Office of the Privacy Commissioner for Personal Data, Hong Kong

International Conference on Big Data from a Privacy Perspective
10 June 2015

Big Data and the Future of Data Protection

Fred H. Cate

Distinguished Professor and
C. Ben Dutton Professor of Law
Director, Center for Information Privacy and Security
Indiana University Maurer School of Law;
Senior Policy Advisor
The Centre for Information Policy Leadership at Hunton & Williams LLP

Sources of “Big Data”

We are witnessing an explosion not only in the volume of personal data being generated, but also in the comprehensiveness and granularity of the records those data create about each of us. “Big data” describes data that are not only large, but also increasingly complete and granular. Sources include:

- Data we generate and reveal either by choice, for example, through social media and email, or through compulsory disclosure, for example, as a condition of banking or traveling.
- Transactional data that result from purchases, health care interactions, communications, etc.
- Data collected by sensors that surround us in our smart phones, tablets, laptops, wearable technologies and even sensor-enabled clothing, RFID-equipped passports, cars, homes, and offices. According to a 2014 study by HP, nine out of ten of the most popular internet-connected devices carry personal data.
- Data calculated or inferred—created, not collected—based on demographic information, census data, and past behavior.

Part of a Larger Data Ecosystem

"Big data" is part of a broader context that includes:

- Ubiquitous surveillance
- Interconnected sensors (i.e., the "Internet of Things")
- Exponential increases in storage capacity and decreases in storage costs
- Dramatic increases in, and widespread distribution of, computational capacity.
- Pervasive networks

Characteristics of “Big Data” Environment

- Greater volume and velocity of data collection and use.
- More data observed by sensors rather than collected from individuals
- More routine or invisible data collection
- More data inferred or calculated.
- Greater data aggregation (or federation) and use by third parties.
- Data reused for unexpected purposes or to identify correlations suggested by the data themselves.

Extraordinary Potential of “Big Data”

According to a 2014 study by Accenture and General Electric, 84% of the companies surveyed believe that big data would “shift the competitive landscape” for their industry within the next year. Some examples:

- Medical research and treatment
- Scientific research
- Fraud detection and prevention
- Product development and design
- National security
- Law enforcement
- Human and national development
- Disaster prevention and response
- Education
- Oversight of institutions and operations
- Human resources
- Marketing

Privacy Challenges Presented by "Big Data"

1. Challenges the **transactional model** of data protection. Consider the OECD Guidelines, adopted in 1980 and reaffirmed in 2013, on which modern privacy laws are based:
    - Collection Limitation Principle—There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
    - Data Quality Principle—Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.
    - Purpose Specification Principle—The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.
    - Use Limitation Principle—Personal data should not be disclosed, made available or otherwise used for purposes other than those specified [above] except: with the consent of the data subject; or by the authority of law.

2. Challenges **notice and consent**.
    - The APEC Privacy Framework, adopted in 2004, provides: “Where appropriate, individuals should be provided with clear, prominent, easily understandable, accessible and affordable mechanisms to exercise choice in relation to the collection, use and disclosure of their personal information.”
    - The EU Data Protection Directive allows consent as a basis for processing, processing sensitive data, and exporting personal data to non-EU countries lacking “adequate” data protection. The pending draft General Data Protection Regulation refers to **“**consent**”** more than 100 times.
    - McDonald and Cranor calculated in 2008 that to read the privacy policies of just the most popular websites would take an individual 244 hours—or more than 30 full working days—each year.
    - The 2014 report by the President’s Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective,* described the “framework of notice and consent” as “unworkable as a useful foundation for policy.” The report stressed that “only in some fantasy world do users actually read these notices and understand their implications before clicking to indicate their consent.”

3. Ignores the value of **unexpected uses** of big data.
    - The President's Council of Advisors for Science & Technology wrote in 2014: "'The notice and consent is defeated by exactly the positive benefits that big data enables: new, non-obvious, unexpectedly powerful uses of data'"
    - Paul Ohm: Big data "thrives on surprising correlations and produces inferences and predictions that defy human understanding . . . . How can you provide notice about the unpredictable and unexplainable?"

4. Challenges **deidentification** and **anonymization**.
    - Latanya Sweeney has shown that 87% of the U.S. population is uniquely identified by date of birth, gender, and 5-digit ZIP Code.
    - Browser choice and font size can provide an accurate, unique online identifier.
    - There are well-publicized examples of Google, Netflix, AOL, and others releasing deidentified data sets only to have the data reidentified within days by researchers correlating them with other data sets.
    - In a world of big data, Cynthia Dwork writes, "'De-identified data' isn't."

Doing Better

1. Focus less on individual consent and more on placing responsibility for data stewardship, and liability for reasonably foreseeable harms, data users
2. Employ a more systemic and well-developed use of risk management
    - Centre for Information Policy Leadership's Project on the Risk-Based Approach to Privacy:

      "Risk management is the process of systematically identifying harms and benefits that could result from an activity. Risk management does not alter rights or obligations, but by assessing both the likelihood and severity of harms and benefits, risk management helps organizations identify mitigation strategies and ultimately reach an optimum outcome that maximizes potential benefits while reducing the risk of harms to that it falls within acceptable limits.

