

Introduction to Data Sharing Rules

Brian Beamish, Commissioner
Renee Barrette, Director of Policy



Information and Privacy
Commissioner of Ontario

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

Peterborough
DataSHARE

September 10,
2019

Our Office

Information and Privacy Commissioner (IPC) provides **independent** review of government decisions and practices on access and privacy

Commissioner is appointed by, and reports to the Legislative Assembly, to ensure **impartiality**

Our Mandate

- *resolve* access to information appeals
- *investigate* privacy complaints – public sector and health
- *research* access and privacy issues
- *comment* on proposed government legislation and programs
- *educate* the public and government on issues of access and privacy

The Legislation

Freedom of Information and Protection of Privacy Act (FIPPA)

- covers 300 provincial institutions including universities and colleges

Municipal Freedom of Information and Protection of Privacy Act (MFIPPA)

- covers 1,200 municipal organizations

Personal Health Information Protection Act (PHIPA)

- covers individuals and organizations involved in the delivery of health care services

New Mandates

Child, Youth and Family Services Act

- effective January 1, 2020
- big step forward for Ontario's child and youth sectors
 - closes a legislative gap for access and privacy
 - promotes transparency and accountability

Anti-Racism Act

- passed June 2017
- requires public organizations in child welfare, education and justice sectors to collect information about Indigenous identity, race, religion and ethnic origin
- includes requirements to protect collected information, de-identify the personal information and publicly report on aggregated information



Ontario's Privacy Laws

Ontario's Privacy Laws

Organizations must:

- follow rules on collection, use, retention, disclosure and disposal of PI
- collect, use or disclose information only for legitimate, limited and specific purposes
- inform individuals how they intend to use and disclose their PI
- implement reasonable measures to ensure security of information

Individuals have the right to file privacy complaints with the IPC



What is PI and PHI?

PI is identifying information about an individual in **recorded form**, such as:

- name, address, sex, age, education, and medical or employment history
- social insurance number

Business information is **not** personal information

PHI is identifying information about an individual in **oral or recorded form** such as information that:

- relates to their physical or mental health
- relates to the provision of health care
- relates to payments or eligibility for health care
- is a health number

Disclosure Rules under *M/FIPPA*

Disclosure of PI among institutions is only **allowed in limited circumstances** such as:

- for the purpose for which it was collected or a **consistent purpose**
- compelling circumstances involving **health or safety**
- to an **officer or employee or agent** of the institution who needs the record in the performance of their duties
- to comply with a **law**
- with **consent**

Limits on disclosure create silos of PI

The Historical Perspective

Concerns about **privacy implications of data integration** existed before *M/FIPPA* 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 Risks of Disclosure for Data Integration

- lack of transparency
- surveillance and **profiling** and inappropriate access of PI
- potential for **discrimination** based on inaccurate data and flawed algorithms
- cybersecurity
- function creep or **unexpected/inconsistent uses** of PI
- replication of massive government databases of linked and identifiable PI

PHIPA – use and disclosure for planning and analysis

MOH can **collect and link PHI** from health care providers for:

- funding, planning or allocating resources
- detecting, monitoring or preventing fraud

Controls to protect privacy include:

- prescribed unit to perform data collection and integration
- de-identification
- IPC review of unit's practices and procedures

Health care providers can also disclose to **prescribed entities** (e.g., CIHI, ICES) for planning and analysis, if the entity has **IPC approved practices** in place

IPC has **strong investigative/audit powers**

Research

PHIPA permits disclosures of PHI without consent for research purposes if the researcher:

- prepares a **research plan** (that meets certain requirements) and
- a **research ethics board** (that meets certain requirements) approves the plan

CYFSA allows MCCSS to collect PI from service providers for a research purpose without consent and it allows prescribed entities to disclose PI for a research purpose without consent if the researcher:

- prepares a **research plan** (that meets certain requirements) and
- a **research ethics board** (that meets certain requirements) approves the plan

Research

M/FIPPA permits disclosures of PI without consent for research purposes to a researcher if:

- disclosure is consistent with the **reasonable expectations**
- it is **necessary** to use identifiable information
- there is **an agreement** in place (that meets certain requirements)



New Data Sharing Rules

Government Organizations

Public sector organizations want to **share, link, and analyze** data to obtain new insights, to support

- policy development
- system planning
- resource allocation
- performance monitoring

Benefits may be compelling

- higher quality evidence
- better public policy
- better use of money
- fraud detection

M/FIPPA does not permit disclosure for these purposes

Amendments to *FIPPA* - Part III.1

Bill 100 received Royal Assent in May 2019 which brought amendments to *FIPPA* – not yet proclaimed

New Part III.1 of *FIPPA* allows **indirect collection of PI from other ministries and agencies** to enable analysis for the purposes of:

