census database.

Producing accurate statistics while protecting privacy and addressing risks from composition is a challenging problem (Dwork et al., 2016). It is a fundamental law of information that privacy risk grows with the repeated use of data, and this applies to any disclosure limitation technique. Traditional SDL techniques—such as suppression, aggregation, and generalization—often reduce accuracy and are vulnerable to privacy loss due to composition.[6] A rigorous analysis of the effect of composition is important for establishing a robust and realistic understanding of how multiple statistical computations affect privacy.

Privacy attacks such as these have underscored the need for privacy technologies that are immune not only to linkage attacks but to any potential attack, *including attacks that are currently unknown or unforeseen*. It is now understood that risks remain even if many pieces of information are removed from a data set prior to release. Extensive external information may be available to potential attackers, such as employers, insurance companies, relatives, and friends of an individual in the data. In addition, ex post remedies, such as simply "taking the data back" when a vulnerability is discovered, are ineffective because many copies of a set of data typically exist; copies may even

[5] Note that the fact that small risks can combine dramatically is a key insight essential to differential privacy. Differential privacy provides a quantification of privacy risk, and provable guarantees with respect to the cumulative risk from successive data releases. Some risk assessment frameworks, such as the Five Safes framework as originally proposed, make an assumption that "many research outputs pose no disclosure risk because of their functional form" (Desai, Ritchie and Welpton, 2016, pg. 13). Traditional disclosure avoidance methods do not provide ways to quantify the accumulation of privacy risk from multiple uses and releases of data.

[6] See Ganta, Kasiviswanathan and Smith (2008). The impression that these techniques do not suffer accumulated degradation in privacy is merely due to the fact that these techniques have not been analyzed with the high degree of rigor that has been applied to differential privacy. For a discussion of privacy and utility with respect to traditional statistical disclosure limitation techniques, see Chen et al. (2009).

persist online indefinitely.[7]

6.1.3 Features of the Differential Privacy Guarantee

Differential privacy is a strong definition (or, in other words, a standard) of privacy in the context of statistical analysis and machine learning, protecting against the threats described above, including those of unknown attacks and cumulative loss. Tools that achieve the differential privacy standard can be used to provide broad, public access to data or data summaries in a privacy-preserving way. Used appropriately, these tools can, in some cases, also enable access to data that could not otherwise be shared due to privacy concerns and do so with a guarantee of privacy protection that substantially increases the ability of the institution to protect the individuals in the data.

With differential privacy, statements about risk are proved mathematically—rather than supported heuristically or empirically. The definition of differential privacy also has a compelling intuitive interpretation: inferring information specific to an individual from the outcome of an analysis preserving differential privacy is impossible, including whether the individual's information was used at all.

Differential Privacy Is a Standard, Not a Single Tool

Differential privacy is a standard which many tools for analyzing sensitive personal information have been devised to satisfy. Any analysis meeting the standard provably protects its data against a wide range of *privacy attacks*, i.e., attempts to learn private information specific to individuals from a data release.[8]

[7] As an example, in 2006 AOL published anonymized search histories of 650,000 users over a period of three months. Shortly after the release, the New York Times identified a person in the release and AOL removed the data from their site. However, in spite of the withdrawal by AOL, copies of the data are still accessible on the Internet today.

[8] The authors distinguish protection against *privacy attacks*, which involves the attacker making use of the intended "advertised" functionality of a data access mechanism, from protection against *security attacks*, which involves an attacker attempt-

Differential Privacy Is Designed for Analysis of Populations, Not Individuals

Differentially private analyses can be deployed in settings in which an analyst seeks to learn about a population. For example, when statistical estimates (such as counts, averages, histograms, contingency tables, regression coefficients, and synthetic data) are computed based on personal information, the privacy of the individuals in the data needs to be protected.

The Differential Privacy Guarantee

It is mathematically guaranteed that the recipient of a data release generated by a differentially private analysis will make essentially the same inferences about any single individual's private information, whether or not that individual's private information is included in the input to the analysis.

The differential privacy guarantee can be understood in reference to other privacy concepts, such as opt-out and protection of personally identifiable information (PII):

- Differential privacy protects an individual's information essentially as if their data were not used in the analysis at all (i.e., as though the individual opted out and the information was not used).
- Differential privacy ensures that using an individual's data will not reveal essentially any PII that is specific to them. Here, *specific* refers to information that cannot be inferred about an individual unless their information is used in the analysis. Information specific to an individual would be considered PII under a variety of interpretations.[9]

ing to exploit unintended implementation vulnerabilities (e.g., by circumventing access control mechanisms). Differential privacy does not generally provide protection against security attacks, which should be addressed using complementary controls like encryption and access control.

[9] For an example of an analysis of this relationship with respect to the Family Educational Rights and Privacy Act's (FERPA) definition of PII, see Nissim et al. (2018).

Differentially Private Analysis Requires the Introduction of Statistical Noise

To achieve differential privacy, carefully crafted random statistical noise must be injected into statistical and machine-learning analyses.[10]

Protecting Privacy Increases the Uncertainty of Results

The introduction of statistical noise to protect privacy necessarily reduces the accuracy of statistical analyses. As the number n of observations in a data set grows sufficiently large, the loss in accuracy due to differential privacy can become much smaller than other sources of error such as statistical sampling error. However, maintaining high accuracy for studies on small or modest-sized data sets (or modest-sized subsets of large data sets) is a challenge. As a consequence, all results computed using tools for differentially private analysis will be approximate. Conversely, any system that produces exact results without any random modifications cannot meet the differential privacy standard.

Preventing Cumulative Privacy Failure Requires a Budget for Privacy Loss, Which in Turn Limits Utility

Every computation leaks some information about the individual records used as input regardless of the protection method used. To prevent cumulative privacy failure, the privacy loss that accumulates over multiple computations must be calculated, tracked, and limited. Differential privacy provides explicit, formal methods for defining and managing this cumulative loss, referred to as the *privacy-loss budget*.

The inevitability of privacy loss implies that there is an inherent tradeoff between privacy and utility as the former degrades with an increase of the latter. Formal frameworks for statistical disclosure limitation

[10] The choice of noise addition technique—whether statistical noise is used to blur individual data points, the output of a computation, or intermediate computations—is a delicate algorithmic question; a variety of noise addition techniques have been developed for differentially private analysis with the purpose of guaranteeing differential privacy while minimizing the overall inaccuracy introduced.

(such as differential privacy) are distinct from traditional, less formal approaches in that formal frameworks quantify this trade-off explicitly: what can be learned about an individual as a result of their private information being included in a differentially private analysis is strictly limited and quantified by a privacy loss parameter, usually denoted *epsilon* (ε). Further, many tools for differentially private analysis are designed to make efficient trade-offs between privacy and utility.

6.1.4 An Illustrative Scenario: Publishing Education Statistics

The scenarios in this section illustrate the types of information disclosures that are addressed when using differential privacy.

Alice and Bob are professors at Private University. They both have access to a database that contains personal information about students at the university, including information related to the financial aid each student receives. To gain access, Alice and Bob were required to undergo confidentiality training and to sign data use agreements restricting the disclosure of personal information obtained from the database.

In March, Alice publishes an article based on the information in this database and writes that "the current freshman class at Private University is made up of 3,005 students, 202 of whom are from families earning over US$350,000 per year." Alice reasons that no individual's personal information will be exposed because she published an aggregate statistic taken over 3,005 people. The following month, Bob publishes a separate article containing these statistics: "201 families in Private University's freshman class of 3,004 have household incomes exceeding US$350,000 per year." Neither Alice nor Bob is aware that they have both published similar information.

A clever student Eve reads both of these articles and makes an observation. From the published information, Eve concludes that between March and April one freshman withdrew from Private University and that the student's parents earn over US$350,000 per year. Eve asks around and is able to determine that a student named John dropped out around the end of March. Eve then informs her classmates that John's parents probably earn over US$350,000 per year.

John hears about this and is upset that his former classmates learned about his parents' financial status. He complains to the university and Alice and Bob are asked to explain. In their defense, both Alice and Bob argue that they published only information that had been aggregated over a large population and does not identify any individuals.

This story illustrates how the results of multiple analyses using information about the same people, when studied in combination, may enable one to draw conclusions about individuals in the data. Alice and Bob may each publish information that seems innocuous in isolation. However, when combined, the information they publish can compromise the privacy of one or more individuals. This type of privacy breach is generally difficult to prevent by Alice and Bob individually, as it is likely that neither knows what information has already been revealed or will be revealed by others in future. This problem is referred to as the problem of *composition*.

Suppose, instead, that the institutional review board at Private University only allows researchers to access student records by submitting queries to a special data portal, which responds to every query with an answer produced by running a differentially private computation on the student records.

In March, Alice queries the data portal for the number of freshmen who come from families with a household income exceeding US$350,000. The portal returns the noisy count of 204, leading Alice to write in her article that “the current freshman class at Private University is made up of 3,005 students, approximately 205 of whom are from families earning over US$350,000 per year.” In April, Bob asks the same question and gets the noisy count of 199 students. Bob publishes in his article that “approximately 200 families in Private University’s freshman class of 3,004 have household incomes exceeding US$350,000 per year.” The publication of these noisy figures prevents Eve from concluding that one student with a household income greater than US$350,000 withdrew from the university in March. The risk that John’s personal information could be uncovered based on these publications is thereby reduced.

This example hints at one of the most important properties of differential privacy: it is robust under composition. If multiple differentially private analyses are performed on data describing the same set of individuals, then the guarantee is that all of the information released will still be differentially private. Notice how this scenario is markedly different from the previous hypothetical in which Alice and Bob do not use differentially private analyses and inadvertently release two statistics that in combination lead to the full disclosure of John’s personal information. The use of differential privacy rules out the possibility of such a complete breach of privacy. This is because differential privacy enables one to measure and bound the cumulative privacy risk from multiple analyses of information about the same individuals.

However, *every* analysis, regardless of whether it is differentially private, results in *some* leakage of information about the individuals whose data are being analyzed, and this leakage accumulates with each analysis. This is true for every release of data, including releases of aggregate statistics. In particular, the example above should not be understood to imply that privacy does not degrade after multiple differentially private computations. In fact, as indicated in Section

6.2.4, privacy risks accumulate with each release or analysis involving an individual's data. For this reason, there is a limit to how many analyses can be performed on a specific data set while providing an acceptable guarantee of privacy. Therefore, measuring privacy loss and understanding quantitatively how risk accumulates across successive analyses are critical. In the context of the example above, measures need to be established, such as restricting the overall number of queries to which researchers may apply to Private University's database.

6.1.5 What Types of Analyses are Performed Using Differential Privacy

Differentially private algorithms are known to exist for a wide range of statistical analyses, such as count queries, histograms, cumulative distribution functions, and linear regression; techniques used in statistics and machine learning, such as clustering and classification; and statistical disclosure limitation techniques, like synthetic data generation, among many others.

Count Queries Differentially private answers to count queries (i.e., estimates of the number of individual records in the data satisfying a specific condition) can be obtained through the addition of random noise (Dwork et al., 2016).

Histograms Differentially private computations can provide noisy counts for data points classified into the disjoint categories represented in histograms or contingency tables (i.e., cross-tabulations) (Dwork et al., 2016).

Cumulative Distribution Function (CDF) There are differentially private algorithms for estimating the entire CDF of a dataset (or the distribution from which it is drawn) (Bun et al., 2015). These algorithms introduce noise that needs to be taken into account when statistics such as median or interquartile range are computed from the estimated CDF.[11]

Linear Regression Differentially private algorithms for linear regression introduce noise in a variety of different ways, and the choice of which algorithm is best will depend on properties of the underlying data distribution (e.g., the amount of variance in the explanatory variables), the sample size, the privacy parameters, and the intended application (Wang, 2018; Alabi et al., 2020).

Clustering Researchers are developing a variety of differentially private clustering algorithms (i.e., algorithms for grouping data points into clusters so that points in the same cluster are more similar to each other than to points in other clusters) (Stemmer and Kaplan, 2018), and such tools are likely to be included in future privacy-preserving tool kits for exploratory analysis by social scientists.

Classification and Machine Learning Theoretical work has shown it is possible to construct differentially private algorithms for a large collection of classification tasks, such as identifying or predicting to which set of categories a data point belongs based on a training set of examples for which category membership is known (Blum et al., 2005; Kasiviswanathan et al., 2011), and subsequent work has developed more practical methods for differentially private machine learning, including deep learning (Abadi et al., 2016).

Synthetic Data Generation Research has shown that in principle it is possible to generate differentially private synthetic data that preserves a vast collection of statistical properties of the original data set.[12] A significant benefit is that once a differentially private synthetic data set is

[11] For data over an ordered domain, a cumulative distribution function depicts for every value x an estimate of the number of data points with a value up to x. For a more in-depth discussion of differential privacy and CDFs, see Muise and Nissim (2016).

