

Best Practices for Protecting Privacy While Conducting Big Data Analytics

David Weinkauf, Ph.D.

Senior Policy and Technology Advisor



Information and Privacy
Commissioner of Ontario

Commissaire à l'information et à la
protection de la vie privée de l'Ontario

6th National Summit
on Data Analytics for
Health Care

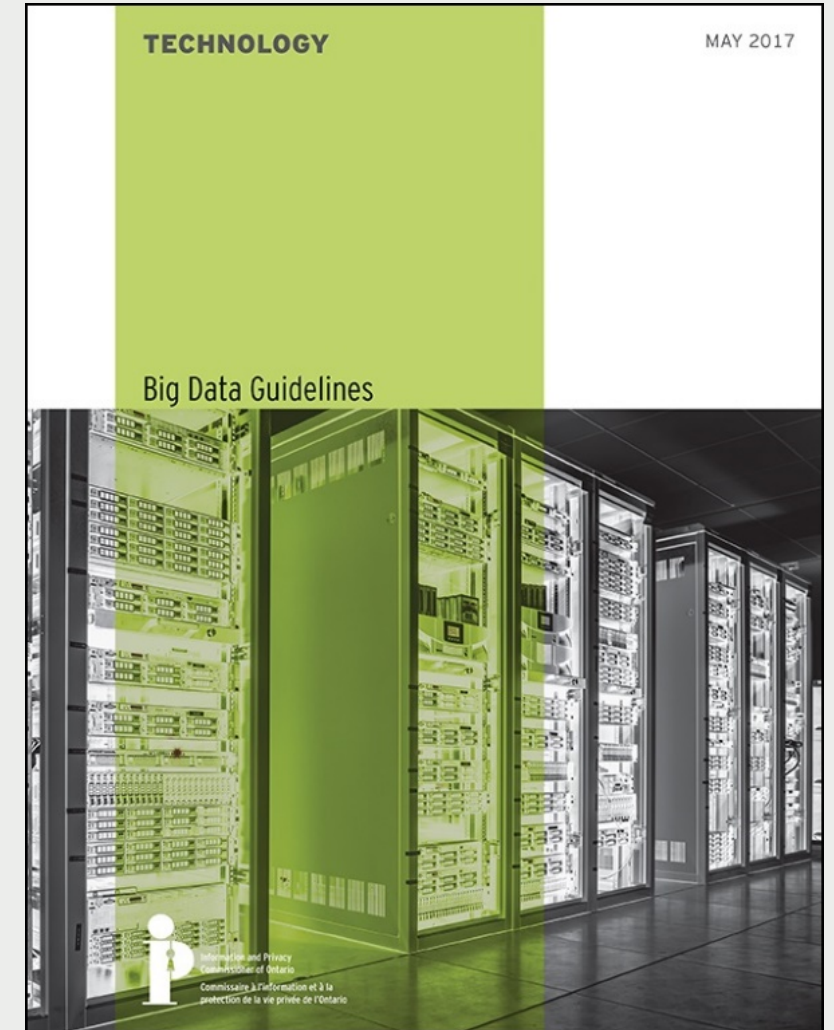
December 4, 2017

Outline

- IPC's "Big Data Guidelines"
- Big data and data minimization
- Big data and discriminatory proxies
- Increased role of REBs
- De-identification
- Questions

IPC's "Big Data Guidelines"

- Released in May 2017
- Designed to inform government institutions of key issues, best practices when conducting big data projects involving personal information (PI)
- Applicable to personal health information (PHI) / health sector
- Divides big data into **four stages**: collection, integration, analysis, profiling
- Each stage raises a number of concerns (14 total)
- Institutions should avoid uses of PI that may be **unexpected, invasive, inaccurate, discriminatory or disrespectful** of individuals



Big Data and Data Minimization

- **Inherent tension** between big data and principle of data minimization
- Big data does not start out with a preconceived rule or hypothesis
- What is now known as “data mining” was originally called “**data fishing**”
- Analyze data first and ask “why” later
- Key question:
 - If the rule or hypothesis to be derived is not known in advance of analyzing the information, how can you select a minimal set of data elements to support or prove it?

