Situating Anonymization Within a Privacy Risk Model

Stuart S. Shapiro
Homeland Security Systems Engineering and Development Institute,™
operated by The MITRE Corporation
Bedford, MA USA
sshapiro@mitre.org

***Abstract*—Privacy risk analysis of complex socio-technical systems suffers from an inadequate risk model that focuses primarily on some form of Fair Information Practice Principles (FIPPs). Anonymization as a privacy risk control suffers from an emphasis on risk of failure, neglecting the circumstances surrounding its selection as a risk control in the first place. By interrelating an enhanced privacy risk model that goes beyond FIPPs and an integrated anonymization framework, the selection and implementation of anonymization as a privacy risk control can be more systematically considered and carried out. The Science and Technology Directorate of the U.S. Department of Homeland Security has sponsored development of both an integrated anonymization framework and an enhanced privacy risk model to support more effective privacy risk management. Both of these are described at a high level and their interoperability illustrated by application to the Google Street View controversy.**

Keywords-anonymization; informational privacy; privacy in socio-technical systems; privacy risk

I. INTRODUCTION

Any risk analysis process must, by definition, be premised on some kind of risk model. Without a risk model to characterize and scope threats, vulnerabilities those threats could exploit, and the likelihood and impact of such exploitation, risk analysis would be wholly ad hoc and idiosyncratic. Privacy Impact Assessment (PIA)—frequently, but not exclusively, focused on informational privacy—generally relies upon a set of privacy principles, sometimes referred to as Fair Information Practice Principles (FIPPs), as the core of its underlying risk model. (This is reflected in the results of a recent survey of international PIA practices [1].) Many versions of such principles, which vary in their scope and specifics but evince a number of common concepts (e.g., minimizing collection of personally identifiable information (PII) to that which is necessary for a stated purpose), have been promulgated by various entities. These include the Organization for Economic Cooperation and Development (OECD), the Canadian Standards Association, and the U.S. Department of Homeland Security (DHS).

Evolving socio-technical systems and the issues they raise have rendered this model inadequate. We are proposing an enhanced privacy risk model [2] that leverages leading-edge privacy scholarship to provide a more sophisticated approach to surfacing and addressing privacy risks in complex socio-technical systems. This new model does not abandon FIPPs, which remain important from an individual rights standpoint, but rather augments them so as to capture normative expressions of privacy, privacy harms beyond violations of privacy principles, and the interaction of systems with their surrounding environments.

In this paper, we aim to explicitly interrelate "anonymization" of PII and our enhanced privacy risk model. We do so by establishing a basis for selecting anonymization as a control for a variety of distinct privacy risks embedded in the model and by defining an integrated anonymization framework to guide implementation following selection. The Cyber Security Division of the DHS Science and Technology Directorate (S&T) has sponsored the development of both the enhanced privacy risk model and the integrated anonymization framework to support more effective privacy risk management.

We take this step in part because of how much has been vested in anonymization as an informational privacy risk control. In a way, anonymization is the quintessential informational privacy risk control, seeing as how it seeks to render PII—the fundamental focus of informational privacy—into something else, something inherently divorced from privacy concerns. Thus, it has been particularly disturbing to see that risk control steadily called into doubt [3, 4, 5].

As we have argued elsewhere [6], we believe a substantial part of this problem results not from the technical frailty of anonymization but from programmatic failings leading to the application of anonymization unsupported by a sufficiently rigorous reasoning process. Our focus here is the selection of anonymization as a risk control in the first place, situating that selection within the context of the enhanced privacy risk model.

The remainder of the paper is organized as follows. In Section II we provide an overview of the enhanced privacy risk

MITRE operates the Homeland Security Systems Engineering and Development Institute for the United States Department of Homeland Security (DHS) under Contract No. HSHQDC-09-D-00001. Preparation of this report was sponsored by DHS under that contract.

model. Section III does the same for the integrated anonymization framework and describes how it interacts with the risk model. As a thought experiment, Section IV applies the resulting construct to Google Street View. Finally, Section V considers potential future work.

