TECH FACTSHEETS FOR POLICYMAKERS

FALL 2020 SERIES

Differential Privacy

AUTHORS

Raina Gandhi (Harvard)

Amritha Jayanti

REVIEWERS

Alexandra Wood, J.D. (Harvard)

Michael B. Hawes (U.S. Census Bureau)

HARVARD Kennedy School
BELFER CENTER
for Science and International Affairs
TECHNOLOGY AND PUBLIC PURPOSE PROJECT

ASH CARTER, TAPP FACULTY DIRECTOR

LAURA MANLEY, TAPP DIRECTOR

The Technology Factsheet Series was designed to provide a brief overview of each technology and related policy considerations. These papers are not meant to be exhaustive.

Technology and Public Purpose Project

Belfer Center for Science and International Affairs
Harvard Kennedy School
79 John F. Kennedy Street, Cambridge, MA 02138

www.belfercenter.org/TAPP

Design and layout by Andrew Facini

Printed in the United States of America

Executive Summary

Differential privacy is a safeguard used to protect an individual's data privacy. It allows for the collection and publication of data patterns and trends, while protecting the privacy of individuals captured in a dataset. Differential privacy is not a tool or method, but rather a criterion or a property that multiple methods can achieve. More specifically, it is a mathematical definition of privacy that quantifies privacy risk. It considers a maximum level of privacy loss, called the *privacy loss parameter*, and manipulates the content of a dataset in order to achieve that level of privacy, while maintaining the utility and accuracy of a dataset.

Differential privacy has clear benefits, particularly that it is robust against a wide range of privacy attacks. Additionally, it is transparent, having the ability to share information about data analyses without negatively affecting individual privacy. It also allows for transparency regarding the margin of error, uncertainty, and other variables that most statistical disclosure limitations do not allow for. As a result, differential privacy has significant potential to enable wide access to data that previously could not be shared, including sensitive data such as medical or financial data.

While these benefits exist, there are also several challenges associated with differential privacy. For example, differential privacy yields better accuracy for larger datasets than it does for smaller ones. There are also no accepted guidelines for determining the privacy loss parameter, and as a result, companies are currently using differential privacy with weaker privacy protection than the research community finds acceptable. There are not many production-ready tools and expert practitioners available in this area yet. Even so, differential privacy has been growing in popularity. From the public sector to the private sector—including organizations like the U.S. Census Bureau, Google, Facebook, Uber, Amazon, and Microsoft—it is being leveraged to protect sensitive data against potential privacy attacks.

However, policy is falling short. Most existing privacy laws and regulations focus on personally identifiable data and rely on legal concepts that have less relevance to differential privacy. There are currently no guidelines, at a national or an international level, on how to safely implement differential privacy (or any other approach to digital privacy for that matter).

There is an urgent need, as well as a tremendous opportunity, to update the privacy framework, specifically to promote investment in and adoption of modern understandings of privacy and new tools for privacy protection, such as differential privacy. The U.S. has the opportunity to lead the way not just on the technology, but also on policy to ensure that differential privacy's challenges are addressed—and its benefits are fully captured.

What is Differential Privacy?

Differential privacy is a rigorous mathematical definition of privacy, which allows for privacy risk to be quantified. It is not a specific tool or a method, but rather a criterion or a property that multiple methods can achieve.

Essentially, differentially private tools introduce "noise," or random information, into a dataset so that it is impossible to tell whether a specific individual's information was used or not. As a result, the output is no longer 100% accurate, but instead is approximately the same as the true value. Thus, differential privacy guarantees that individuals will experience essentially the same privacy risk, whether or not they are included in a differentially private analysis[1]. Further, in the case of a privacy attack, the random noise introduced through different privacy allows for *plausible deniability* by individuals represented in a data set.

Differential privacy enables a quantifiable and provable guarantee of privacy protection. By quantifying privacy risk, differential privacy overcomes the weaknesses of traditional approaches to privacy, such as *data anonymization*, and provides protection against a wide range of data attacks. Additionally, with differential privacy, privacy-preserving strategies can be compared and ranked based on effectiveness.

