ACM US Public Policy Council

COMMENTS ON FEDERAL TRADE COMMISSION PRELIMINARY STAFF REPORT
Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers

RESPONSE FILED BY:
U.S. PUBLIC POLICY COUNCIL OF THE ASSOCIATION FOR COMPUTING MACHINERY

On behalf of the U.S. Public Policy Council (USACM) of the Association for Computing Machinery (ACM) we are submitting the following comments on online consumer privacy in response to the report from the Federal Trade Commission.

With 100,000 members, the Association for Computing Machinery (ACM) is the world's oldest and largest educational and scientific computing society. The ACM U.S. Public Policy Council (USACM) serves as the focal point for ACM's interaction with U.S. government organizations, the computing community, and the U.S. public in all matters of U.S. public policy related to information technology. Our comments are informed by our previously issued statement on Privacy[1] and our response to the Office of Science and Technology Policy on the use of web tracking technologies[2]. Should you have any questions or need additional information, please contact Cameron Wilson, our Director of Public Policy, at 202-659-9711 or at Cameron.Wilson@acm.org.

Introduction

We appreciate and welcome the continued attention that the Federal Trade Commission (FTC) has given to consumer privacy, especially online consumer privacy. Accommodating both individual privacy rights and reasonable commercial needs is a complex task for technologists and policy makers, but one of vital importance. Effective data privacy policies and practices provide benefits to consumers and commercial interests alike. The efforts of the Commission, along with those of the Department of Commerce's Internet Policy Task Force, should enable more nimble data collection and privacy protection in an era of quickly changing technologies.

Before addressing the specific questions the FTC has in Appendix A of the report, we have some general comments and recommendations on how the FTC can effectively craft a consumer privacy framework focused on simplified choice, greater transparency and privacy by design. We strongly encourage the broader adoption and implementation of Fair Information Practice Principles. In addition, two kinds of tools – a dataflow-based lexicon and well-formed privacy risk models - a can help improve the ability of consumers and industry to communicate and understand data practices as well as the purposes for data collection and use. We also support the development of Do Not Track, provided it is: technology neutral; allows for a variety of consumer choices between track all and track nothing; and has an effective enforcement strategy recognizing the global nature of online activity.

A Dataflow-Based Lexicon to Support More Meaningful Choice and Transparency

Because so many Fair Information Practice Principles (FIPPs)—including consent, use limitation, and collection minimization—are affected by the purpose of the system or business process in question, purpose specification deserves special attention. Practical problems observed with notice/awareness and choice/consent require a more rigorous yet concise means of specifying purpose. What is required goes beyond the notion of "commonly accepted practices" and must include a dataflow-based lexicon—a catalog of standard terms that describes how

[1] USACM Privacy Policy Recommendations, http://usacm.acm.org/usacm/Issues/Privacy.htm

[2] http://usacm.acm.org/usacm/PDF/USACM_Web_Tracking_Comments_Final.pdf

personal information flows between different entities in the context of a particular purpose, such as the online purchase of a material object. This lexicon would contain definitions for terms that already appear in the FTC's Personal Data Ecology, as envisioned in Appendix C of the proposed framework. Additionally, the lexicon would specify the what and when of these data flows, and link those flows to specific purposes without disclosing proprietary business practices.

A dataflow-based lexicon should, at a minimum, distinguish terms for common categories of actor (data subject, data broker, etc.) in a personal data ecology, define types of privacy controls, and specify standard data purposes with concise text descriptions that are uniquely indexed, such that each purpose is distinguishable from others. Making purposes distinct involves describing what information will be needed, and for how long it will be needed. By describing data flows and defining both actors and privacy controls, businesses could distinguish themselves and the privacy protections afforded to consumers as extensions to a baseline, industry-wide lexicon. Furthermore, as a catalog of standard purposes, the lexicon would buttress the purpose specification that so many FIPPs depend on and help data subjects understand how their information will be used in a relatively straightforward manner that is consistent across industries. Other, possibly unexpected, uses would appear as new purposes or as extensions to a standard purpose. This representation would highlight any necessary alterations to baseline definitions of terms, such as sensitive and non-sensitive consumer data, which can vary with context. A baseline lexicon that businesses can extend will enable terminology to be more easily adapted to changing technological possibilities — interpretations of terms are then less likely to become outdated or limited in scope. This includes "commonly accepted practices," which could be duly designated or undesignated by whichever appropriate body (the FTC is one possibility) oversees development and maintenance of the lexicon.

