IAPP Canada

Privacy Symposium 2016

iapp

#CPS16

UNTANGLING THE MAZE OF BIG DATA ANALYTICS: PRIVACY, SECURITY AND ETHICS

Ann Cavoukian, Executive Director, Privacy and Big Data Institute, Ryerson University

Melissa Krasnow, CIPP/US, Partner, Dorsey & Whitney

Moderator:

Barbara Yuill, News Director, Intellectual Property, Privacy, Data Security, Tech & Telecom, Bloomberg Law

Ann Cavoukian, Ph.D.

Executive Director
Privacy and Big Data Institute
Ryerson University

285 Victoria Street
Toronto, Ontario
M5B 2K3

Phone: (416) 979-5000 ext. 3138

ann.cavoukian@ryerson.ca

ann.cavoukian@ryerson.ca

twitter.com/AnnCavoukian

Big Data

- **90%** of all data was created within the last 2 years;
- **Big Data** analysis and data analytics promise new opportunities to gain valuable insights and benefits – new predictive modes of analysis;
- But, it will also enable **expanded surveillance**, increasing the risk of unauthorized use and disclosure, on a scale previously unimaginable.

Big Data Technology is Not Foolproof

- *"Despite rampant interest from enterprise leaders and often sizeable investments in Big Data technologies,* ***many programs still sputter or fail completely."***

— Evanta Leadership Network, May 29, 2014.

Big Data Leaves Much to be Desired

Big Data is moving from its inflated expectations phase to a trough of disillusionment.

— Gartner Hype Cycle, April 2014

Quantity Does Not Equal Quality

"But while big data promise much to scientists, entrepreneurs and governments, they are doomed to disappoint us if we ignore some very familiar statistical lessons. There are a lot of small data problems that occur in big data. They don't disappear because you've got lots of the stuff … they get worse!" — David Spiegelhalter, Winton Professor, Cambridge University

— *Big data: are we making a big mistake?* FT Magazine, March 2014.

“Forget Big Data … what is needed is Good Data”

— Barrie McKenna,
The serious economic cost of Canada's data deficit,
Globe and Mail, May 12, 2014

Context is Key

- Performing data analytics on context-free data will only yield correlations (which at times, will be spurious);
- By adding context as a feature in the analytics, we may be able to impute causality – which has the potential to be invaluable in our analyses.

PRIVACY BREEDS INNOVATION: IT DOES *NOT* STIFLE IT!

- The argument that privacy stifles innovation reflects a dated, zero-sum mindset;
- The notion that privacy must be sacrificed for innovation is a false win/lose dichotomy, consisting of unnecessary trade-offs;
- The opposite is true – privacy **drives** innovation – it forces innovators to think creatively to find solutions that will serve multiple functionalities;
- We need to abandon zero-sum thinking and adopt a positive-sum paradigm where both innovation ***and*** privacy may be achieved – we need a new playbook!

Let’s Dispel Some Myths

Privacy ≠ Secrecy

Privacy is *not* about having something to hide

Privacy = Control

Privacy = Personal Control

- User control is critical
- Freedom of choice
- Informational self-determination

Context is key!

