

Maurer School of Law: Indiana University Digital Repository @ Maurer Law

Articles by Maurer Faculty

Faculty Scholarship

2013

Notice and Consent in a World of Big Data

Fred H. Cate

Indiana University Maurer School of Law, fcate@indiana.edu

Viktor Mayer-Schönberger

University of Oxford

Follow this and additional works at: <https://www.repository.law.indiana.edu/facpub>



Part of the [Information Security Commons](#), and the [Privacy Law Commons](#)

Recommended Citation

Cate, Fred H. and Mayer-Schönberger, Viktor, "Notice and Consent in a World of Big Data" (2013). *Articles by Maurer Faculty*. 2662.
<https://www.repository.law.indiana.edu/facpub/2662>

This Article is brought to you for free and open access by the Faculty Scholarship at Digital Repository @ Maurer Law. It has been accepted for inclusion in Articles by Maurer Faculty by an authorized administrator of Digital Repository @ Maurer Law. For more information, please contact wattn@indiana.edu.



LAW LIBRARY
INDIANA UNIVERSITY
Maurer School of Law
Bloomington

Tomorrow's privacy

Notice and consent in a world of Big Data

Fred H. Cate* and Viktor Mayer-Schönberger**

Introduction

Just over four decades ago, the first information privacy statutes were enacted. After intense discussions in North America and Europe, at the end of the 1970s a number of privacy principles emerged under the concept of Fair Information Practices and later became the foundation for the Organisation for Economic Co-operation and Development (OECD) Guidelines on the Protection of Privacy and Transborder Flows of Personal Data adopted in 1980. Those principles, which seek to balance the 'fundamental but competing values' of 'privacy and the free flow of information', form the basis of most privacy legislation around the world. At their core, they require that the processing of personal information be lawful, which in practice means that either the processing is explicitly permissible under law or the individual whose personal data is being processed has—after being informed of the reason, context, and purpose of the processing—given consent.

Intuitively, such an approach makes sense. It empowers individuals so that they—rather than a bureaucratic government agency—can exercise their privacy rights as they see fit. Over the years, and especially in the context of the Internet, this system of 'notice and consent,' originally intended to be only one of multiple ways through which the lawful processing of personal data can take place, has become the dominant mechanism.

Today, almost everywhere that individuals venture, especially online, they are presented with long and complex privacy notices routinely written by lawyers for lawyers, and then requested to either 'consent' or abandon the use of the desired service. That binary choice is not what the privacy architects envisioned four decades ago when they imagined empowered individuals making informed decisions about the

Abstract

- Nowadays individuals are often presented with long and complex privacy notices routinely written by lawyers for lawyers, and are then requested to either 'consent' or abandon the use of the desired service.
- The over-use of notice and consent presents increasing challenges in an age of 'Big Data'.
- These phenomena are receiving attention particularly in the context of the current review of the OECD Privacy Guidelines.
- In 2012 Microsoft sponsored an initiative designed to engage leading regulators, industry executives, public interest advocates, and academic experts in frank discussions about the role of individual control and notice and consent in data protection today, and alternative models for providing better protection for both information privacy and valuable data flows in the emerging world of Big Data and cloud computing.

processing of their personal data. In practice, it certainly is not the optimal mechanism to ensure that either information privacy or the free flow of information is being protected.

Equally challenging is the fact that in the age of 'Big Data,' much of the value of personal information is not apparent at the time of collection, when notice and consent are normally given. Because future uses would require going back to individuals for their amended consent, many future uses that have significant individual and societal benefits might be simply too costly to undertake. Moreover, what used to be a relatively simple relationship between individuals and the

* Editor; Distinguished Professor, C Ben Dutton Professor of Law, and Director of the Center for Applied Cybersecurity Research and the Center on Law, Ethics and Applied Research in Health Information, Indiana University. E-mail: fred@fredhcate.org.

