Data Integration and Big Data In Ontario

Brian Beamish

Information and Privacy Commissioner of Ontario

Access, Privacy and Records and Information Management (RIM) Symposium

October 17, 2016

Our Office

- The Information and Privacy Commissioner (IPC) provides an independent review of government decisions and practices concerning access and privacy
- The Commissioner is appointed by and reports to the Legislative Assembly
- The Commissioner remains independent of the government of the day to ensure impartiality

The Three *Acts*

- The IPC oversees compliance with:
  - *Freedom of Information and Protection of Privacy Act (**FIPPA**)*
  - *Municipal Freedom of Information and Protection of Privacy Act (**MFIPPA**)*
  - *Personal Health Information Protection Act (**PHIPA**)*

Privacy Obligations under *FIPPA*

Collection, use, disclosure rules

No **collection** unless

- authorized by statute
- used for law enforcement or
- necessary to lawfully authorized activity

Must have a legitimate reason for collecting personal information, such as requiring a birth certificate to issue a driver's license

No **use** unless

- purpose collected
- consistent purpose
- written consent

Cannot use information from the birth registry to send out birthday cards

No **disclosure** unless

- consent
- consistent purpose
- comply with legislation
- law enforcement
- health or safety
- compassionate reasons

Video capturing evidence of a crime can be shared with police, even if it contains personal information

Data Integration

- Sometimes known as data linking/linkage or data/computer matching
- Involves the computerized comparison of databases to allow linkages to be made of information
- Technology has changed the landscape
- Where the data integration involves PI, there is a requirement to comply with *FIPPA* and *MFIPPA*

Privacy Challenges of Data Integration

- PI should be collected directly from the individual
- With some exceptions, it should only be used and disclosed for the purpose for which it was collected or a consistent purpose
- The individual to whom the PI pertains has a right to notice of the collection
- The PI used by an institution should not be used unless it is accurate and up to date

The Historical Perspective

- Concerns about the privacy implications of data integration existed before *FIPPA* and *MFIPPA* were proclaimed in force
- 1980 Williams Commission Report on *Freedom of Information and Individual Privacy* stated:

> *"The prospect of greater integration of databases raises, in turn, a number of privacy issues…*
>
> *…it is feared that the use of such dossiers may constitute a form of data surveillance which might operate against the legitimate interests of the individual"*

Privacy and Computer Matching

- In a 1991 report to the Standing Committee on the Legislative Assembly, the IPC recognized the potential benefits of data integration, including:
  - Detection and deterrence of fraud, waste and abuse
  - Improved efficiency and effectiveness of programs
  - Support for evidence based decision-making
- However, IPC also recognized the fundamental tension between data integration and certain basic principles of privacy
- Recommended a task force be created to study appropriate mechanisms to control and monitor data integration within the Ontario government

Big Data Analytics

- Process of running algorithms on integrated data sets to uncover hidden patterns
- Use of these analytics may raise significant privacy and other ethical and fairness issues
- May be used to infer rules that allow for automated decision making (about individuals) and the prediction of future results
- Process works the same regardless of whether analyzed data sets are de-identified or not, although the patterns extracted may differ

Recent Initiatives

- Data integration initiatives differ from past ones
- Purpose is to support policy development, system planning, resource allocation and performance monitoring
- Goal is sharing information about individuals for the purposes of conducting research
- Although not tied to direct service delivery, research may inform future collection and use of PI
- Challenge is to ensure that adequate measures are in place to protect the individuals whose PI is collected, used and disclosed while enabling the initiatives

Privacy Risks of Big Data

- Generation of new PI not collected directly from the individual
- Use of poorly selected data sets that:
  - lack information/are incomplete
  - contain incorrect or outdated information
  - disproportionately represent certain populations
- Incorporation of implicit or explicit biases
- Generation of pseudo-scientific insights that assume correlation equals causation
- Lack of knowledge/transparency regarding the inner “logic” of the system
- *If not designed properly, can result in uses of PI that may be unexpected, invasive and discriminatory*