How it Works

Imagine a database that contains 100 individuals' medical records, and 20 of them have diabetes. A malicious actor wants to know if person A has diabetes, and they have already learned that 19 of the other 99 people viewed in the sample are diabetic. By querying the database and learning that 20 people in the sample are diabetic, the malicious actor learns that person A is diabetic. (This deductive information extraction is called a *differentiated attack*.[2])

Instead, imagine that the database is provided through a differentially private query system. Now, when the actor queries the database, the algorithm returns the truth (20) plus some random noise. The output could be 21, or 25, or 19, based on the random noise. Now even if the malicious actor knew everything about the other 99 people, they cannot say for sure that person A has diabetes. Person A's privacy is protected, since they have essentially the same privacy risk whether they're a part of the dataset or not. This holds for *any* individual and *any* dataset, regardless of how unusual or different an individual's data may be.

1 Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., O'Brien, D., Steinke, T., & Vadhan, S. (2018). Differential Privacy: A Primer for a Non-Technical Audience. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3338027

2 Zhu, T. (n.d.). Explainer: What is differential privacy and how can it protect your data? The Conversation. Retrieved August 5, 2020, from http://theconversation.com/explainer-what-is-differential-privacy-and-how-can-it-protect-your-data-90686

However, imagine that the malicious actor queries the database again and again. Though the output will vary, eventually the malicious actor will realize it is anchored around 20, and they can estimate with a high likelihood that person A has diabetes. Each additional query reveals more information about the sensitive data and thus incurs privacy loss. Thus, risk of a privacy breach increases with repeated queries to a differentially private database.[3] To prevent this, data administrators impose a maximum privacy loss, also known as a **privacy budget**. When the number of queries reaches this threshold, the curator stops answering queries, thus imposing an upper bound on privacy risk.[4]

A **privacy loss parameter** (ε) determines the upper bound for privacy loss. It represents a tradeoff between privacy and accuracy, since this parameter determines how much noise will be added to the dataset. A smaller ε means greater privacy protection, but less accurate output.[5] A large ε results in a more useful analysis, but less privacy protection. If ε is set too high, an analysis can still technically be differentially private but result in an unacceptable level of privacy risk. Choosing ε is the most important decision when applying differential privacy.[6]

The Appeal of Differential Privacy

Difference from Existing Methods. Differential privacy is the first method that provides a provable privacy guarantee about privacy risk, and the only method to quantify that potential privacy loss. Current methods, such as k-anonymity and other statistical disclosure limitation methods, typically rely on generalizing or suppressing data by removing personally identifying information or coarsening the data.

However, many researchers have shown that data that has been de-identified using current methods can often be re-identified by linking them to other publicly available datasets. Latanya Sweeney famously found that three pieces of data—zip code, gender, and date of birth—are enough to uniquely identify 87% of the U.S. population.[7] In 2007, Netflix released 50,000 subscribers' anonymous movie ratings. Researchers were able to re-identify subscribers by using public IMDB ratings, learning not just their movie preferences but also their apparent political and social views.[8]

3 Dwork, C., & Roth, A. (2013). The Algorithmic Foundations of Differential Privacy. Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042

4 Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., ... Vadhan, S. (2018). Differential Privacy: A Primer for a Non-Technical Audience. Vanderbilt Journal of Entertainment & Technology Law, 21, 209–276. https://doi.org/https://doi.org/10.2139/ssrn.3338027

5 Ibid.

6 Ibid.

7 Sweeney, L. (2000). Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3.

8 Narayanan, A., & Shmatikov, V. (2008). 2008 IEEE Symposium on Security and Privacy. In Robust De-anonymization of Large Sparse Datasets. IEEE. https://ieeexplore.ieee.org/document/4531148.

Value of Differential Privacy. Enter differential privacy. Unlike traditional methods, differential privacy is robust against a wide range of privacy attacks, including ones that are currently unknown. It is also robust under composition, meaning that combining the results of multiple differentially private analyses will not reveal individual information.[9] Where traditional methods have to hide how much the data has been transformed, differential privacy is transparent, with no computation or parameters to hide. Users can thus have a sense of how much the data has been transformed through the privacy loss parameter, without negatively affecting privacy.

As a result, differential privacy can enable access to data that typically cannot be shared, such as sensitive medical data that may help researchers better understand disease progression.

It can also help protect individuals from the wide range of data privacy risks, from voice assistants in the home to transaction data to browser data. Privacy breaches have grown increasingly common, and advancing capabilities such as facial recognition, deepfakes, surveillance, and more have only increased the risks. With differential privacy, machine learning and artificial intelligence can be trained on differentially private datasets, providing similar societal value at a fraction of the privacy risk.

