Comments of World Privacy Forum regarding the Draft NIST Special Publication (SP) 800-226, Guidelines for Evaluating Differential Privacy Guarantees

Sent via email to privacyeng@nist.gov

NIST Information Technology Laboratory
Computer Security Research Center
100 Bureau Drive
Gaithersburg MD 20899

25 January 2024

The World Privacy Forum welcomes the opportunity to comment on NIST's Draft Special Publication (SP) 800-226, *Guidelines for Evaluating Differential Privacy Guarantees* (Dec. 11, 2023), https://www.nist.gov/news-events/news/2023/12/nist-offers-draft-guidance-evaluating-privacy-protection-technique-ai-era, and https://csrc.nist.gov/pubs/sp/800/226/ipd.

The World Privacy Forum is a non-partisan 501(c)(3) public interest research group focused on conducting research, analysis, and education in the area of privacy and complex data ecosystems and their governance, including in the areas of identity, AI, health, and others. WPF works extensively on privacy and governance across multiple jurisdictions, including the U.S., India, Africa, Asia, the EU, and additional jurisdictions. For more than 20 years WPF has written in-depth, influential research regarding systemic medical identity theft, India's Aadhaar identity ecosystem —peer-reviewed work which was cited in the landmark Aadhaar Privacy Opinion of the Indian Supreme Court — and The Scoring of America, an early and influential report on machine learning and consumer scores. Most recently, WPF published a report on AI Governance Tools that establishes the beginnings of an evaluative environment for these tools. WPF co-chairs the UN Statistics Data Governance and Legal Frameworks working group, and is co-chair of the WHO Research, Academia, and Technical Constituency. At the OECD, WPF researchers participate in the OECD.AI AI Expert Groups, among other activities. WPF participated in the core group of AI experts that collaborated to write the OECD

Recommendation on Artificial Intelligence, now widely viewed as the leading normative principles regarding AI. WPF research on complex data ecosystems governance has been presented at the National Academies of Science and the Royal Academies of Science. See our reports and other data at World Privacy Forum: https://www.worldprivacyforum.org.

Regarding the NIST Draft Special Publication (SP) 800-226, *Guidelines for Evaluating Differential Privacy Guarantees*, we find the document generally well-written and clear enough for non-technical readers to understand – more or less. None of our criticisms should be taken as detracting from our approval of the quality and clarity of the writing, which is excellent. Further, we take no issue with the goal of establishing guidelines to define and evaluate differential privacy. We agree that there is a role, albeit a bounded one, for differential privacy in protecting personally identifiable information (PII).

The Draft Guidelines state that they seek to help "policymakers, business owners, product managers, IT technicians, software engineers, data scientists, researchers, and academics." [Page 1.] We think that the document's shortcomings relate mostly to the targeted audience outside of those with technical prowess. We do not think the document in its current form would be easily accessible to a number of policymakers. For example, a "query model" and a "privacy loss budget" are well understood terms by technologists working in differential privacy, but these terms may lose their contextually precise meaning for some policymakers, depending on their level of expertise. [Page 4, Pyramid.]

Many of our substantive problems with the Draft relate to the language used in particular areas of the draft. We recognize that some of the objectionable language may reflect existing usage in the technical field of differential privacy. NIST may not be responsible for the existing normative practices and language describing the practices, but we can only comment on the draft as it is. We believe that most of the problems we have articulated in these comments can be ameliorated with thoughtful revisions.

I. Definitional problems in the Draft language regarding contribution of data

Page 6 of the document states a Key Takeaway:

> "Differential privacy promises that the chance of an outcome is about the same whether or not you contribute your data."

We find this a curious statement. Who is the "you" referenced here? Is the promise of differential privacy supposed to reassure a data subject who contributes data? We can set aside the fact that most data subjects will not understand differential privacy. Further, data subjects hardly ever "contribute" their data to anyone. For any given individual, tens of thousands (and perhaps hundreds of thousands) of businesses, organizations, government agencies, and other institutions possess that individual's personally identifiable information. Entire industries – aside from the many academic and other researchers who also utilize PII – exist in whole or in part to collect, compile, process, and/or profit from PII.