[12] See, for example, Blum, Ligett and Roth (2013). Synthetic data are data sets generated from a statistical model estimated using the original data. The records in a synthetic data set have no one-to-one correspondence with the individuals in the original data set, yet the synthetic data can retain many of the statistical properties of

generated, it can be analyzed any number of times, without any further implications for privacy. As a result, synthetic data can be shared freely or even made public in many cases. For example, statistical agencies can release synthetic microdata as public-use data files in place of raw microdata. However, significant challenges remain with respect to both the level of random noise introduced and computational efficiency for general-purpose differentially private synthetic generation in practice, particularly for high-dimensional data.[13]

6.2 How Differential Privacy Protects Privacy

6.2.1 What Does Differential Privacy Protect?

Intuitively, a computation protects the privacy of individuals in the data if the computational output does not reveal any information that is specific to any individual subject. Differential privacy formalizes this intuition as a *mathematical definition*. Similar to showing that an integer is even by proving that it is the result of multiplying some integer by two, a computation is shown to be differentially private by proving it meets the constraints of the definition. In turn, if a computation can be proven to be differentially private, one can rest assured that using the computation will not unduly reveal information specific to a data subject.

To see how differential privacy formalizes this privacy requirement as a definition, consider the following scenario.

the original data. Synthetic data resemble the original sensitive data in format and, for a large class of analyses, results are similar whether performed on the synthetic or original data.

[13] Intuitively, preserving more statistical information (e.g., all entries of a high-dimensional variance-covariance matrix) requires spreading the privacy-loss budget more thinly and thus introducing greater noise. There are much more complex methods that can detect and exploit relationships between the statistics to introduce less noise, but those methods can be computationally infeasible on high-dimensional data.

Researchers have selected a sample of individuals across the US to participate in a survey exploring the relationship between socioeconomic status and health outcomes. The participants were asked to complete a questionnaire covering topics such as where they live, their finances, and their medical history.

One of the participants, John, is aware that individuals have been re-identified in previous releases of de-identified data and is concerned that personal information he provides about himself, such as his medical history or annual income, could one day be revealed in de-identified data released from this study. If leaked, this information could lead to an increase in his life insurance premium or an adverse decision for a future mortgage application.

Differential privacy can be used to address John's concerns. If the researchers only share data resulting from a differentially private computation, John is guaranteed that the release will not disclose anything that is *specific to him* even though he participated in the study.

To understand what this means, consider a thought experiment, which is illustrated in Figure 6.1 and is referred to as *John's opt-out scenario*.[14] In John's opt-out scenario, an analysis is performed using data about the individuals in the study, except that information about John is omitted. His privacy is protected in the sense that the outcome of the analysis *does not depend on his specific information,* because it was not used in the analysis at all.

John's opt-out scenario differs from the scenario depicted in Figure 6.2, referred to as the *real-world* scenario, in which the analysis is based on John's personal information along with the personal information of the other study participants. The real-world scenario involves some potential risk to John's privacy as some of his personal information could be revealed by the outcome of the analysis, because it was used as input to the computation.

[14] Figure 6.1 is reproduced from Wood et al. (2018).

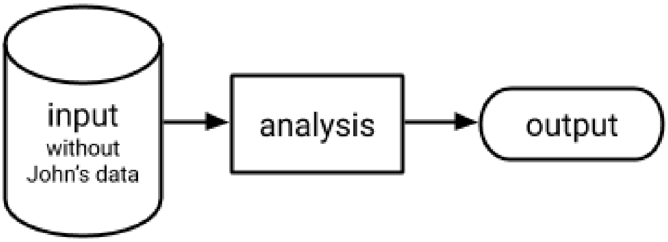

Figure 6.1: John's opt-out scenario

Differential privacy aims to protect John's privacy in the real-world scenario in a way that mimics the privacy protection he is afforded in his opt-out scenario.[15] Accordingly, what can be learned about John from a differentially private computation is (essentially) limited to what could be learned about him from everyone else's data *without his own data being included in the computation*. Crucially, this same guarantee is made not only with respect to John but also with respect to every other individual contributing their information to the analysis.

For a precise description of differential privacy and the mathematics underlying the construction of differentially private analysis, the reader is referred to the literature listed in Appendix C. In lieu of the mathematical definition, this chapter offers a few illustrative examples to discuss various aspects of differential privacy in a way that is intuitive and generally accessible.

6.2.2 Privacy Protection Is a Property of an Analysis—Not a Data Release

Throughout this chapter, we refer to the general concept of an *analysis* that performs a computation on input data and outputs the result (illustrated in Figure 6.2).[16] The analysis may be as simple as deter-

[15] The use of differentially private analysis is *not* equivalent to the traditional use of opting out. On the privacy side, differential privacy does not require an explicit opt-out. In comparison, traditional use of opt-out requires an explicit choice that may cause privacy harms by calling attention to individuals that choose to opt out. On the utility side, there is no general expectation that using differential privacy would yield the same outcomes as adopting the policy of opt-out.

[16] Figure 6.2 is reproduced from Wood et al. (2018).

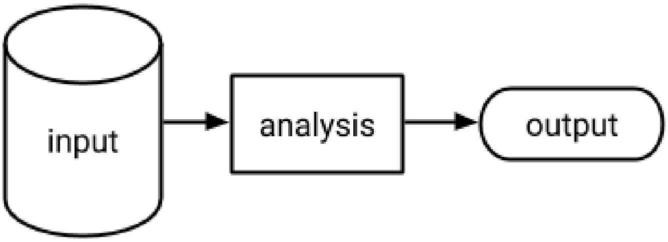

Figure 6.2: An analysis (or computation) transforms input data into some output.

mining the average age of the individuals in the data, or it may be more complex and utilize sophisticated modeling and inference techniques.

We focus specifically on analyses that transform sensitive personal data into an output that can be released publicly. For example, an analysis may involve the application of techniques for aggregating or de-identifying a set of personal data in order to produce a sanitized version of the data that is safe to release. How can the data provider ensure that publishing the output of this computation will not unintentionally leak information from the privacy-sensitive input data?

A key insight from the theoretical computer science literature is that *privacy is a property of the informational relationship between the input and output*, not a property of the output alone.[17] In other words, one can be certain that the output of a computation is privacy-preserving if the computation itself is privacy-preserving. The following examples show why this is the case.

Consider the following statistic: a representative ninth-grade GPA at City High School is 3.5. One might naturally think that this statistic is unlikely to reveal private information about an individual student. However, one needs to know *how* the statistic was computed to make that determination. For instance, if the representative ninth-grade GPA was calculated by taking the GPA of the alphabetically first

[17] This insight follows from a series of papers demonstrating privacy breaches enabled by leakages of information resulting from decisions made by the computation. See, for example, Kenthapadi, Mishra and Nissim (2013). For a general discussion of the advantages of formal privacy models over ad hoc privacy techniques, see Narayanan, Huey and Felten (2016).

student in the school, then the statistic completely reveals the GPA of that student.[18] Alternatively, a representative statistic could be based on average features of the ninth graders in the school—using the most common first name, the most common last name, the average age, and the average GPA to produce "John Smith, a fourteen-year-old in the ninth grade, has a 3.1 GPA." Suppose that coincidentally a student named John Smith subsequently joins the ninth-grade class. Although his name appears in the published statistic, one knows with certainty that the statistic does not reveal private information about him, because it was not based on his student records in any way.

These examples are clearly contrived, and no reasonable analyst would publish either statistic. On a fundamental level, however, the examples demonstrate that when trying to decide whether a data release can be made public, one needs to consider the computation used to produce that release and not the release by itself. Thus, when thinking about privacy in the context of statistical releases, one should think about it as a computational property, especially if the goal is to make rigorous, formal claims about the data. This is one of the properties of differential privacy. If a computation can be proven to be differentially private, the researcher can rest assured that using the computation will not unduly reveal information specific to a data subject. Adopting this formal approach to privacy yields several practical benefits for users, including robustness to auxiliary information, composition, and post-processing, as well as transparency—each discussed in turn below in Section 6.2.3.

6.2.3 Methodology Example: Limiting Privacy Loss from Participation in Research

In the earlier example featuring Professors Alice and Bob at Private University, differentially private analyses add random noise to the

[18] One might object that the student's GPA is not traceable back to that student unless an observer knows how the statistic was produced. However, a basic principle of modern cryptography (known as Kerckhoffs' principle) is that a system is not secure if its security depends on its inner workings being a secret. In this context, it is assumed that the algorithm behind a statistical analysis is public (or could potentially become public).

statistics they produce.[19] This noise masks the differences between the real-world computation and the opt-out scenario of each individual in the data set. This means that the outcome of a differentially private analysis is not exact but an *approximation*. In addition, a differentially private analysis may return different results, even if performed twice on the same data set. Because researchers intentionally add random noise, analyses performed with differential privacy differ from standard statistical analyses, such as the calculation of averages, medians, and linear regression equations.

> Consider a differentially private analysis that computes the number of students in a sample with a GPA of at least 3.0. Say that there are 10,000 students in the sample, and exactly 5,603 of them have a GPA of at least 3.0. An analysis that added no random noise would hence report that 5,603 students had a GPA of at least 3.0.
>
> However, a differentially private analysis adds random noise to protect the privacy of the data subjects. For instance, a differentially private analysis might report an answer of 5,521 students when run on the data; when run a second time on the same information, it might report an answer of 5,586 students.
>
> In a differentially private analysis, the added noise makes every potential answer almost as likely whether John's data are used in the analysis or not. This is done by controlling the likelihood ratio of any answer with John's data included or excluded.

A differentially private analysis might produce many different answers given the same data set. Because the details of a method providing differential privacy can be made public, an analyst may be able to calculate accuracy bounds that show how much an output of the analysis is expected to differ from the noiseless answer.

[19] In other differentially private computations noise may be added to intermediate results of a computation or at the data collection process. The latter is referred to as the *local model* of differential privacy.

An essential component of a differentially private computation is the privacy loss parameter, usually denoted by the Greek letter ε (epsilon). This parameter determines how much noise is added to the computation. Choosing a value for the privacy loss parameter can be thought of as a tuning knob for balancing privacy and accuracy. A lower value for ε corresponds to stronger privacy protection and also a larger decrease in accuracy, whereas a higher value for εcorresponds to weaker privacy protection and also a smaller decrease in accuracy. The following discussion establishes an intuition for this parameter. It can be thought of as limiting how much a differentially private computation is allowed to deviate from the opt-out scenario of an individual in the data.

Consider the opt-out scenario for a certain computation, such as estimating the number of HIV-positive individuals in a surveyed population. Ideally, this estimate should remain exactly the same whether or not a single individual, such as John, is included in the survey. However, ensuring this property *exactly* would require the total exclusion of John's information from the analysis. It would also require the exclusion of Gertrude's and Peter's information in order to provide privacy protection for them. Continuing with this line of argument, one comes to the conclusion that the personal information of *every* surveyed individual must be excluded in order to satisfy that individual's opt-out scenario. Thus, the analysis cannot rely on any person's information and is completely useless.

To avoid this dilemma, differential privacy requires only that the output of the analysis remain *approximately* the same whether John participates in the survey or not. Differential privacy allows for a deviation between the output of the real-world analysis and that of each individual's opt-out scenario. The privacy loss parameter ε quantifies and limits the extent of the deviation between the opt-out and real-world scenarios, as shown in Figure 6.3 below.[20] The parameter ε measures the effect of each individual's information on the output of the analysis. It can also be viewed as a measure of the additional privacy risk an individual could incur beyond the risk incurred in the opt-out scenario.[21]

[20] Figure 6.3 is reproduced from Wood et al. (2018).

[21] ε is a unitless nonnegative quantity measuring probability log-ratio.

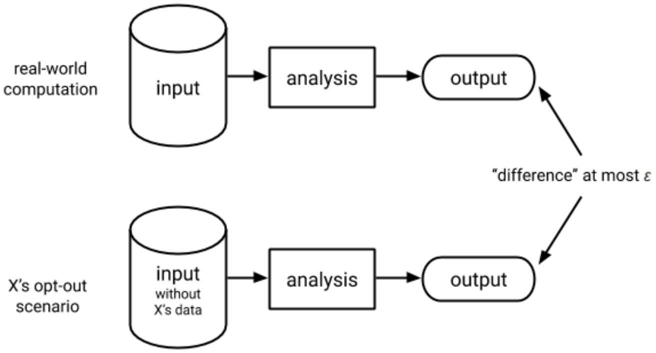

Figure 6.3: Differential privacy. The maximum deviation between the opt-out scenario and real-world computation should hold simultaneously for each individual X whose information is included in the input.

Note that in Figure 6.3 John has been replaced with an arbitrary individual X to emphasize that the differential privacy guarantee is made simultaneously to *all* individuals in the sample, not just John.

Choosing a value for ε can be thought of as tuning the level of privacy protection required. This choice also affects the utility or accuracy that can be obtained from the analysis. A smaller value of ε results in a smaller deviation between the real-world analysis and each opt-out scenario and is therefore associated with stronger privacy protection but less accuracy. For example, when ε is set to zero, the real-world differentially private analysis mimics the opt-out scenario of each individual perfectly. However, as argued at the beginning of this section, an analysis that perfectly mimics the opt-out scenario of each individual would require ignoring all information from the input and accordingly could not provide any meaningful output. Yet when ε is set to a small number, such as 0.1, the deviation between the real-world computation and each individual's opt-out scenario will be small, providing strong privacy protection while also enabling an analyst to derive useful statistics based on the data.

