

WEBINAR**FEBRUARY 11, 2025 | 11 AM - 12 PM EDT**

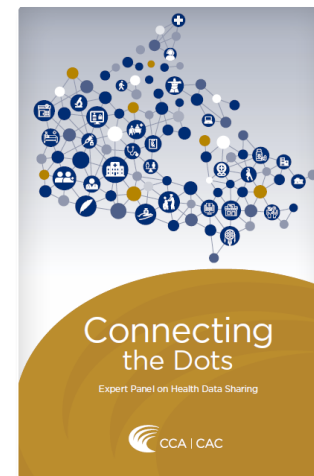
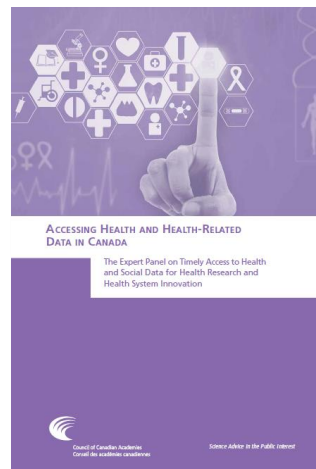
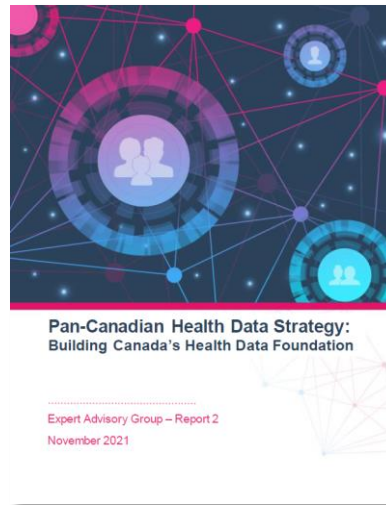
How Canadian Regulators Approach De-Identification: An Interview Study

PRESENTERS

DR. LISA PILGRAM, POSTDOCTORAL FELLOW, UNIVERSITY OF OTTAWA
AND
DR. KHALED EL EMAM, PROFESSOR, UNIVERSITY OF OTTAWA



Multiple reports on health data access



How Canadian Regulators Approach De-Identification: An Interview Study

Lisa Pilgram, MD

Postdoctoral Fellow at the Electronic Health Information Laboratory (Khaled El Emam)



