November 9, 2018

Re: Privacy RFC, Docket No. 180821780-8780-01

Dear Mr. Redi,

The Wikimedia Foundation welcomes and appreciates the opportunity to comment on the NTIA's proposed privacy outcomes and high-level goals. As the non-profit which hosts and operates Wikipedia, protecting user privacy is an important part of achieving our mission to provide free access to knowledge for everyone. Wikipedia readers are motivated by an intrinsic desire to learn, and in October 2018 alone, 3.4 billion Americans turned to Wikipedia for information. Many people also visit to share their knowledge, which is reflected in the nearly 57 million edits that have been made to English Wikipedia in the last year. We believe that privacy is a requirement of intellectual freedom, and that everyone should have the ability to read and write online, without fear of governments or corporations looking over their shoulder. This is why we continually commit to protecting our users' privacy and safety by minimizing the data we collect, allowing pseudonymous editing, and employing HTTPS on all of our sites. We are also transparent about what data we collect, why and how long it is kept, and how we respond to legal demands. **These important protections mean that we hold a unique place as one of the few large internet platforms that do not rely on tracking or sale of user data to generate revenue.**

While it may often seem like an abstract concept, privacy, or the lack thereof, has important consequences for access to knowledge. As we have seen over the last few years, data breaches do occur, even despite best intentions, affecting a wide range of companies and organizations, and exposing sensitive user data on a sometimes massive scale. Even if personal data is kept safe from a breach, law enforcement and government agencies are increasingly engaged in requesting or mandating organizations to share the data they have collected. When people are unsure whether the actions they take online are private or available for others to see and analyze, they will change their behavior. This restricts their individual freedom, and means that the world will miss out on the opportunity to learn from that person's knowledge and experience. The consequences of an over-collection of data can also be harmful to the discovery of knowledge online as well. Although the Wikimedia Foundation collects limited data in an effort to learn how to inspire and engage new readers, many other websites collect data in order to track and control what users see, stifling the user's natural discovery of new information and knowledge. Overall, a lack of privacy hampers both sharing and discovering knowledge online, making it much more difficult to achieve the Wikimedia mission to provide access to knowledge to everyone, everywhere.

We work hard to ensure that we remain current with applicable privacy regulations, but the privacy landscape in the U.S. is currently governed by a patchwork of often conflicting state laws and federal statutes. This type of regulatory confusion often taxes smaller or non-profit organizations, costing valuable time and money that, in the Wikimedia Foundation's case, could better be spent promoting greater access to knowledge online for our billions of readers. The introduction of a federal privacy law which carefully considers the fundamental privacy rights of all internet users, the realities of large companies, and the unique needs of non-profits and small businesses would be a welcome harmonization of the United States's current privacy laws.

In general, we have a few high level suggestions for the NTIA's proposal:

- **First, the proposed approach focuses too much on processes rather than impact on the end-user.** A privacy approach which describes how privacy best practices will look from an everyday user's point of view will not only allow for more varied and competitive privacy practices, but will be more resilient to technological or sociological change over time.

- **Second, as written, many of the key outcomes described in the RfC are not sufficiently clear to encourage meaningful change to currently existing privacy practices.** This is particularly true of larger companies who can more easily afford to operate within the nuances of the law. Adding clarity, or encouraging greater investment in the actual development of best practices, will go a long way to ensuring that everyone invests in compliance with a new, stronger privacy regime in a timely manner.

The Privacy Outcomes

Below, we have addressed each section of the NTIA's proposed privacy outcomes and fundamental goals that could use improvement, and we have also proposed some additions to these categories.

1. Transparency

As mentioned above, Wikimedia is committed to being transparent when dealing with user data. We believe that transparency is a key value in ensuring freedom of expression and access to knowledge. If people are unsure what data is collected about them and how it is used, it can lead certain voices being chilled, particularly when those voices represent vulnerable populations who may have more to fear from their personal information being exposed. This further homogenizes online communication, entrenching existing biases which can then be reflected in projects like Wikipedia. By being transparent about what data is collected, how it is retained, how it will be used, and under what narrow circumstances it might be produced in response to government demands, companies and organizations can help establish trust with their users so they feel safe to participate meaningfully online.

While we appreciate the NTIA's inclusion of transparency as a desired outcome, we believe the approach should be more user-facing. The way to better transparency about data collection is not to prescribe the processes which must be followed, as the NTIA suggests here, but to describe outcomes expected. A person should easily be able to visit a site and find out what data is collected about them and how it is being used, regardless of how that data is ultimately stored or organized. We recommend that this particular outcome focus more on the user experience, rather than specific technologies or practices.