- managing and allocating resources
- planning for the delivery of programs and services provided by the Ontario governments
- evaluating those programs and services

Permits creation of **inter-ministerial and ministry data integration units**

Amendments to *FIPPA* - Part III.1

Other privacy controls include:

- **data minimization**

- PI cannot be collected, used or disclosed if other information will serve the specified purposes
- data integration units cannot collect, use or disclose more PI than is reasonably necessary for the specified purposes

- mandatory **breach reporting**

- IPC has strong **review and order making powers** in relation to Part III.1

Amendments to *FIPPA* - Part III.1

Indirect collection only permitted if certain conditions are met such as:

- PI must be **collected from an institution** as defined by *M/FIPPA* or a prescribed person or entity
- **notice of collection** must be published online and meet certain other requirements
- the minister has determined that there is a **public interest** in collecting the PI

Data standards applicable to all units must be approved by the IPC and in place before any collection can take place

Specific requirements regarding **linking and de-identification** of PI

Restrictions on Personal Health Information

- Only ministry data integration unit **located in Ministry of Health** may collect PHI from health information custodians (HICs)
- Inter-ministerial data integration unit may not collect PHI from HICs unless authorized in regulations
- No collection of **counselling session notes**



De-Identification

Introduction

De-identification is:

- removing information that identifies an individual
- removing information that could be used either alone, or with other information to identify an individual based on what is reasonably foreseeable in the circumstances

In general, privacy laws do not apply to de-identified information

Important to consider:

- cannot reduce the risk of re-identification to zero
- de-identified information can still be stigmatizing and its use and disclosure can impact groups of individuals

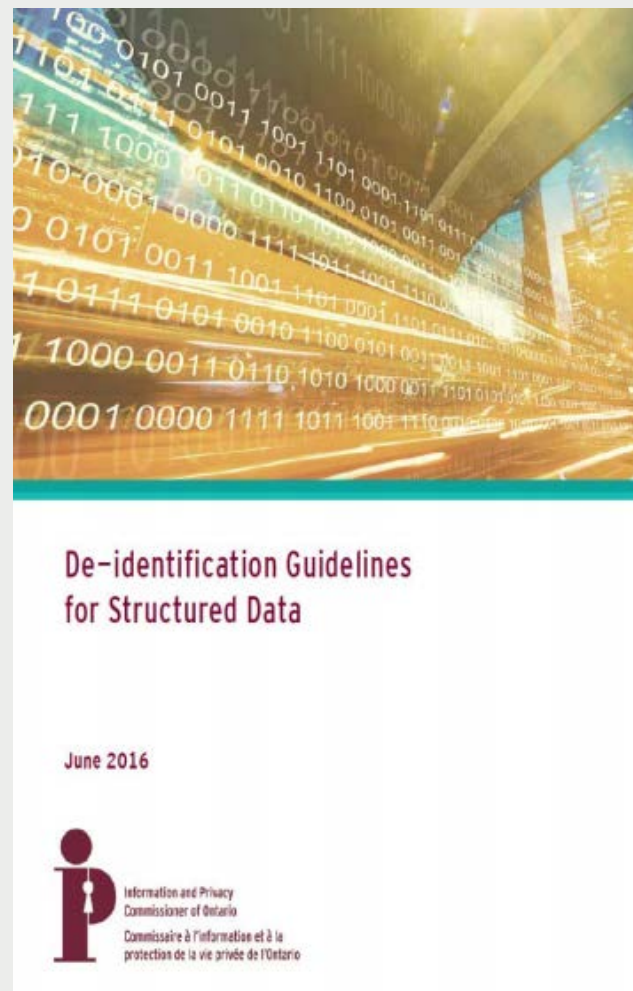
Therefore, organizations must **manage re-identification risk** by taking into account the context for disclosure and proposed use. This can be a complex process

Introduction

- The higher the re-identification risk, the greater the amount of de-identification required; amount of de-identification is **proportional to the level of risk**
- Proportionality helps to preserve **data utility** while protecting individual privacy
- **Governance** is an important aspect of sharing de-identified data sets for example:
 - ongoing risk assessments
 - auditing data recipients
 - transparency
 - accountability
 - training

De-Identification – Guidelines for Structured Data

- IPC released “De-Identification Guidelines for Structured Data” in June 2016
- Risk based approach to de-identification
- **Step-by-step** process
- Process requires consideration of:
 - ✓ release models
 - ✓ types of identifiers
 - ✓ re-identification attacks
 - ✓ de-identification techniques



Scope of IPC's De-Identification Guidelines

- Goal was to produce a **“plain language” guide** to de-identification with straightforward use cases and calculations.
- To enable this, Guidelines were designed to:
 - **err on the conservative side** when it comes to calculating levels of re-identification risk.
 - discuss techniques of **masking, generalization and suppression** only
- Techniques not directly applicable to “high dimensional” datasets (big data)
- In sum, not the full story, but a practical compilation of **core principles and ideas**