The Decade of Privacy by Design

Adoption of "Privacy by Design" as an International Standard

Landmark Resolution Passed to Preserve the Future of Privacy

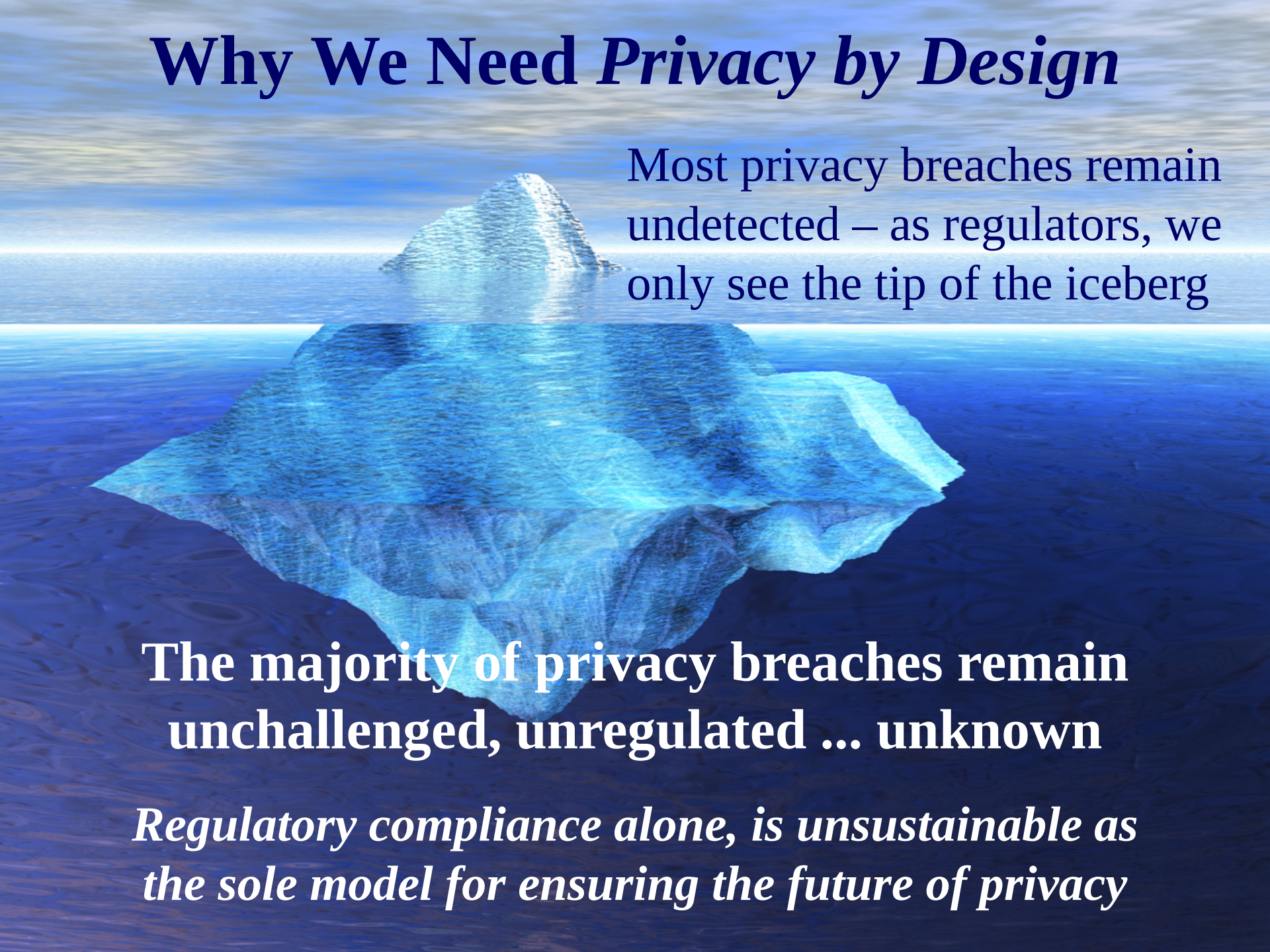
By Anna Ohlden – October 29th 2010 - http://www.science20.com/newswire/landmark_resolution_passed_preserve_future_privacy

JERUSALEM, October 29, 2010 – A landmark Resolution by Ontario's Information and Privacy Commissioner, Dr. Ann Cavoukian, was approved by international Data Protection and Privacy Commissioners in Jerusalem today at their annual conference. The resolution recognizes Commissioner Cavoukian's concept of Privacy by Design - which ensures that privacy is embedded into new technologies and business practices, right from the outset - as an essential component of fundamental privacy protection.

Full Article:

http://www.science20.com/newswire/landmark_resolution_passed_preserve_future_privacy

Why We Need *Privacy by Design*



Most privacy breaches remain undetected – as regulators, we only see the tip of the iceberg

The majority of privacy breaches remain unchallenged, unregulated ... unknown

Regulatory compliance alone, is unsustainable as the sole model for ensuring the future of privacy

Privacy by Design: Proactive in 38 Languages!