Big Data and Discriminatory Proxies

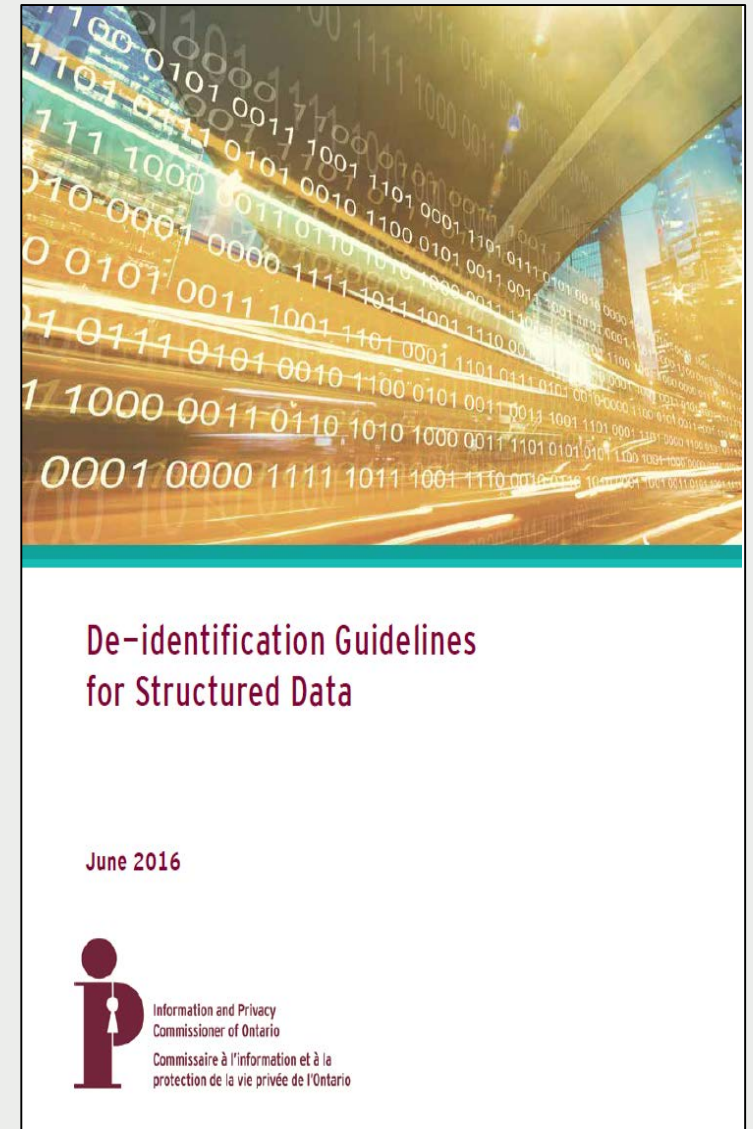
- **Diversity of sources and types** of PHI can create challenges to ensuring there is no unfair treatment or consideration of individuals based on protected attributes
- Analyzed information can contain variables that are not themselves protected but **correlate with protected attributes**
- Key questions:
 - Can a variable correlate with a protected attribute and yet provide non-discriminatory information that is useful to the analysis?
 - What is the role of the variable in the overall analysis?
 - What is the logic of the predictive model?
 - What will the analysis / model be used for?

Increased Role of REBs

- REBs should attempt to achieve “most favourable balance of risks and potential benefits” (TCPS-2)
- In big data projects, this may include **additional considerations** such as:
 - Whether PHI to be collected is reasonably limited
 - Whether outcome of analysis / model result in prohibited discrimination
 - Whether certain data linkages are inappropriate
 - Whether certain types of decisions should be automated
- Big data protocol for REBs is needed

De-Identification

- Although it does not limit the scope of data elements, de-identification **reduces the identifiability of information** to be analyzed
- Protects against theft, loss, unauthorized use and disclosure of PHI
- IPC released “De-identification Guidelines for Structured Data” in June 2016
- Recently won 2017 ICDPPC Award for “Excellence in Research”



HOW TO CONTACT US

Information and Privacy Commissioner of Ontario

2 Bloor Street East, Suite 1400

Toronto, Ontario, Canada M4W 1A8

Phone: (416) 326-3333 / 1-800-387-0073

TDD/TTY: 416-325-7539

Web: www.ipc.on.ca

E-mail: info@ipc.on.ca

Media: media@ipc.on.ca / 416-326-3965