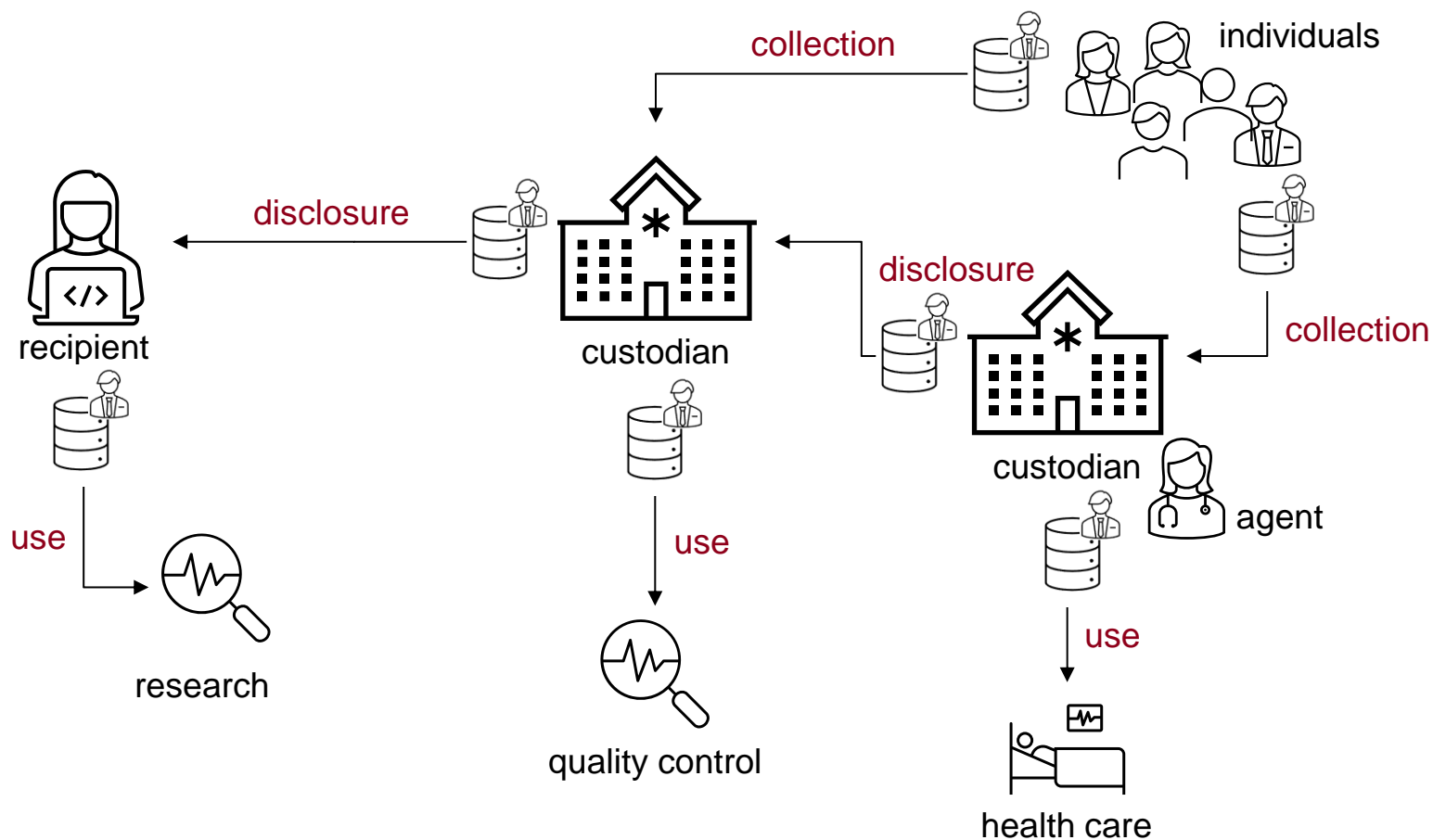
Agenda

1. The Need for De-Identification
 - Perceived and actual risk
 - Regulatory requirements
2. Methodology
3. Privacy Regulators' Perspectives on Regulating
 - De-Identification
 - De-Identified Data
 - AI/ML Models
4. Observations & Recommendations



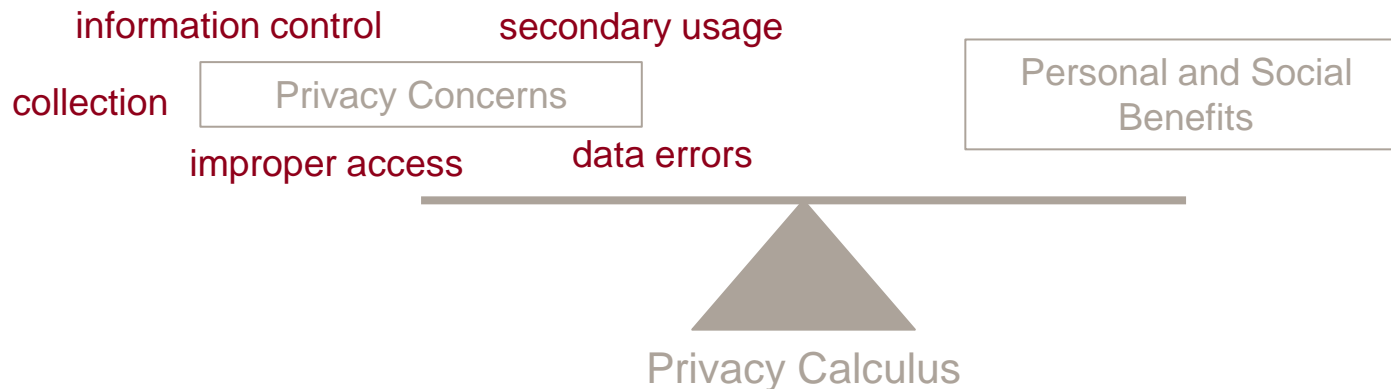
THE NEED FOR DE-IDENTIFICATION

The Life Cycle of Data: Example Health Information



Perceived Risk (i.e., Privacy Concerns) when Sharing Personal Health Information

- Individuals express privacy concerns when sharing personal health information. Concerns decrease with quality of care, technological familiarity, perceived control, self-efficacy and trust.
 - N. Shen et al., “Understanding the patient privacy perspective on health information exchange: A systematic review,” International Journal of Medical Informatics, vol. 125, pp. 1–12, May 2019, doi: 10.1016/j.ijmedinf.2019.01.014.
- Individuals can adopt privacy-protective behaviour.



