**Comments of**

**The Wikimedia Foundation**

**In the Matter of**

*Trade Regulation Rule on Commercial Surveillance and Data Security*

DOCKET ID: FTC-2022-17752

Commercial Surveillance ANPR, R111004

**October 20, 2022**

## Introduction

The Wikimedia Foundation ("the Foundation") submits these comments in response to the Federal Trade Commission's ("FTC") advanced notice of proposed rulemaking ("ANPRM") on Commercial Surveillance and Data Security. The ANPRM asks a number of important questions about current personal data gathering and use practices, harmful impact of certain uses of data, the FTC's authority to protect consumers' privacy through a rulemaking, and other important questions about data security and consumer privacy. The Foundation appreciates the opportunity to submit these comments, highlighting its strong privacy protections and transparency surrounding its data collection practices. We hope these comments will be useful to the FTC as it considers this important rulemaking.

### Statement of Interest

The Foundation is a charitable, nonprofit organization which hosts, and provides the technical infrastructure for twelve (12) online projects dedicated to creating and providing free knowledge to a worldwide audience. As such, the Foundation is not subject to FTC jurisdiction under the Federal Trade Commission Act, 1914 ("FTC Act"), under which the proposed rules would be promulgated. The Foundation, therefore, would likely not be required to comply with any rule issued by the FTC at the conclusion of these proceedings, absent congressional intervention. Nonetheless, Congress is actively debating enacting new consumer privacy legislation, which, if enacted, would apply to the Foundation's practices. Moreover, any rule the FTC does enact will likely become the starting point for any future national privacy rules and legislation. The Foundation is also the host of global websites that collect personal data and has strong privacy practices that are informed by its public interest mission. For these reasons, the Foundation strongly supports this process and offers its views on some of the questions posed in the ANPRM in the hopes that the Foundation's practices may serve as an example of desired outcomes.

### Wikimedia Foundation Background

The most well-known project hosted by the Wikimedia Foundation is Wikipedia, a free and open online encyclopedia, which includes more than 54 million articles in more than 300 languages. Wikipedia is among the most visited websites in the world. Wikipedia is curated, edited and verified by more than 400,000 independent contributors around the world, each of whom is a volunteer, collectively referred to as the "Wikimedia Community." The diversity and openness of

the Wikimedia Community has proved to be an important contributor to the reliability of Wikipedia's content.

The Foundation and the Wikimedia Community are strongly committed to privacy. The goal of the Wikimedia projects is to provide access to freely reusable, objective, and verifiable content that everyone can edit and improve. As the encyclopedia is open to all to contribute to, the Wikimedia community has created policies to ensure the information on Wikipedia is and remains accurate. These include policies which require citation of verifiable secondary sources for every fact included on Wikipedia, and transparency requirements regarding conflicts of interest. Most notably, in the context of privacy, and by way of example, Wikipedia editors for English Wikipedia must comply with the policy on Biographies of Living Persons,[1] which mandates a presumption in favor of privacy when writing about people who are still alive. Contributors who do not comply with Wikipedia's principles may have their edits removed and may be blocked temporarily or permanently by volunteer administrators, the latter being elected by the community of volunteer users. The Wikimedia projects are also subject to a set of guiding principles,[2] which include transparency and accountability. Furthermore, Wikipedia's image use policy[3] also deals with privacy rights. These policies for content on Wikimedia projects work to balance the privacy interests of people whose information might appear on the projects while supporting the projects' goals of increasing the store of all knowledge. The delicate balance between privacy and free expression is one that the Foundation anticipates the FTC will also need to consider in its process. We believe the policies mentioned above will prove helpful examples as the FTC considers what best practices could be.

In addition to favoring privacy for content on the projects that it hosts, the Foundation is committed to actively promoting wide and equitable participation. Privacy is essential to this effort. The logic is simple: storing personally identifying information imperils the privacy of contributors and can discourage some individuals from contributing in the first place. Out of this principle, the Foundation collects very little information about both readers and contributors to Wikimedia projects and also facilitates pseudonymous contributions. The Foundation does not engage in online advertising or have a commercial focus.

[1] Biographies of Living Persons, Wikipedia, https://en.wikipedia.org/wiki/Wikipedia:Biographies_of_living_persons.
[2] Guiding Principles, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Resolution:Wikimedia_Foundation_Guiding_Principles.
[3] Image Use Policy, Wikipedia, https://en.wikipedia.org/wiki/Wikipedia:Image_use_policy#Privacy_rights.

The Foundation is strongly committed to the protection of user privacy in relation to the users of Wikimedia's knowledge projects, including by way of Wikimedia's Privacy Policy,[4] Terms of Use[5] and Human Rights Policy.[6] That user base includes contributors to the Wikimedia projects and donors[7] to the Wikimedia Foundation who are located in the United States of America and around the globe. The links to the above-mentioned policies have been annexed as **Annexure II** to this letter.

With the above as background, we provide the following answers to some of the questions presented in the FTC's ANPRM annexed as **Annexure I**. Please do not hesitate to reach out to Kate Ruane, Lead Public Policy Specialist for the United States, kruane@wikimedia.org, with any questions.

Sincerely,

The Wikimedia Foundation

[4] Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Privacy_policy.
[5] Terms of Use, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Terms_of_Use/en.
[6] Human Rights Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Policy:Human_Rights_Policy.
[7] Wikimedia Donor Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Donor_privacy_policy/en,

WIKIMEDIA
FOUNDATION

## Annexure I

## Answers to ANPRM Questions

**1. Which practices do companies use to surveil consumers?**

**2. Which measures do companies use to protect consumer data?**

**3. Which of these measures or practices are prevalent? Are some practices more prevalent in some sectors than in others?**

**4. How, if at all, do these commercial surveillance practices harm consumers or increase the risk of harm to consumers?**

Essentially all business entities, online and offline, for-profit and not-for-profit, interact with personal information in some way. These entities have a variety of data policies, shaped by their incentives, the products they offer, their size, their regulatory obligations within the United States and abroad, and many other factors. The Foundation anticipates that many commenters will address the practices of for-profit entities in their answers to questions 1-4. The Foundation's intent in answering these questions is to provide an example of privacy practices implemented by a nonprofit organization that values privacy as part of its core mission, does not earn money from selling advertisements to be displayed on any of its platforms, does not permit third parties to track users through Wikimedia projects for any reason, and operates in the public interest.

