Federal Register Notice 87 FR 35250, https://www.federalregister.gov/documents/2022/06/09/2022-12432/request-for-information-on-advancing-privacy-enhancing-technologies, June 9, 2022

Request for Information (RFI) on

Advancing Privacy Enhancing Technologies

Austin, Lisa; Lie, David; Nikolov, Aleksandar; and Papernot, Nicolas

---

Name of the persons filing the comment:

- Dr. Lisa Austin, University of Toronto,
- Dr. David Lie, University of Toronto,
- Dr. Aleksandar Nikolov, University of Toronto, s
- Dr. Nicolas Papernot, University of Toronto,

Respondent type: Academic

We respond inline to a subset of the topics included in the request for information published in the federal register. The topics are indicated in bold below, along with the corresponding topic number included in the original request for information.

1. *Specific research opportunities to advance PETs:* Information about Federal research opportunities that could be introduced or modified to accelerate the development or adoption of PETs. This includes topics for research, hardware and software development, and educational and training programs. This also includes information about specific techniques and approaches that could be among the most promising technologies in this space.

Several techniques have been proposed to train on distributed data with the aim of protecting sensitive data. Most prominently, Federated Learning (FL) allows individual models to be trained on distributed data, where the model gradients, parameters, or other forms of model updates are shared and then aggregated by a central party. This framework achieves data minimization because it does not directly centralize data at the server orchestrating the protocol.

However, in its vanilla form, FL does not provide privacy (in the sense of differential privacy, see response to topic 2 below) since gradients or parameters can still reveal information about the training set of each participating party. The crux of why FL is inherently vulnerable to data reconstruction attacks is that it is designed to provide confidentiality (data does not leave user devices) rather than privacy (output of the computation does not leak sensitive attributes from the users' input). Without additional privacy measures, FL cannot protect users from the central party reconstructing their data.

One key takeaway from this line of work is that privacy-preserving FL, in the presence of an untrusted central party, is not yet practical. Hence, without trust in the central party, FL cannot provide any privacy guarantees to its users. It is therefore important for users to either understand the trust they are placing in the central party when participating in a FL protocol--or to require that new FL protocols be put forward which do not make these implicit assumptions about trust.

We believe that the latter line of research deserves much more attention and support from funding agencies. A promising research direction is to develop new protective measures that can be integrated in the design and deployment of FL to reduce the trust required in the central party. One particular aspect is the addition of noise inside a cryptographic protocol, such as secure multiparty computation, to jointly obtain differential privacy. As of yet, however, known constructions' communication costs are prohibitive so we will investigate new ones.

2. *Specific technical aspects or limitations of PETs:* Information about technical specifics of PETs that have implications for their development or adoption. This includes information about specific PET techniques that are promising, recent or anticipated advances in the theory and practice of PETs, constraints posed by limited data and computational resources, limitations posed by current approaches to de-identification and deanonymization techniques, limitations or tradeoffs posed when considering PETs as well as technical approaches to equity considerations such as fairness-aware machine learning, security considerations based on relevant advances in cryptography or computing architecture, and new or emerging privacy-enhancing techniques. This also includes technical specifications that could improve the benefits or privacy protections, or reduce the risks or costs of adopting PETs.

We argue that defining privacy in terms of de-identification and deanonymization techniques is inadequate because it focuses on modifying personal information rather than the method of analysis of this information. We start by illustrating the challenges inherent in de-identification through examples of how naive approaches can fail. We argue that such failures are inherent when de-identification focuses on the data rather than the analysis procedures. We also give a high-level, non-technical overview of recent frameworks of privacy protection developed by researchers in computer science, AI, and statistics, which, unlike the definition above, are centered around algorithms rather than the data themselves.