      The ultimate goal of risk management, after taking into account those measures that the data user can take to reduce risk, is to create presumptions concerning common data uses so that both individuals and users can enjoy the benefits of predictability, consistency, and efficiency in data protection." (https://www.informationpolicycentre.com/)

- In 2012 the FTC published a report recommending that companies should "implement accountability mechanisms and conduct regular privacy risk assessments to ensure that privacy issues are addressed throughout an organization." FTC Report, *supra* at 30.
- In 2013 the OECD Council of Ministers revised the *OECD Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* to "implement a risk-based approach." In the accompanying Explanatory Memorandum, the drafters noted the "importance of risk assessment in the development of policies and safeguards to protect privacy."
- The draft text of the European Union General Data Protection Regulation focuses significantly on risk management.

3. Place greater focus on *uses* of big data as opposed to the mere collection or retention of data or the purposes for which data were originally collected
   - There is often a compelling reason for personal data to be disclosed, collected, or created.
   - Assessing the risk to individuals posed by data almost always depends upon the context in which they are used. Data used in one context or for one purpose or subject to one set of protections may be both beneficial and desirable, where the same data used in a different context or for another purpose or without appropriate protections may be both dangerous and undesirable.

4. Develop a broad framework of cognizable harms identified through a transparent, inclusive process including regulators, industry, and individuals
    - Reducing or eliminating the harm that personal information can cause to individuals requires a clear understanding of what constitutes "harm" or other undesired impact in the privacy context.
    - Include **tangible injuries** (e.g., financial loss, physical threat or injury, unlawful discrimination, identity theft, loss of confidentiality, and other significant economic or social disadvantage), and **intangible harms** (e.g., damage to reputation or goodwill, or excessive intrusion into private life) and potentially broader **societal harms** (such as contravention of national and multinational human rights instruments).
5. Pay more attention to transparency and redress
    - The one certainty of data analysis is that there will be errors—errors resulting from problems with data matching and linking, erroneous data, incomplete or inadequate algorithms, and misapplication of data-based tools.
    - Whenever big data are used in ways that affects individuals, there must be effective transparency and redress to protect the rights of individuals, enhance the accuracy and effectiveness of big data tools, and create disincentives for deploying tools inappropriately.

5. Reserve notice and choice for where meaningful and effective
    - Reduces the burden imposed on individuals, focuses their attention on data processing activities only where there are meaningful, effective choices to be made, and reduces “notice fatigue.”

Thank you.

fred@fredhcate.org

Sources

- Asia-Pacific Economic Cooperation, *APEC Privacy Framework*, 2004/AMM/014rev1 (2005), at 17.
- Fred H. Cate & Viktor Mayer-Schönberger, *Data Use and Impact Global Workshop*, Center for Applied Cybersecurity Research (2013).
- Fred H. Cate, Peter Cullen & Viktor Mayer-Schönberger, *Data Protection Principles for the 21st Century*, Oxford Internet Institute (2013).
- Fred H. Cate & Viktor Mayer-Schönberger, *Notice and Consent in a World of Big Data*, Microsoft Corporation (2012).
- Fred H. Cate & Viktor Mayer-Schönberger, "Notice and Consent in a World of Big Data," *International Data Privacy Law*, vol. 3, no. 2 at 67 (2013).
- Fred H. Cate, Peter Cullen & Viktor Mayer-Schönberger, *Data Protection Principles for the 21st Century World*, Microsoft Corporation (2014).
- Centre for Information Policy Leadership at Hunton & Williams LLP, *A Risk-based Approach to Privacy: Improving Effectiveness in Practice* (2014).
- Centre for Information Policy Leadership at Hunton & Williams LLP, The Role of Risk Management in Data Protection (2014).
- *Directive 95/46/EC of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data* (Eur. O.J. 95/L281), Preamble, arts. 7(a), 8(2)(a), 26(1)(a).
- Cynthia Dwork, "Differential Privacy: A Cryptographic Approach to Private Data Analysis," in Julia Lane, Victoria Stodden, Stefan Bender, Helen Nissenbaum, eds., *Privacy, Big Data, and the Public Good* 297 (Cambridge 2014).

- Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (2014).
- Viktor Mayer-Schönberger & Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (2013).
- Alecia M. McDonald & Lorrie Faith Cranor, "The Cost of Reading Privacy Policies," *I/S: A Journal of Law and Policy for the Information Society* (2008).
- OECD, *OECD Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*, C(80)58/FINAL, as amended by C92013)79 (2013), 12.
- OECD, *Supplementary Explanatory Memorandum to the Revised Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data* (2013), 30. Paul Ohm, "Changing the Rules: General Principles for Data Use and Analysis," in Julia Lane, Victoria Stodden, Stefan Bender, Helen Nissenbaum, eds., *Privacy, Big Data, and the Public Good* 100 (Cambridge 2014).
- President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective* (2014).
- Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Unofficial Consolidated Draft Text, Oct. 22, 2013.
- Latanya Sweeney, Simple Demographics Often Identify People Uniquely, Carnegie Mellon University, Data Privacy Working Paper 3 (2000).
- U.S. Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers*, Preliminary FTC Staff Report (2010).
- U.S. Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers*, FTC Report (2012).