The lexicon should explicitly and unambiguously distinguish among the diverse set of industrial data purposes - plus government purposes, when those purposes involve the disclosure of consumers' personal information - with the aim of moving toward a national standard for the terms it covers. As a catalog of data practices, this proposed lexicon is similar in nature to the North American Industrial Classification System (NAICS) published by the U.S. Census Bureau. However, it would be smaller, more focused, and more domain-specific than NAICS. In those respects, the lexicon would resemble something like the International Classification of Diseases (ICD) published by the Centers for Disease Control. By indexing data purposes uniquely in a lexicon, purposes can be more easily identified in privacy policies in a reusable, machine-readable fashion and linked to enterprise business practices and information assets to improve accountability. This would make it easier to judge compliance with FIPPs and to monitor data collection practices.

For consumers to make informed choices, terms in the lexicon should differentiate between meaningful variations in the same class of privacy control or data practice. For example, businesses may present consumers with "opt-out" controls that are limited to either third party secondary uses of personal information, or to first party secondary uses, or both. Similarly, the meaning of "online tracking" can include many different things, including: tracking to deliver targeted, third party advertisements; tracking for analytics to improve website navigation; or tracking for website functionality, such as managing shopping carts or other user sessions. Therefore, the lexicon should allow consumers and businesses to uniquely distinguish terms for data practices, privacy controls, and their varieties, in privacy policies and other statements of record. By uniquely indexing terms in the lexicon, web browsers and other relevant applications, as well as privacy-enhancing technologies potentially, can automatically reference and effectively communicate salient information. Moreover, consumers would be better able to make informed choices because a standard lexicon would exist across the multitude of websites with which they interact. However, unlike the Platform for Privacy Preferences Project (P3P), the lexicon would not restrict the types of privacy policies that companies may create, nor does it serve as a platform to match or compare entire privacy policies. Rather, it is a standardized set of purposes and descriptions of privacy controls that concerned individuals, organizations, and regulators can reference and that can be linked to expected practices in the form of data flow descriptions.

Data flow descriptions will add additional context by describing the kind of personal information required by the business for the particular purpose and the points and duration of collection and use. Businesses using the same purpose and privacy control descriptions may still vary in terms of what they collect, when they collect it, and how long they keep it.

As the lead U.S. regulatory authority in information privacy, the FTC would be a logical possibility to create and maintain this standard lexicon to improve transparency and choice. The "personal data ecosystem" in the framework is a step in the right direction, because it defines important stakeholder categories by example. However, a dataflow-based lexicon would go further by accommodating multiple meanings and changing terminology for both data practices (e.g., tracking, linkability) and privacy controls (e.g., opt-out, anonymization).

Privacy Risk Models to Support Privacy by Design

Building privacy into systems and processes requires thorough privacy risk assessment, in the same way that building in security requires thorough security risk assessment. Identified risks can then be mapped to mitigating controls, including privacy-enhancing technologies, which can be integrated into systems and processes as they are developed. Ultimately, privacy risk assessment must be based on a rich privacy risk model that includes norms and harms, in addition to FIPPs. Otherwise, practices that are at odds with reasonable expectations that could result in a variety of harms can nonetheless be perceived as without risk, owing to a nominal compliance with FIPPs. While we appreciate the FTC's recognition of some of these problems, as reflected by its attention to harms as well as FIPPs, a well-formed privacy risk model is needed. Risk models exist in many domains– such as insurance, epidemiology, and financial services – however, the most relevant examples are those from information security. This includes the risk framework developed by the National Institute of Standards and Technology (NIST), which provides a mechanism for determining the necessary levels of system security properties and maps those to defined risk controls. Regardless of the domain, any risk model provides a systematic framework that relates objectives or resources of concern to relevant threats and controls.