** Professor of Internet Governance and Regulation, Oxford Internet Institute, University of Oxford.

processors or users of their personal data has often become complicated as datasets are combined and data processors and users change. That also makes it even harder for individuals to fully grasp the complexity of the situation they are asked to assess. Finally, Big Data is not only big, but also collected and processed so often as to make opportunities to consent an unacceptable burden for most individuals. (To take just one example, the *New York Times* reported in 2012 that one US company that few people have ever heard of engages in more than 50 trillion transactions involving recorded personal data every year.¹)

Taken together, these realities challenge the dominant current privacy mechanism of notice and consent. They can leave individuals' privacy badly exposed, as individuals are forced to make overly complex decisions based on limited information, while data processors can perhaps too easily point to the formality of notice and consent and thereby abrogate much of their responsibility. At the same time, current privacy mechanisms can unduly interfere with the innovation potential of data use. These challenges require a rational reassessment of the privacy landscape, as well as an evaluation of the optimal mix of mechanisms available to protect information privacy in a world that is beginning to realize the latent value of Big Data.

To help foster that reassessment, in 2012 Microsoft sponsored an initiative designed to engage leading regulators, industry executives, public interest advocates, and academic experts in frank discussions about the role of individual control and notice and consent in data protection today, and alternative models for providing better protection for both information privacy and valuable data flows in the emerging world of Big Data and cloud computing.

Between May and August of 2012, Microsoft hosted a series of regional privacy dialogues in Washington, DC, Brussels, Singapore, Sydney, and São Paulo. These dialogues involved a total of 78 participants drawn in almost equal proportions from government, academia, advocacy, and industry (including editors of this journal and members of its editorial board). Each discussion was moderated by a privacy scholar from the region: Professor Fred H. Cate in Washington, DC, Professor Viktor Mayer-Schönberger in Brussels, former Australian Privacy Commissioner Malcolm Crompton in Sydney and Singapore, and Professor Nelson Remolina

Angarita in São Paulo. The discussions followed the Chatham House Rule, under which participants were welcome to use the information learned there, but agreed not to disclose the source or name of the individual or institution involved.

Following the five regional events, in September 2012 Microsoft welcomed more than 70 privacy and data protection experts from government, industry, non-profit organizations, and academia to a global privacy summit in Redmond, Washington. Drawn from 19 countries on five continents, the participants came together to consider the future of data sources and uses and practical steps to enhance privacy protection. Many had participated in the regional discussions, and all received in advance of the summit a summary of the key points from those discussions.

This article briefly summarizes the key points of discussion during the regional privacy dialogues and the global privacy summit. It is based on a report that was reviewed by the participants prior to being released, but neither that report, nor this article, purport to reflect any consensus of the participants or the views of any individual participant or organization, including Microsoft. The complete report, including a list of all participants and their affiliations, is available online.²

The regional privacy dialogues

Despite considerable variety in the five regional discussions in Washington, DC, Brussels, Singapore, Sydney, and São Paulo, there was significant overlap concerning key issues. There was a widely shared sense that notice and consent either have, or are perceived as having, become the dominant means of data protection. Even in countries in which notice and consent are not the primary data protection tools provided by law, they have nevertheless assumed undue importance in policy debates and popular discussions about data protection. As a result, or perhaps as a cause, ensuring individual control over personal data is widely perceived as the goal of data protection and is often highlighted as such by political leaders and commentators.

Further, there was broad general agreement that privacy frameworks that rely heavily on individual notice and consent are neither sustainable in the face of dramatic increases in the volume and velocity of information flows nor desirable because of the burden they

1 Natasha Singer, 'You for Sale: Mapping, and Sharing, the Consumer Genome', *N.Y. Times*, 17 June 2012, at BU1, available at <<http://www.nytimes.com/2012/06/17/technology/axiom-the-quiet-giant-of-consumer-database-marketing.html?pagewanted=all&r=0>> accessed 5 March 2013.

2 The report is available at <<http://download.microsoft.com/download/9/8/F/98FE20D2-FAE7-43C7-B569-C363F45C8B24/Microsoft%20Global%20Privacy%20Summit%20Report.pdf>> accessed 5 March 2013.

place on individuals to understand the issues, make choices, and then engage in oversight and enforcement. In short, ensuring individual control over personal data is not only an increasingly unattainable objective of data protection, but in many settings it is an undesirable one as well.