Key Distinction: Direct Identifiers and Quasi-identifiers

The removal of **direct identifiers**, such as a name, address, and social insurance number, may not be sufficient to manage re-identification risk

Indirect or “**quasi-identifiers**” can be used, either individually or in combination, to re-identify an individual by being linked to other identifiable information. Examples include:

- o gender
- o marital status
- o location
- o diagnosis
- o profession
- o ethnic origin
- o race

De-identification – Best Practices

Techniques:

- remove direct identifiers and quasi-identifiers
- aggregate quasi identifiers such as by changing variables into ranges, (e.g. birthdates can be modified to age ranges)
- removing outliers

IPC recommends that you develop de-identification process in **consultation with an expert** in de-identification



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

Data Sharing Agreements

- DSAs help mitigate the risk of re-identification for non-public releases (not open data)
- DSA should include provisions such as:
 - ✓ defining the **information to be disclosed** and the **purpose**
 - ✓ defining **limited staff who are permitted to access** and use data - “need-to-know”
 - ✓ **confidentiality undertakings** and privacy training for all staff, including external collaborators and subcontractors
 - ✓ destruction of data after a specified **retention period**
 - ✓ **limit disclosure of** data with third parties except where have prior approval
 - ✓ **privacy and security policies are** in place, monitored and enforced
 - ✓ privacy breach protocol is in place
 - ✓ detailed **logging and monitoring systems** implemented
 - ✓ **encrypted protocol** used to electronically transmit data

AOL Search Data


- In 2006, AOL publicly released 20 million Web search queries for over 450,000 users
- Search queries were left unmodified; the only precaution taken was to replace usernames with pseudonyms
- A New York Times reporter identified user No. “4417749”
- Search terms for user “4417749” included:
 - “60 single men”
 - “dog that urinates on everything”
 - “landscapers in Lilburn, Ga”
 - several people with the last name “Arnold”

Lessons Learned

- Ad-hoc approaches to de-identification are problematic:
 - Masking of direct identifiers is not enough; indirect identifiers must be addressed
 - Free-form text may contain identifiable information
 - Rare attributes increase the risk of re-identification
 - Data sharing agreements are important
- Despite repeated claims, above case is not an example of broken de-identification
- Using an **effective de-identification protocol** is essential for managing the risks of disclosing information about individuals

Waterloo's Smart City Challenge Submission

- **Data-sharing platform** that aggregates previously siloed PI and PHI on children and youth to help understand factors contributing to their well-being
- Data sources: regional, provincial and national
- PI to be **de-identified** before sharing
- Waterloo committed to consulting with de-identification experts



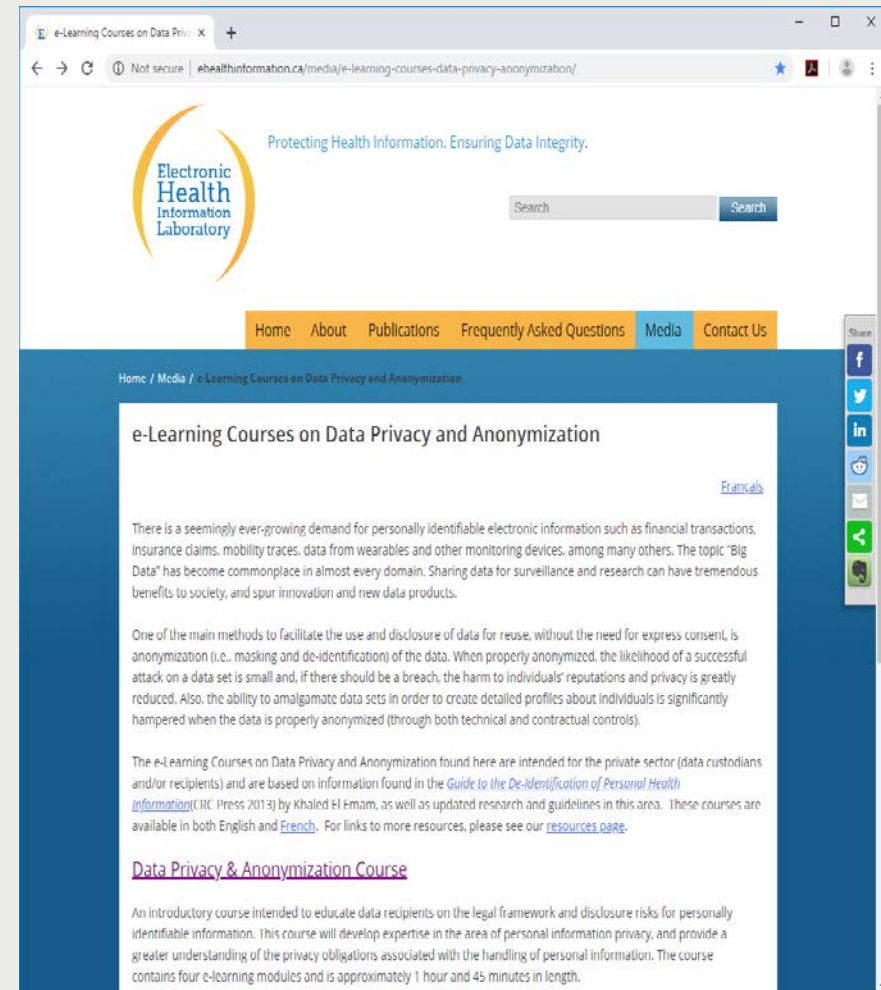
The graphic features a central white diamond shape on a dark blue background with a circuit board pattern. The diamond is divided into four quadrants: top-left (orange wavy pattern), top-right (purple t-shirt with 'Smart Waterloo Region' text), bottom-left (cyan portrait of a woman with glasses), and bottom-right (red photo of a group of people). In the center of the diamond is the Smart Waterloo Region logo, which consists of a 2x2 grid of squares: top-left (white 'S'), top-right (white 'W'), bottom-left (circuit board pattern), and bottom-right (white 'R'). To the right of the 'R' square is the text 'Smart Waterloo Region'.