The Foundation's public interest mandate contributes to a different relationship with the data to which the Foundation has access, how it is used, and how long it is kept. Wikimedia is committed to protecting the human rights of its users, including privacy as a crucial right that enables many others. This commitment is reflected in the Foundation's Privacy Policy,[8] Terms of Use,[9] and Human Rights Policy.[10] In particular, the Foundation is keenly aware that participation in Wikimedia projects can expose users to external threats, particularly in regions of the world where governments may attempt to control the dissemination of information. Each of our privacy-related policies is guided by this awareness and by the idea that the minimization of personal information held by the Foundation can provide an additional layer of safety to the people we serve.

[8] Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Privacy_policy.
[9] Terms of Use, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Terms_of_Use/en.
[10] Human Rights Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Policy:Human_Rights_Policy.

The Foundation practices data minimization in the first instance by collecting very little personal information. Currently, a user does not need to sign up for an account or sign into their account in order to contribute to any of the projects. When people do create accounts, the Foundation does not require them to provide their real names or even an email address to hold an account. However, we do note in our privacy policy that the Foundation can only help recover access to accounts for which a person has provided an email address, and that the Foundation also collects information related to accounts that would be considered personal information, even though real names are not required. The Foundation also collects information as people use the websites, make public contributions, or engage with Foundation staff by email or otherwise provide feedback to the Foundation when requested. The Foundation makes clear that it collects this information in order to understand how the Wikimedia websites are used, so that the Foundation can improve operations and make the websites more useful.

In addition, the Foundation is transparent regarding how, when and with whom it may share[11] personal information. Most importantly, the Foundation never sells personal information, does not earn money from selling advertisements, and does not permit third parties to track users through Wikimedia projects for any reason. The Foundation only shares personal information in a few general ways on a case-to-case basis:

- With consent;
- To fulfill its legal obligations, including in response to a warrant, court order or other valid law or regulation. However, the Foundation strives to notify all individuals whose information is disclosed as a result of response to legal process, except when we are legally barred from doing so;
- To protect the projects, the Foundation, and the community, when it is reasonably necessary to enforce or investigate any violation of the Terms of Use,[12] including by sharing with certain volunteer administrators who have been elected to monitor the projects for vandalism and other signs of abuse;
- With service providers the Foundation engages to help run or improve the Wikimedia projects, which are bound by contract not to use the data in any way other than those permitted by the Foundation; and,
- To understand and experiment in ways that improve the projects, through limited sharing as necessary with open source software developers working to improve and evolve the software that powers Wikipedia.

[11] Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Privacy_policy#when-we-may-share.
[12] Terms of Use, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Terms_of_Use.

The Foundation has extremely short data retention[13] times. Most non-public personal information is deleted, aggregated or de-identified within ninety (90) days if it is retained at all. In addition to being a strong privacy practice, our short data retention window also increases data security, as discussed in more detail below, because it makes Foundation databases less attractive to hackers. Basically, if we do not have it, no one can steal it.

The Foundation is also accountable to the Wikimedia Community for its privacy practices. Our Privacy Policy[14] is thorough and understandable to readers and contributors to the projects. It was developed with community input, and we periodically carry out additional consultation from the Wikimedia Community as we seek to improve privacy practices. We also have a volunteer ombudsperson commission that responds to any privacy-related concerns and interacts with Foundation staff to address privacy-related incidents.

These practices and policies make clear that the Foundation's priority with respect to personal information is privacy and the protection of our community of users in a way that supports the development of the projects we host. We use a combination of data minimization, limited data retention, and transparency regarding the data we collect and how we use it in order to ensure that the community who edits and reads Wikimedia projects can do so freely and safely. We hope that outlining these strong practices will aid the FTC in its task of determining potential rules for limiting commercial surveillance.

**6. Are there some harms that consumers may not easily quantify or measure? Which are they?**

The harms associated with the proliferation of commercial surveillance in public and private spaces, and various sectors of the economy, may not be easily quantified or measured for years, if not decades. The pervasive tracking of people across services and websites for the purpose of targeting them with ever more precise advertising is the primary driver of commercial surveillance. This targeted advertising alone produces harm.[15] One of the most pervasive and detrimental harms of commercial surveillance is the "chilling effect" that the expansion of surveillance—both actual and perceived—has on individuals' behavior and increasing

[13] Data Retention Guidelines, Wikimedia Foundation, https://meta.wikimedia.org/wiki/Data_retention_guidelines.
[14] Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Privacy_policy#when-we-may-share.
[15] Silvia Milano, *Targeted ads aren't just annoying, they can be harmful*, Fast Company (Jul. 31, 2021), https://www.fastcompany.com/90656170/targeted-ads-arent-just-annoying-they-can-be-harmful-heres-how-to-fight-back; Rae Nudson, *When targeted ads feel a little too targeted*, Vox (Apr. 9, 2020), https://www.vox.com/the-goods/2020/4/9/21204425/targeted-ads-fertility-eating-disorder-coronavirus.

self-censorship. Consciously or unconsciously, people will alter what they do, search for, ask, express, and access when they are aware that their activities, habits, and communications may be being monitored. This has harmful effects on their mental health and wellbeing, their rights to privacy, free association, and free expression, including the exchange of free knowledge and information.