Much data activity today happens without the effective knowledge or knowing consent of the data subject. Large categories of PII in the United States do not have legislated privacy protections. U.S. privacy laws cover a small fraction of consumer data or of organizations that process consumer data. Even when there is a privacy law, the data ecosystems and processes

in place today often escape the typically inadequate limits of the law. Further, much PII is sold, rented, traded, or otherwise exchanged by those who hold or process the data. There are additional research uses for PII, which may occur in certain health and other research contexts. In short, PII is not always intentionally "contributed." Further, data that is used with the intent to f effectuate a financial transaction, can be subjected to many downstream uses. It is a difficult argument in today's world to segregate who contributed what data, and with what level of knowingness, except in certain use cases, such as human subject research subject to the Common Rule.[1]

None of these aspects of the privacy landscape in the U.S. is NIST's fault, nor do we expect NIST to effectuate sweeping improvements across all ecosystems. Nevertheless, NIST has an obligation to describe the status of PII in this document with appropriate recognition of the realities of PII processing. We see no justification for talking about an unspecified "you" "contributing" "your data" to anyone. If the "you" is any or all of the PII data processors in the U.S., the PII they hold should not be described as "their" data to be "contributed" as they see fit.

The view of privacy and PII processing reflected in the draft document's statement quoted above may be the most disquieting part of the entire draft. The understanding how how privacy is effectuated in reality in today's ecosystems is central to a policy-level and technical level of understanding of privacy. This point is important enough that we will provide additional background information about this issue.

We note that many deep machine learning and other analytic activities may no longer need to rely on PII in the same way they did 10 or 15 years ago, but nevertheless the outputs can be attributed to an individual, a household, a census block, or a group of people as defined by certain characteristics or activities.

We begin with the factual basis the data broker ecosystem provides to us regarding this problem of articulating privacy as primarily managing PII more effectively. Data brokers are not a new topic. There have been extraordinary reports about data brokers and harms resulting from data brokers for decades now. Chris Jay Hoofnagle, now a Berkeley Law Professor, in 2003 published *Big Brother's Little Helpers.* This report stands as the first major modern reporting of data broker activities.[2] In this report, Professor Hoofnagle documents the myriad ways that the U.S. government relies on data collected by third parties, data that has levels of accuracy that are non-transparent and questionable. The World Privacy Forum has also published multiple reports about data brokers, as well as testifying before Congress about the topic.[3] One report

---

[1] Federal Policy for Protection of Human Subjects (Common Rule), U.S. Department of Health and Human Services, https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html.

[2] Chris Jay Hoofnagle, *Big Brother's Little Helpers: How ChoicePoint and other commercial data brokers collect and package your data for law enforcement*, EPIC, 2003. https://lawcat.berkeley.edu/record/1118906 .

[3] Robert Gellman and Pam Dixon, *Data Brokers and the Federal Government: A new front i the battle for privacy opens*, 30 October 2013. https://www.worldprivacyforum.org/2013/10/report-data-brokers-and-the-federal-government-a-new-front-in-the-battle-for-privacy-opens/ . See also links to Congressional testimony: https://www.worldprivacyforum.org/category/congressional-testimony/.

WPF published was the *Scoring of America* in 2014, which documented individual and group harms from a variety of machine learning (ML) scoring mechanisms.[4]

In the *Scoring of America* report, WPF did not focus on lists of identifiable consumers that data brokers were selling, because that practice was and still is receding. A new form of data brokering was becoming prominent, which was scoring people, and groups of people, and classifying them into categories and types of people, consumers, purchasers, etc. An era of AI and machine learning was coming, and the *Scoring of America* report is a benchmark for the beginning of that era. The report forms a bridge between more analog data broker ecosystems and present-day data broker ecosystems.

The summary of the report states:

> This report highlights the unexpected problems that arise from new types of predictive consumer scoring, which this report terms consumer scoring. Largely unregulated either by the Fair Credit Reporting Act or the Equal Credit Opportunity Act, new consumer scores use thousands of pieces of information about consumers' pasts to predict how they will behave in the future. Issues of secrecy, fairness of underlying factors, use of consumer information such as race and ethnicity in predictive scores, accuracy, and the uptake in both use and ubiquity of these scores are key areas of focus.
>
> The report includes a roster of the types of consumer data used in predictive consumer scores today, as well as a roster of the consumer scores such as health risk scores, consumer prominence scores, identity and fraud scores, summarized credit statistics, among others. The report reviews the history of the credit score – which was secret for decades until legislation mandated consumer access -- and urges close examination of new consumer scores for fairness and transparency in their factors, methods, and accessibility to consumers.