However, an important note is that differential privacy protects against privacy attacks, such as re-identification, record linkage, and differencing attacks. **It does not protect against security attacks**, which seek to gain unauthorized access into a system.[10] This distinguished differential privacy from digital security approaches, such as encryption. It is important to note that the two systems can be used together. It is important to both protect sensitive data using security controls (preventing unauthorized access) and to protect the privacy of individuals once access to the sensitive data is granted to an authorized third party.

Types of Differential Privacy: Curator vs. Local Models

In a **curator model**, a database administrator (the "curator") has access to a database which includes private data. This administrator uses the database to generate differentially private data summaries. This means that the database itself does not satisfy differential privacy, but differentially private analyses run on the data yield differentially private output.

In contrast, a **local model** ensures differential privacy at the point of data collection. Individuals may be required to answer questions about their own data in a differentially private manner. For example, a respondent answering a true/false question may flip a coin: if the coin is heads, they answer truthfully; if the coin is tails, they flip another coin and answer "yes" for heads and "no" for tails.

9 Vadhan, S. (2017). The Complexity of Differential Privacy. In Y. Lindell (Ed.), Tutorials on the Foundations of Cryptography (pp. 347–450). Springer International Publishing. https://doi.org/10.1007/978-3-319-57048-8_7

10 "Differential Privacy: A Primer for a Non-Technical Audience."

By sending the data curator noisy data, local models trust no one—and this is the key distinction between the two models. The curator model assumes the existence of trusted party which collects personal information, produces its outputs by applying differentially private computations; the party is trusted to otherwise protect the data from potential adversaries. In contrast, the local model is a fully distributed model which assumes no trust. In the case of local models, the individual answers are less useful on their own, but are useful in aggregate.

Although local models offer very strong privacy protections, they come at a large cost to statistical accuracy. In application, this means that it is mostly only useful for massive datasets, such as the user bases at large technology companies like Google and Facebook.

Challenges with Differential Privacy

- **Applying strict differential privacy requirements to real-world data.** Many differentially private approaches only apply to particular data or types of data, such as univariate data or categorical data, yet real-world data often comes in other forms. As a result, some advocate for using relaxed differentially private approaches instead, though these present additional privacy risks.[11]

- **No standards for setting privacy loss parameter.** There are no accepted guidelines or frameworks for how to set the privacy loss parameter **ε**, though multiple papers have been written on it. Without standards or best practices, data is likely being shared without adequate privacy protection *or* without an appropriate amount of data utility.[12]

- **Computational complexity.** Several theoretical approaches require significant computational resources. This makes differential privacy challenging and even unfeasible for the average data curator with limited computational resources. This is a particular issue for techniques that seek to release entire data sets.[13]

- **Reduces utility of small datasets.** Differential privacy is well-suited for very large data sets, where the additional noise does not meaningfully affect the data's accuracy or utility. For smaller datasets, the tradeoff between privacy and utility becomes more difficult. Record-level data in particular are difficult to protect meaningfully while still "leaving the data useful for unspecified analytical purposes."[14]

11 Snoke, J., & McKay Bowen, C. (2019, March 1). Differential Privacy: What Is It? Amstat News. https://magazine.amstat.org/blog/2019/03/01/differentialprivacy/.

12 Dwork, C., Kohli, N., & Mulligan, D. (2019). Differential Privacy in Practice: Expose your Epsilons! Journal of Privacy and Confidentiality, 9(2), Article 2. https://doi.org/10.29012/jpc.689

13 "Differential Privacy: What Is It?"

14 Ruggles, S. (2018). IPMUS. University of Minnesota.

Current Applications and Market Landscape

Differential privacy is becoming increasingly popular in practice**.** From the public sector to the private sector, a number of organizations are implementing differential privacy in an attempt to securely leverage and share data insights.