II. Enhanced Privacy Risk Model

As noted above, FIPPs, while essential expressions of privacy due process, are increasingly inadequate as the central constituents of a privacy risk model. First, they are relative with respect to purpose, permitting PII collection and use for essentially any reason, no matter how fundamentally inimical to privacy. Second, they encourage framing of privacy harms purely in terms of principle violations, as opposed to the actual impact on individuals. (Even fairly sophisticated privacy risk analysis frameworks, such as those presented in [7] and [8], end up defining risks in terms of violations, at varying degrees of granularity, of privacy principles.) Finally, they tend to focus almost exclusively on the characteristics of the system at the expense of the characteristics of the surrounding environment and the interactions between the two.

We have developed a privacy risk model that attempts to address these shortcomings by synthesizing the work of Nissenbaum [9] and Solove [10]. Both approaches ground their analyses in privacy problems per se rather than the application of (or failure to apply) general privacy principles. Nissenbaum's contextual integrity heuristic addresses both the absence of norms and the inward focus of FIPPs, while Solove's taxonomy of privacy problems addresses the need to articulate privacy risks in terms of potential harms to individuals. The resulting privacy risk model is situated in a general but slightly tailored risk management framework.

The risk management framework was synthesized by examining risk modeling and management in a variety of domains. The synthetic framework was then slightly adjusted to reflect the fact that privacy risks arise neither exclusively from the system nor exclusively from the surrounding environment. The resulting framework contains the following stages.

A. Characterization

For a socio-technical system, characterization must address both the technology and the surrounding environment. Therefore, the model explicitly addresses both. However, the general context is first established to bound the scope of the analysis and consists of goals or purposes associated with the context, canonical roles of entities directly involved in the pursuit of those goals or purposes, and canonical activities of those entities.

For informational privacy, characterization amounts to charting the information flows and state transitions in terms of actors (data subjects, senders, receivers, and users) or states, attributes (PII), and the conditions that govern those flows and transitions (including relevant FIPPs). To make this more manageable, and also because it enables easier alignment with many PIA processes, this is parsed by information life cycle stage.

B. Vulnerability Identification

Privacy vulnerabilities arise out of disruptions to PII flows and state transitions due to the interaction of technology and environment. Whether these vulnerabilities lead to actual risks will be determined in the next step. First, though, these disruptions must be identified. Having characterized the environment and the technology, we look for conflicts between the two.

C. Risk Identification

Having identified relevant disruptions to PII flows and state transitions (i.e., vulnerabilities), we then identify salient risks by determining what privacy harms could potentially arise out of those disruptions. In doing this, we leverage Solove's taxonomy, which describes sixteen distinct types of privacy harm, each arising within one of four main contexts: information collection, information processing, information dissemination, and invasions.

D. Risk Assessment

Once the relevant risks have been identified, it remains to explicitly tie these to the implicated characteristics. This requires tracing back from the risks to the characteristics (actors or states, attributes, conditions) that engendered them. This allows the identified risks to be assessed in terms of the relevant information flows and state transitions, providing a basis for estimating risk severity and for performing risk management, i.e., selecting control actions for the identified risks.

E. Risk Response Determination

Determining appropriate responses to identified privacy risks is a function of multiple factors, including risk tolerance, available resources, and cost-benefit calculations. No privacy risk management framework can directly prescribe appropriate responses. A framework can, however, offer useful guidance by associating certain types of controls with the risks they typically might address.

Once controls are selected, their effect can be systematically evaluated by reworking the analysis. Adjustments to the characterizations will propagate through vulnerability identification, risk identification, and risk assessment. Through this process, residual risk can be evaluated.

F. Risk Control Implementation

Once the effects of the selected controls have been evaluated and judged acceptable, implementation specifics must be determined. This can take a variety of forms, including tool selection and configuration, process definition and application, and insertion of specific requirements into the system development life cycle.

G. Monitor and Review

Once selected risk controls have been implemented, it remains to monitor and review the situation on an ongoing basis. In particular, if the characteristics of the system or the

environment significantly change, those changes must find their way into an updated analysis.

III. Integrated Anonymization Framework

Despite recent high-profile problems, anonymization continues to be a popular privacy risk control. A variety of supporting commercial tools exist and the HIPAA Privacy Rule continues to encourage it. Most anonymization tools and techniques, though, are aimed at static anonymization of explicitly or implicitly structured textual data. (Explicitly structured data includes data held in a relational database, while implicitly structured data includes natural language, which is often referred to, somewhat inaccurately, as unstructured data.) However, there are other forms of data that could potentially benefit from anonymization. These include graph data, transcripts of oral communication, recorded audio, and photographic and video images.