Because of the subtleties in defining de-identification, it is often instructive to start instead with failures to de-identify data. One naive attempt at de-identifying data is to simply strip the data from personally identifying information, i.e., to remove common identifiers like names, addresses, social insurance numbers, etc. There are many examples of how this approach usually fails, because seemingly non-identifying pieces of information can often be linked together to uniquely identify people. This observation goes back to pioneering work by Latanya Sweeney, who famously identified former Massachussettes Governor William Weld's health records from his date of birth, gender, and zip code.[1] This is known as a "reconstruction attack" (when the attacker reconstructs a data set containing personal information) or a "re-identification attack" (when the attacker can also link the personal information to identifiable individuals). To

[1] Latanya Sweeney. Simple demographics often identify people uniquely. Health 671, 1–34 (2000).

illustrate how it works, we will take a lesser-known and more recent example: the re-identification of the Myki data set by Culnane, Rubinstein, and Teague.[2]

Myki is a smart card-based public transit ticketing system used in the state of Victoria in Australia: travellers tap on when boarding a bus or train, tap off when exiting, and the system registers their trip and charges them accordingly. As part of a datathon in 2018, the Victoria government released a large data set of Myki usage information, which contestants could use to analyze the state's public transit system. The data set contained information about trips: when and where a card was tapped on, when and where it was tapped off, and which bus or train line was used. By way of de-identification, card ID numbers were substituted with random strings, and, of course, no names or addresses of registered users were included. However, if the same card was used for several trips, then the same random string was listed for each trip.

The first step researchers took in breaking the anonymity of this data set feels surprisingly innocuous: they found their own cards in the data set. This is simple: you can look up several times and places where you took a bus. For example, the times you took a bus to work on two consecutive work days, and one weekend trip. If there is only one card in the data set that was used for all of these trips, then this must be your card. This works most of the time because it turns out that only two trips are usually enough to identify a card uniquely.

While identifying your own card scarcely feels like a privacy violation, it enables a much more damaging second step: identifying the card of someone you know. For example, knowing which card in the data set is yours, you can easily identify the card of a coworker. Perhaps the coworker takes the bus home at the same time as you, and maybe you had that one work dinner together. You can check which cards were used with yours at these times, and, again, it turns out that, more often than not, there is a unique match in the data set. Having identified the card of your coworker, you can find out any other trip they have taken: weekend trips, doctor visits, or other excursions that they probably expect to be private.

The researchers also showed that such privacy breaks can be extended further by linking the data set with publicly available information. They cross-referenced the tweets of Anthony Carbines, a member of the Victorian Legislative Assembly, with the data set, and, since Carbines often proudly tweets about the Victoria public transit system when he's boarding a train, they could identify his card as well.

Let us draw several lessons from this attack:

- Anything is potentially personally identifying information, because combining a few innocuous pieces of information often identifies a person uniquely.

- Who does the identification matters. It is likely that whoever is curious about your private information already knows a lot about you. This knowledge can be leveraged to find out

[2] Chris Culnane, Benjamin Rubinstein, Vanessa Teague. Stop the Open Data Bus, We Want to Get Off

even more.

- De-identification is hard in a connected world. Twitter, LinkedIn, and other social networks and websites provide easily accessible information that an attacker can leverage to facilitate re-identifying people in a supposedly de-identified data set.

In many practical scenarios, an adversary will not have direct access to the data but instead to a by-product of the data. This could be, for instance, the output of a database query, or the prediction of a machine learning model trained on the data. One may hope that, because such by-products only contain aggregate statistical information, they are inherently privacy-preserving. This hope, unfortunately, does not bear out in reality.

As a first example, let us take reconstruction attacks from counts. Researchers and official statistics agencies often publish data summaries in the form of tables which contain counts of how many people satisfy some property: these can be summaries of surveys, or voting polls, or census data published by the US Census Bureau. Such counts feel safer, from a privacy perspective, than fine-grained data sets such as the Myki data set. They can, however, still enable an attacker to identify individuals. Imagine, for example, that someone poses two counting queries to the data set: “How many computer science professors at the University of Toronto smoke?” and “How many computer science professors at the University of Toronto who were not born in Bulgaria smoke?” Each count is (probably) a relatively large number, so would not seem like a threat in isolation, but subtracting the second number from the first identifies immediately whether one of the authors of this blog post smokes.