2. Control

Transparency must go hand in hand with control in order to achieve a truly user-centric outcome. Without transparency, a user may remain under-informed about their privacy options. Without control, a user will not be able to exercise these options. Because control is so important, we recommend that the definition of control include more clarity about the desired outcome.

As is, it is unclear what "reasonable" controls would look like, and leaving so much room for interpretation may slow the pace of adoption while the limits of that standard are tested in courts. While we are not advocating for the inclusion of specific required controls or format, a better definition of what is and is not within the bounds of reasonable controls would ensure that platforms are all starting from the same framework. Throughout this comment, we will be encouraging government investment into the development of best practices and we believe this may be an area well-suited to just that. Convening stakeholder and consumer forums to discuss best practices is an advisable non-regulatory step toward discovering what exactly "reasonable" means to different parties and how to reconcile those different definitions.

3. Reasonable Minimization

We strongly believe that the goal should not only be "reasonable" minimization, but simply "minimization." After all, minimization does not mean that no data must be collected, but that what is collected is as little as possible. At the Wikimedia Foundation, we intentionally minimize the data we collect on users in order to encourage free and open participation on our projects. Our volunteers can edit Wikipedia knowing that their edit history will not be tied to information like their name, age, gender, sexual orientation, or race. This is an intentional choice to encourage a diversity of voices on our projects.

We are also concerned about what seems like a false compromise in the data minimization proposal, which says, "Other means of reducing risk of privacy harm . . . can help reduce the need for such minimization." As explained above, data minimization is not simply an additional security measure meant to reduce the risk of unlawful use. It is a conscious choice on the part of websites to encourage speech and participation, and users should be protected from potentially harmful *lawful* uses of data as well.

4. Security

Security has become an increasingly important topic recently as large data breaches at companies have made headlines in the U.S. Unfortunately, codifying specific security standards presents a unique set of challenges. If the standards are too vague, it will do little to incentivize the adoption of robust security. However, if standards are too specific, they will likely grow outdated before they are even adopted.

The NTIA consultation makes reference to “current consensus best practices,” but is unclear who would be setting or legitimizing those best practices. If the NTIA has a specific standard in mind, we ask that they clarify which standards so they can be evaluated. If not, this is another area where we recommend government investment in convening stakeholders regularly to discuss and make recommendations for best practices. While these recommendations may not necessarily result in setting legal standards for determining a company’s compliance with the law, they may be used as persuasive evidence of what practices are currently recommended for robust security.

5. Access and Correction

We have discussed above how important data control and minimization are to users and to the Wikimedia Foundation. However, as an online encyclopedia editable by anyone, we have some unique concerns with this particular privacy outcome. Although Wikipedia users are asked to provide very little data about themselves, Wikipedia does have a significant quantity of publicly available encyclopedic information that may be considered “personal data” about persons that meet Wikipedia's notability guidelines. This is why we feel it is important to clarify that access and correction should be limited to the personal data which has been provided by the user, and not simply all personal data held about someone.

The erasure of all information about a particular person, with no consideration of the context in which that data is held, can lead to a complete loss of important, factual information about notable individuals. On Wikipedia, each language community decides via consensus which topics are notable enough to warrant inclusion in the encyclopedia, but there are also processes in place for articles subject to request corrections and removal of information about them. This combination of consensus-building and administrative oversight over sensitive topics such as biographies is an example of how nuance is important in a system that must strike the balance between two fundamental rights: privacy and freedom of expression. Any regulation which mandates a simpler system, such as giving everyone the unfettered ability to remove all information about themself, has the potential to harm these democratic processes with a “one-size-fits-all” answer. Thus, any regulation which mandates access and correction should carefully balance these controls with fundamental rights like freedom of expression.

6. Risk Management

At the moment, it is unclear what exactly the NTIA seeks to achieve through its risk management outcome. The actual mandate, “Users should expect organizations to take steps to manage and/or mitigate risk…” says nothing about how this expectation could be met by organizations and companies that collect user data. We believe that a better definition of risk management, with examples or guidelines about what activities can increase or mitigate risks, could help to clarify what is meant by this section.

We do, however, appreciate the NTIA’s reference to flexibility in risk management. It is important to recognize that everyone assesses risk differently, and allow for the law to be flexible to those different approaches. We would like to suggest the addition of a similar principle to this particular privacy outcome: proportionality. As we will explain further below, we believe that proportionality differs slightly from flexibility and thus should be incorporated as its own high-level goal. In the case of risk management, this means the NTIA should make the required amount of risk management proportional to how much and what type of data an organization actually collects. This not only allows for greater innovation among small businesses and non-profits, but it also encourages data minimization practices wherever possible to save on associated risk management costs.