Although there are techniques available for manipulating each of these types of data, they do not necessarily directly lead to anonymized data. Anonymizing some forms of data, therefore, requires successive application of multiple methods. Whatever the initial form of the data, it must be manipulated sufficiently to get it to the point at which a sound anonymization technique can be applied. An integrated anonymization framework, therefore, can be structured as a data flow process in which specific manipulations are targeted at specific types of data with the goal of either moving the data closer to a form that is amenable to anonymization or actually anonymizing it. Fig. 1 depicts such a (noncomprehensive) process.

The techniques that effect "anonymization" actually encompass two distinct goals. The primary goal is actual de-identification of information. The secondary goal is to render sensitive attributes less sensitive. In other words, anonymization is more accurately viewed as reducing the ability to associate information with specific individuals. To the extent the implicated characteristics of risks involve identity information and sensitive attributes, anonymization can serve to reduce privacy risk, assuming it is practical.

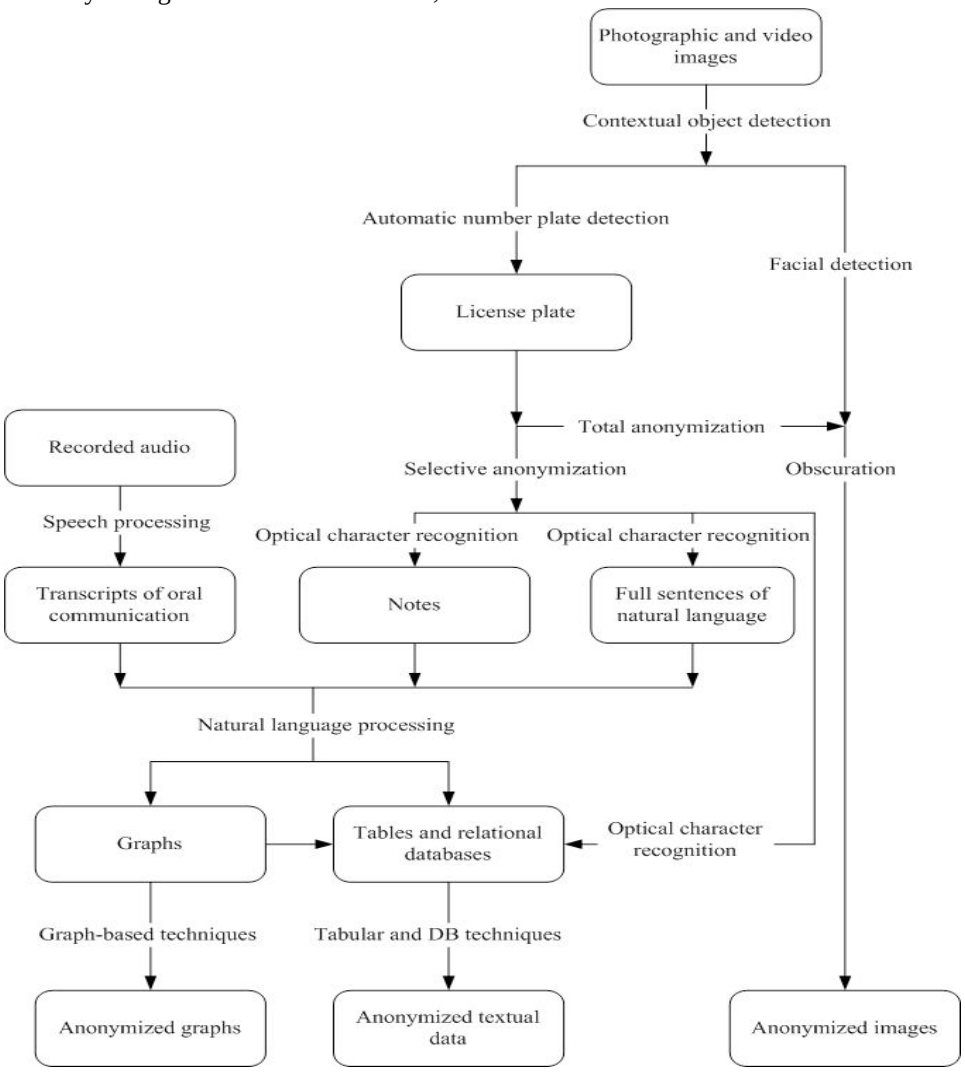

Figure 1. Data Flow Anonymization Process

One of the great benefits of implementing anonymization as a control in the context of the enhanced privacy risk model is that the model provides a straightforward way of assessing the nominal impact of the control. Because anonymization adjusts the characteristics underlying the identified risks, a reassessment can be performed by revising the characteristics in the risk analysis and re-identifying and re-assessing relevant risks. This process, though, is distinct from the process of ensuring that the anonymized data exhibits the desired characteristics and that those characteristics are robust.