The World Privacy Forum defines a consumer score as follows:

> A consumer score that describes an individual or sometimes a group of individuals (like a household), and predicts a consumer's behavior, habit, or predilection. Consumer scores use information about consumer characteristics, past behaviors, and other attributes in statistical models that produce a numeric score, a range of scores, or a yes/no. Consumer scores rate, rank, or segment consumers. Businesses and governments use scores to make decisions about individual consumers and groups of consumers. The consequences can range from innocuous to important. Businesses and others use consumer scores for everything from predicting fraud to predicting the health care costs of an individual to eligibility decisions to almost anything.

It is critical to understand this particular aspect of the evolution of data brokering: it has moved toward scoring consumers in clusters, groups, households, and — yes, sometimes individually. Consumer scoring is already more widespread than most people realize. Thousands of

---

[4] Pam Dixon and Robert Gellman, *The Scoring of America: How secret consumer scores threaten your privacy*, World Privacy Forum, April 2014. https://www.worldprivacyforum.org/2014/04/wpf-report-the-scoring-of-america-how-secret-consumer-scores-threaten-your-privacy-and-your-future/ .

consumer scores exist, perhaps more. How many Americans have them? Almost all do. Minors are less likely to be scored than adults, although they, too can have or influence some consumer scores. For example, household scores often reflect interests and activities of minors.

Among American adults, each individual with a credit or debit card or a bank account is likely to be the subject of one or more scores. Fed by the masses of consumer data now available, consumer scoring is quickly becoming a simple shorthand to make sense of a sea of information. It is not likely that the practice will abate soon, and it would be extremely difficult or impractical for most adult U.S. residents to avoid all scoring.

In 2023, the U.S. Office of the Director of National Intelligence (ODNI) released and declassified a report discussing problems with the U.S. use of commercial data brokers.[5] While this action was the right thing to do, it inadvertently documented the practices that are as of yet still not constrained by appropriate guardrails. The ODNI report is helpful in several respects in determining the contours of the modern data broker ecosystem. The report, *The Office of the Director on National Intelligence Senior Advisory Group Panel on Commercially Available Information*, approved for release 5 June 2023, documents that the U.S. government intelligence community purchased commercially available information, which is described in the report as clearly providing intelligence value. The ODNI report states that commercially available information:

> "…clearly provides intelligence value, whether considered in isolation and/or in combination with other information, and whether reviewed by humans and/or by machines. It also raises significant issues related to privacy and civil liberties. The widespread availability of CAI regarding the activities of large numbers of individuals is a relatively new, rapidly growing, and increasingly significant part of the information environment in which the IC must function." [6]

Traditionally, consumer opt-out mechanisms have been seen as a solution to the problems raised by data risks. However, this technique has become less and less feasible. In January 2023 NIST published the *NIST AI Risk Management Framework*.[7] Examining this framework, it becomes clear that opting out can no longer be considered an effective remedy or mitigation for data broker challenges: the overarching data ecosystem is too complex, has too many layers, and the data is becoming more diffuse.

Solving the problems relating to privacy today must be seen as a total system of both individuals and groups of people, and the solutions must encompass these ideas as part of a larger ecosystem. For example, the total ecosystem of data, technology, and AI-fueled analysis that wraps around these varying systems. The systems of today are stunningly advanced, facilitating analysis of even data that has been de-identified and is not legally defined as PII. To give an

---

[5] *The Office of the Director on National Intelligence Senior Advisory Group Panel on Commercially Available Information*, approved for release 5 June 2023. https://www.odni.gov/files/ODNI/documents/assessments/ODNI-Declassified-Report-on-CAI-January2022.pdf .

[6] Id note 4.

[7] *NIST AI Risk Management Framework and Playbook*, NIST, January 2023. https://www.nist.gov/itl/ai-risk-management-framework .

example of why this is already a non-trivial problem, consider how significant of a challenge it is to keep track of data after it has been replicated, split, and /or fed into algorithmic and machine learning systems. Individual's data or household or census block data might be incorporated into several different intersecting models and data sets, which are then crunched into a score. The score reflecting these groups and households then gets rolled into yet more algorithms and systems and scores. The permutations are extensive, and it is not too much to state that they can be profoundly complex. Differential privacy has a role to play, as do many other technologies. But the landscape is complex, and it will take a wide range and mix of solutions - technical, legal, and policy to create meaningful, sustainable improvements.