Simple conventions for choosing ε have not yet been developed; the current best practice for choosing ε is to explore the trade-off between the choice of ε and the utility provided by an analysis for every application, as well as to consider the potential risks to individuals and the level of risk the data owner might be permitting given their legal, contractual, and ethical obligations. It is expected that as the use of differentially private analyses in real-life applications increases, the accumulated experience will shed light on how to reach a reasonable compromise between privacy and accuracy. As a rule of thumb, however, ε should be thought of as a small number, between approximately $1/100$ and 1.[22]

This chapter has discussed how the privacy loss parameter limits the deviation between the real-world computation and each data subject's opt-out scenario. However, it might not be clear how this abstract guarantee relates to privacy concerns in the real world. Therefore, in this section, a practical interpretation of the privacy loss parameter is discussed as a bound on the financial risk incurred by participating in a study.

Any useful analysis carries the risk that it will reveal information about individuals (which in turn might result in a financial cost). The following example shows that while differential privacy necessarily cannot eliminate this risk, it can guarantee that the risk will be limited by quantitative bounds that depend on ε.

Gertrude, a 65-year-old woman, is considering whether to participate in a medical research study. While she can envision many potential personal and societal benefits resulting from her participation in the study, she is concerned that the personal information she discloses over the course of the study could lead to an increase in her life insurance premium.

For example, Gertrude is apprehensive that the tests she would undergo as part of the research study would reveal that she is

[22] In general, setting ϵ involves making a compromise between privacy protection and accuracy. The consideration of both utility and privacy is challenging in practice and, in some of the early implementations of differential privacy, has led to choosing a higher value for ϵ. As the accuracy of differentially private analyses improves over time, it is likely that lower values of ϵ will be chosen.

predisposed to suffer a stroke and is significantly more likely to die in the coming year than the average person of her age and gender. If such information related to Gertrude's increased risk of morbidity and mortality is discovered by her life insurance company, it will likely increase her premium substantially.

Before she decides to participate in the study, Gertrude wishes to be assured that privacy measures are in place to ensure that her involvement will have a limited effect (if any) on her life insurance premium.

Gertrude's life insurance company may raise her premium based on something it learns from the medical research study, even if Gertrude does not herself participate in the study. The following example is provided to illustrate such a scenario.[23]

Gertrude holds a US$100,000 life insurance policy. Her life insurance company has set her annual premium at US$1,000 (i.e., 1 percent of US$100,000) based on actuarial tables that show that someone of Gertrude's age and gender has a 1 percent chance of dying in the next year.

Suppose Gertrude opts out of participating in the medical research study. Regardless, the study reveals that coffee drinkers are more likely to suffer a stroke than non-coffee drinkers. Gertrude's life insurance company may update its assessment and conclude that as a 65-year-old woman who drinks coffee, Gertrude has a 2 percent chance of dying in the next year. The insurance company decides to increase Gertrude's annual premium from US$1,000 to US$2,000 based on the findings of the study.

In this hypothetical example, the results of the study led to an increase

[23] Figures in this example are based on data from US Social Security Administration (2011).

in Gertrude's life insurance premium, even though she did not contribute any personal information to the study. A potential increase of this nature is likely unavoidable to Gertrude because she cannot prevent other people from participating in the study. This type of effect is taken into account by Gertrude's insurance premium in her *opt-out scenario* and will not be protected against by differential privacy.

Next, consider the increase in risk that is due to Gertrude's participation in the study.

Suppose Gertrude decides to participate in the research study. Based on the results of medical tests performed on Gertrude over the course of the study, the researchers conclude that Gertrude has a 50 percent chance of dying from a stroke in the next year. If the data from the study were to be made available to Gertrude's insurance company, it might decide to increase her insurance premium from US$2,000 to more than US$50,000 in light of this discovery.

Fortunately for Gertrude, this does not happen. Rather than releasing the full data set from the study, the researchers release only a differentially private summary of the data they collected. Differential privacy guarantees that if the researchers use a value of $\varepsilon = 0.01$, then the insurance company's estimate of the probability that Gertrude will die in the next year can increase from 2 percent to at most 2.04 percent, as per the equation:

$$2\% \cdot (1 + 2 \cdot \varepsilon) = 2\% \cdot (1 + 2 \cdot 0.01) = 2.04\%^{a}$$

Thus, Gertrude's insurance premium can increase from US$2,000 to US$2,040, at most. Gertrude's first-year cost of participating in the research study in terms of a potential increase in her insurance premium is at most US$40.

Note that this analysis *does not* imply that the insurance company's estimate of the probability that Gertrude will die in the next year must increase as a result of her participation in the study, nor that if the estimate increases it must increase to 2.04 percent. What the analysis shows is that if the estimate were to increase, it would not exceed 2.04 percent.

Consequently, this analysis *does not* imply that Gertrude would incur an increase in her insurance premium or that if she were to see such an increase it would cost her US$40. What is guaranteed is that if Gertrude should see an increase in her premium, this increase would not exceed US$40.

[a] The approximate calculation provided in this example only holds for small ε, using $e^{2\cdot\varepsilon} \approx 1 + 2 \cdot \varepsilon$. See Table 6.1 for an exact formula.

Gertrude may decide that the potential cost of participating in the research study, US$40, is too high and she cannot afford to participate with this value of ε and this level of risk. Alternatively, she may decide that it is worthwhile. Perhaps she is paid more than US$40 to participate in the study or the information she learns from the study is worth more than US$40 to her. The key point is that differential privacy allows Gertrude to make a more informed decision based on the worst-case cost of her participation in the study.

It is worth noting that should Gertrude decide to participate in the study, her risk might increase even if her insurance company is not aware of her participation. For instance, the study might determine that Gertrude has a very high chance of dying next year, and that could affect the study results. In turn, her insurance company might decide to raise her premium, because she fits the profile of the studied population (even if the company does not believe her data were included in the study). On the other hand, differential privacy guarantees that even if the insurance company knows that Gertrude *did* participate in the study, it can essentially only make inferences about her that it could have made if she had not participated in the study.

One can generalize from Gertrude's scenario and view differential privacy as a framework for reasoning about the increased risk that is incurred when an individual's information is included in a data analysis. Differential privacy guarantees that an individual will be exposed to essentially the same privacy risk regardless of whether their data are included in a differentially private analysis. In this context, think of the privacy risk associated with a data release as the potential harm that an individual might experience due to a belief that an observer forms based on that data release.

In particular, when ε is set to a small value, the probability that an observer will make some inference that is harmful to a data subject based on a differentially private data release is no greater than $1+\varepsilon$ times the probability that the observer would have made that inference without the data subject's inclusion in the data set.[24] For example, if ε is set to 0.01, then the probability of any adverse event to an individual (such as Gertrude being denied insurance) can grow by a multiplicative factor of 1.01 (at most) as a result from participation in a differentially private computation (compared with not participating in the computation).

As shown in the Gertrude scenario, there is also the risk to Gertrude that the insurance company will see the study results, update its beliefs about the mortality of Gertrude, and charge her a higher premium. If the insurance company infers from the study results that Gertrude has probability p of dying in the next year, and her insurance policy is valued at US$ 100,000, this translates into a risk (in financial terms) of a higher premium of $p\times$ US$ 100,000. This risk exists even if Gertrude does not participate in the study. Recall how in the first hypothetical, the insurance company's belief that Gertrude will die in the next year doubles from 1 percent to 2 percent, increasing her premium from US$1,000 to US$2,000, based on general information learned from the individuals who did participate. Also, if Gertrude does decide to participate in the study (as in the second hypothetical), differential privacy limits the change in this risk relative to her opt-out scenario. In financial terms, her risk increases by US$40 at most, since it can be shown that the insurance company's beliefs about her probability of death change from 2 percent to no greater than $2\% \cdot (1+2\cdot\varepsilon) = 2.04\%$,

[24] In general, the guarantee made by differential privacy is that the probabilities differ at most by a factor of $e^{\pm\varepsilon}$, which is approximately $1 \pm \varepsilon$ when ε is small.

when $\varepsilon = 0.01$.[25]

Note that the above calculation requires certain information that may be difficult to determine in the real world. In particular, the 2 percent baseline in Gertrude's opt-out scenario (i.e., Gertrude's insurer's belief about her chance of dying in the next year) is dependent on the results from the medical research study, which Gertrude does not know at the time she makes her decision whether to participate. Fortunately, differential privacy provides guarantees relative to every baseline risk.

> Without her participation, the study results would lead the insurance company to believe that Gertrude has a 3 percent chance of dying in the next year (instead of the 2 percent chance hypothesized earlier). This means that Gertrude's insurance premium would increase to US$3,000. Differential privacy guarantees that if Gertrude had instead decided to participate in the study, the insurer's estimate for Gertrude's mortality would have been at most $3\% \cdot (1 + 2 \cdot \varepsilon) = 3.06\%$ (assuming an ε of 0.01), which means that her premium would not increase beyond $3,060.

Calculations like those used in the analysis of Gertrude's privacy risk can be performed by referring to Table 6.1.[26] For example, the value of ε used in the research study in which Gertrude considered participating was 0.01, and the baseline privacy risk in her opt-out scenario was 2 percent. As shown in Table 6.1, these values correspond to a worst-case privacy risk of 2.04 percent in her real-world scenario. Notice also how the calculation of risk would change with different values. For example, if the privacy risk in Gertrude's opt-out scenario were 5 percent rather than 2 percent and the value of epsilon remained the same, then the worst-case privacy risk in her real-world scenario would be 5 percent.

[25] The reason that the multiplicative factor is $1 + 2 \cdot \varepsilon \approx e^{2 \cdot \varepsilon}$ rather than $1 + \varepsilon \approx e^{\varepsilon}$ is that posterior beliefs can be expressed as a ratio of two probabilities, each of which can change by a factor of at most e^{ε}. The factor of 2 was incorrectly omitted in the original paper (Wood et al., 2018) describing this example.

[26] Table 6.1 corrects a calculation error appearing in the original paper (Wood et al., 2018).

Table 6.1: Maximal change between posterior beliefs in Gertrude's opt-out and real-world scenarios. The notation $A(x')$ refers to the application of the analysis A on the dataset x', which does not include Gertrude's information. As this table shows, the use of differential privacy provides a quantitative bound on how much one can learn about an individual from a computation. The entries in the table are calculated using the formula $q = \min(e^{2\varepsilon}q', 100 - e^{-2\varepsilon}(100 - q'))$, where q' is the posterior belief given $A(x')$ and q is the upper bound on the posterior belief given $A(x)$, both expressed as percentages.

posterior belief given $A(x')$ in %	value of ε					
	0.01	0.05	0.1	0.2	0.5	1
0	0	0	0	0	0	0
1	1.02	1.11	1.22	1.49	2.72	7.39
2	2.04	2.21	2.44	2.98	5.44	14.78
3	3.06	3.32	3.66	4.48	8.15	22.17
5	5.10	5.53	6.11	7.46	13.59	36.95
10	10.20	11.05	12.21	14.92	27.18	73.89
25	25.51	27.63	30.54	37.30	67.96	89.85
50	50.99	54.76	59.06	66.48	81.61	93.23
75	75.50	77.38	79.53	83.24	90.80	96.62
90	90.20	90.95	91.81	93.30	96.32	98.65
95	95.10	95.48	95.91	96.65	98.16	99.32
98	98.04	98.19	98.36	98.66	99.26	99.73
99	99.02	99.10	99.18	99.33	99.63	99.86
100	100	100	100	100	100	100
	maximum posterior belief given $A(x)$ in %					

The fact that the differential privacy guarantee applies to *every* privacy risk means that Gertrude can know for certain how participating in the study might increase her risks relative to opting out, even if she does not know *a priori* all the privacy risks posed by the data release. This enables Gertrude to make a more informed decision about whether to take part in the study. For instance, she can calculate how much additional risk she might incur by participating in the study over a range of possible baseline risk values and decide whether she is comfortable with taking on the risks entailed by these different scenarios.

6.2.4 Strengths of Differential Privacy Over Traditional SDL Approaches

This discussion outlines some of the key features of differential privacy that enable it to overcome the weaknesses of traditional approaches and provide strong protection against a wide range of privacy attacks.

Differential Privacy is Robust to Auxiliary Information

As illustrated by the re-identification attack on the 2010 Decennial Census database described in Section 6.1.2, effective privacy protection requires taking auxiliary information into account. A data provider designing a differentially private data release need not anticipate particular types of privacy attacks, such as the likelihood that one could link particular fields with other data sources that may be available. When using differential privacy, even an attacker utilizing arbitrary auxiliary information cannot learn much more about an individual in a database than they could if that individual's information were not in the database at all.

Currently, differential privacy is the only framework that provides meaningful privacy guarantees in scenarios in which adversaries have access to arbitrary external information. Releases constructed in a differentially private manner provide provable privacy protection against any feasible adversarial attack, whereas de-identification concepts only counter a limited set of specific attacks.