Such a model (or sector-specific models) should incorporate, as appropriate, relevant research touching on privacy risk. This research speaks to issues such as the range of potential privacy harms[3] and the analysis of novel socio-technical systems[4]. Moreover, such a model or models must also be capable of accommodating new or changed contexts and harms, as well as technical developments that invalidate previous assumptions. This should be done in part by explicitly incorporating provisions for technical contingencies, such as data thought to be unlinkable becoming potentially linkable, an issue that has become increasingly prominent.

Prior to 2002, many assumed that two data sets were linkable only if they shared a common index or key in the data sets, such as the Social Security number. With the discovery of k-anonymity[5], we learned that linkability is a context-sensitive construct that depends on which attributes are shared by a population, such as age, gender and zip code. These attributes were shown to uniquely identify most individuals in the U.S. and are collected by many online service providers; thus, data that cannot be linked by a common index can be linked through shared attributes. Five years later, the concept of "linkability" was extended when we learned that inadequate diversity of attribute values in a data set allows us to link specific values to individuals without knowing which record in the data set corresponds to which individual[6]. Due to the fluid and contingent nature of linkability, it is best treated as a context-specific risk rather than as a general property. Like the dataflow-based lexicon, privacy risk models—including provisions addressing linkability—must be revised on an ongoing basis and must also be extensible in order to accommodate atypical situations.

---

[3] See, for example, Daniel J. Solove, *Understanding Privacy,* Cambridge, MA: Harvard University Press, 2008.

[4] See, for example, Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life,* Palo Alto: Stanford Law Books, 2009.

[5] Sweeney, Latanya. "k-anonymity: a model for protecting privacy." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5): 557-570, 2002.

[6] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, Muthuramakrishnan Venkitasubramaniam. "L-diversity: privacy beyond k-anonymity." *ACM Transactions on Knowledge Discovery from Data*, 1(1): 3, 2007.

The dataflow-based lexicon and privacy risk models are intended to provide tools that are flexible and contextual. Indeed, they could even become synergistic with the models, leveraging the lexicon and the lexicon incorporating new definitions in response to revisions and extensions to the models. This can help standardize data management practices (with local flexibility) and can make it easier to adapt to new challenges in data management – whether it is the latest technologies or the next incremental change in processing capability that makes it that much easier to link data.

Specific Questions

We now address some of the questions noted in the staff report about the proposed privacy framework. The questions are in **bold**, with relevant headings from Appendix A in ***bold italics***.

Scope

• Are there practical considerations that support excluding certain types of companies or businesses from the framework – for example, businesses that collect, maintain, or use a limited amount of non-sensitive consumer data?

The dataflow-based lexicon, together with richer privacy risk models, could address what actions would be necessary to ensure effective privacy for consumer data across a broad range of contexts. Precisely because environments, including technical capabilities, are constantly evolving, reasonable grounds for exclusion may change radically in short order. What may be viewed as limited or non-sensitive today may not be so tomorrow. Therefore, we hesitate to endorse any such exclusion and instead urge a sufficiently comprehensive and usable framework.

• Is it feasible for the framework to apply to data that can be "reasonably linked to a specific consumer, computer, or other device"?

To the extent that data "reasonably linked to a…computer, or other device" can also be linked to a consumer, it would be appropriate for the framework to apply to that data. However, as discussed below, linkability is not a straightforward concept and it is unclear that a universal threshold for "reasonably linked" could be established.

• How should the framework apply to data that, while not currently considered "linkable," may become so in the future?

Well-formed privacy risk models can help anticipate future linkability moving forward, especially in conjunction with the dataflow-based lexicon.

• Are there reliable methods for determining whether a particular data set is "linkable" or may become "linkable"?

Linkability is not a binary state; it is a spectrum and the position of any given data set on that spectrum is a function of domain, knowledge resources, motivation, and technical capability, most of which are subject to change. Therefore, we recommend that linkability be treated as a risk-based assessment rather than as a precise determination. A sufficiently rich risk model can provide a systematic and common basis for this assessment.