The discussions also addressed the advent of Big Data, the increasingly ubiquitous nature of data collection and use, and the technological developments that expand our capacity to interconnect, analyse, identify, and extract new and unanticipated value from even old or seemingly worthless data as factors that require new approaches to data protection. A key sentiment expressed in all of the discussions is that those new approaches must shift responsibility away from data subjects towards data users, and towards a focus on accountability for responsible data stewardship, rather than mere compliance while ensuring that expectations and protection of privacy is preserved.

As to additional mechanisms to ensure privacy, one of the most widely discussed alternatives was focusing more attention on the 'use' of personal information rather than on its 'collection', given the increasingly pervasive nature of data collection and surveillance, inexpensive data storage and sharing, and the development of valuable new uses for personal data. Many participants were careful to note that focusing on the use of personal data does not mean that there should not be responsibilities or regulation relating to data collection, nor should a focus on data collection in specific or sensitive circumstances be abandoned. Rather, in most situations, a more practical, as well as sensitive, balancing of valuable data flows and more effective privacy protection is likely to be obtained by focusing more attention on appropriate, accountable use.

Many of the dialogues devoted considerable time to what constitutes a 'use' of personal data, what uses should be permitted or prohibited (or should require some greater form of authorization, for example, specific affirmative consent), and by what standards these determinations should be made. As many participants noted, the failure to build consensus around the standards that data protection laws should implement currently impedes effective regulation and efforts at international harmonization.

There seemed broad agreement that uses should include disclosure, but there was uncertainty about whether uses should include analysis of data within an institution if the data are not used to make a decision or create new information. Similarly, there was nearly universal agreement that the 'harms' or 'impacts' that data protection laws should be designed to avoid must

not only include physical and financial injury but also broader concepts consistent with protecting privacy as a human right—such as reputational or social harm and the chilling effect of surveillance, but there was little consensus as to precisely which other impacts should be included or how they might be determined.

At the same time, while recognizing privacy as a human right and the need to more clearly define impacts, there was recognition of the need to resolve conflicts with other fundamental rights. For example, privacy can be in conflict with the right to engage in free speech or to live in a society free from the threat of terrorism. The quest needs to be for more effective and efficient protection of privacy, not a weakening of the protections that existing frameworks are intended to provide, even if they do not always do so successfully.

In addition to considering how the standards to guide data protection should be determined, the participants also devoted considerable energy to the question of how those standards should be implemented as a matter of both law and individual entity policy. There was broad agreement that implementation should be practical, flexible, and focused on data users ensuring and demonstrating accountability for their responsible use of personal data.

Despite the limits of notice and consent, many participants noted that this mechanism might continue to play a role in the future, even if in a modified form from today. For example, notice may be a key tool for transparency, although this may suggest that disclosure to a regulator or a central, accessible repository might be more efficient than individual notice. Similarly, consent may be necessary for the use of certain types of data or for certain uses of data. Some participants expressed the hope that by reserving notice and consent for more appropriate uses, individuals might pay more attention when this mechanism is used. Recognizing that notice and consent will have continuing value in certain settings also reflects an 'evolutionary' rather than 'revolutionary' approach to updating data protection principles, which many of the participants found desirable. As a result, while all of the dialogues were clear that merely fine-tuning notice and consent will not provide the sort of new approaches to data protection widely thought necessary, this does not suggest that notice and consent should not be improved however possible so that when used, they are more effective.

One key element of responsible data stewardship that emerged at all five events was the need for better security to protect personal data against unintended access, loss, alteration, or disclosure. Any new model

of data protection must ensure a high degree of confidence that personal data will be appropriately protected. While standards for data security are increasingly being implemented around the globe, there was discussion about the need to ensure that security standards remain flexible, given the constantly evolving nature of security challenges, and that they be more substantive than focused on providing notice to people whose data may have been compromised. Again, a key goal of many participants is to shift the responsibility for data protection away from the data subject towards the data user.

A number of discussions touched on the enforcement of data protection laws and policies. Participants agreed that enforcement is a critical element of data protection, but some placed special emphasis on enforcement as a way to transform 'self-regulation' into 'co-regulation', by giving the force of law to institutional or sectoral privacy undertakings that meet minimum requirements. There was also discussion of the extent to which relying on multinational enforcement mechanisms (such as designated lead enforcement agencies, an international enforcement body, or binding arbitration) might help build cross-border accountability and trust while reducing the costs of enforcement and avoiding duplicative enforcement.