Differential Privacy is Robust to Composition

When evaluating privacy risk, it is important to recognize that privacy risk accumulates with each release or analysis involving an individual's data. Under what has come to be called the fundamental law of information recovery, releasing "overly accurate answers to too many questions will destroy privacy in a spectacular way" (Dinur and Nissim, 2003; Dwork et al., 2017; Dwork and Roth, 2014). This is true whether or not any privacy-preserving technique is applied and regardless of the specific privacy-preserving technique in use.[27] A reconstruction attack, such as the reconstruction of the 2010 Decennial Census database presented in Section 6.1.2, is an example of a privacy attack that leveraged composition.

One of the most powerful features of differential privacy is its robustness under composition; in other words, the combination of multiple differentially private analyses preserves differential privacy (Dwork et al., 2016; Ganta, Kasiviswanathan and Smith, 2008). Differential privacy provides provable bounds with respect to the cumulative risk from multiple data releases, and is the only existing approach to do so. Recall that the definition of differential privacy is equipped with a numeric parameter $\varepsilon > 0$ that bounds privacy risk.[28] Furthermore, one can reason about—and bound—the overall privacy risk that accumulates when multiple differentially private computations are performed on an individual's data. As a simple example, imagine that two differentially private computations are performed on data sets containing information about the same individuals. If ε_1 bounds the privacy risk of the first computation and ε_2 bounds the privacy risk of the second computation, then the cumulative privacy risk resulting from these computations is no greater than the risk associated with an aggregate parameter of $\varepsilon_1 + \varepsilon_2$. In other words, the composition of the two differentially private analyses is also a differentially private analysis with privacy risk at most $\varepsilon_1 + \varepsilon_2$. Importantly, no coordination is needed between the two mechanisms for this bound to hold.

The example above is a simple instance illustrating how analysts can bound the total disclosure risk due to multiple differentially private disclosures. Often, better bounds can be achieved via applying a set of tools known as *composition theorems*. The fact that the total dis-

[27] For further discussion see Wood et al. (2018); Altman et al. (2015).

[28] See Section 6.2.3 for further discussion of how ϵ quantifies privacy risk.

closure risk can be bounded—without having mechanisms coordinate their actions—allows for a rigorous management of privacy risks across multiple disclosures and access points. As an example, a registry such as the Epsilon Registry suggested by Dwork, Kohli and Mulligan (2019) can hold information about the value of the privacy parameter ε used in implementations of differentially private data releases and hence serve as a basis for bounding the total disclosure risk.[29]

Differential Privacy is Robust to Post-Processing

It is also important to evaluate whether an approach to privacy that is being considered can be made ineffective through post-processing, i.e., via further analyzing a data release that purports to preserve privacy. For example, Machanavajjhala and Kifer (2015) describe post-processing vulnerabilities for some algorithms that satisfy k-anonymity. The demonstration of a reconstruction attack on the 2010 Decennial Census database presented in Section 6.1.2 is an example of a privacy attack that employed post-processing: while the released data tables purportedly preserved privacy, analyzing the releases enabled the reconstruction of individual respondents' records.

Differential privacy is an example of an approach that is robust to post-processing. To understand what this means, consider a scenario in which an analyst applies a post-processing transformation B on the output of the ε-differentially private analysis A. For instance, after a data publisher adds noise to a collection of statistics using a differentially private tool, they might wish to round the statistics or replace negative statistics with zero before publishing them. In such cases, the resulting analysis $(B \circ A)$ is also ε-differentially the risk to privacy. A data publisher can even share details about the analysis A, the transformation B, and the value of ε without increasing privacy risk. Importantly, the guarantee that $(B \circ A)$ is ε-differentially private holds for any transformation B—even one that is designed with an intention to breach privacy.

[29] The proposal for an Epsilon Registry is intended to be a publicly available bulletin board where firms would disclose information about their deployment of differential privacy. *See* Dwork, Kohli and Mulligan (2019).

Differential Privacy Does Not Rely on Security by Obscurity

Differentially private tools also have the benefit of transparency, as maintaining secrecy around a differentially private computation or its parameters is not necessary. This feature distinguishes differentially private tools from traditional de-identification techniques, which often require concealment of the extent to which the data have been transformed and thereby leave data users with uncertainty regarding the accuracy of analyses on the data. This approach can enable public scrutiny of the privacy-preserving techniques used. Further, the amount of noise added by differential privacy can be taken into account in the measure of accuracy, unlike traditional techniques that keep the information needed to estimate the privacy error secret.

6.2.5 What Does Differential Privacy *Not* Protect?

The following example illustrates the types of information disclosures that differential privacy does not aim to address.

> Ellen is John's friend and knows that he regularly consumes several glasses of red wine with dinner. Ellen learns that a research study had found a positive correlation between drinking red wine and the likelihood of developing a certain type of cancer. Based on the study and her knowledge of John's drinking habits, she might conclude that he has a heightened risk of developing cancer.

It may seem that the publication of the research results enabled a privacy breach by Ellen, as the study's findings helped her infer new information about John's elevated cancer risk of which he himself may be unaware. However, Ellen would be able to infer this information about John regardless of his participation in the medical study (i.e., it is a risk that exists in both John's opt-out scenario and the real-world scenario). Risks of this nature apply to everyone, regardless of whether they shared personal data through the study or not. Differential privacy is a concept specifically designed to allow for studies such as in this example. Therefore, differential privacy does not guarantee that *no* information about John can be revealed. The use of differential privacy only protects the information that is *specific* to him, i.e., information about John that cannot be inferred unless an analysis received his personal information as part of the input.

This and similar examples demonstrate that any useful analysis carries a risk of revealing some information about individuals. However, such risks are largely unavoidable. In a world in which data about individuals are collected, analyzed, and published, John cannot expect better privacy protection than is offered by his opt-out scenario, because he has no ability to prevent others from participating in a research study or to prohibit a release of public records. Moreover, the types of information disclosures enabled in John's opt-out scenario often result in individual and societal benefits. For example, the discovery of a causal relationship between red wine consumption and elevated cancer risk can inform John about possible changes he could make in his habits that would likely have positive effects on his health.

6.3 Aligning Risks, Controls, and Uses: Where Is the Use of Differential Privacy Appropriate?

This section discusses factors to take into account when evaluating whether differential privacy is an appropriate tool to be applied within a specific context, as well as factors in determining whether differential privacy should be deployed alone, in combination with other controls, or as part of a tiered access system. As an overview, Table 6.2 provides some of the key factors that weigh in favor of, or against, an appropriate use of differential privacy. For example, use cases involving statistical analysis of a population or large groups and the possibility of significant and lasting informational harms to individuals weigh heavily in favor of the adoption of differential privacy.

Table 6.2: Considerations when deciding whether to use differential privacy for a particular use case

Use cases where DP is more likely to be appropriate	Use cases where DP is not appropriate	Use cases where DP is challenging
• Informational harm derives from making inferences about individuals or small groups • Intended use is statistical analysis of population or large groups • Sensitivity of information is high • Information and analyses are highly structured • Datasets are large • Types of analyses to be conducted are known in advance • Composition effects are important • Release of (low-dimensional) synthetic data is acceptable or preferred	• Informational harm derives from making inferences about large groups • Intended use is individual inference or individual intervention • Intended control is purpose limitation • Intended control is computation limitation[1] • Datasets are very small (e.g., less than a few dozen observations)	• Supporting data linking • Supporting data cleaning • Estimating complex statistical models efficiently • Datasets are small (e.g., dozens to thousands of observations)[2] • Differentially private analysis not yet available • Intended output is high-dimensional synthetic data

[1] A control on computation is designed to "limit the direct operations that can be meaningfully performed on data. Commonly used examples are file-level encryption and interactive analysis systems or model servers. Emerging approaches include secure multiparty computation, functional encryption, homomorphic encryption, and secure public ledgers, eg blockchain technologies." (Altman et al., 2018).
[2] For a real-world example, see the Opportunity Atlas case study presented in Section 6.4.2.

To help guide a systematic analysis of the relevant factors within a specific use case, this discussion follows a framework for selecting privacy controls based on a systematic analysis of harm, informational risk, and intended analytic uses as presented by Altman et al. (2015).

6.3.1 Selecting Privacy Controls Based on Harm and Informational Risk: A Framework

Altman et al. (2015) propose a framework for selecting reasonable and appropriate privacy and security measures that are calibrated to the intended uses, threats, harms, and vulnerabilities associated with a specific research activity.[30] For applying this framework in practice, Altman et al. (2015) recommend a life-cycle approach to decomposing the factors at each information stage, including the collection, transformation, retention, access and release, and post-access stages. A diagram from Altman et al. (2015) illustrating a partial conceptualization of this framework is reproduced in Figure 6.4. The x-axis represents the sensitivity of the information, or the maximum level of expected harm to an individual in the data resulting from uncontrolled use of the data. The y-axis represents the post-transformation identifiability, or the potential for others to learn about individuals based on the inclusion of their information in the data. Examples range from data sets containing direct or indirect identifiers to data shared using expertly applied rigorous disclosure limitation techniques backed by a formal mathematical proof of privacy (e.g., user-level differential privacy with a low value of ε).

These factors—the level of expected harm from uncontrolled use of the data and the post-transformation identifiability of the data—suggest minimum privacy and security controls that are appropriate in a given case, as shown by the shaded regions in Figure 6.4. The subsets of controls within each region illustrate some possible combinations of controls from the more comprehensive set of procedural, economic, educational, legal, and technical controls (some of which are covered in other chapters of this Handbook). For data associated with only negligible or minor and fleeting harms, the use of differential privacy without any additional controls may be appropriate, but for more significant and lasting or even life altering harms, notice and consent mechanisms as well as terms of service may also be required. Obtain-

[30] In this framework, evaluating the intended uses of the data involves an assessment of the types of uses or analytic purposes intended by each of the relevant groups of data users and how privacy controls implemented at each stage enable or restrict such uses. An evaluation of the threats involves assessing potential adverse circumstances or events that could cause harm to a data subject as a result of the inclusion of that subject's data in a specific data collection, storage, use, or release. Privacy harms are injuries sustained by data subjects as a result of the realization of a privacy threat, and privacy vulnerabilities are defined as characteristics that increase the likelihood that threats will be realized. See Altman et al. (2015).

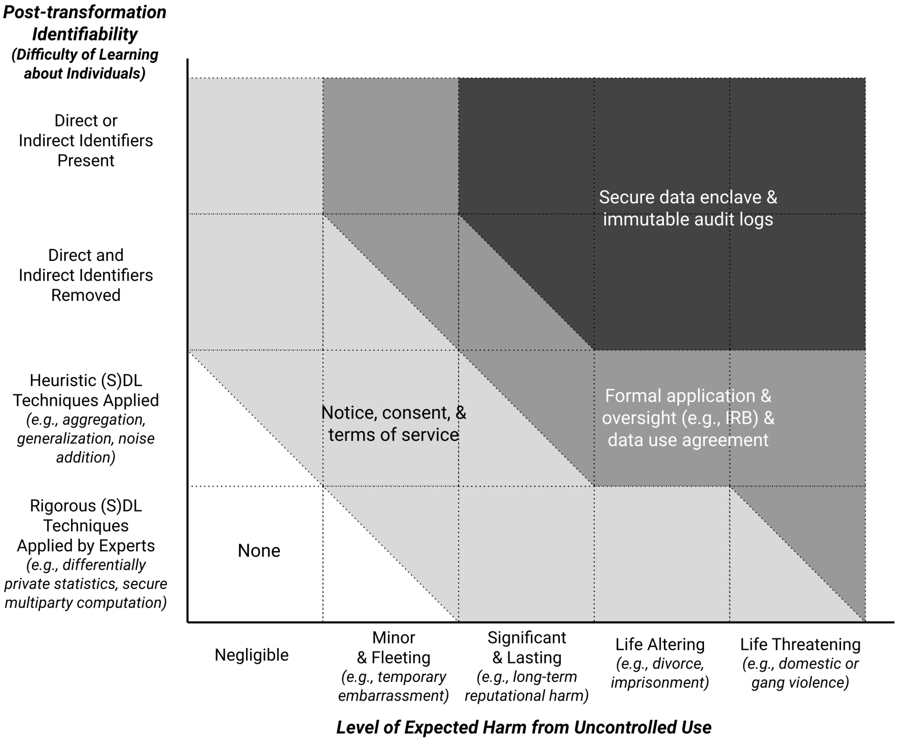
Post-transformation Identifiability
(Difficulty of Learning about Individuals)
Direct or Indirect Identifiers Present
Direct and Indirect Identifiers Removed
Heuristic (S)DL Techniques Applied
(e.g., aggregation, generalization, noise addition)
Rigorous (S)DL Techniques Applied by Experts
(e.g., differentially private statistics, secure multiparty computation)
Secure data enclave & immutable audit logs
Formal application & oversight (e.g., IRB) & data use agreement
Notice, consent, & terms of service
None
Negligible
Minor & Fleeting
(e.g., temporary embarrassment)
Significant & Lasting
(e.g., long-term reputational harm)
Life Altering
(e.g., divorce, imprisonment)
Life Threatening
(e.g., domestic or gang violence)
Level of Expected Harm from Uncontrolled Use

Figure 6.4: Calibrating privacy and security controls

ing consent is particularly important when using data for secondary uses not initially disclosed to the data subjects or when the selected value of ε is large. For data associated with potentially life-threatening harms, a formal application and oversight process, such as an institutional review board or restricted data access committee, together with a data use agreement may be necessary. As Figure 6.4 illustrates, in many cases, the use of differential privacy allows data analysis projects to be carried out safely with fewer additional privacy and security controls than would be required with other approaches.