Private companies' surveillance of people online amplifies governmental surveillance of people online, and the knowledge of surveillance carried out by both private and government actors contribute to these pernicious chilling effects. For instance, in 2016, Jonathan W. Penney of York University in Toronto released the first original empirical study of the regulatory chilling effects associated with online government surveillance. The study, "Chilling Effects: Online Surveillance and Wikipedia Use,"[16] successfully quantified the impact of such surveillance on Wikipedia users and articles, and web traffic data more generally, resulting from the June 2013 NSA PRISM surveillance program revelations. Penney chose to focus on Wikipedia because it is an "essential source of information and knowledge online" and an "important public tool in promoting collective understanding, decision-making, and deliberation." Therefore, as he argues, any demonstrated chilling effect on Wikipedia users has broader implications for the global free knowledge movement and democratic processes.

The analysis demonstrated a statistically significant, immediate decline in traffic for privacy-sensitive Wikipedia articles after June 2013, as well as a statistically significant long-term trend shift of decreasing traffic to such articles. In other words, empirical evidence showed a long-term chilling effect on accessing article content raising privacy concerns.[17]

More recently, Penney noted[18] that while the study focused on the impacts of a program enacted in the wake of the "War on Terror" and that the U.S. government defended as a necessary "anti-terrorism" measure, surveillance-related chilling effects are not limited to war or national security-related events. Discussions about chilling effects increased in the lead up to and in the wake of the Supreme Court's decision in *Dobbs v. Jackson Women's Health*,[19] particularly

[16] Jon Penney, *Chilling Effects: Online Surveillance and Wikipedia Use*, 31 Berkeley Tech, L.J. Vol. 1, 117 (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2769645.
[17] *Id*. at 147-48; 152-53.
[18] Benjamin Powers, *The ACLU and NSA May Soon Square Off At the Supreme Court - About Wikipedia*, Grid.News (Sept. 27, 2022), https://www.grid.news/story/technology/2022/09/27/the-aclu-and-the-nsa-may-soon-square-off-in-the-supreme-court-over-wikipedia/.
[19] *Dobbs v. Jackson Women's Health*, ___ U.S. ___, No. 19-1392 (2022), https://www.supremecourt.gov/opinions/21pdf/19-1392_6j37.pdf.

among civil liberties[20] and pro-choice advocacy groups.[21] Many of these discussions emphasized the free expression-related implications of the collection and intentional or inadvertent disclosure[22] of consumer-level browsing,[23] location,[24] and communications data that may reveal whether a person is pregnant[25] or seeking an abortion, or helping someone who is,[26] thus opening them up to potential legal liability in at least 15 states.[27] In at least one example, a Nebraska woman was criminally charged[28] for helping her 17-year-old daughter obtain an abortion after state investigators obtained their private Facebook messages discussing the termination and the instructions that came with the medication. It is worth considering how these looming restrictions can impact the free exchange of knowledge and the reliability of articles on projects like Wikipedia. If people are afraid to access information about abortion or are restricted from communicating accurate information about an important medical procedure, like abortion, the reliability of information about these procedures that is publicly available, including on Wikpedia, could suffer. Ensuring that people are able to share information privately and free from commercial surveillance will preserve free knowledge and protect human rights.

**10. Which kinds of data should be subject to a potential trade regulation rule? Should it be limited to, for example, personally identifiable data, sensitive data, data about**

---

[20] Summer Lopez and Nadine Farid Johnson, *Why SCOTUS Abortion Ruling is a Disaster for Free Expression*, The Daily Beast (Jul. 17, 2022), https://www.thedailybeast.com/why-scotus-abortion-ruling-is-a-disaster-for-free-expression.
[21] Kade Crockford and Nathan Wessler, *Impending Threat of Abortion Criminalization Brings New Urgency to the Fight for Digital Privacy*, ACLU (May 17, 2022), https://www.aclu.org/news/privacy-technology/impending-threat-of-abortion-criminalization-brings-new-urgency-to-the-fight-for-digital-privacy.
[22] Heather Kelly, Tatum Hunter and Danielle Abril, *Seeking an Abortion? Here's how to avoid leaving a digital trail*, Wash. Post (Aug. 12, 2022), https://www.washingtonpost.com/technology/2022/06/26/abortion-online-privacy/.
[23] Raquel Maria Dillon, *Google workers sign petition asking company to protect people's abortion search data*, NPR (Aug, 18, 2022), https://www.npr.org/2022/08/18/1118051812/google-workers-petition-abortion-data.
[24] Alfred Ng, *'A uniquely dangerous tool': How Google's data can help states track abortions*, Politico (Jul. 18, 2022), https://www.politico.com/news/2022/07/18/google-data-states-track-abortions-00045906.
[25] Press Release, Supreme Court Abortion Ruling "Devastating" for Women's Right to Privacy, Cen. Dem. Tech. (June 24, 2022), https://cdt.org/press/cdt-supreme-court-abortion-ruling-devastating-for-womens-right-to-privacy/.
[26] David S. Cohen, Greer Donley, and Rachel Rebouché, *The Harshest Abortion Restrictions are Yet to Come*, The Atlantic (Jul. 11, 2022), https://www.theatlantic.com/ideas/archive/2022/07/pro-life-legal-strategies-abortion/661517/.
[27] New York Times, *Tracking the States Where Abortion is Now Banned*, NY Times (updated Oct. 13, 2022), https://www.nytimes.com/interactive/2022/us/abortion-laws-roe-v-wade.html.
[28] Assoc. Press, *Nebraska woman charged with helping daughter have abortion*, Politico (Aug. 9, 2022), https://www.politico.com/news/2022/08/09/nebraska-woman-charged-with-helping-daughter-have-abortion-00050763.

**protected categories and their proxies, data that is linkable to a device, or non-aggregated data? Or should a potential rule be agnostic about kinds of data?**

The Foundation favors a broad definition of "personal information." Under the Foundation's Privacy Policy,[29] we consider personal information to be any information a user provides to us and any information that Wikimedia collects that could be used to identify a person. We take this broad view because studies[30] have[31] shown[32] that it can take very few data points to trace that information to a particular individual. It is also quite difficult to discern which personal information is "sensitive." Some people may not find their gender identity, sexual orientation or race to be sensitive information; others may. Rather than attempting to delineate which information is sensitive and which is "not," Wikimedia prefers a broad definition of personal information accompanied by strong data minimization requirements and use restrictions that prevent harmful uses of personal information. An example of a harmful use would be platforms using personal information to discriminate against people in their ability to access a service on the basis of their race, gender, gender identity, sexual orientation or membership in any other protected class. We recommend the FTC follow the same course and should define personal information to include all data that is linked or reasonably linkable to an individual or device.