U.S. Census Bureau

One of the largest-scale implementations of differential privacy, and one with serious political and economic implications, was conducted by the **U.S. Census Bureau** for the 2020 Decennial Census.[15] Following their 2020 Census, the Bureau published reflections and observations on a "number of questions about the proper balance between privacy and accuracy in official statistics, the prioritization of certain data uses over others, and the future of statistical offices and their data products."[16]

The Census Bureau also uses approximate differential privacy for commuter data called OnTheMap and uses differential privacy for the Post-Secondary Employment Outcomes data and Opportunity Atlas on childhood social mobility.[17]

Big Tech & Other Private Sector Actors

Large technology companies such as Microsoft, Apple, Uber, and Google have been investing significantly, embedding differential privacy within their products and investing in ongoing research and development.

- **Microsoft.** Microsoft uses differential privacy "to protect user privacy in several applications, including telemetry in Windows, advertiser queries in LinkedIn, suggested replies in Office, and manager dashboards in Workplace Analytics."[18]

- **Apple.** Apple uses local differential privacy on iPhones to improve features such as: QuickType suggestions, emoji suggestions, lookup hints, Safari, and health type usage.[19] However, researchers suggest that Apple's privacy loss parameter is set too high to offer adequate protection.[20]

15 Hawes, M. 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

16 Ibid.

17 "Differential Privacy: What Is It?"

18 Manager, S. B. P. P., Bird, S., & *, N. (2020, June 18). *Introducing the new differential privacy platform from Microsoft and Harvard's OpenDP*. Open Source Blog. https://cloudblogs.microsoft.com/opensource/2020/05/19/new-differential-privacy-platform-microsoft-harvard-opendp/.

19 Apple. Apple Differential Privacy Technical Overview.

20 Tang, J., Korolova, A., Bai, X., Wang, X., & Wang, X. (2017). Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12. *Arxiv*, *Preprint*. https://arxiv.org/pdf/1709.02753.pdf.

- **Google.** Google implemented a local differential privacy method in Chrome called RAPPOR (Randomized Aggregatable Privacy-Preserving Ordinal Response) and uses differentially private methods in Google Fi and Google Map's business busy hours and dish popularity.[21,22] Google claims the differential privacy system in Chrome has an epsilon of 2 in most cases, and a lifetime ceiling of 8 to 9.[23] Google also uses differential privacy to produce COVID-19 mobility reports[24] and for its open source library for machine learning, TensorFlow[25].

- **Uber.** Uber uses approximate differential privacy.[26] They have created an open-source tool in collaboration with UC-Berkeley that reduces the computational resources associated with implementing differential privacy. This technique, called elastic sensitivity, enables them to efficiently calculate how sensitive a query is.[27]

- **Facebook.** Facebook uses differential privacy to share data with researchers on election misinformation, movement of individuals affected by the Australian bushfires, and movement of individuals during COVID-19.[28,29]

- **Amazon.** Amazon has tested differential privacy in analyses of customer-provided textual data.[30,31]

- **LinkedIn**. LinkedIn has tested differential privacy to provide marketing analytics based on members' data.[32]

- **Snapchat.** uses differential privacy to train machine learning models.[33]

- There are also a number of small and large private sector companies investing in differential privacy. **Privitar**, a London-based enterprise software startup, claims to have developed a differential privacy product called Privitar Lens.[34]

21 Greig, J. (2019, September 9). *Google hopes to protect users with open source differential privacy library*. TechRepublic. https://www.techrepublic.com/article/google-hopes-to-protect-users-with-open-source-differential-privacy-library/.

22 *Enabling developers and organizations to use differential privacy*. Google Developers Blog. (2019, September 5). https://developers.googleblog.com/2019/09/enabling-developers-and-organizations.html.

23 Fingas, R. (2017, September 15). *Apple's 'differential privacy' still collects too much specific data, study says*. AppleInsider. https://appleinsider.com/articles/17/09/15/apples-differential-privacy-still-collects-too-much-specific-data-study-says.

24 Google. (2020). COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility/.

25 *Introducing TensorFlow Privacy: Learning with Differential Privacy for Training Data*. The TensorFlow Blog. (2019, March 6). https://blog.tensorflow.org/2019/03/introducing-tensorflow-privacy-learning.html.

26 "Differential Privacy: What Is It?"

27 Uber Privacy & Security. (2018, March 13). *Uber Releases Open Source Project for Differential Privacy*. Medium. https://medium.com/uber-security-privacy/differential-privacy-open-source-7892c82c42b6.

28 Hutchinson, A. (2020, June 3). *Facebook Outlines New Differential Privacy Framework to Protect User Information in Shared Datasets*. Social Media Today. https://www.socialmediatoday.com/news/facebook-outlines-new-differential-privacy-framework-to-protect-user-inform/579167/.