More sophisticated versions of this kind of attack were introduced by Dinur and Nissim,[3] who showed that, even if some noise is added to the counts, they can still be used to reconstruct private information about most individuals in a data set. Recently, researchers showed these attacks are much more than a theoretical threat, and can be successfully carried out against Diffix, a system specifically designed to answer counting queries while protecting privacy [CN20]. Moreover, the US Census Bureau has conducted its own reconstruction experiments and concluded that reconstructing sensitive microdata from the 2010 Decennial Census is possible, at least theoretically.[4] On a high-level, reconstruction attacks are possible whenever the attacker can get accurate enough answers to many uncorrelated counting queries, i.e., counting questions that ask about properties that do not overlap much. (A precise mathematical condition was identified by Muthukrishnan and one of the authors of this response.[5])

What this points to is that, rather than focusing on making a particular set of data private (e.g., through de-identification by removing personally identifying information), the scientific community has discovered that making an analysis technique (or algorithm) private provides more meaningful guarantees. Seeing a data set in isolation makes it hard, if not impossible, to

[3] Irit Dinur, Kobbi Nissim. Revealing information while preserving privacy. PODS 2003.
[4] Simson Garfinkel, John Abowd, Christian Martindale. Understanding Database Reconstruction Attacks on Public Data. ACMQueue, Vol. 16, No. 5 (September/October 2018): 28-53.
[5] Aleksandar Nikolov, S. Muthukrishnan, Optimal Private Halfspace Counting via Discrepancy. STOC 2012.

decide whether the data are successfully de-identified. As we observed already, this depends on what an attacker trying to re-identify individuals knows, and what additional information may be available from other sources. Without knowing who may try to do the re-identification, and what information they possess, or which individuals or what new information they are interested in, we cannot decide if a data set is safe for publication. If we know, however, the method (i.e., the algorithm) through which the data was analyzed to produce some by-product of it, such as a table of counts or a machine learning model, then we can actually make guarantees that hold against any possible attacker, with any kind of side information. This insight was developed initially in work by Dwork, McSherry, Nissim, and Smith,[6] who introduced a seminal framework for private data analysis known as differential privacy. Differential privacy has seen an increasing number of recent adoptions, both in industry (by Google, Apple, Facebook, among others) and by official statistics agencies, most notably the US Census Bureau starting with the 2020 Decennial Census. This shift away from de-identification and the associated focus on the dataset also requires a shift away from transferring the modified dataset: the privacy guarantee now comes from the method of analysis so the data must be kept secure and access to it restricted. To obtain access to the dataset, one would have to specify (a) what analysis they are interested in running on the data and (b) how they will ensure that this analysis is privacy-preserving. For example, (b) could take the form of a proof of differential privacy for the algorithm that is used to analyze the data.

Indeed, as we mentioned, differential privacy defines privacy with respect to an algorithm used to analyze data, rather than with respect to the data themselves.

Informally, the definition can be described using a thought experiment. Imagine that we execute the algorithm on a data set, and we also execute it on the data set with the data of one individual modified. To be differentially private, the algorithm must behave almost identically in these two scenarios, and this must happen for all data sets and all modifications to the data of any one individual. For example, a differentially private algorithm must have the property that if it produces a particular machine learning model from a data set that includes your data, then it is almost as likely to produce the same model if your data were removed or changed completely. This means that an adversary observing or using the model output by this algorithm is unable to tell whether your data (or any other person's data) were used to train the model at all. Then, whatever information an attacker may learn about you from observing the algorithm's output could have also been learned even if your data were never seen by the algorithm.

Differentially private algorithms have the property that they reveal statistical information about the data, such as, for example, correlations between genetic factors and medical conditions, but do not reveal information about particular individuals, such as whether they have some medical condition. This is because statistical information does not depend crucially on any one person's data, while private personal information does. We should note that the actual technical definition is quantitative, and the level of privacy protection is tunable, allowing us to trade privacy off for the machine learning model's accuracy, or to query answers produced by the algorithm.

[6] Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. TCC 2006

4. *Specific regulations or authorities that could be used, modified, or introduced to advance PETs:* Information about Federal regulations or authorities that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes privacy-related rulemaking authorities under the Office of Management and Budget, the Federal Trade Commission, and financial regulatory bodies, as well as acquisition regulations under the Federal Acquisition Regulations. This also includes the Federal authority to set procedures for agencies to ensure the responsible sharing of data. This also covers hiring authorities to recruit Federal employees with expertise to advance PETs, as well as acquisition authorities (*e.g.,* Other Transaction Authority) to procure PETs for development.