7. Accountability

While external accountability is an important effect of any privacy legislation, this is another area which needs significantly more clarity. We are hesitant to support a statement which solely mandates “external accountability” without explaining in detail just who will be administering that accountability, how, and what the potential consequences are. The process of determining just how and by whom companies and organizations should be held accountable should be open and transparent and include all relevant stakeholders.

The High-Level Goals

We would like to take this opportunity to highlight a few of the goals we feel are most aligned with the Wikimedia Foundation’s policy priorities, and which we believe will make the greatest impact on privacy without sacrificing freedom of expression and other rights online.

First, we agree that harmonization should be one of the primary goals of any NTIA privacy framework. Currently, privacy laws in the United States exist in a patchwork of state and federal law that can sometimes even contradict itself, particularly in the area of data breach notification. Having a single standard for certain privacy issues, including data breach notifications, will help bring clarity to U.S.-based organizations, and allow for important company resources and expertise to be devoted to other efforts to ensure people’s privacy.

Second, we support the NTIA’s risk and outcome based approach to privacy legislation, rather than a compliance model that prescribes specific practices. This will allow organizations the flexibility to design systems how they see fit, but still creates specific predictable and intuitive privacy outcomes for consumers that are similar across platforms. As addressed above, we believe that more specificity is needed in describing the exact outcomes desired in order make any privacy laws enforceable, and investment in the development of best practices is one important way to do so. The more best practices are researched, discussed, and developed, the closer we come to having a standard against which to compare existing practices to look for user harm.

Thus, we also support the suggestion that the U.S. Government should encourage research and development of products and services that improve privacy protections. While privacy may finally be receiving attention in national and international headlines, there remains a dearth of current consumer awareness of what concrete steps may be taken to protect privacy. Unbiased studies into what users may actually want or need from their privacy controls will help inform how these controls develop, and what baseline best practices can be set in the future. However, this research is not going to materialize without incentivization, particularly monetary incentivization. The government should be actively investing or making the way easier for independent research groups working on privacy tools and research.

What's Missing

Trust. The NTIA's proposed outcomes and high-level goals are focused on user outcomes, but one additional high-level goal we think should be included is user trust. Trust may be hard to quantify, but makes a useful category to encompass many privacy-preserving actions on behalf of both companies and governments. Some of the activities which encourage trust include: transparency around data collection, retention, rental/sale, and surveillance practices; a historical record of data minimization; and actions taken to secure data and ensure that breaches are handled in a safe manner. As companies are incentivized to and do adopt trust-building practices, it follows those whose speech may have been chilled previously may feel more comfortable participating online, leading to the representation of a larger variety of perspectives, which is important to projects like Wikipedia.

Proportionality. The idea of proportionality is already embodied in certain concepts within the NTIA's existing proposed framework through the mention of "flexibility" in both the privacy outcomes and high level goals. However, we would like to draw a distinction between flexibility and proportionality that we feel would make proportionality a complementary but not duplicative goal. Flexibility in a regulation ensures that everyone has the ability to address privacy concerns in a way that is most intuitive to them. This is an important goal and should remain in the NTIA's framework. However, proportionality in a regulation ensures that the burdens are not equally placed on every actor, despite vast differences in their operations. The standard that applies to a small organization that collects demographic data about a few hundred clients to improve their services should not be the same as the one that applies to a company that collects sensitive information about age, gender, sexual orientation, and race about millions of people every day. Having a baseline which everyone must meet will ensure some measure of user privacy, but meaningful protection for privacy rights online also means ensuring that the requirements and consequences of a law are proportional to the gravity of the potential privacy violations at stake. The NTIA should ensure that as data collection and use scales, so do the burdens of consent, transparency, and protection of data.

Conclusion

We thank you for considering the Wikimedia Foundation's perspective on privacy as a part of this consultation. As an organization dedicated to providing free access to knowledge for everyone, we view privacy as an essential protection which allows freedom of expression to thrive online. This is why we are excited to see what type of privacy guidelines the NTIA will propagate, and why we believe our input will be especially helpful. Privacy laws that will be truly impactful for internet users must be clear, focus on user experience, and be backed up with investment in research and best practices from the government. The NTIA's framework must also incorporate the concepts of trust and proportionality, which ensure that regulations treat organizations differently depending on their relative track records, size, and data collection practices. With these additional suggestions in mind, the NTIA's privacy framework has the potential to be a good starting point to create some much needed harmonization in the U.S. privacy landscape.

Sincerely,
Eileen B. Hershenov
General Counsel, Wikimedia Foundation