II. Substantive definitional problems regarding inferential disclosure: groups, households, families, and other collections of individuals' privacy interests are not acknowledged

On page 8, the draft proposes a new definition for inferential disclosure:

> "…access to a statistical database should not enable one to learn anything about an individual that could not be learned without that individual's data."

Our concern here is with language and an apparent limited concept of privacy that this statement suggests. Privacy is not just about individuals. Groups, households, families, neighborhoods, and other collections of individuals have privacy interests that go beyond the interest of any given individual member. Admittedly, group privacy concepts are not sharply defined or easily addressed, but those concepts exist and must be recognized. We acknowledge and noted the areas of the draft that discusses systemic bias against groups in section 3.3.1. That discussion is appropriate and welcome, but we note that the discussion in 3.3.1 addresses the *use of data* rather than the *fundamental notion of the rights and interests* that privacy protects. It is this that we are referring to.

Differential privacy can be quite brittle in actual implementations in a number of ways — including bias effects, as discussed in the Draft, and appropriately so. This became a well-studied issue during the 2020 U.S. Census.[8] This is a complex topic, but also it is a critically important point, and we believe this idea needs to be acknowledged in the report at the definitional level. To adopt a definition of privacy as involving PII without acknowledging other critically important aspects of privacy relating to the rights of groups of people that exist would have potential normative implications which would be problematic for effectuating privacy going forward as our data and technical ecosystems move into a more advanced AI era. This issue is important enough that we are providing additional background information here.

First, germane to these comments is the question of the existence of privacy of groups of people, and the rights of those groups to establish group privacy norms. Privacy is often seen in

---

[8] Michael B. Hawes, *Implementing differential privacy: Seven lessons from the 2020 United States Census,* HDSR, Issue 2.2 Spring 2020. https://doi.org/10.1162/99608f92.353c6f99. See also, 2020 Decennial Census: Processing the CountL Disclosure avoidance modernization, US Census Bureau, 2020. https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html.

terms of individual data rights, such as the individual right to deletion, the individual right to not be included in certain data sets, and so forth. While the conception of privacy as an individual right is currently ascendant in terms of legislation today,[9] conceptions of privacy as a group or community-based privacy right are emerging as well, and can be found articulated, for example, in the U.S. Indigenous Data Sovereignty Network.[10] This is part of a broader global trend toward understanding and incorporating expressions of group privacy in government discussions of data collection, analysis, use, and privacy.

For example, the articulation of group and community-based consent and privacy can also be found in Māori approaches to privacy, which have been incorporated by the New Zealand Privacy Commissioner and other parts of the government. [11] These ideas and approaches can also be seen in the *First Nations Principles of OCAP*, which establishes how First Nations' data and information will be collected, protected, used, or shared in Canada.[12] Of note, the United Nations has published the United Nations Declaration on the Rights of Indigenous Peoples, which sets forth Indigenous peoples' right to participate in decision-making in matters which would affect their rights, through representatives chosen by them in accordance with their own procedures, as well as to maintain and develop their own Indigenous decision-making institutions.[13] In some cases, these procedures may involve group conceptions of privacy.

The NIST draft definitional language in this instance does not acknowledge the presence of these group-based forms of privacy thought as a right, or generally acknowledge other needs of

---

[9] For example, a European-influenced articulation of individual privacy may be seen in OECD's Recommendation on Privacy (the Fair Information Practice Principles). A full articulation of the European approach may be seen in Directive 95/46/EC, 1995 O.J. (L 281) 31 and in the current EU General Data Protection Regulation.

[10] U.S. Indigenous Data Sovereignty Network, https://usindigenousdatanetwork.org/resources/ . See also: Maggie Walter, Tahu Kukutai, Stephanie Russo Carroll and Desi Rodriguez-Lonebear, editors, Indigenous Data Sovereignty and Policy, Routeledge: London, October 2020. https://doi.org/10.4324/9780429273957 .