**Smart Waterloo
Region Proposal**

📷 🐦 @SmartWatRegion
#smartwr #bestcommunity4kids
www.smartwr.ca

Online Resources on De-Identification

- The Electronic Health Information Laboratory (based in Ottawa) has developed **two e-learning courses** on de-identification
 - one for institutions and one for researchers
- Excellent resources covering the legal and regulatory framework around privacy, and how to properly de-identify data
- See: <http://www.ehealthinformation.ca/media/e-learning-courses-data-privacy-anonymization/>



The screenshot shows a web browser window displaying the Electronic Health Information Laboratory website. The page title is "e-Learning Courses on Data Privacy and Anonymization". The website header includes the logo for the Electronic Health Information Laboratory and the tagline "Protecting Health Information. Ensuring Data Integrity." Below the header is a navigation menu with links for Home, About, Publications, Frequently Asked Questions, Media, and Contact Us. The main content area features the title "e-Learning Courses on Data Privacy and Anonymization" and a "Français" link. The text describes the growing demand for personally identifiable electronic information and the importance of de-identification. It also mentions that the e-Learning Courses on Data Privacy and Anonymization are intended for the private sector and are based on information found in the "Guide to the De-Identification of Personal Health Information" (CIC Press 2013) by Khaled El-Fam, as well as updated research and guidelines in this area. The courses are available in both English and French. A link to "Data Privacy & Anonymization Course" is provided, along with a brief description of the course content.

CANON

Network of private, public and health care sector privacy specialists working to promote use of de-identification as a tool to support privacy

<https://deidentify.ca/>



Welcome

The Canadian Anonymization Network ("CANON") is an informal network, comprised of data custodians from across private, public and health sectors, whose primary purpose is to promote anonymization as a privacy-respectful means of leveraging data for innovative and beneficial purposes.

Co-founded by [AccessPrivacy](#) and [Privacy Analytics](#), CANON has quickly grown to include some of the largest data custodians from private, public and health institutions across the country, all with the common goal of promoting anonymization as a means of leveraging responsible use of data for economic and socially beneficial purposes.



Privacy Impact Assessment (PIA)

What are PIAs

- PIA refers to a process/approach for **identifying and analyzing privacy risks** when changing or developing programs or systems
- A good PIA analysis provides senior management and program and system designers with sufficient information to **reduce, mitigate or avoid different types of privacy risks**
- Help avoid need for re-design, delays and risk of project cancellation
- **Demonstrates due diligence** in the event of a privacy complaint

PIA Guide

- Intended for *FIPPA* and *MFIPPA* institutions
- Simplified **4-step methodology** with tools
- Evaluates privacy risks throughout PI's **lifecycle**
- Can be used as a basis for developing internal PIA policies and procedures



Planning for Success: Privacy Impact Assessment Guide



Information and Privacy
Commissioner of Ontario
Commissaire à l'information et à la
protection de la vie privée de l'Ontario



Guidance Materials

TECHNOLOGY

MAY 2017

Big Data Guidelines



De-identification Guidelines for Structured Data

June 2016



Planning for Success: Privacy Impact Assessment Guide



Our Open Door Policy

- Any public institution or agency considering programs which may impact privacy can approach IPC for advice
- Most privacy challenges can be addressed through collaboration
- Privacy protections can be developed and can be implemented
- It is best to address privacy concerns from the outset
- Success depends on involvement of other agencies and stakeholders

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