Perceived Risk (i.e., Privacy Concerns) when Sharing Personal Health Information

- Individuals express privacy concerns when sharing personal health information. Concerns decrease with quality of care, technological familiarity, perceived control, self-efficacy and trust.
 - N. Shen et al., “Understanding the patient privacy perspective on health information exchange: A systematic review,” *International Journal of Medical Informatics*, vol. 125, pp. 1–12, May 2019, doi: 10.1016/j.ijmedinf.2019.01.014.
- Individuals can adopt privacy-protective behaviour. This can be the withholding of information from health professionals, reluctance to use beneficial but data intensive technologies or unwillingness to disclose information for research.
 - A. Kharlamov, R. Hohmann, and G. Parry, “Data sharing decisions: Perceptions and intentions in healthcare,” *Strategic Change*, vol. 32, no. 6, pp. 223–237, 2023, doi: 10.1002/jsc.2558
 - E.-M. Schomakers, C. Lidynia, and M. Ziefle, “Listen to My Heart? How Privacy Concerns Shape Users’ Acceptance of e-Health Technologies,” in *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct. 2019, pp. 306–311. doi: 10.1109/WiMOB.2019.8923448

Actual Risk when Sharing Personal Health Information

- Data breaches in the health care sector are often due to hacking or IT incidents.
 - “U.S. Department of Health & Human Services - Office for Civil Rights.” Accessed: Dec. 27, 2024. [Online]. Available: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf.
- The health care sector is increasingly a target for malicious attacks.
 - S. Alder, “Healthcare Ransomware Attacks Continue to Increase in Number and Severity,” The HIPAA Journal. Accessed: Nov. 08, 2024. [Online]. Available: <https://www.hipaajournal.com/healthcare-ransomware-attacks-2024/>.
- Historically, data has been misused with devastating consequences.
 - D. M. Luebke and S. Milton, “Locating the victim: An overview of census-taking, tabulation technology and persecution in Nazi Germany,” IEEE Annals of the History of Computing, vol. 16, no. 3, pp. 25-, 1994, doi: 10.1109/MAHC.1994.298418
 - M. Abella, “The United States’ and Japan’s Immigration Dilemmas in Comparative Perspective,” American Behavioral Scientist, vol. 56, no. 8, pp. 1139–1156, Aug. 2012, doi: 10.1177/0002764212441779
 - Anderson M, Seltzer W. Use and Misuse of the United States Census: The Role of Data in the Incarceration of Japanese Americans During World War II. Springer Nature Switzerland; 2023. doi:10.1007/978-3-031-38619-0

When and Why We De-Identify

- To enable data use and disclosure that fosters research, supports the development of innovations and solutions, and drives technological progress while
 - addressing privacy concerns and building trust
 - aligning with ethical principles
 - preventing potential harm from information disclosure
 - adhering to legal requirements (use and disclosure may not be permitted otherwise)

Use and Disclosure of Information

- The key question is identifiability.
 - Identifiable (or personal) information is typically regulated in respective privacy legislation.
 - Non-identifiable (or non-personal) information is typically outside the scope of privacy regulation and thereby subject to fewer obligations.
 - Non-identifiable (or non-personal) information may still be subject to other legislation.

Definitions of Identifiability are Vague

Health Insurance Portability and Accountability Act of 1996

"(6) INDIVIDUALLY IDENTIFIABLE HEALTH INFORMATION.--The term 'individually identifiable health information' means any information, including demographic information collected from an individual, that--

"(A) is created or received by a health care provider, health plan, employer, or health care clearinghouse; and

"(B) relates to the past, present, or future physical or mental health or condition of an individual, the provision of health care to an individual, or the past, present, or future payment for the provision of health care to an individual, and--

"(i) identifies the individual; or

"(ii) with respect to which there is a reasonable basis to believe that the information can be used to identify the individual.



Definitions of Identifiability are Vague

Health Insurance Portability and Accountability Act (HIPAA)

"(6) INDIVIDUALLY IDENTIFIABLE HEALTH INFORMATION.--The term 'individually identifiable health information' means any information, including demographic information, that--

"(A) is created or received by a health care provider, health plan, employer, or clearinghouse; and

"(B) relates to the past, present, or future physical or mental health, the provision of health care to an individual, or the past, present, or future health care to an individual, and--

"(i) identifies the individual; or

"(ii) with respect to which there is a reasonable basis to believe that the information can be used to identify the individual.



Art. 4 GDPR Definitions

For the purposes of this Regulation:

- (1) 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;



Definitions of Identifiability are Vague

Health Insurance Portability and Accountability Act

"(6) INDIVIDUALLY IDENTIFIABLE INFORMATION" means information that--

"(A) is created or received by a clearinghouse; and

"(B) relates to the provision of health care to an individual;

"(i) identifies the individual;

"(ii) with respect to the individual, identifies the individual.