1. *English*
2. *French*
3. *German*
4. *Spanish*
5. *Italian*
6. *Czech*
7. *Dutch*
8. *Estonian*
9. *Hebrew*
10. *Hindi*
11. *Chinese*
12. *Japanese*
13. *Arabic*
14. *Armenian*
15. *Ukrainian*
16. *Korean*
17. *Russian*
18. *Romanian*
19. *Portuguese*
20. *Maltese*
21. *Greek*
22. *Macedonian*
23. *Bulgarian*
24. *Croatian*
25. *Polish*
26. *Turkish*
27. *Malaysian*
28. *Indonesian*
29. *Danish*
30. *Hungarian*
31. *Norwegian*
32. *Serbian*
33. *Lithuanian*
34. *Farsi*
35. *Finnish*
36. *Albanian*
37. *Catalan*
38. *Georgian*

Positive-Sum Model: *The Power of "And"*

Change the paradigm
from a zero-sum to
a "positive-sum" model:
Create a win-win scenario,
not an either/or (vs.)
involving unnecessary trade-offs
and false dichotomies …

replace "vs." with "and"

Privacy by Design:
The 7 Foundational Principles

1. ***Proactive*** not ***Reactive***:
   Preventative, not Remedial;
2. Privacy as the ***Default*** setting;
3. Privacy ***Embedded*** into Design;
4. ***Full*** Functionality:
   Positive-Sum, not Zero-Sum;
5. End-to-End **Security**:
   Full Lifecycle Protection;
6. Visibility **and** Transparency:
   Keep it **Open**;
7. Respect for User Privacy:
   Keep it **User-Centric**.

PbD

www.privacybydesign.ca

Privacy by Design

The 7 Foundational Principles

Ann Cavoukian, Ph.D.
Information & Privacy Commissioner
Ontario, Canada

Privacy by Design is a concept I developed back in the 90's, to address the ever-growing and systemic effects of Information and Communication Technologies, and of large-scale networked data systems.

Privacy by Design advances the view that the future of privacy cannot be assured solely by compliance with regulatory frameworks; rather, privacy assurance must ideally become an organization's default mode of operation.

Initially, deploying Privacy-Enhancing Technologies (PETs) was seen as the solution. Today, we realize that a more substantial approach is required — extending the use of PETs to PETS *Plus* — taking a positive-sum (full functionality) approach, not zero-sum. That's the "*Plus*" in PETS *Plus*: positive-sum, not the either/or of zero-sum (a false dichotomy).

Privacy by Design extends to a "Trilogy" of encompassing applications: 1) IT systems; 2) accountable business practices; and 3) physical design and networked infrastructure.

Principles of *Privacy by Design* may be applied to all types of personal information, but should be applied with special vigour to sensitive data such as medical information and financial data. The strength of privacy measures tends to be commensurate with the sensitivity of the data.