[11] See for example, the *Model Development Lifecycle* (MDL) which implements the *Algorithm Charter of Aotearoa New Zealand*; the MDL also specifically supports the use of the Te Ao Māori frameworks, item 22 p. 13, available at: https://www.msd.govt.nz/documents/about-msd-and-our-work/work-programmes/initiatives/phrae/mdl-governance-guide-for-effective-operational-algorithm-decision-making.pdf . See also Te Mana Raraunga, *Māori Data Audit Tool*, Maori Sovereignty Network, https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/59152b7db8a79bdb0e64424a/1494559615337/Māori+Data+Audit+Tool.pdf. See also: Lauren Skogstad, *Whose Artificial Intelligence?* Design Assembly, https://designassembly.org.nz/2023/05/08/whose-artificial-intelligence-reflecting-on-the-intersection-of-ai-and-te-ao-maori/ .

[12] First Nations Information Governance Centre, *The First Nations Principles of OCAP,* https://fnigc.ca/ocap-training/ .

[13] *See: United Nations Declaration on the Rights of Indigenous Peoples,* United Nations, General Assembly Res 61/295 art. 18 (Sept. 13, 2007) http://www.un-documents.net/a61r295.htm.

privacy for groups of people. The discussion of differential privacy is focused largely (but not solely) on PII. Group privacy is factually important to recognize in the definition; there is unambiguous documentation of the existence of these policies in practice today. WPF respectfully requests that NIST include the idea of privacy of groups of people and emerging group privacy rights in its draft. Acknowledging the emerging thought and practices around group privacy in a socio-technical context would be very beneficial.

III. Definitional problem regarding privacy risks arising from "problematic data collection"

On page 7, the draft states:

> "Privacy risks arise from problematic data actions, which are actions taken on data that could cause an adverse effect for individuals."

Privacy risks arise may arise from problematic data actions, this is correct. Some specific data types may be more likely to cause an adverse effect for individuals. However, this language excludes the risks to privacy today that arise from the mere collection, recording, or existence of PII as well as data that may not be legally defined as PII. Privacy risks arise from data processing activities that are legitimate and fully within the scope of data collection. Privacy risks arise whether or not actions (problematic or otherwise) taken on data could cause an adverse effect for individuals.

We see this notion implemented in the European Union's General Data Protection Regulation, which allows for collection of personal data only "for specified, explicit and legitimate purposes." See Article 5(1)(b). Privacy risks arise at the first step of any PII processing and not only when there is an identifiable adverse effect for data subjects. We do not know whether the statement quoted from the Draft p. 7 above suggests that the adverse effect be known or identified in advance. We note that in many instances, adverse effects of data processing are not foreseeable in advance, in part because many privacy affecting activities are unknown to data subjects.

A common defense for commercial data processors caught using PII in an unwanted manner is that there is no identifiable adverse effect on data subjects. In crude terms, that argument is: "We can do anything we want with your PII unless you can prove monetary harm in a court of law." Only in the U.S. is that argument colorable. We reject the notion, and we reject the statement in the draft that ties privacy to adverse effects. To be clear, privacy is a legitimate interest of a data subject independently of whether there is an actual, apparent, or possible adverse effect. NIST's approach to privacy as reflected in this section of draft is starting from the wrong place. We suspect this is inadvertent. We encourage NIST to revise this definition so that privacy risks are more precisely characterized. Not all privacy problems have arisen from intentionally taken data actions. Certainly, this is true in some notorious cases. But not in all cases.

IV. Definitional problems regarding framing privacy in terms of limiting disclosures

We make a similar language objection to a later statement also on page 7.

> "Privacy can also be framed in terms of limiting different kinds of disclosures…"

We agree that disclosures of PII are a major concern of privacy policy. We do not, however, agree that all of privacy can or should be reduced to controlling disclosure of PII. Here, we recall the statement of facts in Section I of this comment, noting the actions that advanced analysis and data sharing in today's ecosystems have had on privacy. We also note that even when there are strict sectoral regulations in place regarding disclosure of information, such as with HIPAA, the federal health privacy rule, there are meaningful loopholes that still allow substantial and lawful disclosure of highly sensitive data.[14] It is a very difficult concept to pin privacy on limiting different kinds of disclosures. At the OECD, there is a privacy approach being discussed called "Data Free Flow with Trust."[15] WPF does not comment here on the merits or non-merits of this approach; we mention it here to note that multiple countries and a significant multilateral institution are attempting to grapple with the substantive reality that limiting disclosures is a very difficult task.