Art. 4 GDPR

Definitions

Personal Information Protection and Electronic Documents Act

S.C. 2000, c. 5

Assented to 2000-04-13

personal health information, with respect to an individual, whether living or deceased, means

- (a) information concerning the physical or mental health of the individual;
 - (b) information concerning any health service provided to the individual;
 - (c) information concerning the donation by the individual of any body part or any bodily substance of the individual or information derived from the testing or examination of a body part or bodily substance of the individual;
 - (d) information that is collected in the course of providing health services to the individual; or
 - (e) information that is collected incidentally to the provision of health services to the individual.
- (renseignement personnel sur la santé)

natural person
directly or
indirectly
identification number,
biometric data,
physical,
mental, or
cultural person;



Definitions of Personal Information can Vary Within one Country

Personal Health Information Protection Act, 2004

[S.O. 2004, CHAPTER 3](#)

SCHEDULE A

Personal health information

4 (1) In this Act,

“personal health information”, subject to subsections (3) and (4), means identifying information about an individual in oral or recorded form, if the information,

- (a) relates to the physical or mental health of the individual, including information that consists of the health history of the individual's family,
- (b) relates to the providing of health care to the individual, including the identification of a person as a provider of health care to the individual,
- (c) REPEALED: 2020, c. 13, Sched. 3, s. 8 (7).
- (c.1) is a plan that sets out the home and community care services for the individual to be provided by a health service provider or Ontario Health Team pursuant to funding under section 21 of the *Connecting Care Act, 2019*,
- (d) relates to payments or eligibility for health care, or eligibility for coverage for health care, in respect of the individual,
- (e) relates to the donation by the individual of any body part or bodily substance of the individual or is derived from the testing or examination of any such body part or bodily substance,
- (f) is the individual's health number, or
- (g) identifies an individual's substitute decision-maker. 2004, c. 3, Sched. A, s. 4 (1); 2007, c. 8, s. 224 (6); 2007, c. 10, Sched. H, s. 2; 2020, c. 13, Sched. 3, s. 8 (7, 8).

Definitions of Personal Information can Vary Within one Country

Personal Health Information Protection Act, 2004

[S.O. 2004, CHAPTER 3](#)

SCHEDULE A

Personal health information

4(1) In this Act

Québec

© Québec Official Publisher

Updated to October 1 2024
This document has official status.

chapter P-39.1

ACT RESPECTING THE PROTECTION OF PERSONAL INFORMATION IN THE PRIVATE SECTOR

2. Personal information is any information which relates to a natural person and directly or indirectly allows that person to be identified.

1993, c. 17, s. 2; 2021, c. 25, s. 102.

(e) information
(renseignemen

(f) is the individual's health number, or

(g) identifies an individual's substitute decision-maker. 2004, c. 3, Sched. A, s. 4 (1); 2007, c. 8, s. 224 (6); 2007, c. 10, Sched. H, s. 2; 2020, c. 13, Sched. 3, s. 8 (7, 8).

Definitions of Personal Information can Vary Within one Country

Personal Health Information Protection Act, 2004

BILL NO. 89

(as passed, with amendments)



2nd Session, 61st General Assembly
Nova Scotia
59 Elizabeth II, 2010

(r) "personal health information" means identifying information about an individual, whether living or deceased, and in both recorded and unrecorded forms, if the information

- (i) relates to the physical or mental health of the individual, including information that consists of the health history of the individual's family,
- (ii) relates to the application, assessment, eligibility and provision of health care to the individual, including the identification of a person as a provider of health care to the individual,
- (iii) relates to payments or eligibility for health care in respect of the individual,
- (iv) relates to the donation by the individual of any body part or bodily substance of the individual or is derived from the testing or examination of any such body part or bodily substance,
- (v) is the individual's registration information, including the individual's health-card number, or
- (vi) identifies an individual's substitute decision-maker;



Personal health inform
4(1) In this Act

Québec

© Québec Official Publisher

chapter P-39.1

ACT RESPECTING THE PROTECTION OF PER SECTOR

2. Personal information is any information wh
indirectly allows that person to be identified.

1993, c. 17, s. 2; 2021, c. 25, s. 102.

(e) information
(renseignemen

(f) is the individual's hea

(g) identifies an individual's substitute decision-maker. 2004, c. 3, Sched. A, s. 4 (1); 2007, c. 8, s. 224 (6); 2007, c. 10, Sched. H, s. 2; 2020, c. 13, Sched. 3, s. 8 (7, 8).