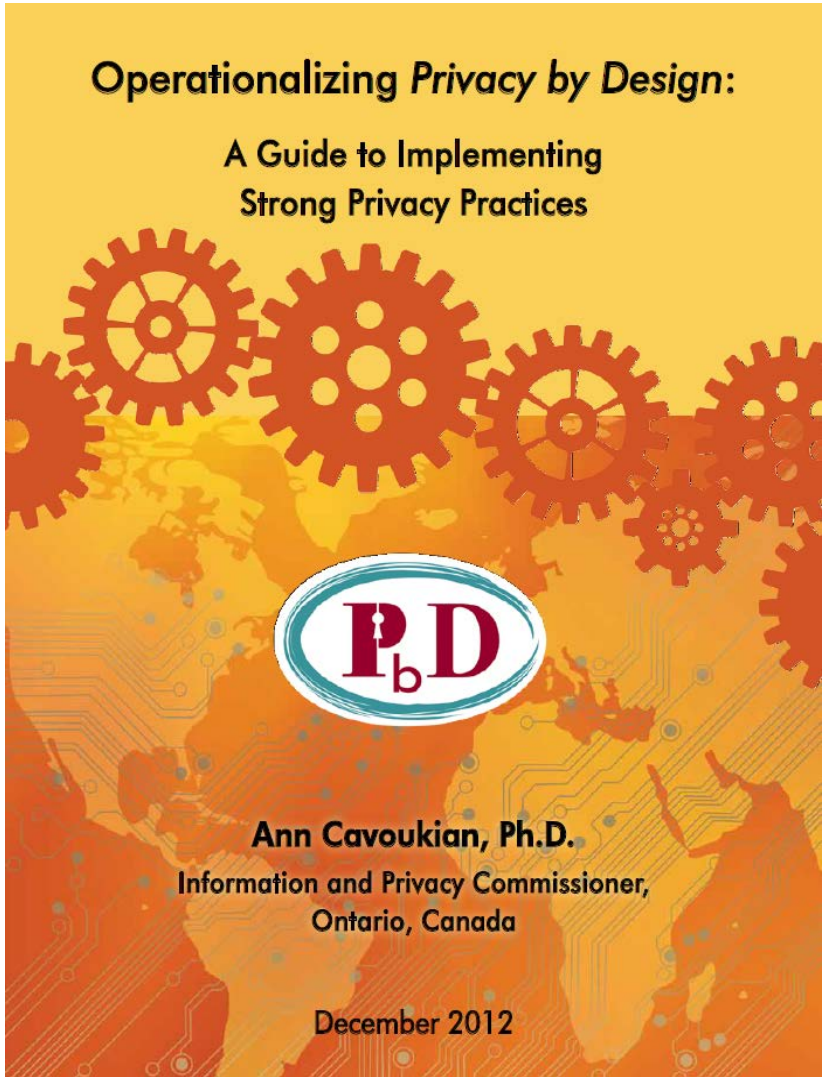
The objectives of *Privacy by Design* — ensuring privacy and gaining personal control over one's information and, for organizations, gaining a sustainable competitive advantage — may be accomplished by practicing the following 7 Foundational Principles *(see over page)*:

www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf

Operationalizing *Privacy by Design*

9 PbD Application Areas

- CCTV/Surveillance cameras in mass transit systems;
- Biometrics used in casinos and gaming facilities;
- Smart Meters and the Smart Grid;
- Mobile Communications;
- Near Field Communications;
- RFIDs and sensor technologies;
- Redesigning IP Geolocation;
- Remote Home Health Care;
- Big Data and Data Analytics.



Privacy by Design in the Age of Big Data

June 8, 2012

Ann Cavoukian, Ph.D.
Information & Privacy Commissioner
Ontario, Canada

Jeff Jonas
IBM Fellow
Chief Scientist, IBM Entity Analytics

Big Privacy:

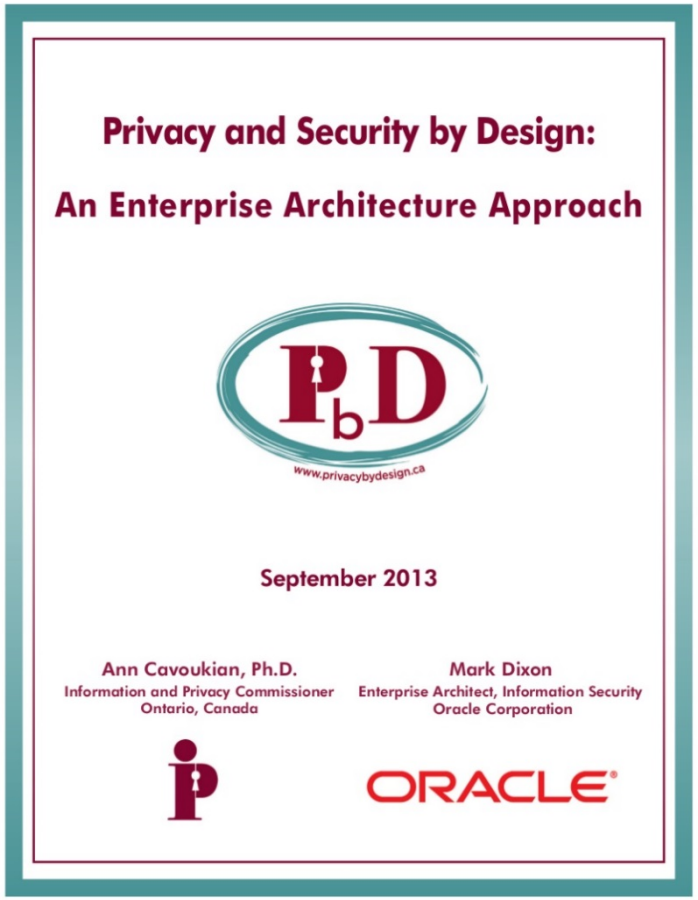
Bridging Big Data and the Personal Data Ecosystem Through *Privacy by Design*