V. Unit of Privacy

We understand the draft's use of the *unit of privacy*. It is a useful concept, and we understand that the draft has used it appropriately and carefully. However, we are looking at this from a policy perspective, and we think this phrase in a policy context could be problematic if not cabined with precision and in a differential privacy context. We are not prepared to argue that the manner in which the concept behind these words is applied is wrong. At a policy level, however, privacy does not come in units. A clarification that a "unit of privacy" is a technical term of art with specificity to meaning in the differential privacy context and not to generalized policy will be important to include. It would not be a good outcome to see the term "unit of privacy" appear in general legislation without any reference to the specificity of its meaning in the differential privacy context.

VI. The challenges of addressing self-inflicted privacy harms

We next turn to the models of privacy discussion in section 4.2, where we find this statement:

> "All of the models assume that the data subjects are trusted because differentially private systems are designed to protect the data subjects from the other parties, and there is no incentive for data subjects to cause privacy harms to themselves."

---

[14] See for example, Bob Gellman, Pam Dixon, John Fanning, and Dr. Lewis Lorton, *A Patient's Guide to HIPAA,* FAQ 51-68. World Privacy Forum, 3rd edition, 2019. https://www.worldprivacyforum.org/2019/03/hipaa/. See also FAQ addendum on reproductive health and lawful HIPAA disclosures available at: https://www.worldprivacyforum.org/2022/07/hipaa-and-reproductive-health-a-companion-faq-to-the-patients-guide-to-hipaa/ .

[15] *Moving forward on data free flow with trust: New evidence and analysis of business experiences*, OECD Digital Economy Papers, 27 April 2023. https://www.oecd.org/digital/moving-forward-on-data-free-flow-with-trust-1afab147-en.htm .

We wish this assumption were true. Whether or not individuals have an "incentive to cause privacy harms to themselves," individuals cause privacy harms to themselves regularly, sometimes inadvertently, and, in many cases, understandably. Individuals share their data but rarely read privacy policies or terms of service. Individuals disclose personal information without knowing how or if the information will be shared or used to affect their lives. Individuals check boxes (or fail to uncheck pre-checked boxes) that determine their rights. Individuals often share their personal data even though the sharing is against their own interests. We could go on listing ways that individuals can lose control over their PII or otherwise cause privacy harms to themselves, but the point should be clear. Individuals can and do cause privacy harms to themselves, both knowingly and otherwise. In many cases, they do so because they have an incentive to take an action that directly or indirectly results in privacy harms.

We are unable to suggest how to factor the realities of the world into the threat models that the draft defines. We understand that models can have utility even if flawed in some way. Nevertheless, our concern is the unrealistic view of how the real world works. A model so far removed from reality is of questionable value. As a solution, we suggest that the draft consider the inclusion of the reality that many consumers do indeed suffer from self-inflicted privacy harms.[16]

VII. The Central Model description needs a more robust evidentiary basis

We have an objection to and concerns regarding the articulation of the Central Model in the draft, which states in 4.2.1:

> The key component of the **central model** is a trusted data curator. Each individual submits their sensitive data to the data curator, who stores all of the data in a central location (i.e., on a single server). The data curator is trusted in that users assume that they will not look at the sensitive data directly, will not share it with anyone, and cannot be compromised by any other adversary.

We On p. 38, the draft further notes:

> When evaluating a differential privacy guarantee, the most important consideration is where the threat model's trust assumptions match reality. For example, in the c**entral model** of differential privacy (described in Sec. 4.2.1), the curator must be trusted. If the central model is used with an untrustworthy curator, then the differential privacy guarantee breaks down because the curator may simply release the sensitive data to the public.

---

[16] Consider for example the large numbers of people who post personal health information to social media, including photos of digital imaging, and even post-surgical notes. While informative to a social network of friends, this kind of posting causes enormous harm in that it removes HIPAA and certain confidentiality protections of the data that has been disclosed. This can have meaningful downstream consequences. See: Robert Gellman, *Legal Analysis: Why many PHRs threaten confidentiality,* World Privacy Forum, 2008. https://www.worldprivacyforum.org/2008/02/report-personal-health-records-why-many-phrs-threaten-privacy/.