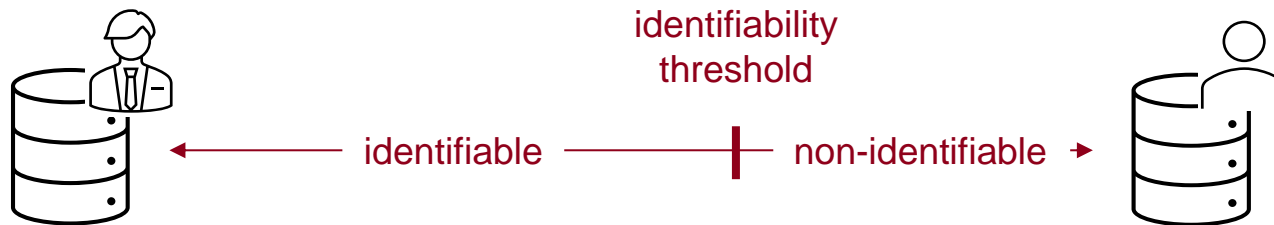


Non-identifiability Typically Allows for a (Very Small) Residual Risk

- Jurisdictions have different requirements to consider a dataset as non-identifiable.
 - precise: e.g. HIPAA Safe Harbor
 - contextual: e.g. GDPR
- Non-identifiability typically allows for a (very small) residual risk.
- G7 data protection and privacy authorities aim at a more consistent understanding.*

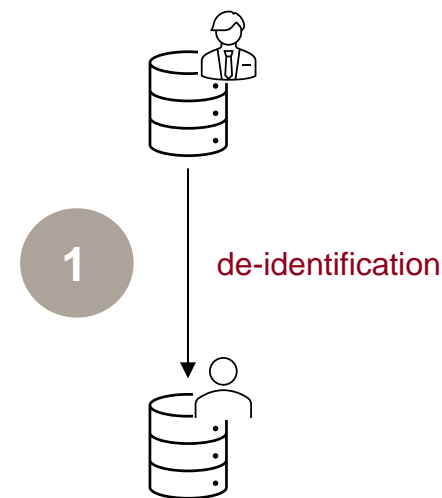
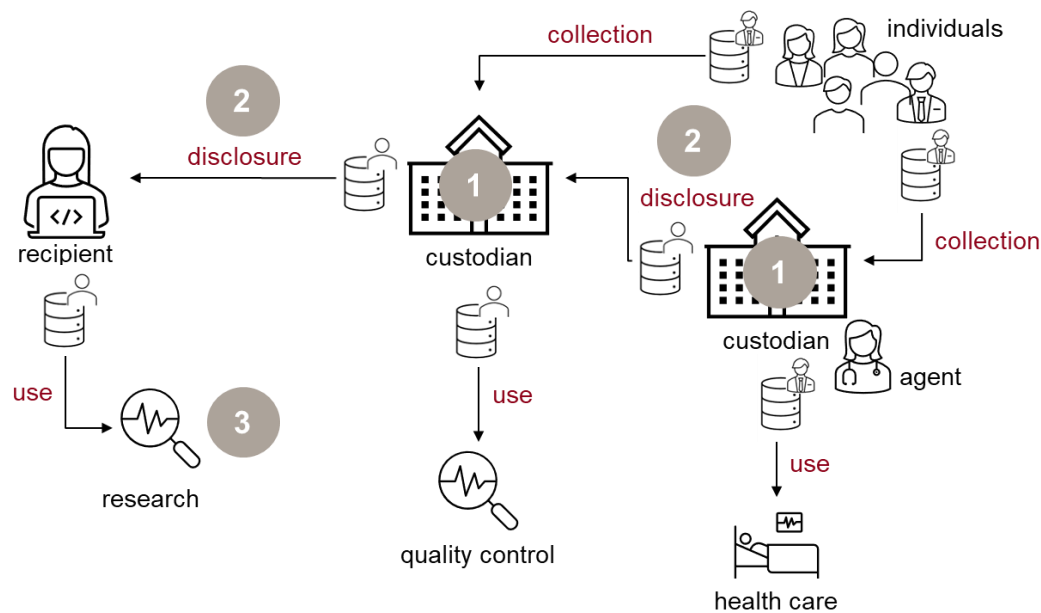
* see: https://www.priv.gc.ca/en/opc-news/news-and-announcements/2024/de-id_20241011/

De-Identification Can Result in Data that is Below an Identifiability Threshold



- Information can be anywhere along the identifiability spectrum.
- In this presentation
 - De-identification = process of producing de-identified data of an identifiability below a threshold.
 - In practice, de-identification is a non-binary process and can be of any degree.