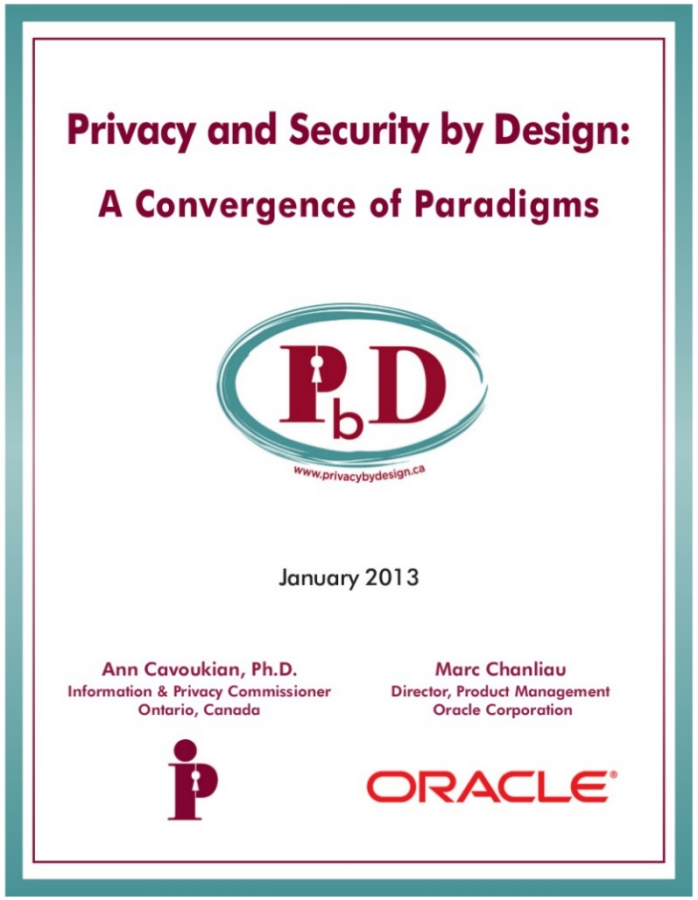
December 2013

Ann Cavoukian, Ph.D.
Information and Privacy Commissioner
Ontario, Canada

Drummond Reed
Co-Founder and CEO
Respect Network

Privacy and Security by Design



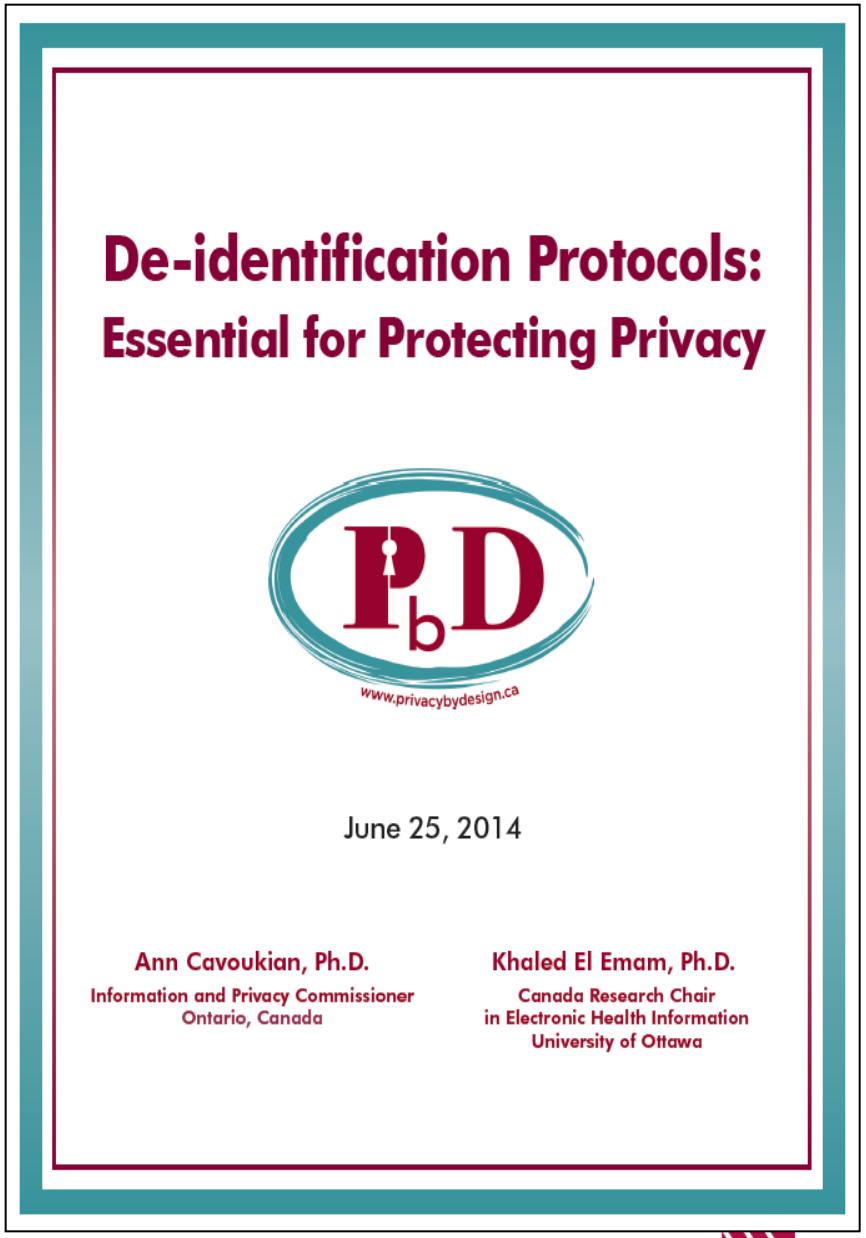


"Privacy is just as Big as Big Data. The tools exist to systemically protect personal information and bring about the benefits of Big Data. Together we can ensure that Big Data and 'Big Privacy' can both be accomplished to enable win-win scenario."

— Commissioner Cavoukian

"There are considerable risks in abandoning de-identification efforts, including the fact that individuals and organizations may simply cease disclosing de-identified information for secondary purposes, even those seen to be in the public interest."

— Commissioner Cavoukian



The Argument that Privacy Stifles Big Data Innovation Reflects a Dated, Zero-Sum Mindset

1. **Big Data and privacy are *not* mutually exclusive:**
   - Data is one of the most valuable assets of any organization;
   - Privacy is about *personal* information;
   - Consumer demands are creating additional pressures;
2. **Proactive privacy drives innovation:**
   - It is entirely possible to achieve privacy in the Big Data era, while also using data analytics to unlock new insights and innovations to move an organization forward;
3. **Innovation and privacy: You *can* have it all:**
   - Organizations will continue to apply data analytics to Big Data in order to advance their strategic goals and better serve their customers.