Discussions at all five events addressed the need for greater harmonization and interoperability in data protection across national borders in a way that does not lead to a 'race-to-the-bottom, lowest-common-denominator' result. Harmonization and interoperability have assumed even greater importance with the growth of cloud computing and e-commerce, which often involve instant flows of data across geographic boundaries. One advantage of developing new approaches to data protection based on widely shared twenty-first-century standards of appropriate data stewardship is that it might well lead to more consistent privacy protection across borders and greater harmonization of privacy practices and obligations. Moreover, greater harmonization and interoperability are potentially effective tools for maximizing the scarce resources available for data protection and enforcement, and for ensuring that individuals enjoy commensurate levels of privacy protection—and that their rights can be vindicated affordably and easily—no matter where they travel, browse, or shop, or where their data are stored.

It is always risky to try to summarize rich and varied discussions among talented privacy professionals, but two themes seemed to dominate most of the discussions at all five locations: that society should, to the greatest extent possible, shift responsibility for protecting privacy

from the individual data subject to the data user, and that the tools used to do so should be as flexible, efficient, practical, interoperable, and sensitive to competing values and realities as possible to achieve responsible data stewardship.

The global privacy summit

The global privacy summit began with presentations on innovative new uses of personal data by Craig Mundie, chief research and strategy officer at Microsoft; Leroy Hood, president of the Institute for Systems Biology; Kush Parikh, senior vice president of business development at Inrix; and Kenn Cukier, data editor at the *Economist*, but the bulk of the summit was spent in interactive discussions reflecting on the themes identified in the regional workshops and considering how best to address them in practice.

The participants were able to respond to speakers, pose questions, and interact with one another not only face to face, but also using an interactive tool that allowed participation by everyone and permitted every idea to be captured (anonymously and with consent). To facilitate maximum engagement, discussions took place as a single large group as well as in seven smaller groups, in all cases with professional facilitators and rapporteurs. Every effort was made to ensure not only that all voices were heard and all interjections included, but also that the discussion progressed toward a practical and useful outcome.

After the summit, participants had another two weeks to review the presentations and documents online and add additional comments to the record. In addition, participants had the opportunity to review and comment on the draft report on which this article is based. The remainder of this article seeks to capture the major themes that emerged during the summit. Given the breadth and depth of the discussions, this article is necessarily selective, and while it is based on the discussions in Redmond, it does not purport to reflect a consensus view of the participants.

Significant challenges

Participants identified a number of privacy challenges in the near future, but five broad themes emerged:

- There was considerable concern about the need for greater public awareness of privacy issues, increased transparency about the uses of personal data, and more effective education about privacy and the valuable uses of personal data. There was broad agreement that even if data protection systems come to

rely less on notice and consent, practical and ethical considerations require that the public and other key stakeholders (including policy-makers, regulators, the press, and business) be more informed about data processing activities and the benefits and risks of those activities.

- The pressing need for increased standardization, consistency, and interoperability across data protection laws and practices was emphasized by many participants. While participants also recognized that there are distinctive national and cultural aspects of privacy and stressed the continuing role of national data protection laws, there was a widely shared belief that individuals, societies, and data users can benefit from greater consistency and interoperability across national systems.
- A number of participants noted the values in tension with privacy and therefore stressed the importance of 'balance' when protecting privacy and the need to restore the balance between privacy and the free flow of information reflected in the OECD Privacy Guidelines. For example, one common refrain was the importance of not suppressing innovation with overly restrictive privacy laws. Also, closely related to the importance of balance was the perceived need for flexibility in data protection regimes, especially in light of rapidly changing technologies and applications, evolving expectations of privacy, and recognition that different types and uses of personal data will inevitably need to be treated differently.
- One theme that became increasingly prevalent as the discussion focused on practical steps for moving forward was the need for clear, specific terms and definitions. Participants noted that as privacy protection increasingly focuses on permitting or preventing certain uses of information or guarding against certain risks or side effects of data use, the more important it is to define uses, risks, and other terms clearly and concretely. This is obviously a challenge and may well be in tension with other themes, but it is nevertheless important to heed.
- The final prevalent theme was the need for an updated or enhanced framework for protecting personal data. The OECD Privacy Guidelines, on which most modern data protection laws are based, was crafted more than 30 years ago, before the advent of the World Wide Web, cloud computing, smart phones, or Big Data. While the guidelines seem

to have continuing relevance, they are no longer adequate as a guide for twenty-first-century data protection or as the basis for greater interoperability among national data protection regimes.