Best Practices

- Legislative authority to collect, use and disclose PI within and among institutions
- Independent review process to govern projects including PIAs, TRAs and research ethics
- Transparency of approved projects
- Secure process for linking PI
- Requirement to de-identify PI after linking
- Delete the linked data once the research is complete

Additional Safeguards

- Prohibit the use of sensitive categories of PI
- Verify or ensure the accuracy and non-bias of the results in an independent manner
- Provide notice to affected individuals
- Allow affected individuals to challenge or respond to the results

Governance and Oversight

- Accountability frameworks for data integration and big data analytics should involve senior staff with authority to monitor and provide effective oversight
- Projects should engage experts in human rights, research ethics, privacy and de-identification

Digital and Big Data Literacy

- Develop clear policies setting out:
  - The administrative, technical and physical safeguards in place to secure the data
  - The nature of the privacy, human rights and research ethics review to be conducted on projects:
- Ensure that staff receive training on the policies and systems and permissible collections, uses and disclosure of PI

Reform of *FIPPA* and *MFIPPA*

- Need principled based legislation governing data linking and big data analytics which could include the following safeguards:
  - Creation of a data institute or institutes with expertise in privacy, human rights and ethical issues involved with data integration and analytics
  - Requirements for data minimization
  - Privacy impact assessments and threat risk assessments
  - Mandatory breach notification and reporting to the IPC and the affected individuals
  - Order-making and audit powers for the IPC

PHIPA Offers Model

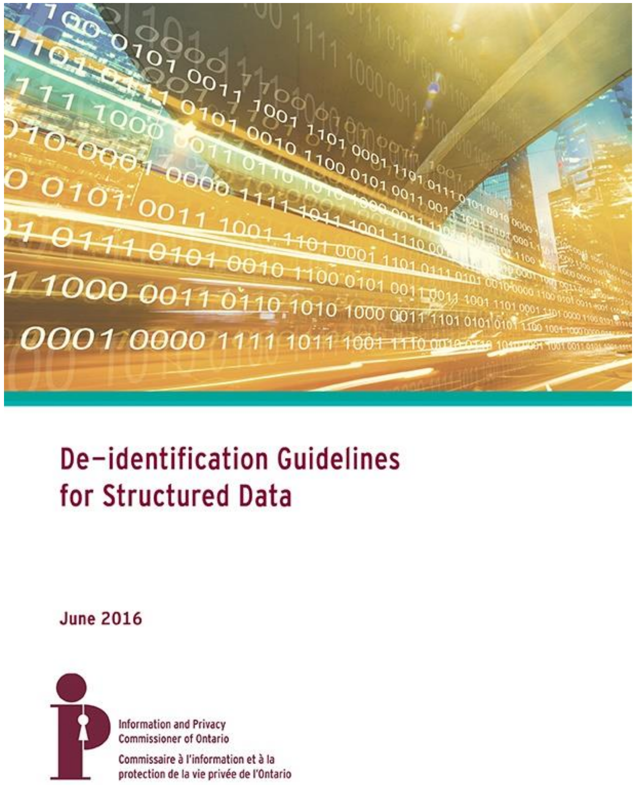
- *Personal Health Information Protection Act* (Section 47) can serve as a model for achieving some of these goals
- Disclosure for analysis of health system
- Section 47 sets out important requirements for data sharing:
  - Creation of health data institute
  - Strong oversight by IPC
  - De-identification
  - Secure policies and procedures

Summary

- Ensure you have the authority to collect and disclose
- De-identification protects against the disclosure of individuals’ identities, but not against other “big data” harms
- Be aware of “data fundamentalism”—i.e., the belief that correlation always implies causation and numbers always represent objective truth
- Individuals affected by automated decision-making have important rights

De-identification

- “De-identification” - the removal of PI from a record or data set
- Outlines a risk-based, step-by-step process to assist institutions in de-identifying data sets containing PI



De–identification Guidelines
for Structured Data

June 2016

Information and Privacy
Commissioner of Ontario

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

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