29 Herdağdelen, B. A., Herdağdelen, A., & Dow, A. (2020, June 3). *Protecting privacy in Facebook mobility data during the COVID-19 response*. Facebook Research. https://research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response/.

30 Diethe, T. (2020, January 23). *Preserving privacy in analyses of textual data*. Amazon Science. https://www.amazon.science/blog/preserving-privacy-in-analyses-of-textual-data.

31 Feyisetan, O., Balle, B., Drake, T., & Diethe, T. *Privacy- and Utility-Preserving Textual Analysis via Calibrated Multivariate Perturbations*. Amazon Science. https://www.amazon.science/publications/privacy-and-utility-preserving-textual-analysis-via-calibrated-multivariate-perturbations.

32 Rogers, R., Subramaniam, S., Peng, S., Durfee, D., Lee, S., Kancha, S. K., ... Ahammad, P. (2020). LinkedIn's Audience Engagements API: A Privacy Preserving Data Analytics System at Scale. *Arxiv*. https://arxiv.org/pdf/2002.05839.pdf.

33 Pihur, V., Korolova, A., Liu, F., Sankuratripati, S., Yung, M., Huang, D., & Zeng, R. (2018). Differentially-Private "Draw and Discard" Machine Learning. *Arxiv*. https://arxiv.org/pdf/1807.04369.pdf.

34 *Privitar Authors Review of Differential Privacy for the Office for National Statistics' Quality Review*. Privitar. (2020, October 30). https://www.privitar.com/press-releases/privitar-authors-review-of-differential-privacy-for-the-office-for-national-statistics-quality-review/.

Additionally, there are a few open-source libraries to allow for collaborative and accessible tools to implement differential privacy standards across various companies, organizations, and project. For example, **Google** offers a differential privacy tools library; and **SmartNoise** also offers an open-source toolkit.[35]

Current Governance and Regulation

Currently, there are no regulations or governance schemes in place that specifically address differential privacy techniques. There are, though, related or relevant privacy-focused laws:

Related U.S. Governance and Regulation

- **13 U.S. Code § 9 – Information as confidential; exception[36].** Title 13 Section 9 requires the protection of confidential information by the U.S. Census Bureau.
- **Confidential Information Protection and Statistical Efficiency Act (CIPSEA)[37].** CIPSEA establishes unifo rm confidentiality protections for information collected for statistical purposes by U.S. statistical agencies. It allows some data sharing between the Bureau of Labor Statistics, Bureau of Economic Analysis, and Census Bureau.
- **Family Educational Rights and Privacy Act (FERPA)[38].** FERPA, through the U.S. Department of Education, sets policy around the privacy of student education records.
- **Fair Credit Reporting Act (FCRA)[39].** FCRA sets policy, through the U.S. Federal Trade Commission, for the privacy of data held by consumer reporting companies.
- **Health Insurance Portability and Accountability Act (HIPAA)[40].** HIPAA sets policy, through the U.S. Department of Health and Human Services' Office for Civil Rights, regarding healthcare data, defines a set of national standards for protection of certain health information.

35 *SmartNoise*. http://www.smartnoise.org/.

36 Legal Information Institute. *13 U.S. Code § 9 - Information as confidential; exception*. Legal Information Institute. https://www.law.cornell.edu/uscode/text/13/9.

37 Bureau of Labor Statistics, Confidential Information Protection and Statistical Efficiency. https://www.bls.gov/bls/cipsea.pdf.

38 US Department of Education (ED). (2020, December 15). *Family Educational Rights and Privacy Act (FERPA)*. Home. https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html.

39 *Fair Credit Reporting Act*. Federal Trade Commission. (2020, March 4). https://www.ftc.gov/enforcement/statutes/fair-credit-reporting-act.

40 Centers for Disease Control and Prevention. (2018, September 14). *Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. Centers for Disease Control and Prevention. https://www.cdc.gov/phlp/publications/topic/hipaa.html.

Related International Regulation

- **EU General Data Protection Regulation (GDPR)[41].** GDPR governs the processing of personal data of individuals in the EU for professional or commercial activities, whether by individuals, companies, or organizations.