**11. Which, if any, commercial incentives and business models lead to lax data security measures or harmful commercial surveillance practices? Are some commercial incentives and business models more likely to protect consumers than others? On which checks, if any, do companies rely to ensure that they do not cause harm to consumers?**

When examining the impact of business models on commercial surveillance practices, the Foundation would like to emphasize the constructive impact that nonprofit, public interest business models can have on consumer privacy awareness and choice. Public interest organizations that incorporate robust privacy protections into their mission can help to educate consumers on their options when it comes to privacy, while providing vital services free from commercial surveillance. As a nonprofit organization, the Foundation is able to make plans for the Wikimedia projects and their operations based on core values rather than profit motives. Our

[29] Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Privacy_policy#when-we-may-share.
[30] Natasha Lomas, *Researchers spotlight the lie of anonymous data*, TechCrunch (Jul. 24, 2019), https://techcrunch.com/2019/07/24/researchers-spotlight-the-lie-of-anonymous-data/.
[31] Larry Hardesty, *Privacy Challenges*, MIT News (Jan. 29, 2015), https://news.mit.edu/2015/identify-from-credit-card-metadata-0129.
[32] Stuart Thomson and Charlie Warzel, *One Dataset, Zero Privacy*, NY Times (Dec. 19, 2019), https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html.

commitments to keep the WIkimedia projects free from advertisements and to minimize the data collected on our readers and contributors are not just enabled by our nonprofit business model, but essential to it as well. Through public education and transparency around our privacy practices, the Foundation is able to better advance our goal of access to free knowledge because our readers and contributors understand they can be more secure in their browsing and editing on Wikimedia projects. Strong privacy rules that take lessons from the practices of the public interest focused internet will incentivize the prioritization and protection of online communities, leading to a better internet overall.

**18. To what extent should trade regulation rules distinguish between different age groups among children ( *e.g.,* 13 to 15, 16 to 17, etc.)?**

Enacting broadly-applicable trade regulations, which would require all covered entities to divide users by age and limit access to content accordingly, would force platforms to collect personally identifiable information (PII), including age and identity information, on each and every user in order to comply with the regulations. If such a requirement applied to Wikipedia, for example, this could require collecting more personal data than currently collected. Furthermore, a requirement that entities collect and hold more data would appear to be antithetical to the overarching goals of protecting privacy and creating accountability for commercial surveillance that underpin these proceedings.

At a minimum, if the FTC is to proceed with the possibility of requiring entities to make distinctions on the basis of age, the agency must take into account the different types of platforms—with unique functions, features, policies, governance structures, and content moderation models—when enacting these regulations. If it fails to do so, the FTC would be damaging community-governed platforms and those that collect minimal user data by design. If such a requirement could be imposed upon the Wikimedia Foundation, it would require a platform like Wikipedia to collect more user data than is currently—or has historically been—collected. This goes against data minimization principles and raises significant privacy concerns.

The FTC must also consider that children have rights to privacy and free expression,[33] separate and apart from those of their guardians and families. Their privacy and free expression rights counsel against incentivizing practices like “age-gating,” which would effectively force platforms

[33] Richard Gaines, *Wave of Bills Banning Topics in US Schools Threaten Free Knowledge,* Wikimedia Foundation (Apr. 12, 2022), https://medium.com/wikimedia-policy/wave-of-bills-banning-topics-in-us-schools-threatens-human-rights-and-free-knowledge-ccd3efdf420e.

to collect additional personal data and to erect barriers to information, infringing upon minor users' rights to privacy and freedom of expression, including the rights to seek out and access information. As a community-governed platform, if the Foundation and the Wikimedia projects were covered under such proposed regulations, neither the Foundation nor Wikimedia Community would find such levels of personal data collection and censorship acceptable, and would be deeply troubled by the notion of barring anyone from access to reliable information and knowledge.

The Foundation is currently in the process of conducting a Child Rights Impact Assessment (CRIA) to better understand and mitigate the risks that minors face on Wikimedia projects, while still protecting those users' rights to privacy and freedom of expression. Projects like Wikipedia can have an affirmative impact on children's rights because minor users are able to exercise their freedom of expression, participate in self-governance, share their opinions, and access information to inform their own opinions about social and political topics. The Foundation understands online interactions can have negative consequences, such as potentially being exposed to harassment, harmful content, or inappropriate communication. As an organization that is committed to upholding the human rights of its users, the Foundation is committed to mitigating these risks. This CRIA will help to identify mitigations that are consistent with our community-led model.

Every online platform is different. Enforcing a uniform set of rules would have serious unintended consequences to many of them. Rather than regulate platforms broadly and uniformly, the FTC should encourage platforms to be responsible actors. Platforms should be encouraged to take the time to self-reflect and conduct their own impact assessments in order to evaluate the platform-specific risks and opportunities, find ways to mitigate those risks while enhancing the positive and affirmative impacts, and produce more tailored recommendations to protect children's rights.

**30. Should the Commission pursue a Section 18 rulemaking on commercial surveillance and data security? To what extent are existing legal authorities and extralegal measures, including self-regulation, sufficient? To what extent, if at all, are self-regulatory principles effective?**

The Foundation supports the FTC's pursuit of a Section 18 rulemaking. Comprehensive data privacy rules in the United States are long overdue. The Foundation would also support federal legislation on this issue. In the absence of the enactment of new strong legislation, the creation of rules using existing authority is a desirable step toward protecting people's privacy.