The Life Cycle of De-Identified Data: Example Health Information



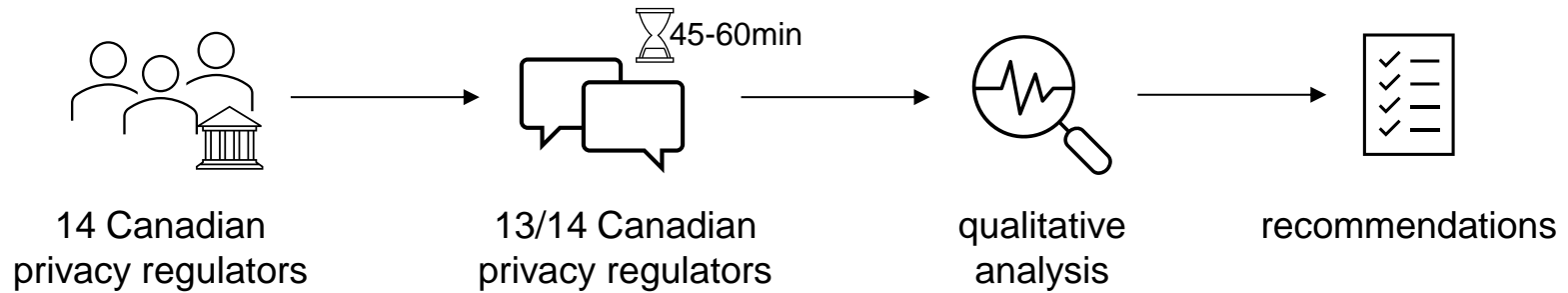
Regulation of De-Identified Information

- 1 Should the process of de-identification be regulated or overseen?
- 2 Should the de-identified data be regulated or overseen?
- 3 Should AI/ML models trained on personal or de-identified data be regulated or overseen?

Khaled El Emam, Anita Fineberg, Elizabeth Jonker, Lisa Pilgram, Perspectives of Canadian privacy regulators on anonymization practices and anonymized information: a qualitative study, *International Data Privacy Law*, 2024; ipae017, <https://doi.org/10.1093/idpl/ipae017>

METHODOLOGY

Interview Study



REGULATORS' PERSPECTIVES

JOURNAL ARTICLE

**Perspectives of Canadian privacy regulators on
anonymization practices and anonymized
information: a qualitative study** 

Khaled El Emam , Anita Fineberg, Elizabeth Jonker, Lisa Pilgram

International Data Privacy Law, ipae017, <https://doi.org/10.1093/idpl/ipae017>

Published: 18 December 2024



1

Should the process of de-identification be regulated or overseen?

- De-identification = using (or processing) personal data: regulated by privacy legislation
- Four opinions on the need of consent for de-identification
 - De-identification is an explicitly permitted discretionary use of personal data.
 - The case of legitimate interest with beneficial aspects for the data subject can be made for de-identification.
 - In certain conditions, no consent is required for de-identification (e.g. if de-identification is carried out internally, if de-identified data is used for initial purposes)
 - There must be explicit consent for de-identification.

→ The majority would not require a specific consent for (properly applied) de-identification but legal uncertainty remains.



1

Should the process of de-identification be regulated or overseen?

How to ensure proper de-identification?

So, I think that some of the control probably should be [...] how to do it properly so that there's consistent expectations [...].

[...] assuming that they did a good job of de-identifying the data, and of course [...] that often is the key question here. Everybody knows what standards we're trying to meet. It's whether they're actually met in a given case or not, that's often the real issue.

[...] one of the tools that could help in bringing the high general legal standard [downwards], would be codes of practice approved by the public body [...].

It's not easy to do, but that's what I would imagine would be happening in health and other fields, is that standards, templates and ultimately audits are checking up to make sure things are being done the way they're supposed to be done.

2

Should the de-identified data be regulated or overseen?

- Perspectives varied:
 - De-identified data falls outside the scope of privacy regulation.
 - Proportionate obligations for de-identified data apply.
 - Remaining ethical considerations must be addressed.
 - Attempts to re-identify or attack de-identified data and the malicious use of de-identified data should be prohibited.

→ The majority expressed that some sort of (proportionate) regulation is necessary to account for ethical considerations.

2

Should the de-identified data be regulated or overseen?

What are ways to oversee de-identified data?

I think we need to be looking at it from a population perspective as well. I don't think you can ignore that fact. So, I definitely think that it's important to be having ethical reviews of the use of this data, particularly where it's in the public domain.

We could say, 'provided it's for any socially beneficial purpose', but then that doesn't get us very far. Some decision maker [...] would still have to decide [...].

Transparency absolutely should happen, but it needs to be meaningful transparency.

[...] they have that right to know [...] I think, [it] is helpful, and there should always be that transparency [with] citizens.