Altman et al. (2015) note that the design of a real-world data management plan should consider a wide range of available interventions and incorporate controls at each stage of the lifecycle, including the post-access stage, and not be limited to the choices of controls illustrated in Figure 6.4. "[A]lthough the data transformation and release stages typically attract the most attention, threats and vulnerabilities arising from other lifecycle stages should not be ignored. For example, privacy risks may be present at the collection stage if the data collection process could be observed by an adversary; data retained in long-term storage are vulnerable to unintended breaches; and, increasingly in a big data world, external, independent publication of auxiliary information may create new or unanticipated privacy risks long into the post-access stage" (Altman et al., 2015). Further, one should note that some of the regions in Figure 6.4 are divided by a diagonal line; these areas correspond to situations in which an actor could decide between different choices based on factors related to the intended uses of the data or existing institutional or contractual requirements. It is also important to observe that the recommendations reflected in this diagram may differ from current practice. For example, Altman et al. (2015) argue that data that have been de-identified using simple redaction or other heuristic techniques should in many cases be protected using additional controls.

6.3.2 Considerations When Deciding Whether to Use Differential Privacy

As summarized in Section 6.3.1, differential privacy fits into a broader framework of privacy and security controls that should be applied across the information life cycle to appropriately mitigate risks of informational harm. Within a coherent set of information controls, differential privacy's primary role is as a formal criterion for disclosure control that ensures limitations on types of *inferences* that can be made about individuals and small groups based on the outputs of computations. In other words, implementations of differential privacy (especially in the curator model as discussed and contrasted with other models for differential privacy in Appendix A) modify summary information before it is published in order to prevent others from learning any information that is unique and specific to any individual who was part of the group being summarized.

In the context of designing a secure and private information system, differential privacy is used as part of a collection of controls aimed at mitigating informational harm while enabling some types of information uses. Differential privacy is usually neither sufficient protection on its own nor uniquely necessary—and in some cases differential privacy may simply not be appropriate for the intended use.

Three considerations are critical in deciding whether to use differential privacy: (1) how are recipients of protected information intending to use it, and how well do differentially private analyses support these intended uses; (2) what is the nature and degree of informational risk to be mitigated, and are there serious harms that could arise from learning about individuals; and (3) what complementary and alternative controls are available for protecting the data? Each of these questions is discussed in turn below.

How Well Does Differential Privacy Fit the Intended Uses of the Data?

Evaluating the intended uses of the data involves answering a series of sub-questions, including (a) what level of inference is intended; (b) what types of questions, queries, or models must be supported; and (c) how much accuracy is needed?

What Level of Inference is Intended? Differential privacy is a standard that was designed to support statistical analysis of populations or large groups yet prevent inferences about (and thus interventions targeted to) individuals and very small groups. Consider, for example, the collection, analysis, and sharing of public health information related to the COVID-19 pandemic. Differentially private analyses can be applied in tasks such as estimating the extent to which large communities adhere to social distancing, measuring the efficacy of infection rate reduction measures like social distancing and masks, identifying large disease clusters, and selecting and fitting statistical models of disease transmission.[31] If performed with differential privacy these analyses would yield valuable and meaningful statistics while providing strong

[31] See, e.g., Google's COVID-19 Community Mobility Reports (Aktay et al., 2020), https://www.google.com/covid19/mobility (accessed 2020-12-17).

protection for the privacy of individual medical results, locations, social encounters, etc. If analysis at the individual level is desired (e.g., to identify specific individuals for testing or quarantine) disclosure control methods other than differential privacy should be used. Researchers who intend to prevent certain types of learning about large groups, such as information that could be used to discriminate on the basis of protected group status, should be aware of limitations; while differential privacy protects information that is specific to groups consisting of a small number of individuals, the use of differential privacy alone does not provide protection against group-level inference for larger groups.

What Types of Questions, Queries, or Models Must be Supported? In theory, with the exception of learning about individuals or small groups, differential privacy could be used to compute any form of answer for any purpose, as it is a constraint on inference, not on purpose or computation (Altman et al., 2018). And in practice, as outlined in Section 6.1.5, a large number of analyses can be performed with differential privacy guarantees.

However, there are some limitations on the current understanding of how to perform certain classes of tasks privately (e.g., the use of differential privacy in analyzing records of textual data is currently limited); how to measure the accuracy or utility of protected results; and how to optimize the privacy versus utility trade-off. Even where algorithms to perform specific calculations are known, robust software that implements these methods may not yet be available. Generally, differentially private tools limit both the *number* and *form* of analyses that are possible. Most differentially private tools that provide interactive access to data by design support a limited range of model specifications or statistical operators. For example, a particular tool may allow one to pose queries that can be expressed in terms of counts on definable subsets of the data set (which allows for contingency tables and hence fitting logistic regression models) but not to run any arbitrary statistical model. Similarly, an analyst can apply any model to a non-interactive, synthetically generated data set, but only a limited range of models will return accurate or useful results. Further, it is generally more difficult to apply differential privacy if the methods used by analysts are qualitative, unstructured, or do not lend themselves to rigorous mathematical definitions. Certain queries, such as estimating the number of individuals with specific attributes, are quite straightforward. However, in-depth data cleaning is difficult to define in a sufficiently formal way to apply differential privacy protections to the process.

Appendix C lists currently available software tools for differentially private computation. In general, these tools support a wide range of summary tabulations and summary statistics, the generation of synthetic data sets for some forms of multivariate analysis, and selected applications such as geospatial or location-based analysis. If the intended analyses fall outside of the capabilities of existing tools, one should anticipate that considerably more effort will be required to deploy an effective system in order to support such analyses. This is the case even if the core algorithms for those calculations are already known. Those following this approach should engage experts in differential privacy as part of the design and deployment process.

What is the Required Level of Accuracy? Differential privacy provides a quantifiable trade-off between privacy and utility (or accuracy). The amount of noise that differential privacy needs to introduce for a *single count query* is on the order of $1/\varepsilon$ in which ε is the privacy-loss parameter. At minimum, the data set being analyzed must have at least $1/\varepsilon$ observations to obtain meaningful results. For most analyses, however, the size of the data set must be much larger than $1/\varepsilon$ to obtain useful results, and how much larger will depend on a number of factors including how many statistics are being calculated, the complexity of the statistical model, the dimensionality of the data, and the particular differentially private algorithm being used. Thus, it is difficult to provide a rule of thumb. In practice, one can run experiments on nonsensitive synthetic or public data as a way to evaluate the accuracy of a tool or algorithm for a given application ahead of time. (Using experiments on the *sensitive* data to select an algorithm or set its parameters may leak information that violates differential privacy.)

When operating within the framework of existing tools, one should plan to test that outputs remain useful for the intended purposes. There are many different measures of utility and, even if an algorithm does a good job at trading off between utility and privacy, the utility loss for a particular use case may be quite different than the average loss.

What Is the Nature and Degree of Informational Risk to be Mitigated?

Another factor to consider when deciding whether to adopt differential privacy is the nature and degree of informational risk to be mitigated. Figure 6.4 illustrates an approach to conceptualizing whether differential privacy is a suitable control to use given different levels of harm associated with uncontrolled use of a particular data set. Some of the relevant questions to consider involve the sensitivity of the information and the potential for risks to accumulate with multiple releases of information about the same individuals or groups of individuals.

How Sensitive are the Data? When evaluating informational risk, consider the sensitivity of the information or its potential to cause harm to individuals, groups of individuals, or society at large. Generally, information should be treated as sensitive when it reveals information specific to an individual (even partially or probabilistically and possibly in combination with other information) and such inference is likely to cause significant harm to an individual, group, or society.[32] Informational harms "may occur directly as the result of a reaction of a data subject or third parties to the information, or indirectly as a result of inferences made from information" (Altman et al., 2015). Appli-

[32] For an extended discussion and framework for assessing information sensitivity, see Altman et al. (2015).

cable laws[33] and institutional policies[34] may provide some guidance regarding sensitivity, but data may be sensitive and have the potential to cause harm, even if the data do not include categories of information traditionally considered sensitive (Altman et al., 2015). Other key factors increasing informational risk include the number of independent attributes associated with each subject in the data, the scope of intended analytic uses, the number of individuals included in the data, and the size and diversity of the population observed (Altman et al., 2018). Risks can also grow due to characteristics related to time, such as an increase in the amount of time between collection and analysis, in the period of time over which data are collected, and in the frequency of collection (Altman et al., 2018).

Does Composition of Multiple Releases Pose a Significant Threat? Privacy risk inevitably grows as more computations are released. Differentially private protection mechanisms have the advantage that risk composes predictably and slowly across multiple releases. In contrast, when information is released through other mechanisms, multiple releases could result in sudden and catastrophic loss of privacy.

Absent formal protection mechanisms, it is not possible to definitively assess composition risks ex ante. As general guidance, composition effects are of greatest ex ante concern under the following conditions: (a) data are collected from the same individuals by uncoordinated data controllers, (b) releases are updated frequently, (c) many releases are performed over time, (d) releases are high-dimensional, or (e) prior

[33] See, e.g., Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (2016) OJ L119/1, Article 9 (providing that the "[p]rocessing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited," unless one of the delineated exceptions in Paragraph 2 of the Article applies).

[34] See, e.g., Harvard University Information Security, Handout—Research Data Security Levels with Examples, https://security.harvard.edu/handout-research-data-security-levels-examples (accessed 2020-12-17).

releases cannot be reliably recalled.[35]

Alternatively, if the data controller is aware of all potential auxiliary information, it could attempt to assess the cumulative privacy risk post-computation but prior to release. Or, if harm to individuals is readily detected, the data controller could purchase insurance to compensate such harm ex post. These caveats notwithstanding, in the modern information environment, composition risks are generally substantial and ex post formal protections are typically infeasible.

What Complementary and Alternative Controls are Available for Protecting the Data?

As illustrated in Figure 6.4, various controls can be complementary to differential privacy. Some examples include contractual approaches for enforcing purpose restrictions, vetting and oversight of analysts for the purpose of privacy budget allocation, and encryption and other information security restrictions on private databases, especially if now exposed to a different set of users through a publicly available differentially private interactive query mechanism. Other tools may be used as an alternative for purposes that differential privacy does not support, such as the role that access via a secure data enclave can play as part of a tiered access system.

Further, a single mode of access will generally not be appropriate for the needs of all users. Different communities of users seek answers to different questions and may have different quality and accuracy requirements even when addressing the same question. It is therefore essential to understand end user usages of inferences and their implied utility and quality criteria (as discussed in Appendix A). An analyst should take these factors into account in particular when allocating the privacy budget across analyses and when selecting the specific interactive and static publication mechanisms to be included.

Tiered access will generally be necessary to accommodate a wide range of desired uses of the data. For a given set of data, access may be made

[35] For discussions of how data privacy risks accumulate, see Altman et al. (2018); Fluitt et al. (2019).

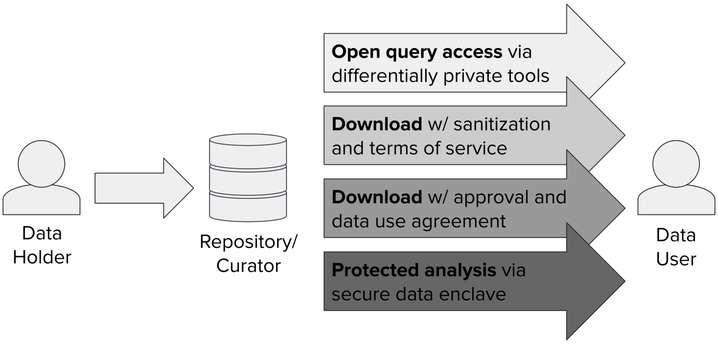
Open query access via differentially private tools
Download w/ sanitization and terms of service
Download w/ approval and data use agreement
Protected analysis via secure data enclave
Data Holder
Repository/ Curator
Data User

Figure 6.5: An example of a tiered access model

available to different categories of users through different modes of release. Figure 6.4 demonstrates how controls can be selected at each tier. For example, data associated with potential harms that are only minor and fleeting could be released to the public after traditional statistical disclosure limitation techniques, such as aggregation and generalization, have transformed the data. Users who seek to obtain the full data set, including direct and indirect identifiers, would be required to submit an application to an institutional review board or other oversight body, and their use would be subject to the terms of a data use agreement. This approach makes it possible to calibrate data releases to the risk profile of a data set as well as specific uses intended by different data users. Figure 6.5 provides an example of such a tiered access model (see also Sweeney, Crosas and Bar-Sinai, 2015; Crosas, 2019).