Privacy principles for the twenty-first century

To help focus thinking and facilitate practical outcomes, participants at the summit worked in small groups to consider how the OECD Privacy Guidelines might be updated in light of the themes that emerged in the regional discussions. To aid in their task, participants received a draft version of revised principles reflecting the regional discussions.³

Collection Limitation Principle

With respect to what is meant by 'personal data,' in the Collection Limitation Principle, '[t]here shall be limits to the collection of personal data,' there is a growing awareness that many forms of previously unidentifiable data might become personally identifiable in a world of Big Data and advanced analytics. Applying the Collection Limitation Principle to all data seems unworkably broad, but to limit it to data already recognized as 'personal' seems too narrow. This theme continued in the discussion of many of the other principles.

In addition, the requirement in the original OECD principle that data be collected, 'when appropriate,' with the 'knowledge or consent of the data subject,' seems to ignore the reality of the extraordinary volume of data that is generated today through routine activities and transactions and near-ubiquitous sensors (such as surveillance cameras, location monitoring by smart phones, and embedded computers in cars and other devices). Often, knowledge or consent of data collection in these situations is either non-existent or likely to be so vague as to be meaningless.

Data Quality Principle

This principle seemed to strike most participants as useful and relevant, with one possible exception. The language 'to be used' in the principle, '[p]ersonal data should be relevant to the purposes for which they are to be used,' could be interpreted as suggesting that the determination as to relevance might need to be made only at the time of collection, with an eye toward intended use. This, of course, is inconsistent with the world of Big Data in which new uses for data are discovered over time. The principle might be more consistent with twenty-first-century reality and offer better, continuing protection for personal privacy if the

3 That draft version is available at <[http://download.microsoft.com/download/9/8/F/98FE20D2-FAE7-43C7-B569-C363F45C8B24/Microsoft%](http://download.microsoft.com/download/9/8/F/98FE20D2-FAE7-43C7-B569-C363F45C8B24/Microsoft%20Global%20Privacy%20Summit%20Report.pdf)

[20Global%20Privacy%20Summit%20Report.pdf](http://download.microsoft.com/download/9/8/F/98FE20D2-FAE7-43C7-B569-C363F45C8B24/Microsoft%20Global%20Privacy%20Summit%20Report.pdf)> accessed 5 March 2013.

words 'to be' were removed and if relevance were evaluated for each use at the time of the use.

Purpose Specification Principle

Many participants found the Purpose Specification Principle to be largely inconsistent with the ways in which data are used today and will be used in the future, and it is the one principle that many participants suggested might be omitted entirely or at the very least dramatically reshaped. There seemed to be a broadly shared sentiment that of course there should be limits on the uses of data, but that those limits need not necessarily be linked to the purposes for which the data were originally collected. Use limits are discussed further under the next principle.

Use Limitation Principle

Many participants considered the Use Limitation Principle, originally adopted in 1980, unworkably narrow in the twenty-first century because it restricts the uses of data to those required by law or to which the individual has consented. The Use Limitation Principle threatens medical research, fraud prevention, identity verification, credit worthiness assessment, and many other valuable uses of data, not to mention valuable benefits from Big Data.

The harder question, as many summit participants noted, is what should replace this outdated principle. There was considerable discussion about restructuring the principle to permit all uses of data other than those meeting certain criteria. But determining those criteria proved quite difficult. Seemingly everyone agreed that uses that cause financial or physical injury to the data subject or that involve discrimination or some other illegal act would clearly be prohibited, but efforts to move beyond the obvious prohibitions rapidly ran into the twin objections of being vague and of potentially ignoring distinct cultural sensibilities. As was the case with the regional dialogues, there seemed a broad willingness to go further than just restricting uses likely to cause 'harms', but far less agreement emerged on whether prohibited uses should extend to those causing reputational injury (even more controversial if the data are true) or those causing apprehension or discomfort on the part of the data subject but not otherwise violating the law.