3

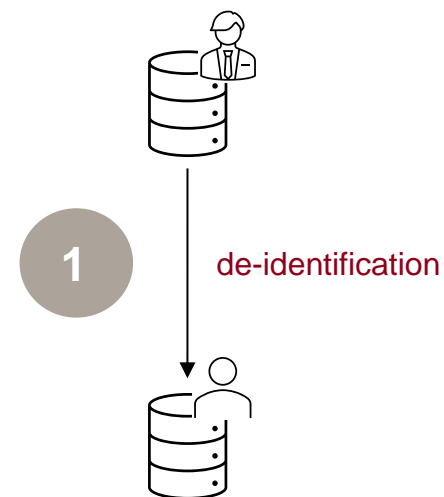
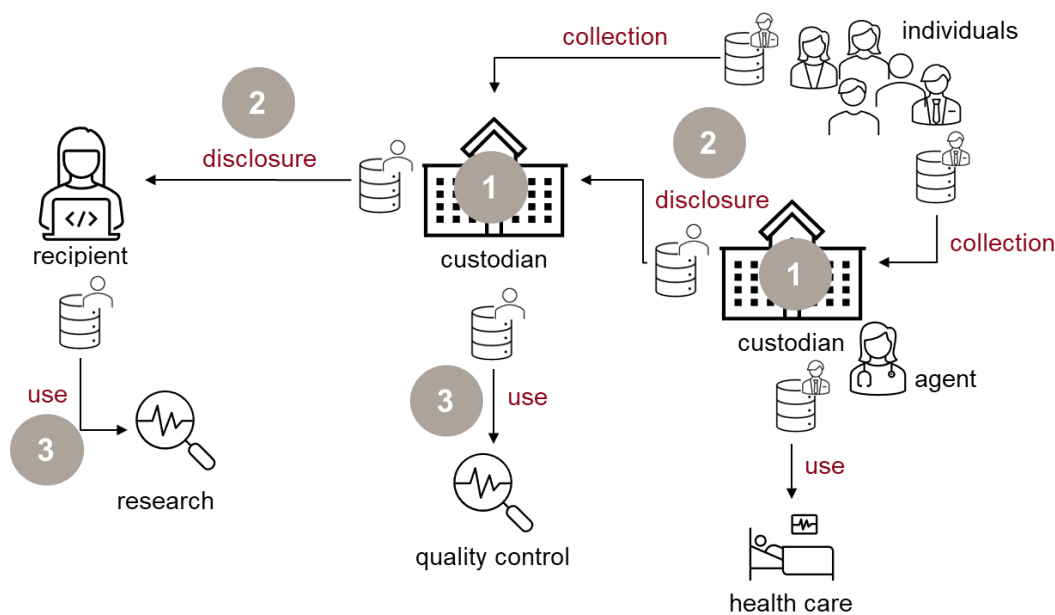
Should AI/ML models trained on personal or de-identified data be regulated or overseen?

- It depends:
 - If personal information can be recovered from a model, then the model must be treated as such.
 - If the de-identified training data was deemed non-identifiable, then regulation depends on whether or not such data requires regulation.
 - There are ethical considerations about the downstream harm of AI/ML models.

→ Regulators view the AI/ML regulation as dependent on the regulation of their training data. If this is de-identified data, proportionally fewer requirements apply.

OBSERVATIONS & RECOMMENDATIONS

The Life Cycle of De-Identified Data: Example Health Information



Research Questions

- 1 Should the process of de-identification be regulated or overseen?
- 2 Should the de-identified data be regulated or overseen?
- 3 Should AI/ML models trained on personal or de-identified data be regulated or overseen?

1

Should the process of de-identification be regulated or overseen?

The majority would not require a specific consent for (properly applied) de-identification but legal uncertainty remains.



1

Should the process of de-identification be regulated or overseen?

The majority would not require (applied) de-identification but

Personal Health Information Protection Act, 2004	Ontario
Use	
Permitted use	
37 (1) A health information custodian may use personal health information about an individual,	
<ul style="list-style-type: none"> (a) for the purpose for which the information was collected or created and for all the functions reasonably necessary for carrying out that purpose, but not if the information was collected with the consent of the individual or under clause 36 (1) (b) and the individual expressly instructs otherwise; (b) for a purpose for which this Act, another Act or an Act of Canada permits or requires a person to disclose it to the custodian; (c) for planning or delivering programs or services that the custodian provides or that the custodian funds in whole or in part, allocating resources to any of them, evaluating or monitoring any of them or detecting, monitoring or preventing fraud or any unauthorized receipt of services or benefits related to any of them; (d) for the purpose of risk management, error management or for the purpose of activities to improve or maintain the quality of care or to improve or maintain the quality of any related programs or services of the custodian; (e) for educating agents to provide health care; <li style="border: 1px solid red;">(f) in a manner consistent with Part II, for the purpose of disposing of the information or modifying the information in order to conceal the identity of the individual; (g) for the purpose of seeking the individual's consent, or the consent of the individual's substitute decision-maker, when the personal health information used by the custodian for this purpose is limited to the name and contact information of the individual and the name and contact information of the substitute decision-maker, where applicable; (h) for the purpose of a proceeding or contemplated proceeding in which the custodian or the agent or former agent of the custodian is, or is expected to be, a party or witness, if the information relates to or is a matter in issue in the proceeding or contemplated proceeding; (i) for the purpose of obtaining payment or processing, monitoring, verifying or reimbursing claims for payment for the 	

1

Should the process of de-identification be regulated or overseen?