6.3.3 Regulatory and Policy Compliance

Statistical agencies, companies, researchers, and others who collect, process, analyze, store, or share data about individuals must take steps to protect the privacy of the data subjects in accordance with various laws, institutional policies, contracts, ethical codes, and best practices. In some settings, tools that satisfy differential privacy can be used to analyze and share data while both complying with legal obligations and providing strong mathematical guarantees of privacy protection for the individuals in the data (Nissim et al., 2018). Indeed, differentially private tools provide privacy protection that is more robust than that provided by techniques commonly used to satisfy regulatory requirements for privacy protection.

That said, privacy regulations and related guidance do not directly answer the question of whether the use of differentially private tools is sufficient to satisfy existing regulatory requirements for protecting privacy when sharing statistics based on personal data. This issue is complex because privacy laws are often context dependent, and there are significant gaps between differential privacy and the concepts underlying regulatory approaches to privacy protection. Different regulatory requirements are applicable depending on the jurisdiction, sector, actors, and types of information involved. As a result, data sets held by an organization may be subject to different requirements. In some cases, similar or even identical data sets may be subject to different requirements when held by different organizations. In addition, many legal standards for privacy protection are to a large extent open to interpretation and therefore require a case-specific legal analysis by an attorney.

Other challenges arise as a result of differences between the concepts appearing in privacy regulations and those underlying differential privacy. For instance, many laws focus on the presence of personally identifiable information (PII) or the ability to identify an individual's personal information in a release of records. Such concepts do not have precise definitions, and their meaning in the context of differential privacy applications are especially unclear. In addition, many privacy regulations emphasize particular requirements for protecting privacy when disclosing individual-level data, such as removing PII, which are arguably difficult to interpret and apply when releasing aggregate statistics. While in some cases it may be clear whether a regulatory standard has been met by the use of differential privacy, in other cases—particularly along the boundaries of a standard—there may be considerable uncertainty.

Regulatory requirements relevant to issues of privacy in computation rely on an understanding of a range of different concepts, such as PII, de-identification, linkage, inference, risk, consent, opt-out, and purpose and access restrictions. The definition of differential privacy can arguably be interpreted to address these concepts while accommodating differences in how they are defined across various legal and institutional contexts (Wood et al., 2018). For instance, when differential privacy is used, it can be understood as ensuring that using an individual's data will not reveal essentially any PII specific to them.[36] Differential privacy arguably addresses record linkage in the following sense. Differentially private statistics provably hide the influence of every individual (even small groups of individuals). Although linkage has not been precisely defined, linkage attacks seem to inherently result in revealing that specific individuals participated in an analysis. Because differential privacy protects against learning whether an individual participated in an analysis, it can therefore be understood to protect against linkage. Furthermore, differential privacy provides a robust guarantee of privacy protection that is independent of the auxiliary information available to an attacker. Indeed, under differential privacy, even an attacker utilizing arbitrary auxiliary information cannot learn much more about an individual in a database than they could if that individual's information were not in the database at all.

The foregoing interpretations of the differential privacy guarantee can be used to demonstrate that in many cases a differentially private mechanism would prevent the types of disclosures of personal information that privacy regulations have been designed to address. Moreover, differentially private tools often provide privacy protection that is more robust than that provided by techniques commonly used to satisfy regulatory requirements for privacy protection. However, further research is needed to develop methods for proving that differential privacy satisfies legal requirements, and setting the privacy loss parameter ε based

[36] Note that the reference to "using an individual's data" in this statement means the inclusion of an individual's data in an analysis, and the use of the term "specific" refers to information that is unique to the individual and cannot be inferred unless the individual's information is used in the analysis. Furthermore, the use of the word "essential" in the statement "will not reveal essentially any PII specific to them" means that, compared with an opt-out scenario where no information specific to an individual is leaked, some small leakage of such information (inevitably) occurs. The parameter ϵ bounds this leakage.

on such requirements is needed.[37] In practice, data providers should consult with legal counsel when considering whether differential privacy tools—potentially in combination with other tools for protecting privacy and security—are appropriate within their specific institutional settings.

6.4 Case Studies

Differential privacy is a relatively new concept, first presented in the theoretical computer science literature in 2006 and now seeing early stages of application in real-world settings. This section provides short case studies on three implementations of differential privacy: the 2020 Decennial Census, the Opportunity Atlas, and the Dataverse Project. This discussion focuses on describing aspects of the context in which these differentially private solutions were developed, as well as the design choices that were made with respect to the relevant contextual factors.

This selection of case studies, though limited by the small number of practical implementations of differential privacy to date, aims to reflect a range of different scenarios. The first case study involves a national statistical agency publishing statistical data products from a census, the second involves a team of researchers developing a web-based visualization tool for exploring sensitive administrative data analyzed as part of a research study, and the third describes the functionalities of a general-purpose differential privacy tool being developed for use by data providers and analysts who do not have expertise in differential privacy. Although none of these examples directly describe sharing data from sub-national agencies, they carry real-world lessons relevant to employing differential privacy in such contexts.

Each of the case studies reflects one point in the space of design factors discussed in Section 6.3 and Appendix A. These factors are summa-

[37] For an extended discussion of the gaps between legal and computer science definitions of privacy and a demonstration that differential privacy can be used to satisfy an institution's obligations under FERPA, see Nissim et al. (2018).

rized in Table 6.3. The remainder of this section expands upon critical features of each case and their implications.

6.4.1 The 2020 Decennial Census

In September 2017, the US Census Bureau announced its decision to deploy differential privacy in the disclosure avoidance mechanism for the 2020 Decennial Census (Garfinkel, 2017). This decision was motivated in part by the composition effects revealed by a reconstruction attack on the 2010 Census data release (see Section 6.1.2) and the confidentiality and data publication mandates that bind the US Census Bureau.[38]

In many ways, the data from the US Decennial Census is an excellent fit for differential privacy. Compared to most survey data, it is low-dimensional (i.e., only asks a few questions of each respondent) and the sample size is very large (minimizing the relative impact of the noise added for differential privacy). These features normally would allow for a straightforward application of standard differentially private algorithms (e.g., those which add independent noise to each cell of different cross-tabulations). However, there are a number of other features of the Decennial Census data products that have created challenges and debate among stakeholders over the transition to differential privacy (Garfinkel, Abowd and Powazek, 2018; Hawes, 2020; boyd, 2020).

First, these data products have a long history of being used for a vast and diverse range of applications, such as apportioning seats in the US House of Representatives, redistricting, funding allocations, provision of local emergency resources, and social science research. To minimize the impact on data users and the software they use, the Census Bureau has decided to produce differentially private data products that have the same form as the traditional products and consist of tables that are

[38] Specifically, the US Constitution mandates the Decennial Census (U.S. Const. art. 1, 2.), and it is carried out by the US Census Bureau, bound by Title 13 of the US Code, which prohibits Census Bureau employees from “mak[ing] any publication whereby the data furnished by any particular establishment or individual under this title can be identified” (13 U.S.C. § 9(a)(2)).

Table 6.3: Design choices in case study implementations of differential privacy

	2020 Decennial Census	The Opportunity Atlas	Dataverse repositories
Risks & Sensitivity	**Sensitivity:** Data subject to stringent statutory protections. Trust in confidentiality critical to collecting sensitive information from respondents. **Risks:** Concerns about composition effects and reconstruction attacks motivated adoption of DP.	**Sensitivity:** Data subject to stringent statutory protections. **Risks:** Prior methods of de-identification and redaction judged not to sufficiently mitigate risk.	**Sensitivity:** General-purpose system designed to support analyses of data of varying degrees of sensitivity. **Risks:** Vary by data source. DP provides stronger mechanism to mitigate risk than pre-deposit redaction and deidenfication.
Tiered Access Controls	Part of a tiered access system that has historically included custom tabulations service for institutional clients; and Research Data Centers for access by vetted individuals to private data.	Original data sources remain available to vetted users through federal Restricted Data Center mechanism.	Part of a tiered access model that also supports access to private data with vetting and restricted license.
Trust & Publication Models	Curator model, based on prior data collection design, with cleaning before DP applied. Focus on non-interactive publication of tables.	Curator model applied to previously collected data, with cleaning and linkage (between Census and IRS data) before DP-like methods applied.	Curator model, based on previously collected and deposited data. Supports both non-interactive releases of summary statistics and interactive queries.
Budget Allocation	Must allocate budget and optimize accuracy for broad range of current and future analyses.	Budget analysis focused on balancing privacy vs. societal utility, leading to choice of a rather large epsilon.	Provides recommended choices of epsilon based on sensitivity of data. Choice to allow per-analyst budgets requires semi-trusted and accountable analysts.
Estimating Uncertainty	Adopting DP has made noise addition explicit, whereas data users had previously treated Census tables as if they have no error.	Designed to produce uncertainty estimates (taking privacy noise into account) together with quantities of interest, and estimates also calculated in a DP-like manner.	Important to expose uncertainty estimates from noise due to privacy, both before and after release.
Granularity	Focused both on individuals and households, as appropriate to data measurement design	Focused on individuals.	Determined by data depositor.

exactly consistent with an underlying synthetic data set (rather than a collection of noisy statistics that would be produced by a standard differentially private algorithm), along with other information that needs to be published exactly (e.g., the state population totals). This required the design of custom differentially private algorithms by experts at the Bureau (Garfinkel, Abowd and Powazek, 2018; Abowd et al., 2019).

Second, the sources of error in the Decennial Census data products (in particular, disclosure avoidance) have historically not been made explicit and have been largely ignored by data users. Differential privacy is transparent about its noise addition and thus creates concern among stakeholders for the potential impact on their applications. Reconstruction attacks (Dinur and Nissim, 2003) tell us that the data products cannot be simultaneously accurate for all possible uses and maintain privacy, leaving the Bureau with the challenging problems of deciding which users and uses to prioritize for accuracy and then optimizing the algorithm and its privacy-loss budget allocation accordingly. To this end, the Bureau published a Federal Register Notice (Bureau of the Census, 2018) to understand what aspects of their data products were most important for data users and also released a series of demonstration products showing the impact of potential versions of their differentially private algorithms on past Decennial Censuses.[39]

Referring to some of the other design choices discussed in Appendix A, the plans for the 2020 Decennial Census are utilizing a curator model (with the US Census Bureau as the trusted curator) with a noninteractive publication model corresponding with the pre-existing data collection and dissemination design. However, historically, access to data from the Decennial Census has not been limited to the public-use products discussed above but have also been made available through other means, including a custom tabulation service for institutional clients and Federal Statistical Research Data Centers for access by vetted individuals. Thus, the planned use of differential privacy fits within an existing tiered access system. It remains to be seen whether and how

[39] See United States Census Bureau, https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/2020-census-data-products/2020-das-updates.html (accessed 2020-12-17).

interactive differential privacy will play a role in subsequent accesses to data from the 2020 Census. Similar to past Census disclosure avoidance systems, the planned algorithm is to be applied after data cleaning edits are performed (Garfinkel, 2017). It will enforce privacy at the granularity of individuals as well as at the granularity of households for publications that are based on household characteristics.

Consider the application of differential privacy to the Decennial Census in contrast with another data product from the US Census Bureau— namely, the Post-Secondary Employment Outcomes (PSEO) data (Foote, Machanavajjhala and McKinney, 2019). This data product includes estimates of the cumulative distribution function of earnings for different subsets of the national student population, based on linking college transcripts with Longitudinal Employer-Household Dynamics (LEHD) data. In contrast with the Decennial Census products, this was a new product first released in 2018, so there was no history of entrenched data use that constrained the form of the data release. As a result, it was possible to employ standard differentially private algorithms (namely, binning the earnings within each subset and adding noise to the counts in each bin). Note that the linking of transcript data with LEHD data is done prior to the application of the differentially private algorithm. The PSEO release used a privacy-loss parameter of $\varepsilon = 1.5$ (US Census Bureau Center for Economic Studies, n.d.).

6.4.2 The Opportunity Atlas

The Opportunity Atlas is a web-based visualization tool for exploring social mobility data. It was published as the result of a collaboration between the US Census Bureau, Harvard University, and Brown University (Chetty et al., 2018). The database contains data relevant to understanding children's economic outcomes in adulthood for every Census tract in the United States. Researchers and policymakers can use the Opportunity Atlas to understand how individuals' prosperity or poverty is rooted in the neighborhoods in which they grew up and how interventions can be targeted in certain neighborhoods to help more children rise out of poverty.

The Opportunity Atlas is based on data about over 20 million children and their parents, compiled from multiple statistical and administrative data sources. Census data sources include the 2000 and 2010 Decennial Censuses and the American Community Survey. Administrative data sources included de-identified data from IRS income tax returns and data on students receiving Federal Pell Grants, obtained from the US Department of Education's National Student Loan Data System.

Raj Chetty and John Friedman, Director and Co-Director of the Opportunity Insights research team, respectively, developed the privacy protection mechanism for the Opportunity Atlas in consultation with the US Census Bureau and the Harvard University Privacy Tools Project (Chetty and Friedman, 2019). Consistent with the US Census Bureau's broader efforts to modernize its approach to disclosure limitation (as discussed in Section 6.4.1) and the legal protections for both Census and IRS data,[40] the Opportunity Atlas was produced using a method inspired by differential privacy.

