October 23, 2019

Katie MacFarland
National Institute of Standards and Technology
100 Bureau Drive, Stop 2000
Gaithersburg, MD 20899

Via Electronic Mail

Re: NIST Privacy Framework: Comments on Preliminary Draft

New America's Open Technology Institute appreciates the opportunity to comment on the National Institute of Standards and Technology's (NIST) preliminary draft of the privacy framework.[1] NIST has clearly worked hard to create a thorough privacy framework that can be used in conjunction with the cybersecurity framework, allowing organizations to create a more holistic internal privacy and security program. Encouraging organizations to think more proactively about privacy protections is a good thing.

The framework asks at the outset, "does this preliminary draft … adequately define outcomes that … strengthen individuals' privacy protections?" While some areas of the framework are promising, we have identified the following weaknesses: (1) organizations (companies) should not play the role of identifying what level of risk individuals can accept, however, if organizations ultimately end up playing that role in the final framework, NIST needs to be clear about the types of risks that need to be evaluated; (2) the framework should emphasize data minimization as the most effective way to reduce individual risk overall, including at the collection stage; and (3) to the extent de-identification is allowed under the framework, the subject is inadequately discussed and ultimately will lead to significant privacy harm because of the likelihood of re-identification.

I. NIST's treatment of risk is insufficient to protect individual privacy

NIST's approach to risk in the framework is insufficient to protect individual privacy. Organizations should not decide the level of privacy risk that individuals are willing to accept. If

[1] Preliminary Draft, NIST Privacy Framework (Sept 6. 2019), https://www.nist.gov/sites/default/files/documents/2019/09/09/nist_privacy_framework_preliminary_draft.pdf ("framework").

NIST retains the risk-based approach for the final framework, it should make clear what types of risks an organization must consider.

a. Organizations should not determine the level of risk that individuals are willing to accept

Placing organizations in charge of assessing what level of privacy risk is acceptable to individuals will likely increase risks to individuals and likely cause individual harm. Throughout the proposed framework, NIST refers to privacy risks in both the individual and organizational sense. It states that "individuals may not be able to understand the potential consequences for their privacy as they interact with systems, products, and services. Organizations may not fully realize the consequences either."[2] The framework goes on to place the entire privacy risk assessment burden—for individuals and organizations—on organizations that use the framework. For instance, "[o]nce an organization can identify the likelihood of any given problem arising from the data processing … it can assess the impact should the problematic data action occur."[3] Further, "[o]rganizations may choose to respond to privacy risk in different ways, depending on the potential impact to individuals and resulting impacts to organizations" and the framework suggests that organizations should "minimize … risk[s] to an acceptable degree."[4] Organizations should not make all such decisions on behalf of individuals.

Individuals have different privacy risk tolerances that organizations may not recognize or appreciate. NIST contemplates an organization could "minimize … risk[s] to an acceptable degree," but there is no indication of how an organization would determine that a risk is acceptable to individuals or which individuals should be considered or how. Organizations, especially sizeable organizations, may have large user bases who will have different risk tolerances that cannot always be accounted for. For example, Target did not recognize the potential risk to individual customers of using data it collected to target baby-related ads; but in one situation, Target's practice took away a customer's ability to decide when to disclose her pregnancy.[5] Equifax did not appreciate the privacy risk associated with storing the Social Security Numbers and credit card numbers of millions of individuals before that data was breached.[6] Netflix did not understand that releasing millions of supposedly anonymized movie ratings would lead to the re-identification of many of those in the dataset, and provide highly personal information about customers, including a lesbian mother who had not chosen to publicly disclose her sexual orientation.[7] Grindr shared HIV status, GPS data, phone ID, and email of its users with third parties with little regard for the risks associated with

---

[2] Framework, at 4.
[3] Framework, at 7.
[4] *Id.*
[5] Gus Lubin, *The Incredible Story of How Target Exposed A Teen Girl's Pregnancy*, Bus. Insider (Feb 16, 2012), https://www.businessinsider.com/the-incredible-story-of-how-target-exposed-a-teen-girls-pregnancy-2012-2.
[6] Allen St. John, *Equifax Data Breach: What Consumers Need to Know*, Consumer Reports (Sept. 21, 2017), https://www.consumerreports.org/privacy/what-consumers-need-to-know-about-the-equifax-data-breach.
[7] Julianne Pepitone, *5 Data Breaches: From Embarrassing to Deadly*, CNN (Dec. 14, 2010), https://money.cnn.com/galleries/2010/technology/1012/gallery.5_data_breaches/index.html.

sharing such personal, intimate information.[8] Even situations that seem more benign can be problematic. For example, Overstock broadcast over a public Facebook account a person's engagement ring purchase based on data from the Facebook Beacon, and it notified all his friends, including his then-partner.[9] In all of these situations, the organization did not provide sufficient privacy safeguards to match the level of risk that the individuals would have chosen, and harm resulted. Making organizations solely responsible for determining individual privacy risks will lead to more harm.