• What technical measures exist to "anonymize" data and are any industry norms emerging in this area?

The "anonymization" of data (we prefer the more precise term "de-identification") is a highly contextual process, and must address the same issues of contingent risk as linkability. What may be accepted industry practice for de-identification can be rendered obsolete by environmental changes, including increased availability of other knowledge sources and improvements in technical capability. De-identification should be viewed as a discrete

privacy risk control that will not necessarily obviate the need for additional controls, and may not even be particularly sensible for a given use case.

Companies should promote consumer privacy throughout their organizations and at every stage of the development of their products and services

Incorporate substantive privacy practices

• Should the concept of "specific business purpose" or "need" be defined further and, if so, how?

Specific business purposes will change over time and new purposes will emerge as new needs are defined and technology enables new capabilities. As previously noted, the specification of business purposes is fundamental to implementing FIPPs, including prevention of unauthorized secondary uses. We propose that the dataflow-based lexicon serve as a common reference point for purposes, which can be extended as new purposes are envisioned and put into practice. This will allow the FTC, consumers, and businesses to define and understand new purposes in the context of existing practices.

Companies should simplify consumer choice

Commonly accepted practices

• Is the list of proposed "commonly accepted practices" set forth in Section V(C)(1) of the report too broad or too narrow?

The list of commonly accepted practices is broad, as demonstrated by the varied examples in each practice description provided by the FTC. However, this breadth begs the question of the appropriate level of definitional granularity for particular practices. The dataflow-based lexicon would enable more precise meanings and therefore promote consistency of interpretation and expectation.

• Are there practices that should be considered "commonly accepted" in some business contexts but not in others?

The FTC provides compelling examples of practices that appear to be commonly accepted in some contexts and not others (e.g., obtaining a consumer's address for delivering a product). The lexicon could be used as a means to document and designate which practices are viewed as commonly accepted for a type of service or business function, enabling businesses to avoid burdening consumers with excessive and indiscernible choice. However, definitions must be sufficiently specific to prevent substantive divergences within a given practice.

Practices that require meaningful choice

General

• Should the method of consent be different for different contexts?

In principle, users should have to opt-in to any collection and use of their personal information. As a purely practical matter, though, risk-based reasoning can serve to usefully guide consent requirements. Contexts characterized by greater privacy risk should tend toward more explicit and affirmative consent mechanisms while more implicit and/or disaffirming mechanisms may be a defensible approach for less risky contexts, assuming

that a suitable risk model has been used. As with other aspects of the framework, we urge a move toward flexible and robust risk analysis.

• What additional consumer protection measures, such as enhanced consent or heightened restrictions, are appropriate for the use of deep packet inspection?

Deep Packet Inspection (DPI) involves analyzing the data contained in each packet, rather than just the packet header. DPI has been used to assist in network traffic management, but a complete and detailed behavioral profile can be constructed to generate targeted advertising or for even more invasive purposes. As such, it presents serious privacy risks; these risks are inevitably magnified when transparency and choice are absent. In contrast to cookie-based approaches, customers have no way to easily "opt out" of DPI-based behavioral profiling. .While the information provided by DPI might enhance the user's online experience, this technology is essentially silent or invisible. This context would require that DPI-based behavioral profiling have very explicit and affirmative consent requirements. However, DPI for network management purposes would not require additional consumer protection measures, assuming it is implemented in an appropriate manner. A useful way to distinguish between DPI for behavioral profiling and DPI for network management is to emphasize dataflow-based purpose specification, as in the proposed lexicon.

Special choice for online behavioral advertising: Do Not Track

• How should a universal choice mechanism be designed for consumers to control online behavioral advertising?

The appeal of Do Not Track reflects an emerging understanding that simple opt-in and opt-out choices for consent are inadequate for the very different kinds of online data collection and use that consumers may or may not be willing to accept, depending on the circumstances. As indicated above, tracking takes place for a variety of different purposes. However, it also takes place in a variety of different online contexts. Tracking in a social networking context is different from tracking in an e-commerce context and both are different from tracking consumption of news and opinion. The privacy risks differ, which means the selection and implementation of privacy risk controls will differ as well, whether Do Not Track or something else is used. Furthermore, the technologies that support tracking will change over time. For all of these reasons, it will be important to establish a definition of Do Not Track that is technology neutral and allows for a variety of consumer choices between tracking everything and tracking nothing.