Self-regulatory measures are, essentially, what is currently in place within most of the United States to protect people from commercial surveillance. The current state of the market suggests that self-regulatory measures are not sufficient to protect consumers' personal information, privacy, and interests because businesses are able to generate profit through commercial surveillance that often outweighs any incentive to create or obey strong self-regulatory regimes. There needs to be a set framework for and means of enforcement to back up the regulatory program, including: (a) clear delineations of which entities are bound and which are not; (b) set procedures for amendments and additions to the rules; (c) penalties for first and subsequent offenses; and, (d) processes for evaluating, adjudicating, and resolving claims of noncompliance. Without this, certain industry players will avoid complying voluntarily and thereby gain a competitive edge over those who are compliant. Moreover, commercial surveillance vendors from other countries—e.g., Israel, China—will have little to no incentive to voluntarily comply with toothless regulations in the U.S.

32. **Should, for example, new rules require businesses to implement administrative, technical, and physical data security measures, including encryption techniques, to protect against risks to the security, confidentiality, or integrity of covered data? If so, which measures? How granular should such measures be? Is there evidence of any impediments to implementing such measures?**

The Foundation believes that effective organizational data security measures are critically important for protecting the privacy of data subjects. These measures must be organization-wide, encompassing both administrative and technical procedures, and including encryption. They must not only be technically sound, but must also recognize the inherently sociotechnical nature of data security, and may require paradigm shifts with regard to who is trusted by default to access sensitive user data.

A large part of the Foundation's data security ethos is based on the concept of security through transparency, as opposed to security through obscurity. At its core, this philosophy boils down to clear communication and public accountability. As a nonprofit organization with a user base comprising hundreds of millions of people, the Foundation is strongly incentivized to make policies about privacy[34] and data retention[35] public and understandable to readers and contributors to the projects. The Foundation publishes source code[36] and in-depth

[34] Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Privacy_policy.
[35] Data Retention Guidelines, Wikimedia Foundation, https://meta.wikimedia.org/wiki/Data_retention_guidelines.
[36] Wikimedia, Github, https://github.com/wikimedia.

documentation[37] of existing systems (including descriptions of private tables), as well as hosts forums for policy and technical debate.[38]

To put it simply: we work in public. Besides the cryptographic keys and passwords that manage access to our systems, there are very few secrets.

This set of values has major positive implications for the Foundation's data security program. We transparently implement access control mechanisms, strong SSH (i.e., SecureShell) keys, DDoS (i.e., distributed denial-of-service) protection services, and related technologies to keep our platform secure. External actors can verify[39] that we are not collecting more information than we say we are, that we are following our data retention policy by deleting data after ninety (90) days, and that only trusted community members have access to sensitive databases.

By being transparent and establishing provenance for decision-making, we ensure a safer and more accountable place for everyone's data. We hope that by explaining these principles and practices, we can help the internet become a safer and more accountable place too.

**33. Should new rules codify the prohibition on deceptive claims about consumer data security, accordingly authorizing the Commission to seek civil penalties for first-time violations?**

The Foundation believes that new rules ought to codify the prohibition on deceptive claims about consumer data security, and authorize the FTC to seek civil penalties for first-time violations. Industry norms are very important in this realm: when organizations handling sensitive data make deceptive claims about their data security practices, it has an impact on the data ecosystem more broadly. Deceptions, omissions of truth, and misleading marketing statements affect consumer confidence when they are uncovered, and a lack of disincentives against these sorts of behaviors create a perverse race to the bottom for industry actors.

The Foundation, as a mission-driven nonprofit, has a completely different set of incentives in its structure from most organizations that handle personal data. The Foundation's goal is not to use data to make a profit: rather, our goal is to improve the projects, and to better support the creation and exchange of free knowledge. This fact informs a slate of transparent policies and

[37] Wikitech, Wikimedia Foundation, https://wikitech.wikimedia.org/wiki/Main_Page.
[38] Meta-Wiki, Wikimedia Foundation, https://meta.wikimedia.org/wiki/Main_Page.
[39] Data deletion and sanitization, Wikitech, https://wikitech.wikimedia.org/wiki/Analytics/Systems/Cluster/Data_deletion_and_sanitization

processes as well as secure software and hardware practices that build trust with the communities that develop the projects. Hence, penalties for first-time violations, if authorized by the FTC, would change the landscape of incentives that has so far led to much consumer harm. The Foundation supports creating market incentives to induce companies to act in the interest of data security, and believes that other organizations that collect, store, and use personal data ought to support these measures as well.

**37. How do companies collect consumers' biometric information? What kinds of biometric information do companies collect? For what purposes do they collect and use it? Are consumers typically aware of that collection and use? What are the benefits and harms of these practices?**

The Foundation strongly supports the creation of robust protections for individuals' biometric information. We will leave the particulars of how best to enforce those protections to other organizations. We write in response to this question to offer a word of caution regarding the definition of biometric information. The Foundation does not collect biometric information at this time; however, we do host a large amount of content, including photographs and voice recordings, on Wikimedia Commons.[40] This Foundation-hosted project is one of the largest repositories of open images, sound, and other media in the world. These media contain information that could be used to generate biometric information or could be considered biometric information if the definition of biometric information is written imprecisely.

To ensure that increased protections can be given to biometric information while preserving the free exchange of information, the Foundation recommends that the FTC take cognizance of these facts and exclude photographs, videos, voice recordings, handwritten text, and other similar information from the definition of biometric information. Definitions that accomplish this goal already exist both in U.S. legislative proposals[41] and in existing U.S. state law.[42] We encourage the FTC to refer to these laws and proposals as it considers how to define biometric information.

**43. To what extent, if at all, should new trade regulation rules impose limitations on companies' collection, use, and retention of consumer data? Should they, for example, institute data minimization requirements or purpose limitations, i.e., limit**

[40] Wikimedia Commons, https://commons.wikimedia.org/wiki/Main_Page.
[41] S. 4400, National Biometric Privacy Act (116th Cong. 2020), https://www.congress.gov/bill/116th-congress/senate-bill/4400/text.
[42] Biometric Information Privacy Act, 740 ILCS 14, https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57.