In some instances, the type of potential harm may be so serious that a risk/benefit analysis is not appropriate and the data processing should not be allowed. One such example is the use of data for discriminatory purposes or where there is a disparate impact on a protected class. Discrimination based on a protected class is a serious and unacceptable risk in any circumstance, yet organizations using the framework could determine that discrimination is unlikely when it in reality it is not always capable of making that determination. Moreover, secondary use of data is itself a privacy risk to individuals, and that practice should be restricted entirely rather than allowing an organization to make its own determination as to the risk it poses to individuals.[10]

Subjective privacy risks are least likely to be considered, if they are considered at all, by organizations because the framework makes no mention of them. Professor Ryan Calo wrote in 2011 that privacy harms or risks can be categorized as "subjective" and "objective."[11] He defined "subjective" harms broadly as "the perception of unwanted observation."[12] The feeling of being watched, on its own, can constitute a privacy harm.[13] Pervasive monitoring can be even more harmful to individuals for the same reason, but the framework does not contemplate constant monitoring as a potential individual risk. As a result, organizations could engage in this type of "unwanted observation" through collection and use of data from all its users (to, for instance, sell

---

[8] Azeen Ghorayshi & Sri Ray, *Grindr Is Letting Other Companies See User HIV Status and Location Data*, Buzzfeed (Apr. 2, 2018),
https://www.buzzfeednews.com/article/azeenghorayshi/grindr-hiv-status-privacy.

[9] Julianne Pepitone, *5 Data Breaches: From Embarrassing to Deadly*, CNN (Dec. 14, 2010),
https://money.cnn.com/galleries/2010/technology/1012/gallery.5_data_breaches/3.html

[10] *See* Comments of Open Technology Institute to the Federal Trade Commission, at 7-9,
https://newamericadotorg.s3.amazonaws.com/documents/Comments_of_New_Americas_OTI_Consumer_Privacy.pdf.

[11] "Objective privacy harms are those harms that are external to the victim and involve the forced or unanticipated use of personal information." M. Ryan Calo, *The Boundaries of Privacy Harm*, 86 Indiana L.J. 1132, 1148 (2011), *available at*
https://www.repository.law.indiana.edu/cgi/viewcontent.cgi?article=1016&context=ilj.

[12] *Id.* at 1144.

[13] Neil M. Richards, *The Dangers of Surveillance*, Harv. L. Rev. Symposium,
https://harvardlawreview.org/2013/05/the-dangers-of-surveillance ("surveillance is harmful because it can chill the exercise of our civil liberties…. A second special harm that surveillance poses is its effect on the power dynamic between the watcher and the watched. This disparity creates the risk of a variety of harms, such as discrimination, coercion, and the threat of selective enforcement, where critics of the government can be prosecuted or blackmailed for wrongdoing unrelated to the purpose of the surveillance.")

advertising), and the framework would not cause them to weigh such monitoring as an individual harm, much less a harm that individuals may not want to endure.[14]

In addition to not fully accounting for subjective risks, organizations have incentives to downplay individual privacy risks to allow for more data collection and use, and companies have in general vastly expanded their data collection practices.[15] The draft framework fails to provide a method to account for this dynamic and weight factors appropriately. Data is often a source of revenue, especially for organizations that rely on behavioral advertising,[16] and the fewer restrictions placed on collecting and using that data, the greater the ability for companies to monetize it. These incentives have long plagued protection of privacy in the United States. The Federal Trade Commission (FTC) is empowered to hold organizations responsible for broken promises under its deception authority, but organizations themselves decide what promises to make to accomplish whatever end they prefer; in many cases, that end is maximum data collection for maximum monetization. In the same way, organizations using the framework would be deciding the level of acceptable privacy risk to individuals, but the framework fails to provide a mechanism to rebalance these incentives to ensure organizations will appropriately measure privacy risks.

Where an organization is unlikely to experience consequences for their actions, the framework fails to require organizations to give proper weight to individual risks. The framework assumes that the organization "experience[s] impacts such as noncompliance costs, customer abandonment of products and services, or harm to its external brand reputation or internal culture."[17] In markets with one dominant player, that is not always true; harm to the brand from privacy violations may be limited because individuals lack the ability to vote with their feet or wallets. For example, Facebook's social media platform lacks robust competitors, and in the wake of the Cambridge Analytica scandal, did not see significant user churn in the United States. Goldman Sachs stated that the social network *gained* 7% more users in April 2018 as per the year before.[18] It

---

[14] While some may argue that the individual simply should not use the service, there may not be many options. See discussion below of services that lack robust competition, like Facebook.