— Commissioner Cavoukian,
Using Privacy by Design to achieve Big Data Innovation Without Compromising Privacy

Concluding Thoughts

- Privacy risks are best managed by proactively embedding the principles of *Privacy by Design* – prevent the harm from arising – avoid the data breach;
- Focus on prevention: It is much easier and far more cost-effective to build in privacy, up-front, rather than after-the-fact;
- Abandon zero-sum thinking – embrace doubly-enabling systems: Big Data ***and*** Big Privacy: Yes, we can;
- Get smart – lead with *Privacy – by Design*, not privacy by chance or, worse, *Privacy by Disaster!*

Melissa Krasnow, CIPP/US

Partner
Dorsey & Whitney

Phone: (612) 492-6106
Cell: (312) 350-1082
krasnow.melissa@dorsey.com

twitter.com/melissakrasnow

4 STEPS IN THE LIFE CYCLE OF BIG DATA

- Collection
- Compilation and consolidation
- Data mining and analytics
- Use

U.S. Federal Trade Commission (FTC) Data Brokers Report addresses collection, compilation and consolidation and data mining and analytics and Big Data Report addresses use

TRANSPARENCY AND CHOICE

"Nomi Technologies, offered consumers two ways to exercise choices not to be tracked, but failed to provide one of those choices. In another context, we took action against a popular flashlight app for failing to disclose to consumers that it was providing geolocation information to ad networks. The message from these cases is that, regardless of where you are in the data chain, you have responsibilities with respect to transparency and choice."

Protecting Privacy in the Era of Big Data, Remarks of FTC Chairwoman Edith Ramirez at https://www.ftc.gov/system/files/documents/public_statements/671661/150610era_bigdata.pdf

PRIVACY BY DESIGN AND DATA MINIMIZATION

"….practice privacy by design, which includes considering privacy issues at every stage of product development.

As part of privacy by design….strive to assess….collection practices and, to the extent practical, collect only the data….need[ed] and properly dispose of the data as it becomes less useful."

FTC Data Brokers Report at https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

PRIVACY BY DESIGN

"Such an exercise is integral to a privacy-by-design approach and helps ensure that the company has given thought to its data collection practices on the front end by asking questions such as what types of data it is collecting, to what end, and how long it should be stored."

FTC Internet of Things Report at https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

SECURITY BY DESIGN

"…. companies should build security into their devices at the outset, rather than as an afterthought. As part of the security by design process, companies should consider: (1) conducting a privacy or security risk assessment; (2) minimizing the data they collect and retain; and (3) testing their security measures before launching their products."

FTC Internet of Things Report at https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

ACCESS CONTROL

"Correctly implementing such (and other) advanced access control policies requires a very good understanding of:

- Today's increasingly complex security policy requirements and how they impact technical access control implementation
- The impact of more and more complex IT environments, such as cloud, IoT etc. on access policy
- The available advanced access control approaches with their benefits and (complexity) challenges
- Approaches and processes to manage advanced access policies despite the complexity and dynamicity
- Understanding of which advanced access controls are most suitable for which use case (e.g. enterprise, big data, cloud, IoT)"

Source: http://www.tripwire.com/state-of-security/featured/access-control-in-2016-what-you-need-to-know/

ANONYMIZATION AND DE-IDENTIFICATION

What does anonymization mean?

"….there is comparatively little known about the underlying science of de-identification….Given the growing interest in de-identification, there is a clear need for standards and assessment techniques that can measurably address the breadth of data and risks…."

NIST De-Identification of Personal Information at http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf

DE-IDENTIFIED AND ANONYMOUS DATA

"When a company states that it maintains de-identified or anonymous data, the [FTC] has stated that companies should (1) take reasonable steps to de-identify the data, including by keeping up with technological developments; (2) publicly commit not to re-identify the data; and (3) have enforceable contracts in place with any third parties with whom they share the data, requiring the third parties to commit not to re-identify the data. This approach ensures that if the data is not reasonably de-identified and then is re-identified in the future, regulators can hold the company responsible."