The complexity of the techniques used for and the privacy risks presented by online behavioral tracking makes a single, simple technical control like a white list, persistent cookie, or special server header insufficient. There needs to be enforceable means on both the client side and the server side of online interactions in order for a Do Not Track option to be effective. Enforcement of Do Not Track will be at least as important, if not more so, as any technical options made available to consumers or companies. Browser vendors are already working on settings and/or plug-ins to assist consumers in avoiding unwanted data collection (though current offerings do not represent a comprehensive means for consumers to protect against all possible data collection). However, it will be relatively easy for a website to hand-off data collection to a third-party outside of the U.S. and circumvent the enforcement capacity of the U.S. government. How the FTC will deal with this and other international aspects of enforcing a Do Not Track policy needs to be addressed before the policy is implemented.

Companies should increase the transparency of their data practices

Improved privacy notices

• What is the feasibility of standardizing the format and terminology for describing data practices across industries, particularly given ongoing changes in technology?

We believe it is feasible to standardize terminology based on nominal flows of personal information; however, a centralized authority such as the FTC must be responsible for developing and maintaining the resulting lexicon, either by establishing a committee and/or by soliciting public comment on its own proposals. The FTC should also take steps to encourage use of and adherence to the lexicon by businesses and consumers. Potential approaches include incentive-based strategies such as some kind of safe harbor from relevant enforcement actions.

• Should companies increase their use of machine-readable policies to allow consumers to more easily compare privacy practices across companies?

Mandating that companies increase their use of machine-readable policies without proven technologies that consumers will use can be unnecessarily burdensome to industry. Machine-readable policies can help consumers to more easily compare privacy practices across companies; however, we urge a broad interpretation of "policies." By maintaining and encouraging use of common terminology representing defined practices, and tying these terms to machine-readable identifiers, new approaches to reading and comparing privacy practices can be explored and their value to consumers can be demonstrated.. The proposed lexicon moves in this direction.

Privacy

USACM Policy Recommendations

Background

Current computing technologies enable the collection, exchange, analysis, and use of personal information on a scale unprecedented in the history of civilization. These technologies, which are widely used by many types of organizations, allow for massive storage, aggregation, analysis, and dissemination of data. Advanced capabilities for surveillance and data matching/mining are being applied to everything from product marketing to national security.

Despite the intended benefits of using these technologies, there are also significant concerns about their potential for negative impact on personal privacy. Well-publicized instances of personal data exposures and misuse have demonstrated some of the challenges in the adequate protection of privacy. Personal data -- including copies of video, audio, and other surveillance -- needs to be collected, stored, and managed appropriately throughout every stage of its use by all involved parties. Protecting privacy, however, requires more than simply ensuring effective information security.

The U.S. Public Policy Council of the Association for Computing Machinery (USACM) advocates a proactive approach to privacy policy by both government and private sector organizations. We urge public and private policy makers to embrace the following recommendations when developing systems that make use of personal information. These recommendations should also be central to any development of any legislation, regulations, international agreements, and internal policies that govern how personal information is stored and managed. Striking a balance between individual privacy rights and valid government and commercial needs is a complex task for technologists and policy makers, but one of vital importance. For this reason, USACM has developed the following recommendations on this important issue.

Recommendations

MINIMIZATION

1. Collect and use only the personal information that is strictly required for the purposes stated in the privacy policy.

2. Store information for only as long as it is needed for the stated purposes.

3. If the information is collected for statistical purposes, delete the personal information after the statistics have been calculated and verified.

4. Implement systematic mechanisms to evaluate, reduce, and destroy unneeded and stale personal information on a regular basis, rather than retaining it indefinitely.

5. Before deployment of new activities and technologies that might impact personal privacy, carefully evaluate them for their necessity, effectiveness, and proportionality: the least privacy-invasive alternatives should always be sought.