[15] Big Data, a Tool for Inclusion or Exclusion?, FTC Report (Jan. 2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf ("While companies historically have collected and used information about their customer interactions to help improve their operations, the expanding use of online technologies has greatly increased the amount of consumer data that flows throughout the economy. In many cases, when consumers engage digitally—whether by shopping, visiting websites, paying bills, connecting with family and friends through social media, using mobile applications, or using connected devices, such as fitness trackers or smart televisions—companies collect information about their choices, experiences, and individual characteristics. The analysis of this consumer information is often valuable to companies and to consumers, as it provides insights into market-wide tastes and emerging trends, which can guide the development of new products and services. It is also valuable to predict the preferences of specific individuals, help tailor services, and guide individualized marketing of products and services.").

[16] Becky Chao & Eric Null, *Paying for our Privacy*, Open Technology Institute Report (Sept. 17, 2019), https://www.newamerica.org/oti/reports/paying-our-privacy-what-online-business-models-should-be-limits.

[17] Framework, at 7.

[18] Jake Kanter, *The backlash that never happened: New data shows people actually increased their Facebook usage after the Cambridge Analytica scandal*, Bus. Insider (May 20, 2018), https://www.businessinsider.com/people-increased-facebook-usage-after-cambridge-analytica-scandal-2018-5.

also appears the "#DeleteFacebook backlash never really arrived."[19] Thus, while trust in Facebook may have dipped, as did its stock price[20] (at least temporarily[21]), Facebook users largely remained on the platform and Facebook escaped the privacy scandal relatively unscathed. Facebook and companies like it have no incentive to strongly protect against individual privacy risks.

Similarly, data brokers are unlikely to experience consequences from customers leaving their service because individuals are not data brokers' customers. Data brokers do not collect data directly from individuals, but from organizations that collect it from individuals. They then repackage and sell the individuals' data to other organizations. This business practice is one that presents significant risks to individuals and constitutes a secondary use of data. If a data broker experiences a data breach, or violates an individual's privacy, those individuals have little recourse. The Equifax breach, where Equifax disclosed personal information of people who did not know Equifax had that information, is one example of such a data broker privacy violation where individuals could not vote with their feet.[22]

For these reasons, organizations should not hold sole responsibility for determining how much privacy risk individuals are willing to accept, or at the very least, the framework should provide a mechanism for organizations to better assess how individuals measure privacy risk.

b. Should NIST retain its risk-based approach, it should clarify the types of risks organizations should assess

Even the most well-intentioned organizations will not predict every risk that would be relevant to individuals, and therefore, should NIST retain its risk-based approach, it should very clearly describe the full range of risks it expects organizations to assess. Currently, NIST devotes a single sentence to describing the types of individual risks that organizations should take into account: "The problems individuals can experience as a result of data processing can be expressed in various ways, but NIST describes them as ranging from dignity-type effects such as embarrassment or stigmas to more tangible harms such as discrimination, economic loss, or physical harm."[23] Leaving the definition of risks nearly entirely to the organizations adopting the framework is very likely to lead to those organizations defining risk narrowly, ignoring risks, or undercutting the importance of certain risks for certain individuals, as described above.

---

[19] *Id.*

[20] Rupert Neate, *Over $119bn wiped off Facebook's market cap after growth shock*, Guardian (July 26, 2018), https://www.theguardian.com/technology/2018/jul/26/facebook-market-cap-falls-109bn-dollars-after-growth-shock.

[21] Jacob Sonenshine, *Facebook is trading at its best level since the Cambridge Analytica data scandal*, Markets Insider (May 7, 2018), https://markets.businessinsider.com/news/stocks/facebook-stock-price-best-level-since-cambridge-analytica-data-scandal-2018-5-1023663283.

[22] Allen St. John, *Equifax Data Breach: What Consumers Need to Know*, Consumer Reports (Sept. 21, 2017), https://www.consumerreports.org/privacy/what-consumers-need-to-know-about-the-equifax-data-breach ("Unlike a credit card company or retailer, consumers generally don't choose to do business with credit-reporting firms. Instead, those companies gather information on consumers as part of their business.").

[23] Framework, at 6.