FTC Internet of Things Report at https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

FTC BIG DATA REPORT: QUESTIONS FOR LEGAL COMPLIANCE

1. If you compile big data for others who will use it for eligibility decisions (credit, employment, insurance, housing, government benefits, etc.), are you complying with the accuracy and privacy provisions of the Fair Credit Reporting Act ("FCRA")?
2. If you receive big data products from another entity that you will use for eligibility decisions, are you complying with the provisions applicable to users of consumer reports?
3. If you are a creditor using big data analytics in a credit transaction, are you complying with the requirement to provide statements of specific reasons for adverse action under Equal Credit Opportunity Act ("ECOA")? Are you complying with ECOA requirements related to requests for information and record retention?

FTC BIG DATA REPORT: QUESTIONS FOR LEGAL COMPLIANCE (CON'T)

4. If you use big data analytics in a way that might adversely affect people in their ability to obtain credit, housing, or employment: are you treating people differently based on a prohibited basis, such as race or national origin?
5. Do your policies, practices, or decisions have an adverse effect or impact on a member of a protected class, and if they do, are they justified by a legitimate business need that cannot reasonably be achieved by means that are less disparate in their impact?

FTC BIG DATA REPORT: QUESTIONS FOR LEGAL COMPLIANCE (CON'T)

6. Are you honoring promises you make to consumers and providing consumers material information about your data practices?
7. Are you maintaining reasonable security over consumer data?
8. Are you undertaking reasonable measures to know the purposes for which your customers are using your data?

FTC Big Data Report at https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

FTC BIG DATA REPORT, SECTION 5 OF THE FEDERAL TRADE COMMISSION ACT AND BIG DATA ANALYTICS

"Section 5 of the Federal Trade Commission Act….prohibits unfair or deceptive acts or practices….Companies engaging in big data analytics should consider whether they are violating any material promises to consumers—whether that promise is to refrain from sharing data with third parties, to provide consumers with choices about sharing, or to safeguard consumers' personal information—or whether they have failed to disclose material information to consumers. In addition, companies that maintain big data on consumers should take care to reasonably secure consumers' data. Further, at a minimum, companies must not sell their big data analytics products to customers if they know or have reason to know that those customers will use the products for fraudulent or discriminatory purposes. The inquiry will be fact-specific, and in every case, the test will be whether the company is offering or using big data analytics in a deceptive or unfair way."

FTC Big Data Report at
https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

FTC BIG DATA REPORT: QUESTIONS TO ASK REGARDING DISCRIMINATORY HARMS

To maximize benefits and limit discriminatory harms, companies should consider the following questions:

1. How representative is your data set?
2. Does your data model account for biases?
3. How accurate are your predictions based on big data?
4. Does your reliance on big data raise ethical or fairness concerns?

FTC Big Data Report at https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf

“EQUAL OPPORTUNITY BY DESIGN”

“Using the principle of ‘equal opportunity by design’ and grounding engineering with sound ethical and professional best practices will also help mitigate discriminatory results over time and increase inclusion.

Just as in other areas, programmers and data scientists may inadvertently or unconsciously design, train, or deploy big data systems with biases. Therefore, an important factor in implementing the ‘equal opportunity by design’ principle is engaging with the field of ‘bias mitigation’ to avoid building in the designers’ biases that are an inevitable product of their own culture and experiences in life….

‘Data fundamentalism’—the belief that numbers cannot lie and always represent objective truth—can present serious and obfuscated bias problems that negatively impact people’s lives."

Source: Executive Office of the President’s Big Data Report on Algorithmic Systems, Opportunity, and Civil Rights at https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

RECENT DEVELOPMENTS

U.S. Federal Communications Commission seeks comment on a proposed framework for ensuring that consumers have the tools they need to make informed choices about how their data is used and when it is shared by their broadband providers

"We note that edge providers, data brokers, and other entities in the Internet ecosystem also collect, process, retain, and distribute large quantities of sensitive consumer data. Should we consider the restrictions, or lack thereof, that are currently placed on edge providers or other entities in crafting rules for broadband providers?"

Source: Federal Communications Commission Notice of Proposed Rulemaking at http://transition.fcc.gov/Daily_Releases/Daily_Business/2016/db0401/FCC-16-39A1.pdf