**companies from collecting, retaining, using, or transferring consumer data beyond a certain predefined point? Or, similarly, should they require companies to collect, retain, use, or transfer consumer data only to the extent necessary to deliver the specific service that a given individual consumer explicitly seeks or those that are compatible with that specific service? If so, how? How should it determine or define which uses are compatible? How, moreover, could the Commission discern which data are relevant to achieving certain purposes and no more?**

44. **By contrast, should new trade regulation rules restrict the period of time that companies collect or retain consumer data, irrespective of the different purposes to which it puts that data? If so, how should such rules define the relevant period?**

The Foundation, following the ethos of data minimization, supports the imposition of limitations on how organizations collect, use, and retain consumer data. Our organizational policies mandate minimization both in the collection and retention of data. Due to the open-source, public access-focused nature of Wikimedia projects, we also default to making much of our data available to the public after it has been aggregated and de-identified, allowing any entity to download, reuse or remix it to their liking.

This set of policies creates an incentive structure that rewards the responsible handling of consumer data while impeding possibilities for data breaches or misuse. Because of these values, we cannot collect or retain too much data. In addition, it makes it easier to create data applications using public data rather than private data, since any data that might have been sold to other parties prior to aggregation and de-identification is already public and freely available.

By imposing limitations on how organizations can collect, use, and retain data, including considering restrictions on the period of time that companies can collect or retain data, the FTC can create market-wide incentives for the responsible handling of consumer information. Because the FTC has this power, the Foundation supports restricting the collection of data to only what is reasonably necessary to provide a particular service, as well as retention time limits for consumer data. We note that there will be complexity when attempting to distinguish which data are reasonably necessary to provide a service and when additional uses of that data are permissible because the answer to those questions will depend heavily on the nature of the service and the nature of the additional uses of the data. In its efforts to build clarity around newly announced standards, we encourage the FTC to draw distinctions between uses of data that improve services in the public interest and those that target and track users for the sole

purpose of generating profit. We note specifically that companies' desire to increase their profits often leads to the imperative of collecting and using more data than necessary to provide the services people want, which compounds risks to privacy. Strong data minimization requirements will be crucial to reducing those risks.

**47. To what extent would data minimization requirements or purpose limitations protect consumer data security?**

Data minimization requirements are a useful—and seemingly underutilized—tool for protecting consumer data security. In the case of the Foundation, data minimization has two prongs: 1) limiting the amount and kinds of sensitive data that may be collected from our services; and, 2) only keeping most raw personal data for ninety (90) days before it is aggregated and de-identified.

For example, it is a common practice on many websites to place first-party tracking cookies on users' browsers to log their activity for analytics purposes. The Foundation has judged that this practice opens our users to unwarranted surveillance, and as such does not use first-party tracking cookies to track the page-viewing history of non-signed-in users. The information that is indeed collected in the course of a typical web interaction (i.e., IP address and UserAgent) is only kept for ninety (90) days before it is aggregated and deleted.

These strictures are a key aspect of the Foundation's data security approach. Since we are not sitting on decades of in-depth data that can be tied to individual users, our datasets are less attractive to hackers or other people—insiders or outsiders—who might want to compromise the data security of our users. They cannot hack, leak or misuse what we do not have.

**64. To what extent, if at all, does Section 230 of the Communications Act, 47 U.S.C. 230, bar the Commission from promulgating or enforcing rules concerning the ways in which companies use automated decision-making systems to, among other things, personalize services or deliver targeted advertisements?**

Strong intermediary liability protections, like those afforded by Section 230 (47 U.S.C. § 230), are vital to Wikimedia projects' existence. This is so even with the protections the First Amendment to the U.S. Constitution also affords to community-driven projects. Section 230 allows communities to edit projects like Wikipedia, and establish and enforce their own content moderation rules in a distributed way with safety and security top of mind, rather than focusing on avoiding liability risk. It is important, therefore, that whatever action the FTC takes to protect privacy, the agency also considers the limits Section 230 may place upon its authority, because

those limitations are critical to maintaining the balance between free expression online and protecting privacy.

Section 230 generally protects interactive computer service providers, particularly online platforms that serve as conduits for the expression of their users, from liability based upon the content that they publish when that content was created by those users.[43] However, Section 230 does not shield online platforms from liability for the content they create themselves, the content they materially contribute to creating, or for their own conduct. These limitations to the scope of Section 230's shield are important because they ensure that Section 230 fulfills its purpose of encouraging free expression online without functioning as a "get out of jail free" card for online platforms.

In the context of the use of automated decision making systems, the scope of section 230 is the same as it is in any other context. If the claim seeks to impose liability upon an online platform because of the content the platform is hosting, then Section 230 should serve to shield that platform from liability.[44] For instance, many express legitimate concern that automated decision making systems "amplify" harmful content, like hate speech, terrorist content, and disinformation.[45] Section 230 would likely bar any claims for relief for such automated "amplification" because the basic objection is to the offensive nature of the content itself. And that is to say nothing of the fact that the First Amendment would likely bar any liability claims for such content, offensive as it may be, in any case.