There are many risks that exist along the spectrum between what the framework describes as "tangible" or "dignity" related. But the framework must do more than simply point to the spectrum, it should be more detailed. For a broad definition of risk, the Intel model privacy bill provides an extensive list:

(A) Direct or indirect financial loss or economic harm;
(B) Physical harm;
(C) Psychological harm, including anxiety, embarrassment, fear, and other demonstrable mental trauma;
(D) Significant inconvenience or expenditure of time;
(E) Negative or harmful outcomes or decisions with respect to an individual's eligibility for rights, benefits or privileges in employment (including, but not limited to, hiring, firing, promotion, demotion, compensation), credit and insurance (including, but not limited to, denial of an application or the granting of less favorable terms), housing, education, professional certification, or the provision of health care and related services;
(F) Stigmatization or reputational harm;
(G) Disruption and intrusion from unwanted commercial communications or contacts;
(H) Price discrimination;
(I) Effects on an individual that are not reasonably foreseeable, contemplated by, or expected by the individual to whom the personal data relate that are nevertheless reasonably foreseeable, contemplated by, or expected by the covered entity assessing privacy risk, that significantly—
    (i) alter that individual's experiences;
    (ii) limit that individual's choices;
    (iii) influence that individual's responses; or
    (iv) predetermine results or outcomes for that individual; or
(J) Other demonstrable adverse consequences that affect an individual's private life, including private family matters, actions, and communications within an individual's home or similar physical, online, or digital location, where an individual has a reasonable expectation that personal data will not be collected, observed, or used.[24]

While OTI does not endorse all aspects of the Intel model bill and we would suggest some modifications to Intel's list, this list is a helpful model in spelling out specific risks within the framework's spectrum. Identifying specific risks in the framework will help guide organizations as they assess the risks that they and individuals take on as a result of particular data practices. Without such identification, there is little hope that organizations will assess risk broadly; organizations following the framework may *only* assess, very narrowly, "embarrassment or stigmas [or] discrimination, economic loss, or physical harm." Such an assessment will leave out many real risks that individuals (and organizations) take on and will lead to individual harm.

[24] Innovative and Ethical Data Use Act of 2019, Intel Draft Privacy Bill, at 6-7, https://usprivacybill.intel.com/wp-content/uploads/IntelPrivacyBill-05-25-19.pdf.

II. The framework should emphasize data minimization as the most effective way to reduce individual risk overall

A cornerstone of privacy protections is data minimization. OTI included arguments for including data minimization in the framework in comments it filed with NIST in January 2019.[25] The framework includes some mention of minimization, but does not make clear that, on an ongoing and forward-looking basis, organizations should minimize the amount of data they collect, or that they should assess privacy risks before deciding whether to collect new types of data. The framework should emphasize data minimization as a core privacy practice of any organization that wants to protect individual privacy.

By collecting more data from individuals, organizations necessarily increase the privacy risk to individuals and the risks that the organizations themselves take on. Many organizations build vast databases of information about individuals that interact with their services. The more data an organization collects, the greater the risk of data breaches. Further, organizations have incentives to monetize the data in ways that users may never have anticipated--as discussed above, potentially to discriminate, or for secondary uses like sale to a data broker that could then be used to impact, for instance, a credit score.[26]

The framework should strongly encourage organizations to avoid or get out of that cycle of collect, store, and monetize. The framework should emphasize in its IDENTIFY-P category that companies need to develop a thorough understanding of why they are collecting particular data, and similarly what risks those data practices entail (making sure that organizations are taking into account all potential risks involved). The framework should make clear that not only should organizations make these assessments for data they already collect, but they should also conduct such an analysis before deciding to collect a new type of data. Under the minimization principle, which should be incorporated into this part of the framework, organizations should attempt to minimize the amount of data they collect and ensure that they only collect information necessary to their businesses. Similarly, as part of the CONTROL-P category, companies must critically analyze their policies and ensure that they are processing data only in ways that are mission critical.

Therefore, OTI suggests NIST remove the words "relevant and" from the subcategory CT.DP-P6. Organizations that follow the framework should limit data processing only to those activities that are "necessary for" the organization's system and objectives. The term "relevant" is too vague. An organization could justify almost any data processing as "relevant" because earning revenue will always be "relevant" to an organization's business, and having access to proprietary data can be used to earn revenue. But NIST should challenge organizations to think beyond earning revenue and think about the individual risks presented by endless data collection. Further,

---

[25] The prior NIST comments referenced OTI comments filed with the National Telecommunications and Information Administration, available at https://newamericadotorg.s3.amazonaws.com/documents/OTI_NTIA_Comments_11.9.pdf.