The majority would not regulate (or apply) de-identification

Personal Information Protection and Electronic Documents Act^{federal}

S.C. 2000, c. 5

Assented to 2000-04-13

Use without knowledge or consent

(2) For the purpose of clause 4.3 of Schedule 1, and despite the note that accompanies that clause, an organization may, without the knowledge or consent of the individual, use personal information only if

- (a)** in the course of its activities, the organization becomes aware of information that it has reasonable grounds to believe could be useful in the investigation of a contravention of the laws of Canada, a province or a foreign jurisdiction that has been, is being or is about to be committed, and the information is used for the purpose of investigating that contravention;
- (b)** it is used for the purpose of acting in respect of an emergency that threatens the life, health or security of an individual;
- (b.1)** the information is contained in a witness statement and the use is necessary to assess, process or settle an insurance claim;
- (b.2)** the information was produced by the individual in the course of their employment, business or profession and the use is consistent with the purposes for which the information was produced;
- (c)** it is used for statistical, or scholarly study or research, purposes that cannot be achieved without using the information, the information is used in a manner that will ensure its confidentiality, it is impracticable to obtain consent and the organization informs the Commissioner of the use before the information is used;
- (c.1)** it is publicly available and is specified by the regulations; or
- (d)** it was collected under paragraph (1)(a), (b) or (e).

1

Should the process of de-identification be regulated or overseen?

The majority would not require a specific consent for (properly applied) de-identification but legal uncertainty remains.

- Provide clarity about de-identification as permitted use of personal data without consent.

1

Should the process of de-identification be regulated or overseen?

The majority would not require a specific consent for (properly applied) de-identification but legal uncertainty remains.

- Provide clarity about de-identification as permitted use of personal data without consent.
- Develop Code(s) of Practice for proper de-identification.
 - Enforceable set of guidelines or rules with a certification or audit mechanism
 - Approved by a public body
 - Based on (inter)national resources, e.g. ISO/IEC 27559:2022
 - Common generic principles with tailored guidelines for certain industries



2

Should the de-identified data be regulated or overseen?

The majority expressed that some sort of (proportionate) regulation is necessary to account for ethical considerations.



Be explicit about transparency as an obligation.

- Transparency that is meaningful and easy to understand
- Building up trust with data subjects

2

Should the de-identified data be regulated or overseen?

The majority expressed that some sort of (proportionate) regulation is necessary to account for ethical considerations.

- Be explicit about transparency as an obligation.
- Address the moral gap for data without ethical oversight.
 - Ethical considerations are orthogonal to identifiability.
 - Ethical review is already good practice or even mandatory (e.g., TCPS) in the research context.

2

Should the de-identified data be regulated or overseen?

The majority expressed that some sort of (proportionate) regulation is necessary to account for ethical considerations.

- Be explicit about transparency as an obligation
- Address the moral gap for data without ethical oversight.
- Reduce obligations for properly de-identified data.
 - Incentive for implementing and improving de-identification

3

Should AI/ML models trained on personal or de-identified data be regulated or overseen?

Regulators view the AI/ML regulation as dependent on the regulation of their training data. If this is de-identified data, proportionally fewer requirements apply.



See recommendations for the regulation of de-identified data

Limitations

- **Understanding of terminology:** Throughout our study, we defined the term de-identification when using it. However, de-identification, anonymization and pseudonymization are differently used across jurisdictions which may have resulted in different interpretations of the questions.
- **Generalizability:** Interviews were conducted among privacy regulators in Canada. Conclusions may not be transferrable to other jurisdictions.
- **Qualitative analysis:** The analysis was informed by grounded theory, but researchers' presumptions could have influenced the process.

Read more in

Khaled El Emam, Anita Fineberg, Elizabeth Jonker, Lisa Pilgram, Perspectives of Canadian privacy regulators on anonymization practices and anonymized information: a qualitative study, *International Data Privacy Law*, 2024; ipae017, <https://doi.org/10.1093/idpl/ipae017>





Questions?

Lisa Pilgram, MD

Postdoctoral Fellow at the Electronic Health Information Laboratory (Khaled El Emam)