While we have developed a process for governing the anonymization of PII with particular properties, this supports only a single segment in Fig. 1. While techniques are available to support a number of the other segments, their amenability to a programmatic approach consistent with what we have already defined remains to be seen. To the extent that the various data flow segments in Fig. 1 demand distinct programmatic as well as technical approaches, the potential for successful and cost-effective anonymization becomes more limited.

IV. Revisiting the Google Street View Privacy Controversy

We are using Google Street View, a feature of Google Maps that provides 360-degree street-level photographic views of various geographic locations, as an example of the application of the enhanced privacy risk model, and therefore employ it here as well. Here, though, we take our pre-existing privacy risk analysis as the starting point and demonstrate how anonymization could be systematically and appropriately considered and applied as a risk mitigation. In so doing, we end up in more or less the same place as Google Street View eventually did (with some cross-national variations), but in an arguably much less haphazard and reactive fashion.

A. Risk Response Determination

Our original analysis stopped after the risk assessment phase. Apropos of our interest in anonymization as a privacy risk control, we start here with risk response determination.

Anonymization, like any other privacy risk control, is better suited to mitigating some risks than others. Table I relates anonymization to privacy risks based on Solove's taxonomy. Risks that were identified for Street View are italicized. Each of those risks is either largely or somewhat amenable in principle to anonymization as a risk control. When one examines the implicated characteristics in Table II, which include faces and vehicle license plates (person identifiers) in almost every case, it becomes evident that anonymization would be a highly suitable risk mitigation, relevant factors such as technical feasibility and resources permitting.

This point can be systematically verified and residual risk assessed by revisiting and appropriately revising the original analysis. Due to space limitations, we omit the step-by-step revisions to the analysis and focus on the resulting changes to the characteristics implicated by the identified risks.

TABLE I. Applicability of Anonymization to Privacy Risks (Street View Risks Italicized)

Privacy Risk	Applicability of Anonymization
Surveillance	To the extent the surveillance is information-based (including digital photos/video), anonymization could mitigate the risk.
Interrogation	The nature of this risk is such that it cannot be mitigated by anonymization.
Aggregation	Anonymization can mitigate this risk by making it impossible to associate discrete pieces of information with the same individual. However, if aggregation per se must be performed, pseudonymity can maintain linkability of the information while still mitigating risk to the individual. Further mitigation might be obtained by reducing the information contained in the attributes being aggregated.
Identification	Anonymization directly mitigates this risk.
Insecurity	Anonymization can mitigate this risk by reducing the information being protected and/or the ability of others to associate the information with specific individuals.
Secondary Use	Anonymization can mitigate this risk by reducing the information being used and/or its linkage to an identifiable individual. However, substantial residual risk may remain regardless of the extent to which the data has been de-identified if secondary use may affect the individual as a member of an identifiable group.
Exclusion	Anonymization can mitigate this risk through de-identification. However, de-identification is seldom absolute; therefore, individuals likely will retain a stake in their information.
Breach of Confidentiality	This risk is grounded in trust relationships; therefore, anonymization would not be a particularly effective mitigation.
Disclosure	Anonymization can mitigate this risk by reducing the information disclosed and/or the ability of others to associate the information with specific individuals.
Distortion	Anonymization can mitigate this risk by reducing the information being used and/or its linkage to an identifiable individual. However, because the harm arises in part from inaccuracy of the information, the mitigation obtained from information reduction may be very limited.
Exposure	To the extent the exposure is information-based (including digital photos/video), anonymization could mitigate the risk.
Increased Accessibility	Anonymization can indirectly mitigate this risk by reducing the information being rendered more accessible.
Blackmail	Anonymization can mitigate this risk by reducing the information available and/or its linkage to an identifiable individual.
Appropriation	This risk is grounded in identity; therefore, anonymization can mitigate the risk through de-identification.
Intrusion	The nature of this risk is such that it cannot be mitigated by anonymization.
Decision Interference	The nature of this risk is such that it cannot be mitigated by anonymization.