Based on the evidence, WPF does not view the central model as being a leading choice for privacy protection. In our field research regarding centralized models, we have found and documented meaningful, substantial, structural privacy issues with many central models.[17] We understand that not all central models are the same. However, there has been much discussion of the merits and demerits of central models overall. The evolving global consensus is that central models often introduce structural issues that require meaningful mitigation, such as federation, among other mitigations.

Because data can rarely be stored in a central without any risk of compromise, we encourage NIST to address these data and risk factors in its initial or primary description of the model. We also encourage NIST to consider the group of ISO standards on Big Data Reference Architecture, which may offer helpful definitions, elements, and approaches that may be assistive here.[18]

VIII. Differential privacy guarantees

On page 45, we took note of this language:

> The certification of differential privacy guarantees is particularly important given the challenge of communicating these guarantees to non-experts. A thorough certification process would provide non-experts with an important signal that a particular system will provide robust guarantees without requiring them to understand the details of those guarantees.

We agree that it would be nice if non-experts understood "guarantees" relating to privacy. However, we are not sure that anything in the privacy realm can be *guaranteed*, as there are too many people, processors, transfers, data systems, technologies, legalities, and threat models and actors involved.

More important than adding to the increasingly large numbers and length of unread notices is meaningful accountability structures and mechanisms. Individuals are not likely to read notices routinely, but they (or their lawyers) might read notices when they believe that their interests have been harmed, and they want to find a way to protect themselves. We suggest that users of differential privacy, including those who are sources of PII, who process PII, and who use PII be held accountable for how they process PII. Notices that do not describe or point to accountability measures are less useful.

---

[17] Pam Dixon, *A Failure to Do No Harm: India's Aadhaar biometric ID program and its inability to protect privacy in relation to measures in Europe and the U.S., S*pringer Nature, Health Technology. DOI 10.1007/s12553-017-0202-6. http://rdcu.be/tsWv. Open Access via Harvard-Based Technology Science: https://techscience.org/a/2017082901/.

[18] ISO / IEC Information Technology - Big Data Reference Architecture. 20547-1-5, 2018-2020. Available at: https://www.iso.org/advanced-search/x/title/status/P,U/docNumber/20547/docPartNo/docType/0/langCode/ics/currentStage/true/searchAbstract/true/stage/stageDateStart/stageDateEnd/committee/sdg .

A statute would be the best way to provide accountability,[19] but we recognize that NIST cannot create statutory remedies. It can, however, prompt processors to agree to contractual terms that explicitly state the obligations of all parties and create remedies for those parties and for individuals whose privacy may be affected by a breach of contractual terms. A model contractual agreement covering activities involving the creation, use, and transfer of data subject to differential privacy processes would be most valuable.

When present, contracts should provide remedies for data subjects and should also define the responsibilities of all parties and thereby encourage use of methods that have potential to address needs of data subjects and data users. Everyone benefits when the terms of PII use and transfer are clear, specific, and well understood. We note again that groups of people also have privacy interests and in some cases rights. We suggest that NIST's effort here include a model contract for all data processors who have a role in differential privacy. To put it succinctly, rights, remedies, and accountability are better for data subjects and everyone else than signals.

IX. Conclusion

WPF again thanks NIST for the opportunity to comment on the draft guidelines. We recognize that most of our comments are on the policy level. We hope that our comments are constructive and assistive, as we expect policymakers to use the NIST document to understand the technology and the associated policy considerations regarding differential privacy techniques. Our goal is to assist and provide a view of these issues earned from our research and time working on these issues.

We support the knowledgeable use of differential privacy, and we recognize the skill and elegance of the paper in regards to the technical aspects. However, the policy considerations we have articulated in these comments can often be invisible to those who have not experienced the reality of the impact of certain application of differential privacy in use cases across various sectors, groups, and geographies. Our hope is that we have managed to share our experience in a constructive way.

We stand ready to assist and are happy to help with additional information or use cases.

Respectfully submitted,

Pam Dixon, Executive Director
World Privacy Forum

---

[19] See Robert Gellman, *The Deidentification Dilemma: A Legislative and Contractual Proposal*, 21 Fordham Intellectual Property, Media & Entertainment Law Journal 33 (2010), http://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1277&context=iplj.