Linkage, analysis, and disclosure avoidance were performed in Census facilities. There was a single set of analyses to perform to generate the Opportunity Atlas, and a privacy budget was not reserved for future analyses. They ran simple linear regressions on the data from the Census Bureau and IRS in order to predict child income rank from parent income rank in each Census tract, broken down by race, gender, and other variables. This created challenges for a differentially private solution, as the sample sizes were small (on the order of tens, hundreds, and thousands), and there was sometimes a very small variance in the explanatory variable. However, despite these challenges, the Opportunity Atlas achieved good results using a differential privacy–inspired method. In terms of accuracy, this approach performed better than

[40] The raw data from the Census Bureau is protected by Title 13 of the United States Code, which prohibits "mak[ing] any publication whereby the data furnished by any particular establishment or individual under this title can be identified" (13 U.S.C. § 9(a)(2)). Pursuant to Title 26, the IRS shares federal tax returns and return information with the Census Bureau for statistical purposes, and the Census Bureau is prohibited from disclosing such tax return information except in "a form which cannot be associated with, or otherwise identify, directly or indirectly, a particular taxpayer" (26 U.S.C. 6103(j)(4)).

some traditional statistical disclosure limitation techniques. Indeed, the researchers found that traditional count suppression would have caused them to miss strong relationships that relied on small counts (e.g., between teenage birth rates for black women and the proportion of single parents in Census tracts) (Chetty and Friedman, 2019). The Opportunity Atlas also includes uncertainty estimates (standard errors), which are also calculated in a differential privacy–inspired manner.

Chetty and Friedman suggest selecting the privacy-loss parameter (ε) using the framework of Abowd and Schmutte (Abowd and Schmutte, 2019), equating the marginal societal benefit of increased accuracy with the marginal cost due to reduced privacy. Given the small sample sizes of the Opportunity Atlas and the importance of accurate data for policymaking, the Opportunity Atlas used (with approval of the Census Bureau Disclosure Review Board) a value of ε that is significantly larger than is typically considered in the differential privacy literature. Specifically, they used $\varepsilon = 8$ for each of several statistics published for each demographic group within a tract.

The Chetty-Friedman method is a general technique, in that it applies to many different statistical estimators (not just simple linear regression). However, it is not formally differentially private, and its privacy properties rely on the same analysis being carried out on many different cells (e.g., many Census tracts as in the Opportunity Atlas). For the specific case of simple linear regression, subsequent work has developed formally differentially private methods that are competitive with the Chetty-Friedman method, and thus may be applied even to releases that do not have the cell structure of the Opportunity Atlas (Alabi et al., 2020).

6.4.3 Dataverse Repositories

The Harvard University Privacy Tools Project[41] and the OpenDP initiative[42] have been developing a vision for how differential privacy can be incorporated into research data repositories like Dataverse, ICPSR, and Dryad to help human-subjects researchers safely share and analyze sensitive data (Gaboardi et al., 2016; The OpenDP Team, 2020). Although these solutions have not yet been deployed at the time of this Handbook, software to support the projects are under active construction and may be available for use in the near future. Thus, this section outlines how differential privacy might fit into some of the ways that research data repositories are used, employing a lightly edited extract from the OpenDP whitepaper (The OpenDP Team, 2020). For concreteness, the text is written as specific to using OpenDP software in Dataverse repositories but can be generalized to other repositories and underlying differential privacy software.

Dataverse (King, 2007; Crosas, 2011, 2013; King, 2014), developed at Harvard's Institute for Quantitative Social Science (IQSS) in 2006, enables researchers to share their data sets with the research community through an easy-to-use, customizable web interface, keeping control of, and gaining credit for, their data while the underlying infrastructure provides robust support for good data archival and management practices. Dataverse has been installed and serves as a research data repository in more than fifty institutions worldwide.

Dataverse repositories (like most general-purpose data repositories) currently have little support for hosting privacy-sensitive data. Data sets with sensitive information about human subjects were supposed to be "de-identified" before deposit. Unfortunately, as discussed in Section 6.1.2, research in data privacy starting with (Sweeney, 1997) has demonstrated convincingly that traditional de-identification does not provide effective privacy protection. The current alternative to open data sharing in repositories is that researchers depositing a data set (*data depositors*) declare their data set restricted: the data set would

[41] Harvard University Privacy Tools Project, http://privacytools.seas.harvard.edu (accessed 2020-12-17).

[42] OpenDP, http://opendp.io/ (accessed 2020-12-17).

not be made available for download, and the only way for other researchers to obtain access would be through contacting the data depositor and negotiating terms on an ad hoc basis. This approach is also unsatisfactory, as it can require the continued involvement of the data depositor, the negotiations can often take months, and thus it impedes the ability of the research community to verify, replicate, and extend work done by others.

OpenDP can enable Dataverse to offer additional ways to access sensitive data as illustrated by the following use cases.

1. Enabling variable search and exploration of sensitive data sets deposited in the repository

Dataverse already automatically calculates variable summary statistics (counts, min/max, means, etc.) when a tabular file is deposited. These summary statistics for each variable can be viewed using the Data Explorer tool, even without downloading or accessing the data file. As OpenDP is integrated with Dataverse, a data depositor should be able to generate a differentially private (DP) summary statistics metadata file using an OpenDP user interface. To do this, the data depositor would select "Generate DP Summary Statistics" after the tabular data file is ingested in Dataverse, launching the OpenDP interface. Then they would select the privacy-loss parameter for their data file, and OpenDP would create the differentially private summary statistics file and Dataverse would store the newly created metadata file associated with the sensitive tabular data file. Once the data set is published, an end user would be able to view the summary statistics of the sensitive data file using the Data Explorer tool without ever accessing or downloading the actual data file.

2. Facilitating reproducibility of research with sensitive data sets

At least a third of the data sets deposited in Dataverse are replication data and code associated with a published scholarly paper. With OpenDP, data depositors or owners could create a differentially private release on a sensitive data set, which could be used to computationally reproduce the results of the published paper while protecting the privacy of the original data set. In this case, like in Use Case 1 above, a data depositor would select a privacy-loss parameter through the OpenDP user interface and use OpenDP's statistical query interface to select and run the statistics of choice to create the appropriate replication release. The differentially private replication release file would be made available in the data set and end users would be able to download it, while the original sensitive data set would be protected and not accessible by end users except through the existing processes as above.

3. Enable statistical analysis of sensitive data sets accessible through the repository

For additional flexibility, the data depositor of a sensitive data set could allow for any researcher (end user) to be able to run any statistic available through the OpenDP interface. In this case, the data depositor would configure the allocation of privacy-loss budgets through the OpenDP interface before releasing the data set. Once the data set is published, an end user would be able to click "explore" for the sensitive data file, and the OpenDP statistical query interface would open. The user would not have access to the original sensitive data file but would be able to run the statistics of their choice—up to the point that the established privacy-loss budget allows.

Referring to some of the other design choices discussed in Appendix A, the vision outlined above fits into the curator model of differential privacy, as researchers depositing data in the repository have typically already been trusted to collect the sensitive data. It is part of a tiered access model meant to augment rather than replace the existing methods of accessing restricted data. Use Cases 1 and 2 involve noninteractive releases, whereas Use Case 3 allows for interactive queries. Many of the other key choices associated with implementing differential privacy are left to the data depositor, who cannot be expected to have expertise in differential privacy. Thus, the software tools must provide a clear user interface to guide the depositor in their decisions. There should be a tutorial on the concepts of privacy loss, privacy–accuracy trade-offs, and budgeting, including recommended choices of privacy-loss parameter ε according to different categories of data and sensitivity. The depositor should also be guided in defining the granularity of privacy appropriate for their data and the trade-offs between offering per-analyst budgets for interactive queries versus a global budget for all queries. Domain knowledge will be required of the depositor (and the analyst in Use Case 3) in deciding which statistics to release and which ones to prioritize for accuracy. For the research use cases described above, it will be important that the differentially private analyses offered provide uncertainty estimates whenever possible.

CHAPTER 6

About the Authors

Micah Altman is a social and information scientist at MIT's Center for Research in Equitable and Open Scholarship. Dr. Altman conducts research in social science, information science and research methods – focusing on the intersections of information, technology, privacy, and politics; and on the dissemination, preservation, reliability, and governance of scientific knowledge. Dr. Altman has authored over 100 scholarly works. This work has been recognized with the Pizzigati Prize from the Tides Foundation, the Brown Democracy Award, awards from professional organizations, citations by the U.S. Supreme Court, and coverage by numerous local and national media organizations.

Kobbi Nissim is McDevitt Chair in the department of Computer Science, Georgetown University and affiliated with Georgetown Law. Nissim's work is focused on the mathematical formulation and understanding of privacy. His work from 2003 and 2004 with Dinur and Dwork initiated rigorous foundational research of privacy and in 2006 he introduced differential privacy with Dwork, McSherry and Smith. Nissim was awarded the Caspar Bowden Award for Outstanding Research in Privacy Enhancing Technologies in 2019, the Gödel Prize In 2017, the IACR TCC Test of Time Award in 2016 and in 2018, and the ACM PODS Alberto O. Mendelzon Test-of-Time Award in 2013.

Salil Vadhan is the Vicky Joseph Professor of Computer Science and Applied Mathematics at the Harvard John A. Paulson School of Engineering & Applied Sciences. He is Lead PI on the Harvard Privacy Tools Project and Co-director of the OpenDP software project. Vadhan's research in theoretical computer science spans computational complexity, cryptography, and data privacy. His honors include a Harvard College Professorship, a Simons Investigator Award, a Guggenheim Fellowship, and a Gödel Prize.

Alexandra Wood is a fellow at the Berkman Klein Center for Internet & Society at Harvard University and a senior researcher contributing to the Harvard Privacy Tools Project. Her research explores new and existing regulatory frameworks for data privacy and their relationship to approaches to privacy emerging from other fields. She also contributes to the development of new legal instruments, analytical frameworks, and policy recommendations to better support the sharing and use of research data while preserving privacy, utility, transparency, and accountability. She currently serves as an advisory board member for the Privacy Engineering Section of the International Association of Privacy Professionals and in 2019 she received the Caspar Bowden PET Award for Outstanding Research in Privacy Enhancing Technologies.

Disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of their funders.

Acknowledgements

Work of Kobbi Nissim, Salil Vadhan, and Alexandra Wood was partially supported by the US Census Bureau under cooperative agreement No. CB16ADR0160001. Salil Vadhan was also partially supported by a Simons Investigator Award. The authors thank their coauthors on those papers, as well as Jean-François Couchot and Jean-Marc Gervais for their inquiries that led us to uncover a mistake in the calculation of Table 6.1; Mercè Crosas, Raj Chetty, and John Friedman for their comments on the case studies; Marco Gaboardi and Michael Hay for help with software tools; and Simson Garfinkel, Steven Glazerman, Michel José Reymond, Jayshree Sarathy, Anja Sautmann, Lars Vilhuber, and the members of the Bridging Privacy Definitions Working Group for their review comments on earlier drafts.

References in Chapter 6

Abadi, Martín, Andy Chu, Ian J. Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. "Deep learning with differential privacy." 308–318. ACM. https://doi.org/10.1145/2976749.2978318.

Abowd, John. 2019. "Stepping-up: The Census Bureau Tries to Be a Good Data Steward in the 21st Century, presentation at the simons institute for the theory of computing." https://simons.berkeley.edu/talks/tba-30.

Abowd, John M., and Ian M. Schmutte. 2019. "An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices." *American Economic Review*, 109(1): 171–202. https://doi.org/10.1257/aer.20170627.

Abowd, John, Robert Ashmead, Simson Garfinkel, Dan Kifer, Philip LeClerc, Ashwin Machanavajjhala, Brett Moran, William Sexton, and Pavel Zhuravlev. 2019. "Census TopDown algorithm: Differentially private data, incremental schemas, and consistency with public knowledge." https://github.com/uscensusbureau/census2020-das-2010ddp/blob/master/doc/20191020_1843_Consistency_for_Large_Scale_Differentially_Private_Histograms.pdf.

Aktay, Ahmet, Shailesh Bavadekar, Gwen Cossoul, John Davis, Damien Desfontaines, Alex Fabrikant, Evgeniy Gabrilovich, Krishna Gadepalli, Bryant Gipson, Miguel Guevara, Chaitanya Kamath, Mansi Kansal, Ali Lange, Chinmoy Mandayam, Andrew Oplinger, Christopher Pluntke, Thomas Roessler, Arran Schlosberg, Tomer Shekel, Swapnil Vispute, Mia Vu, Gregory Wellenius, Brian Williams, and Royce J Wilson. 2020. "Google COVID-19 community mobility reports: Anonymization process description (version 1.0)." https://arxiv.org/abs/2004.04145v1.

Alabi, Daniel, Audra McMillan, Jayshree Sarathy, Adam Smith, and Salil Vadhan. 2020. "Differentially private simple linear regression." https://arxiv.org/abs/2007.05157.