While the risks remain, their severity decreases. Removing identifiers such as faces and license plate numbers does not necessarily render a given participant unidentifiable to all possible viewers. There may still be information contained in the images—clothing or vehicles whose appearance will be recognizable, for example—that could enable identification by those in a position to do so. However, anonymization does make identification more difficult. Furthermore, depending on their activities, some individuals may still experience feelings of embarrassment or concern, despite the anonymization. In general, though, the risks have been greatly reduced from what they originally were and constitute residual risk, which may or may not require additional responses.

Non-obvious person identifiers are by definition difficult to selectively remove in an automatic fashion. Similarly, any sensitivity attaching to participant behavior, signage, or street numbers will also be idiosyncratic; systematically blurring all of it, leaving aside any issues of technical feasibility, would probably significantly detract from the utility of Street View. However, supporting requests either before or after the fact to blur or remove specific information would allow residual risk to be addressed on an exception basis.

It is worth noting that, over time, Google has responded to Street View privacy concerns by implementing precisely these kinds of controls: automatic blurring of faces and license plates and the ability of individuals to request blurring of salient objects (persons, vehicles, buildings), as well as the outright removal of images with inappropriate content (e.g., nudity) [11]. However, these controls have evolved over time in response to various privacy controversies ignited by Street View.

B. Risk Control Implementation

Having made the decision to apply anonymization as a risk control by obscuring obvious person identifiers—faces and license plate numbers—implementation must now be addressed. For this, the integrated anonymization framework can be leveraged to establish a roadmap for accomplishing the necessary transformations. To do this, the relevant data must be construed in a way that makes it amenable to the anonymization data flow process depicted in Fig. 1.

In the case of Street View, this is straightforward. The relevant data is, by definition, contained in photographic images. Therefore, "photographic and video images" is the appropriate entry point into the process. Since the intention is to anonymize both faces and license plates specifically, contextual object detection would be used to identify these within the images.

TABLE II. STREET VIEW PRIVACY RISK AND RESIDUAL RISK ASSESSMENT

Life Cycle Stage	Disruptions	Relevant Privacy Risks	Implicated Characteristics	Implicated Characteristics of Residual Risks
Collection	Visual information is comprehensive and many-to-one	Exclusion (no consent, possibly no awareness)	Faces, vehicle license plates, participant behaviors, location identifiers (e.g., signs, street address); momentary	Participant behaviors, location identifiers (e.g., signs, street address); momentary
Processing	Visual and locational information are formally linked	Aggregation (placement of participants and things at a specific location)	Faces, vehicle license plates, participant behaviors, location identifiers (e.g., signs, street address)	Participant behaviors, location identifiers (e.g., signs, street address)
Use	Linked information available for much wider purposes	Secondary use, exclusion (no limitation or control of arbitrary uses)	Internet accessibility	Internet accessibility
Disclosure	Linked information available to much more extensive audience	Insecurity (criminal or other targeting of specific locations), disclosure, exposure, increased accessibility (broad access to comprehensive, location-specific visuals of participants and activities)	Internet accessibility; faces, vehicle license plates, participant behaviors, location identifiers (e.g., signs, street address)	Internet accessibility; participant behaviors, location identifiers (e.g., signs, street address)
Retention	Linked information stored by organization	Secondary use, exclusion (persistent availability of information)	Faces, vehicle license plates, participant behaviors, location identifiers (e.g., signs, street address)	Participant behaviors, location identifiers (e.g., signs, street address)
Destruction	Linked information persists as dictated by organization	Secondary use, exclusion (participants cannot request removal)	Faces, vehicle license plates, participant behaviors, location identifiers (e.g., signs, street address)	Participant behaviors, location identifiers (e.g., signs, street address)

While the detected faces can be directly anonymized through obscuration (blurring), the concern with the license plates is the number rather than the plate per se, so either selective or total anonymization must be chosen. In this case, the objective is the latter. Therefore, detected license plates can be directly anonymized through blurring as well. Both automated facial detection [12] and automated license plate detection [13] are available technologies that can be deployed in support of this process. (Technologies supporting detection of other specific types of objects in photographic and video images can be easily accommodated within the framework in a similar manner.)