However, if the claim seeks to impose liability upon an online platform for its own conduct and not for the content of the published material, Section 230 should not pose a bar to liability. A good example of this distinction occurs in the civil rights context, but could be extended to anticompetitive conduct and other unfair and deceptive practices that might be achieved through an automated system. In the civil rights context, plaintiffs have sued online platforms like Facebook for excluding people from viewing job, housing, and credit opportunities on the basis of race, gender or other protected characteristics. The liability claims are not based upon the content of any of the advertisements. Rather, the claims argue[46] that the platforms made

[43] Valerie C. Brannon and Eric N. Holmes, *Section 230: An Overview*, Cong. Research Serv. Rep. R46751 (Apr. 7, 2021) https://crsreports.congress.gov/product/pdf/R/R46751.
[44] Daphne Keller, *Amplification and its Discontents*, Knight Foundation (June 8, 2021), https://knightcolumbia.org/content/amplification-and-its-discontents.
[45] *Id.*
[46] *See* Brief of The American Civil Liberties Union Foundation, Free Press, The Lawyers Committee for Civil Rights Under Law, and the National Fair Housing Alliance, *Vargas v. Facebook*, No. 21-16499 (9th Cir. filed Jan. 1, 2022). https://www.freepress.net/sites/default/files/2022-01/2022.1.26_Vargas_Amicus_Brief.pdf .

decisions regarding the target audience for ads for housing, employment, and credit opportunities, and in doing so made decisions based on gender, race, other protected characterstics or close proxies for those characteristics, and that those decisions that the platforms made resulted in discrimination in the availability of opportunities for jobs, credit or housing.

The Foundation encourages the FTC to be cognizant of this distinction as it proceeds with this rulemaking regarding commercial surveillance. Section 230 may be a barrier to liability when the claim is based upon content provided by a user of the service, but it should not be a bar to any claim based upon the discriminatory, anticompetitive or otherwise unfair or deceptive conduct of any only platform.

> **73. The Commission invites comment on the effectiveness and administrability of consumer consent to companies' commercial surveillance and data security practices. Given the reported scale, opacity, and pervasiveness of existing commercial surveillance today, to what extent is consumer consent an effective way of evaluating whether a practice is unfair or deceptive? How should the Commission evaluate its effectiveness?**
>
> **74. In which circumstances, if any, is consumer consent likely to be effective? Which factors, if any, determine whether consumer consent is effective?**

Consumer consent is not an effective way of evaluating whether a surveillance or data use practice is unfair or deceptive. True consent to such practices requires that consumers have the time and expertise to read all their options under a privacy policy and contemplate all the various ways their data could be collected and used. It is an incorrect assumption, however, to think that consumers have the time and expertise in question. Recognizing that if a consumer were to read all of the privacy policies they encountered each year, it would amount to scrutinizing many hundreds—if not thousands—of pages of legal language, it becomes evident that consumer consent to company practices regarding their data is largely underinformed.

Consumer consent is unlikely to serve as approval of surveillance and data security practices as fair and, more likely, is only an action taken by the consumer because they are seeking to access a service or product without thorough comprehension of the use of their data. Given the amount of companies collecting and using information, as well as the fact that users are not negotiating or expecting such use of their information, meaningful consent often cannot be given. Circumstances in which consumer consent may be effective are when the request for consent is discrete and asks permission before using data in ways that otherwise would not

have been expected. Consent should be paired with data minimization, narrowing the scope of the data collected, the amount of data collected, and the retention of such data, all of which serve to reduce harm.

As mentioned, the Foundation collects and shares personal information with user consent. That consent is paired with data minimization and transparent practices that are comprehensive to users. This allows the consent to be as informed as possible, as evidenced by our community's feedback and inputs.

As our Privacy Policy[47] notes, we keep information related to a user's use of the Wikimedia projects confidential except as stipulated in the Policy itself. We identify the information we collect and are transparent about when we share user information. We are clear about how we collect and use information about people. With data minimization, limiting our collection of information solely for an immediate and necessary purpose and the minimal amount necessary to fulfill the user's purposes, we can naturally dampen the risks of data breaches, deception, and manipulation.

**80. Have opt-out choices proved effective in protecting against commercial surveillance? If so, how and in what contexts?**

The Foundation supports strong privacy protections that put people first. Regulatory systems cannot rely on consumer consent, whether opt-out or opt-in, to create strong privacy protections.[48] Most people, even seasoned privacy professionals, simply do not have the time or the ability to consider each and every potential use of their personal information and decide whether they want to permit the use or exercise an option not to permit the use. We have seen this phenomenon of consent-fatigue in the implementation of the European Union's General Data Protection Regulation (GDPR), which many websites have determined requires opt-in consent for websites to use "cookies." This has led to an endless series of pop-ups on websites asking for consent to use cookies, and an equally endless series of people mindlessly clicking the "I consent" button without reading the accompanying disclosures.[49]

Wikimedia instead supports a regulatory regime that relies primarily on data minimization, use restrictions, retention time limitations, and transparency to create strong privacy protections. A regime that requires entities to collect as little data as possible at the outset, forbids entities from

[47] Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Privacy_policy.
[48] Clare Park, *How Notice and Consent Fails to Protect Our Privacy*, Open Tech. Institute (March 23, 2020), https://www.newamerica.org/oti/blog/how-notice-and-consent-fails-to-protect-our-privacy/.
[49] Matt Burgess, *The Tyranny of GDPR popups and websites failing to adapt*, Wired (Aug. 29, 2018), https://www.wired.co.uk/article/gdpr-cookies-eprivacy-regulation-popups.

using data in ways that are clearly harmful or discriminatory, requires entities to delete or deidentify data as soon as reasonably practicable, and ensures robust transparency surrounding all other practices and uses of data will provide a strong basis for privacy protections. Within those basics as a foundation for the regulations, we believe there could be both opt-in and opt-out regimes that could be effective for supporting privacy protections.

For example, under California's Consumer Privacy Act, Californians will soon have the ability to universally opt-out of the sale of their personal information. The FTC operates the National Do Not Call Registry, which, for years, has allowed people to place their phone numbers on the list, making it illegal for telemarketers to call their numbers except under limited exceptions. Both of these regimes represent effective approaches to privacy protection within their contexts. The Foundation believes that an opt-out option could be effective at protecting privacy, as long as the basic first principles of data minimization, retention limits, use restrictions, and transparency are met as well.

**83. To what extent should the Commission consider rules that require companies to make information available about their commercial surveillance practices? What kinds of information should new trade regulation rules require companies to make available and in what form?**

The Foundation supports taking into account the different kinds of platform governance models when and if it serves to develop rules requiring companies to make information available about their commercial surveillance practices. Although the Foundation is a nonprofit, our community-led content moderation model is used by for-profit actors as well.