In general, regulations and authorities should contribute or encourage exploratory demonstrations of the benefits of PETs. It can be difficult for researchers to demonstrate the benefits of a technology because the current regulations do not allow them to deploy a prototype of it on relevant data–while at the same time said regulations will not be updated unless there is a demonstration of the benefits of said technology. For instance, it is difficult to demonstrate the benefit of differential privacy for certain applications if current regulations mandate that data which is handled in these applications be de-identified. This is because differential privacy promises that an analysis can be made privacy-preserving without any form of de-identification. Hence, by forcing researchers to evaluate differential privacy on de-identified data, we prevent them from demonstrating the full potential of differential privacy. Put another way, we should ask: what new forms of data analysis can differential privacy enable? Rather than asking: how does differential privacy compare to de-identification. Regulations and authorities here could help in several ways, including by providing access to “sandboxes” in which PETs can be experimented with to motivate more significant changes to the law.

5. *Specific laws that could be used, modified, or introduced to advance PETs:* Information about provisions in U.S. Federal law, including implementing regulations, that could be used, modified, or introduced to accelerate the development or adoption of PETs. This includes provisions, safe harbors, and definitions of use, disclosure, safeguards, and breaches. Information may also include comments on how to advance PETs as part of new or proposed legislation, such as that which would create a National Secure Data Service. Information may also include comments on State law or on international law as it applies to data sharing among international entities.

In general, the regulatory environment needs to make three changes: 1) a shift in its approach to risk, 2) the development of technical standards, and 3) the creation of regulatory frameworks

for data intermediaries. We outline these three shifts below, but do not comment on the specifics of how these might be implemented within the US legal landscape.

1) Shifting the Approach to Risk

While the question of whether a person is "identifiable" from some mass of information is central to many privacy concerns – and often whether data processing is regulated or not – it does not catch the many ways in which contemporary data processing creates privacy vulnerabilities. For example, an individual might not be identifiable in any *single* database, but the accumulation of information about this specific individual across *multiple* databases potentially increases what can be known about this individual. Waiting for that risk to materialize into a "reasonably foreseeable" risk of identifiability (or whatever is the applicable legal standard in a particular jurisdiction) is to address responsible data analysis too late in the data pipeline.

This is why we disagree with risk-based approaches that focus *only* on identifiability. Consider the analogy of speeding on a highway: I might not be wronging any specific individual when I speed, but my behaviour is risky and we regulate it through imposing speed limits in order to reduce this risk. Similarly, my data processing might not identify any specific individual, but my methods might still be risky and we need to regulate to reduce this broad risk—rather than focus on identifiability only.

The way to do this is to ensure that all organizations that process "information about persons" utilize reasonable practices to minimize privacy risks. By privacy risks, we mean here the processing of information about persons in a manner that increases the risk of identifying an individual or inferring information about a specific individual. As we have already outlined above in our response to question 2, mitigating privacy risks should not focus on manipulating the data (the de-identification approach), but should instead involve focusing on the algorithms and the computing environment more generally.

Although this shift in approach to privacy risks would require regulating the processing of "information about persons" rather than regulating the processing of PII, it would not necessarily have to dramatically shift other aspects of privacy regulation. For example, Canada has proposed new privacy legislation that would regulate de-identified, but not anonymous, information but would exempt de-identified information from some of the obligations that pertain to PII (including consent in a number of contexts). We agree with this approach although we are disappointed that the proposed legislation continues to focus on de-identification and fails to adopt the broad privacy risks minimization principle we advocate for here.

2) Development of Technical Standards

There is a strong need for the development of international standards regarding re-identification risks. We outline two components of this: regulatory and technical.