V. Conclusions and Future Work

Useful synergies are possible when combining an integrated anonymization framework with an enhanced privacy risk model. While there are invariably practical limitations to the applicability of anonymization as a privacy risk control, it can add value under certain circumstances. The enhanced risk model supports the systematic analysis necessary to identify those circumstances and to confirm the projected efficacy of what is proposed. We plan to continue refining the risk model and formally interrelating other kinds of privacy controls.

The use of anonymization as a privacy risk control begs an increasingly important question: just what qualifies as PII in a networked world in which rich sources of auxiliary data present myriad opportunities for linking all manner of information to identifiable individuals? One set of responses to high-profile anonymization failures has been to question the very utility of PII as a designation [14]. However, as others have argued, some concept of PII is necessary in order to bound and focus informational privacy laws and regulations [15]. This also holds true for privacy risk modeling and privacy risk controls that deal with informational privacy.

In what we have described, we take it for granted that some operational notion of PII exists that supports identification and manipulation of relevant attributes. However, we too recognize the difficulties with the current conception of PII as information that directly or indirectly identifies an individual or is linked or linkable to an individual. As part of our ongoing refinement of the enhanced risk model, we aim to bring a new perspective to bear on this problem, one grounded in cognitive studies of category construction and use.

A workable conception of PII based on deep analysis of PII as a category (or possibly multiple related categories) could ameliorate at least some of the definitional issues while supporting greater analytical precision in the enhanced privacy risk model. Moreover, such a conception would not necessarily destructively ricochet among anonymization techniques, as these techniques infrequently assume much that is definitive about the nature of PII. Situated within an enhanced privacy risk model, an explicit notion of PII would appropriately propagate to anonymization and other controls. Therefore, we aim to redefine PII not for the sake of anonymization or any other specific risk control, but for the sake of our ability to appropriately assess informational privacy risk.

Acknowledgment

MITRE colleagues Lisa Mitchell, Aaron Powell, and David Weitzel contributed to the work on which this paper is based, which was overseen by Karyn Higa-Smith in the Cyber Security Division at DHS S&T.

References

[1] D. Wright, K. Wadhwa, P. De Hert, and D. Kloza, "A Privacy Impact Assessment Framework for data protection and privacy rights," European Commission Directorate General Justice, JLS/2009-2010/DAP/AG, September 21, 2011.

[2] S. Shapiro, "Privacy risk modeling beyond Fair Information Practice Principles," unpublished.

[3] S. Hansell, "AOL removes search data on vast group of web users," New York Times, August 8, 2006.

[4] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in Proc. of the 2008 IEEE Symposium on Security and Privacy, pp. 111-125, 2008.

[5] P. Ohm, "Broken promises of privacy: Responding to the surprising failure of anonymization," UCLA Law Review, vol. 57, pp. 1701-1777, August 2010.

[6] S. Shapiro, "Separating the baby from the bathwater: Toward a generic and practical framework for anonymization," in Proc. of the 2011 IEEE International Conference on Technologies for Homeland Security, 2011.

[7] S. Spiekermann and L. Cranor, "Engineering privacy," IEEE Trans. on Software Engineering, vol. 35, pp. 67-82, January/February 2009.

[8] M. Oetzel, S. Spiekermann, I Grüning, H. Kelter, and S. Mull, "Privacy Impact Assessment Guideline," Bundesamt für Sicherheit in der Informationstechnik, 2011.

[9] H. Nissenbaum, Privacy in Context: Technology, Policy, and the Integrity of Social Life. Palo Alto: Stanford Law Books, 2009.

[10] D. Solove, Understanding Privacy. Cambridge: Harvard University Press, 2010.

[11] http://maps.google.com/intl/en_us/help/maps/streetview/privacy.html, accessed January 24, 2012.

[12] R. Gross and L. Sweeney, Towards real-world face de-identification. Biometrics: Theory, Applications, and Systems, pp. 1-8, 2007.

[13] L. Keilthy, ANPR system performance, Parking Trend International, 2008.

[14] A. Narayanan and V. Shmatikov, "Myths and fallacies of personally identifiable information," Communications of the ACM, vol. 53, pp. 24-26, June 2010.

[15] P. Schwartz and D. Solove, "The PII problem: Privacy and a new concept of personally identifiable Information," New York University Law Review, vol. 86, p. 1814, 2011.