CONSENT

6. Unless legally exempt, require each individual's explicit, informed consent to collect or share his or her personal information (opt-in); or clearly provide a readily-accessible mechanism for individuals to cause prompt cessation of the sharing of their personal information, including when appropriate, the deletion of that information (opt-out). (NB: The advantages and disadvantages of these two approaches will depend on the particular application and relevant regulations.)

7. Whether opt-in or opt-out, require informed consent by the individual before using personal information for any purposes not stated in the privacy policy that was in force at the time of collection of that information.

OPENNESS

8. Whenever any personal information is collected, explicitly state the precise purpose for the collection and all the ways that the information might be used, including any plans to share it with other parties.

9. Be explicit about the default usage of information: whether it will only be used by explicit request (opt-in), or if it will be used until a request is made to discontinue that use (opt-out).

10. Explicitly state how long this information will be stored and used, consistent with the "Minimization" principle.

11. Make these privacy policy statements clear, concise, and conspicuous to those responsible for deciding whether and how to provide the data.

12. Avoid arbitrary, frequent, or undisclosed modification of these policy statements.

13. Communicate these policies to individuals whose data is being collected, unless legally exempted from doing so.

ACCESS

14. Establish and support an individual's right to inspect and make corrections to her or his stored personal information, unless legally exempted from doing so.

15. Provide mechanisms to allow individuals to determine with which parties their information has been shared, and for what purposes, unless legally exempted from doing so.

16. Provide clear, accessible details about how to contact someone appropriate to obtain additional information or to resolve problems relating to stored personal information.

ACCURACY

17. Ensure that personal information is sufficiently accurate and up-to-date for the intended purposes.

18. Ensure that all corrections are propagated in a timely manner to all parties that have received or supplied the inaccurate data.

SECURITY

19. Use appropriate physical, administrative, and technical measures to maintain all personal information securely and protect it against unauthorized and inappropriate access or modification.

20. Apply security measures to all potential storage and transmission of the data, including all electronic (portable storage, laptops, backup media), and physical (printouts, microfiche) copies.

ACCOUNTABILITY

21. Promote accountability for how personal information is collected, maintained, and shared.

22. Enforce adherence to privacy policies through such methods as audit logs, internal reviews, independent audits, and sanctions for policy violations.

23. Maintain provenance -- information regarding the sources and history of personal data -- for at least as long as

the data itself is stored.

24. Ensure that the parties most able to mitigate potential privacy risks and privacy violation incidents are trained, authorized, equipped, and motivated to do so.

USACM does not accept the view that individual privacy must typically be sacrificed to achieve effective implementation of systems, nor do we accept that cost reduction is always a sufficient reason to reduce privacy protections. Computing options are available today for meeting many private sector and government needs while fully embracing the recommendations described above. These include the use of de-identified data, aggregated data, limited data sets, and narrowly defined and fully audited queries and searches. New technologies are being investigated and developed that can further protect privacy. USACM can assist policy-makers in identifying experts and applicable technologies.

Association for Computing Machinery (ACM)

With 100,000 members worldwide, the Association for Computing Machinery is the world's largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the computing profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

About the ACM U.S. Public Policy Council

The ACM U.S. Public Policy Council (USACM) serves as the focal point for ACM's involvement with U.S. government organizations, the computing community and the U.S. public in all matters of U.S. public policy related to information technology. Supported by ACM's Washington, D.C., Office of Public Policy, USACM responds to requests for information and technical expertise from U.S. government agencies and departments, seeks to influence relevant U.S. government policies on behalf of the computing community and the public, and provides information to ACM on relevant U.S. government activities. USACM also identifies potentially significant technical and public policy issues and brings them to the attention of ACM and the community. USACM publishes a monthly newsletter, the ACM Washington Update, which reports on activities in Washington that may be of interest to those in the computing and information policy communities, and highlights USACM's involvement in many of these issues. USACM is actively engaged in number of public policy issues of critical importance to the computing community.

For more information about USACM, please contact the ACM Office of Public Policy at (202) 659-9711 or see http://www.acm.org/usacm/.