On the regulatory side, there is an increasingly international push to embrace "privacy by design" (PbD). Generally speaking, PETs are considered too narrow and technical to be a full solution to the complexities of PbD. Some of those complexities include the need to comply with

multiple laws, the often open-ended nature of some legal obligations, the need to account for how general legal norms get interpreted and re-interpreted.[7] Advocates of PbD often point to the need for a broad approach that embraces technical measures but also work processes, management structures, etc.[8] That said, we believe that technical standards that address one aspect of privacy – the risk of identifying an individual or inferring information about a specific individual – are possible. Managing this risk is common to most privacy laws across different jurisdictions, even if those laws may address this risk differently. For example, jurisdiction A might require a low level of risk when dealing with "sensitive" information whereas jurisdiction B might not require that. Having technical standards for measuring and addressing this risk is important even if other aspects of PbD might remain jurisdiction-specific and require more holistic approaches.

On the technical side, the creation of strong standards requires a shift away from de-identification as the approach to managing privacy risks. As we outlined in our answer to question 2, it is difficult to measure the risk of re-identification when focusing on data in isolation, as it depends upon knowing what the attacker might know as well as what other information is available (which also changes with time). However, approaches like differential privacy can give more accurate guarantees.

3) Data Intermediaries

Ensuring that privacy risks – understood here as the risk of identifying an individual or inferring information about a specific individual – are mitigated when analyzing information about persons requires a high level of technical expertise. Organizations may wish to analyze information in a privacy-protective way but see the costs of this as prohibitive. We think there is great scope for the development of data intermediaries that could manage this for organizations. The EU has recently signaled a move in this direction with the introduction of the *Data Governance Act*. Some of the interest in Europe lies in creating intermediaries that could help individual data subjects better manage their right to control their PII.[9] However, other models focus on reducing the legal and technical complexities involved in data-sharing, particularly in relation to managing re-identification risks.[10] Creating a regulatory environment that provides clarity regarding issues of liability and oversight is important to create the right incentives for the successful creation of these intermediaries.[11]

Data intermediaries could also provide additional options for managing some of the risks involved in the use of some PETs. For example, as we outlined in our answer to question 1,

[7] Bert-Jaap Koops & Ronald Leenes, "Privacy Regulation Cannot Be Hardcoded. A Critical Comment on the 'Privacy by Design' Provision in Data-Protection Law" (2014) 28:2 Intl Rev L Comp & Tech 159.
[8] Anver Levin, "Privacy by Design by Regulation: The Case Study of Ontario" (2018) 4 Can J Comp & Contemp L 115 at 118.
[9] Sylvie Delacroix and Neil D. Lawrence, "Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance" *International Data Privacy Law*, Volume 9, Issue 4, November 2019, Pages 236–252.
[10] Lisa M. Austin and David Lie, "Safe Sharing Sites" (2020) 94 NYU L Rev 581.
[11] Lisa M. Austin and David Lie, "Data Trusts and the Governance of Smart Environments: Lessons from the Failure of Sidewalk Labs' Urban Data Trust" (2021) 19 *Surveillance and Society* 255.

privacy-preserving FL with an untrusted central party is currently infeasible. Technical solutions to this can be developed but, in addition, the creation of trusted intermediaries that could manage this risk is an alternative pathway.

7. *Risks related to PETs adoption:* Identification of risks or negative consequences resulting from PETs adoption as well as policy, governance, and technical measures that could mitigate those risks. This includes risks related to equity for underserved or marginalized groups, the complexity of implementation and resources required for adoption, as well as from conceptual misunderstandings of the technical guarantees provided by PETs. This also includes recommendations on how to measure risk of PETs adoption and conduct risk-benefit analyses of use.

We highlight two risks associated with the adoption of PETs. First, PETs remain highly technical and evaluating the guarantees they provide often requires an understanding of statistical analysis. For instance, there is a lot of confusion between the technical notions of confidentiality and of differential privacy (see above our response to topic 1). This in turn means that end users need to place a lot of trust in companies handling their data because (a) they are not certain which guarantee is needed to protect their data against the risks they are concerned about and (b) even if they know which guarantee is needed they may be unable to assess how strong the guarantee is and/or should be. There is room here to develop technical standards that are endorsed by legislators and regulators to provide more transparency to end users and make enforcing privacy more actionable for both regulators and companies handling data.