Altman, Micah, Alexandra Wood, David O'Brien, Salil Vadhan, and Urs Gasser. 2015. "Towards a Modern Approach to Privacy-Aware Government Data Releases." *Berkeley Technology and Law Journal*, 1967. https://doi.org/10.2139/ssrn.2779266.

Altman, Micah, Alexandra Wood, David R O'Brien, and Urs Gasser. 2018. "Practical approaches to big data privacy over time." *International Data Privacy Law*, 8(1): 29–51. https://doi.org/10.1093/idpl/ipx027.

Blum, Avrim, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. 2005. "Practical privacy: the SuLQ framework." 128–138. ACM. https://doi.org/10.1145/1065167.1065184.

Blum, Avrim, Katrina Ligett, and Aaron Roth. 2013. "A learning theory approach to noninteractive database privacy." *Journal of the ACM*, 60(2): 12:1–12:25. https://doi.org/10.1145/2450142.2450148.

boyd, danah. 2020. "Balancing data utility and confidentiality in the 2020 US Census." Data & Society. https://datasociety.net/library/balancing-data-utility-and-confidentiality-in-the-2020-us-census/ (accessed 2020-12-15).

Bun, Mark, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. 2015. "Differentially Private Release and Learning of Threshold Functions." *IEEE Computer Society*. https://doi.org/10.1109/FOCS.2015.45.

Bureau of the Census, Department of Commerce. 2018. "Soliciting feedback from users on 2020 census data products." *Federal Register*, 83(139). https://www.federalregister.gov/documents/2018/07/19/2018-15458/soliciting-feedback-from-users-on-2020-census-data-products.

Calandrino, Joseph A., Ann Kilzer, Arvind Narayanan, Edward W. Felten, and Vitaly Shmatikov. 2011. ""You Might Also Like:" privacy risks of collaborative filtering." 231–246. IEEE Computer Society. https://doi.org/10.1109/SP.2011.40.

Chen, Bee-Chung, Daniel Kifer, Kristen LeFevre, and Ashwin Machanavajjhala. 2009. "Privacy-preserving data publishing." *Foundations and Trends® in Databases*, 2(1–2): 1–167. https://doi.org/10.1561/1900000008.

Chetty, Raj, and John N. Friedman. 2019. "A Practical Method to Reduce Privacy Loss When Disclosing Statistics Based on Small Samples." *Journal of Privacy and Confidentiality*, 9(2). https://doi.org/10.29012/jpc.716.

Chetty, Raj, John N. Friedman, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2018. "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility." National Bureau of Economic Research Working Paper 25147, https://doi.org/10.3386/w25147.

Crosas, Mercè. 2011. "The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data." *D-lib Magazine*, 17(1): 2. https://doi.org/10.1045/january2011-crosas.

Crosas, Mercè. 2013. "A Data Sharing Story." *Journal of eScience Librarianship*, 1(3): 7. https://doi.org/10.7191/jeslib.2012.1020.

Crosas, Mercè. 2019. "Dataverse, DataTags, and a decade building a widely-used data repository platform." https://securelysharingdata.com/whitepapers.html (accessed 2020-12-15).

de Montjoye, Yves-Alexandre, César A. Hidalgo, Michel Verleysen, and Vincent D. Blondel. 2013. "Unique in the Crowd: The privacy bounds of human mobility." *Scientific Reports*, 3(1): 1376. https://doi.org/10.1038/srep01376.

Desai, Tanvi, Felix Ritchie, and Richard Welpton. 2016. "Five Safes: Designing data access for research." https://uwe-repository.worktribe.com/output/914745 (accessed 2020-01-30).

Dinur, Irit, and Kobbi Nissim. 2003. "Revealing information while preserving privacy." 202–210. ACM. https://doi.org/10.1145/773153.773173.

Dwork, Cynthia, Adam Smith, Thomas Steinke, and Jonathan Ullman. 2017. "Exposed! A survey of attacks on private data." *Annual Review of Statistics and Its Application*, 4(1): 61–84. https://doi.org/10.1146/annurev-statistics-060116-054123.

Dwork, Cynthia, and Aaron Roth. 2014. "The algorithmic foundations of differential privacy." *Foundations and Trends® in Theoretical Computer Science*, 9(3-4): 211–407. https://doi.org/10.1561/0400000042.

Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam D. Smith. 2016. "Calibrating noise to sensitivity in private data analysis." *Journal of Privacy and Confiden-*

tiality, 7(3): 17–51. https://doi.org/10.29012/jpc.v7i3.405.

Dwork, Cynthia, Nitin Kohli, and Deirdre Mulligan. 2019. "Differential privacy in practice: Expose your epsilons!" *Journal of Privacy and Confidentiality*, 9(2). https://doi.org/10.29012/jpc.689.

Fluitt, Aaron, Aloni Cohen, Micah Altman, Kobbi Nissim, Salome Viljoen, and Alexandra Wood. 2019. "Data protection's composition problem." *European Data Protection Law Review*, 5(3). https://doi.org/10.21552/edpl/2019/3/4.

Foote, Andrew David, Ashwin Machanavajjhala, and Kevin McKinney. 2019. "Releasing earnings distributions using differential privacy: Disclosure avoidance system for post-secondary employment outcomes (PSEO)." *Journal of Privacy and Confidentiality*, 9(2). https://doi.org/10.29012/jpc.722.

Gaboardi, Marco, James Honaker, Gary King, Kobbi Nissim, Jonathan Ullman, and Salil P. Vadhan. 2016. "PSI (ψ): a private data sharing interface." *CoRR*, abs/1609.04340. http://arxiv.org/abs/1609.04340.

Ganta, Srivatsava Ranjit, Shiva Prasad Kasiviswanathan, and Adam D. Smith. 2008. "Composition attacks and auxiliary information in data privacy." 265–273. ACM. https://doi.org/10.1145/1401890.1401926.

Garfinkel, Simson. 2016. "De-Identifying Government Datasets (2nd Draft)." National Institute of Standards and Technology NIST Special Publication (SP) 800-188 (Draft). https://csrc.nist.gov/publications/detail/sp/800-188/draft (accessed 2021-02-01).

Garfinkel, Simson L. 2017. "Modernizing disclosure avoidance: Report on the 2020 disclosure avoidance subsystem as implemented for the 2018 end-to-end test (continued)." https://www2.census.gov/cac/sac/meetings/2017-09/garfinkel-modernizing-disclosure-avoidance.pdf (accessed 2020-12-15).

Garfinkel, Simson L., John M. Abowd, and Christian Martindale. 2019. "Understanding database reconstruction attacks on public data." *Communications of the ACM*, 62(3): 46–53. https://doi.org/10.1145/3287287.

Garfinkel, Simson L., John M. Abowd, and Sarah Powazek. 2018. "Issues encountered deploying differential privacy." *WPES'18*, 133–137. New York, NY, USA:Association for Computing Machinery. https://doi.org/10.1145/3267323.3268949.

Harris-Kojetin, Brian A., Wendy L. Alvey, Lynda Carlson, Steven B. Cohen, Steve H. Cohen, Lawrence H. Cox, Robert E. Fay, Ronald Fecso, Dennis Fixler, Gerald Gates, Barry Graubard, William Iwig, Arthur Kennickell, Nancy J. Kirkendall, Susan Schechter, Rolf R. Schmitt, Marilyn Seastrom, Monroe G. Sirken, Nancy L. Spruill, Clyde Tucker, Alan R. Tupek, G. David Williamson, and Robert Groves. 2005. "Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology." U.S. Federal Committee on Statistical Methodology Research Report. https://nces.ed.gov/FCSM/pdf/spwp22.pdf (accessed 2020-12-15).

Hawes, Michael B. 2020. "Implementing differential privacy: Seven lessons from the 2020 United States Census." *Harvard Data Science Review*, 2(2). https://doi.org/10.1162/99608f92.353c6f99.

Kasiviswanathan, Shiva Prasad, Homin K. Lee, Kobbi Nissim, Sofya Raskhod-

nikova, and Adam D. Smith. 2011. "What can we learn privately?" *SIAM Journal on Computing*, 40(3): 793–826. https://doi.org/10.1137/090756090.

Kenthapadi, Krishnaram, Nina Mishra, and Kobbi Nissim. 2013. "Denials leak information: Simulatable auditing." *Journal of Computer and System Sciences*, 79(8): 1322–1340. https://doi.org/10.1016/j.jcss.2013.06.004.

King, Gary. 2007. "An introduction to the dataverse network as an infrastructure for data sharing." *Sociological Methods & Research*, 36(2): 173–199. https://doi.org/10.1177/0049124107306660.

King, Gary. 2014. "Restructuring the Social Sciences: Reflections from Harvard's Institute for Quantitative Social Science." *PS: Political Science & Politics*, 47(1): 165–172. https://doi.org/10.1017/S1049096513001534.

Machanavajjhala, Ashwin, and Daniel Kifer. 2015. "Designing statistical privacy for your data." *Communications of the ACM*, 58(3): 58–67. https://doi.org/10.1145/2660766.

Muise, Daniel, and Kobbi Nissim. 2016. "Differential Privacy in CDFs." Harvard University Presentation. https://privacytools.seas.harvard.edu/files/dpcdf_user_manual_aug_2016.pdf (accessed 2020-12-15).

Narayanan, Arvind, and Vitaly Shmatikov. 2008. "Robust de-anonymization of large sparse datasets." 111–125. IEEE Computer Society. https://doi.org/10.1109/SP.2008.33.

Narayanan, Arvind, Joanna Huey, and Edward W. Felten. 2016. "A Precautionary Approach to Big Data Privacy." In *Data Protection on the Move*. Vol. 24, , ed. Serge Gutwirth, Ronald Leenes and Paul De Hert, 357–385. Dordrecht:Springer Netherlands. https://doi.org/10.1007/978-94-017-7376-8_13.

Nissim, Kobbi, Aaron Bembenek, Alexandra Wood, Mark Bun, Marco Gaboardi, Urs Gasser, David R O'Brien, Thomas Steinke, and Salil Vadhan. 2018. "Bridging the gap between computer science and legal approaches to privacy." *Harvard Journal of Law & Technology*, 31(2): 687–780. https://privacytools.seas.harvard.edu/publications/bridging-gap-between-computer-science-and-legal-approaches-privacy.

Ramachandran, Aditi, Lisa Singh, Edward Porter, and Frank Nagle. 2012. "Exploring re-identification risks in public domains." 35–42. https://doi.org/10.1109/PST.2012.6297917.

Stemmer, Uri, and Haim Kaplan. 2018. "Differentially private k-Means with constant multiplicative error." 5436–5446. http://papers.nips.cc/paper/7788-differentially-private-k-means-with-constant-multiplicative-error.

Sweeney, Latanya. 1997. "Weaving technology and policy together to maintain confidentiality." *The Journal of Law, Medicine & Ethics*, 25(2-3): 98–110. https://doi.org/10.1111/j.1748-720X.1997.tb01885.x.

Sweeney, Latanya, Mercè Crosas, and Michael Bar-Sinai. 2015. "Sharing Sensitive Data with Confidence: The Datatags System." *Journal of Technology Science*. https://techscience.org/a/2015101601/ (accessed 2020-12-15).

The OpenDP Team. 2020. "The OpenDP white paper." https://projects.iq.harvard.edu/files/opendp/files/opendp_white_paper_11may2020.pdf (accessed 2020-12-15).

The Statutes at Large of the United States of America. 1909. "An Act To provide for the Thirteenth and subsequent decennial censuses." Public Law 61-2. https://www.loc.gov/law/help/statutes-at-large/61st-congress/c61.pdf.

US Census Bureau Center for Economic Studies. n.d.. "US Census Bureau Center for Economic Studies Publications and Reports Page." https://lehd.ces.census.gov/data/pseo_experimental.html (accessed 2020-12-16).

US Social Security Administration. 2011. "Actuarial Life Table: Period Life Table." http://www.ssa.gov/oact/STATS/table4c6.html (accessed 2020-12-15).

Wang, Yu-Xiang. 2018. "Revisiting differentially private linear regression: optimal and adaptive prediction & estimation in unbounded domain." 93–103. AUAI Press. http://auai.org/uai2018/proceedings/papers/40.pdf (accessed 2020-12-15).

Willenborg, Leon, and Ton De Waal. 1996. *Statistical disclosure control in practice.* Vol. 111, Springer Science & Business Media.

Wood, Alexandra, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David R. O'Brien, Thomas Steinke, and Salil Vadhan. 2018. "Differential Privacy: A Primer for a Non-Technical Audience." *Vanderbilt Journal of Entertainment and Technology Law*, 21(1). http://www.jetlaw.org/journal-archives/volume-21/volume-21-issue-1/differential-privacy-a-primer-for-a-non-technical-audience/ (accessed 2019-04-23).

Appendix

A discussion of different technical approaches to disseminating data with differential privacy and key design choices, the implications of differential privacy for data collection, use, and dissemination, and a list of selected tools and resources for implementing differential privacy protections can be found in the Online Appendix at admindatahandbook.mit.edu/book/v1.0/diffpriv.html#diffpriv-appendix