Security Safeguards Principle

There was widespread agreement that security is a key component of privacy and many participants suggested that security should be the subject of greater attention and enforcement as a practical matter.

Openness Principle

Similarly, there was broad agreement that openness or transparency is a critical element of any data protection system. A number of participants expressed the view that greater transparency should be required if there are fewer opportunities for consent or if personal data can be lawfully collected without consent. This was also seen by some as a key opportunity to help educate data subjects, perhaps by requiring as part of the Openness Principle that information about data subjects' legal rights, and ways to exercise them, be made available along with information about data processing activities.

Individual Participation Principle

The OECD Individual Participation Principle sparked considerable discussion—not because there seemed to be any opposition to the concept, but rather because there seemed a tangible risk that the principle could generate significant burdens for both data processors and data subjects in a world of Big Data. Three broad strands of concern were evident in the discussion.

First, some participants noted that with the exponential growth of data collected about individuals, responding to requests for access to such data could be prohibitively expensive and seemingly of little value if the data are not being used for any significant purpose. Inaccurate data, for example, might be very significant if they are being used to determine eligibility for a job or a government benefit, but they might be less relevant as a small part of a vast dataset being used to determine normal spending patterns for fraud detection purposes. One response to this concern would be to focus some or all of the legal obligations created under this principle on data that are used, or are likely to be used, for some significant activity or in a manner affecting a legally protected right, as suggested in the discussion version.

Second, some participants worried about the burden this principle would place both on individuals, who might find it difficult to even know all of the parties with access to their data, and on organizations, which in many cases, due to the distributed nature of data, would have difficulty verifying the identity of individuals and determining which data pertain to them. Some also expressed concern about issues that would arise if data that are being used in a de-identified format must be re-identified purely to respond to a general access request.

Third, participants expressed concern that the principle may be too narrow. By focusing on 'individual participation' rather than on fundamental fairness in data processing, the principle as written in 1980 might exacerbate the concern that in a world of Big Data individuals

are being asked to do more, when it should be data processors that bear the burden of responsible data stewardship. However, discussions about ways to improve or expand this principle raised concerns about vagueness and the difficulty of defining key terms such as 'fair.'

Accountability Principle

Participants broadly supported this principle, although some felt that the original version does not go far enough. A number of participants noted that the original appears to focus on compliance when, in fact, it should focus more on responsible data stewardship and the broad mechanisms that data processors can use to ensure compliance and demonstrate it to regulators and the public, as the discussion version does.

Conclusion

These concerns and thoughts reflect the major streams of a rich and far-ranging debate. It appeared obvious to the participants that the unique challenges presented by Big Data, as well as the complexity and multitude of social interactions online, have created an urgent need to adjust information privacy regulations, and the principles that underlie them, to meet the needs of a new era. The goal is to make privacy protections more effective and efficient as well as to revisit the balance

between privacy and information flows in a world of not only vastly more data, but also more and rapidly changing, valuable uses of that data.

In important ways, participants were in agreement on the general direction of such adjustments. Much more, of course needs to be done—and swiftly, given the pace of technological change—to ensure that individuals remain protected and data processors embrace their responsibilities while innovation is not artificially constrained. Looking at how 'use' can be defined, and acceptable uses delineated from harmful ones, is an evident next step. Continuing the dialogue about how the OECD Privacy Guidelines might be updated to respond to dramatic technological changes and the challenges of an information-based economy is another. The recently launched initiative of the World Economic Forum, 'Rethinking Personal Data', which is premised on the conviction that basic data protection principles, while not flawed, 'do not work in today's world', also promises to help inform and stimulate global discussions about appropriate data protection. Those discussions are vital if we are to protect both privacy and progress in the twenty-first century.⁴

doi:10.1093/idpl/ipt005

4 World Economic Forum, *Unlocking the Economic Value of Personal Data: Balancing Growth and Protection* 3 (2012), available at <http://www3.weforum.org/docs/WEF_IT_UnlockingValueData_BalancingGrowthProtection_SessionSummary.pdf> accessed 5 March 2013; see also <<http://www.weforum.org/issues/rethinking-personal-data>>; see

also World Economic Forum, *Rethinking Personal Data: Strengthening Trust* (2012), available at <http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf> accessed 5 March 2013.