Related International Governance

- **Guidelines Governing the Protection of Privacy & Transborder Flow of Personal Data[42].** These guidelines are an internationally agreed upon set of privacy principles set up through OECD.

Challenges with Regulation

Significant effort has gone into creating laws and frameworks on privacy. However, these approaches often rely on concepts such as personally identifiable information and identification, limiting the scope to which these regulations address differential privacy. Finally, regulatory protections "typically extend only to personally identifiable information; information not considered personally identifiable is not protected."[43] These concepts are not precisely defined.

As previously mentioned, there are no best practices or guidelines for how to appropriately set the privacy loss parameter for any given application. Setting the privacy loss parameter is arguably a policy- and context-based decision; there need to be best practices developed by stakeholders, including policymakers and researchers, as research and implementation progresses.

41 *Official Legal Text*. General Data Protection Regulation (GDPR). (2019, September 2). https://gdpr-info.eu/.

42 *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. OECD. https://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm.

43 "Differential Privacy: A Primer for a Non-Technical Audience."

Public Purpose Considerations

There are several public purpose considerations regarding differential privacy that are important to consider:

- **Data Utility & Accuracy.** The main concern with differential privacy is the tradeoff between data utility and individual privacy. If the privacy loss parameter is set to favor utility, the privacy benefits are lowered (less "noise" is injected into the system); if the privacy loss parameter is set to favor heavy privacy, the accuracy and utility of the dataset are lowered (more "noise" is injected into the system). It is important for policymakers to consider the tradeoffs posed by differential privacy in order to help set appropriate best practices and standards around the use of this privacy preserving practice, especially considering the diversity in organizational use cases.

  It is worth noting, though, that decreased accuracy and utility is a common issue among all statistical disclosure limitation methods and is not unique to differential privacy. What is unique, however, is how policymakers, researchers, and implementers can consider mitigating against the risks presented through this tradeoff.

- **Data Privacy & Security.** Differential privacy provides a quantified measure of privacy loss and an upper bound and allows curators to choose the explicit tradeoff between privacy and accuracy. It is robust to still unknown privacy attacks. However, it encourages greater data sharing, which if done poorly, increases privacy risk. Differential privacy implies that privacy is protected, but this depends very much on the privacy loss parameter chosen and may instead lead to a false sense of security.[44] Finally, though it is robust against unforeseen future privacy attacks, a countermeasure may be devised that we cannot predict.

44 Dwork, C., Kohli, N., & Mulligan, D. (2019). Differential Privacy in Practice: Expose your Epsilons! *Journal of Privacy and Confidentiality*, *9*(2), Article 2. https://doi.org/10.29012/jpc.689

Appendix:
Key Questions for Policymakers

Efficacy & Performance Standards

- How should the privacy loss parameter be set? What are best practices or guidelines?
- Who should create the standards? How could a working group represent cross-sector perspectives?
- What is the minimum acceptable level of privacy?
- Should this vary by the types of data (e.g. medical, browser history, smart data)?
- If so, what criteria need to be considered to determine the level of privacy necessary to release different types of data (for example, 15-min electric smart meter data vs. monthly electric data)?

Individual Security & Risk Communication

- What sort of review should be done before any differentially private but sensitive datasets are made available? Who should be involved?
- How can researchers be involved to check whether the data is still useful?
- How can communities be involved in the decision to share their data? Individuals with the greatest privacy risk?
- How do we determine who faces the greatest risk?
- How do we compare and address individuals with high *risk* versus high potential of *harm*?
- How will the research community ensure that research is appropriately mentioning the level of accuracy expected in results from differentially private analyses?
- What are best practices for notifying individuals about a privacy breach?

Accessibility & Adoption

- What types or sources of data are the highest opportunity (greatest benefit and least risk)?
- What margin of error are we comfortable with when making decisions based on differentially private analyses? How should we balance the tradeoff between accuracy and privacy?

- In what scenarios is relaxing the requirements of differential privacy acceptable? What level of relaxation is appropriate? For example, how strictly differentially private should a dataset or query be if it contains medical information?

Governance & Oversight

- How can the regulatory/political community best keep up with the academic or technological community?
- What legal concepts should underpin privacy, instead of identification and personally identifiable information?
- If there are minimum standards for differential privacy or for privacy loss parameters, who should enforce them? How?
- How will we know when privacy has been breached on differentially private datasets?