Commercial surveillance is the business of collecting, analyzing, and profiting from information about people. Mass surveillance has heightened the risks and stakes of data breaches, deception, manipulation, and other abuses.

To use Wikipedia as an illustrative example to demonstrate the risks of rules that are too prescriptive or heavy-handed, the community-led content moderation model of Wikipedia involves the possible sharing of data between community members so that members can engage in platform management. The Wikipedia platform in this instance should not be required to disclose the names of online community members in transparency reports or access requests because that could endanger the privacy of those individuals. As has been discussed above, privacy is an essential element of creating a safe online environment. The FTC should thus ensure that disclosure requirements do not inadvertently create additional privacy risks (e.g., by requiring disclosure of the names of certain community members or platform users charged with

enforcing some of the platform's rules), or incentivize companies to collect information that they would otherwise not have collected.

However, a way in which rules can be written to reduce the risks that have come into play from commercial surveillance is by enforcing transparency. To use the Wikipedia model as an example once more: our policies and practices are publicly available on the Wikimedia projects, and so are the edit histories of the formulation and iteration of each policy. The Foundation also publishes biannual transparency reports[50] that detail each individual request we receive from governments, individuals, and organizations to disclose information about our users or to delete or alter content on all of our projects.

**84. In which contexts are transparency or disclosure requirements effective? In which contexts are they less effective?**

Disclosure requirements and efforts to increase transparency are effective when the disclosures are clear and concise and they are combined with effective data minimization requirements, use restrictions, and accountability mechanisms that do not require people to rely entirely on a company's disclosures to protect their privacy rights. More detailed disclosures could accompany the clear and concise summary to aid more sophisticated observers in more deeply understanding a company's data practices and holding them accountable when their practices seem to diverge from their disclosures. Ultimately, however, effective transparency should help a person understand quickly what will happen to their data and how their rights with respect to the data can be exercised.

As above, such efforts are less effective when they force companies to collect information that they previously did not, including, for example, users' age, name or other information that is not necessary to provide the requested service, or to retain information for longer periods than they would have previously. Not only is collecting unnecessary information an onerous obligation, especially for small to mid-sized tech companies who may have limited resources, it also means there is more data to be put at risk of abuse and leakage. Lastly, efforts to mandate transparency and disclosure requirements are especially less effective when they force companies to hand over personal information that endanger the privacy of individuals.

**85. Which, if any, mechanisms should the Commission use to require or incentivize companies to be forthcoming? Which, if any, mechanisms should the**

[50] Transparency Report, Wikimedia Foundation (Jul.–Dec. 2021), https://wikimediafoundation.org/about/transparency/2021-2/.

**Commission use to verify the sufficiency, accuracy, or authenticity of the information that companies provide?**

The Foundation supports transparency requirements as a mechanism to incentivize companies to be forthcoming. Nevertheless, the Foundation also urges caution that any such requirements take into account the diversity of the internet ecosystem. The unbalanced way in which mechanisms could fall upon small to medium-sized technology organizations as compared to big technology organizations, or the unbalanced way they could fall onto community volunteers of community-led operation models should be taken into account. Firstly, there is a potential that such mechanisms would have a large impact on much of the internet ecosystem, and may be built in such a way that large technology companies that can afford to upgrade and/or reconfigure their systems would be able to keep up whereas others that cannot afford to do so would be unable to do so. This would drive down the diversity of the internet ecosystem and encourage the consolidation of power amongst a select few big technology companies. Secondly, like our concern above, the collection of additional data may lead to the potential abuse of that data by bad faith actors who are aware of the mandates to collect certain information. Information is safer if not collected. Thirdly, any mechanism that requires the collection—or the longer retention—of more data than is currently collected has the potential to endanger the privacy of individuals.

In relation to the specific types of mechanisms that may present risks, the limiting or tying together of any mechanism to a specific technology presents several risks: (1) the technology may become outdated; (2) certain uses or designs of the technology, such as artificial intelligence that is trained on biased and/or inaccurate datasets or used to achieve illegal ends, may be unscrupulous and not transparent; and, (3) the technology may be expensive to implement.

**89. To what extent should trade regulation rules, if at all, require companies to explain (1) the data they use, (2) how they collect, retain, disclose, or transfer that data, (3) how they choose to implement any given automated decision-making system or process to analyze or process the data, including the consideration of alternative methods, (4) how they process or use that data to reach a decision, (5) whether they rely on a third-party vendor to make such decisions, (6) the impacts of their commercial surveillance practices, including disparities or other distributional outcomes among consumers, and (7) risk mitigation measures to address potential consumer harms?**

The Foundation supports trade regulation rules disclosing all of the above. Specifically, the Foundation supports the formalization of transparency requirements and having companies provide regular and public updates to their data handling practices. Transparency allows for accountability and should be built into how companies operate. For example, the data used by a company and how such data is collected, retained, disclosed or transferred could be explained in a privacy policy. An example of such is the Wikimedia Foundation's Privacy Policy,[51] which explains to users the types of information collected, how it is collected, how we use the information received from users, under what circumstances the information may be shared, how we protect personal information, and how long we keep data. The Foundation also supports impact assessments and tracking aggregate use of any kind of commercial surveillance practices. Although the Wikimedia Foundation does not participate in commercial surveillance practices, we explain our practices and decisions in our policies and in our bi-annual Transparency Report.[52]

[51] Privacy Policy, Wikimedia Foundation, https://foundation.wikimedia.org/wiki/Privacy_policy.
[52] Transparency Report, Wikimedia Foundation (Jul.–Dec. 2021), https://wikimediafoundation.org/about/transparency/2021-2/.

## Annexure II

**Relevant Wikimedia Foundation and Wikimedia Community Policies and Reports**

- Terms of Use
- Human Rights Policy
- Privacy Policy
- Donor Privacy
- Biographies of Living Persons Policy
- Image Use Policy
- Transparency Report