[26] Yael Grauer, *What Are 'Data Brokers,' and Why Are They Scooping Up Information About You?*, Vice (Mar. 27, 2018), https://www.vice.com/en_us/article/bjpx3w/what-are-data-brokers-and-how-to-stop-my-private-data-collection.

organizations should discipline themselves to limit data collection to only data that is necessary to their service and objectives.

III. The framework does not properly address risks associated with de-identification of data

In several places, the framework appears to imply that data processing is acceptable if the organization does not associate data with an individual. For instance, subcategory CT.DM-P1 states that "[d]ata are processed in an unobservable or unlinkable manner," CT.DP-P2 states that "[d]ata are processed to limit the identification of individuals," and the framework also lists "[d]isassociability" as an engineering goal in Appendix D ("Enabling the processing of data or events without association to individuals or devices"). To the extent these aspects of the framework are intended to allow for the free use of de-identified data, OTI opposes such free use.

De-identified data (or "anonymized" data) can often be re-identified. Re-identification poses tangible threats to privacy, in part because once it has been disclosed, data cannot be taken back, and as time goes on, the chances of re-identification increase.[27] It is well known that perfect anonymization does not exist.[28] Supposedly de-identified datasets from medical records, search queries,[29] social network data,[30] genetic information,[31] geolocation data,[32] and taxi-cab history[33] have all been used to specifically identify individuals.[34] It is also possible to re-identify individuals from purportedly anonymous cell-phone location records[35] and telephone metadata.[36] As the amount of publicly-available data grows (including through unauthorized disclosure like a data breach), the amount of data that can be used to re-identify "de-identified" data also grows, and therefore the risks to individuals grow concomitantly. Thus, even with the best intentions of those de-identifying data, re-identification can pose serious risks for data subjects.

---

[27] Arvind Narayanan *et al.*, *A Precautionary Approach to Big Data Privacy*, at 5 (Mar. 15, 2015), http://randomwalker.info/publications/precautionary.pdf ("Precautionary Approach").

[28] Ira S. Rubenstein & Woodrow Hartzog, *Anonymization and Risk*, 91 Washington L.Rev. 703, 704 (2016), *available at* http://digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1589/91WLR0703.pdf?sequence=1&isAllowed=y.

[29] *See* Michael Barbaro and Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. Times (Aug. 9, 2006), http://www.nytimes.com/2006/08/09/technology/09aol.html.

[30] Ratan Dey, Yuan Ding, and Keith W. Ross, *The High-School Profiling Attack: How Online Privacy Laws Can Actually Increase Minors' Risk*, at 1 (2013), https://www.petsymposium.org/2013/papers/dey-profiling.pdf.

[31] Melissa Gymrek *et al.*, *Identifying Personal Genomes by Surname Inference*, 339 Science 321, 321-24 (2013).

[32] Philippe Golle & Kurt Partridge, *On the Anonymity of Home/Work Location Pairs, Pervasive Computing, Seventh International Conference*, Nara Japan (May 11-14, 2009), https://crypto.stanford.edu/~pgolle/papers/commute.pdf.

[33] Vijay Pandurangan, *On Taxis and Rainbows: Lessons from NYC's improperly anonymized taxi logs*, Medium (June 21, 2014), https://medium.com/@vijayp/oftaxis-and-rainbows-f6bc289679a1.

[34] Narayanan, *Precautionary Approach*.

[35] Yves-Alexander de Montjoye *et al.*, *Unique in the Crowd: The Privacy Bounds of Human Mobility*, Sci. Rep. (Mar. 25, 2013), https://www.nature.com/articles/srep01376.

[36] Jonathan Mayer *et al.*, *Evaluating the Privacy Properties of Telephone Metadata*, Proceedings of the National Academy of Sciences (Mar. 1, 2016), https://doi.org/10.1073/pnas.1508081113.

Failure to account for that risk could pose real harms for individuals who may not want to have their de-identified data in a dataset when those individuals could be re-identified. This is especially true where the individual knows there is already extensive publicly-available information about him or her, thus increasing the chance of re-identification. NIST's framework does not grapple with these issues. As a result, companies that rely on the framework may (potentially inadvertently) create significant privacy risks for individuals.

Conclusion

The framework represents a step forward in ensuring that organizations provide protections for individual privacy, but there are still some aspects of the framework that should be improved. OTI urges that NIST take the time to approach risk, minimization of collection, and de-identification with extra care to prevent organizations that use the framework from facilitating individual privacy harm.

Respectfully submitted,

/s/

Eric Null
Sharon Bradford Franklin
New America's Open Technology Institute
740 15th St NW, Suite 900
Washington, DC 2005