De-identification: Updated Guidance from the Ontario IPC

Brenda McPhail, PhD

Senior Technology & Policy Advisor

Information and Privacy Commissioner of Ontario

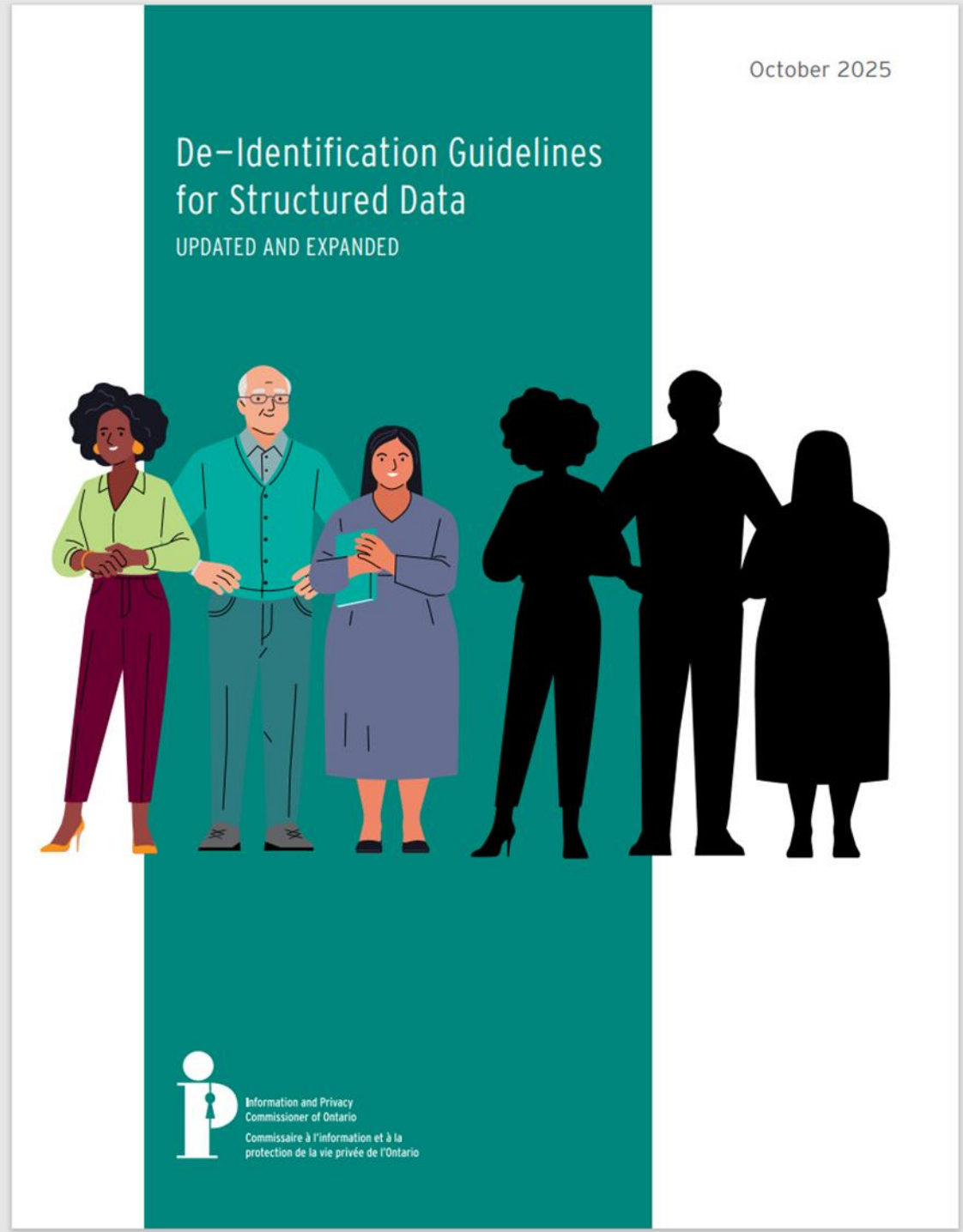
Commissaire à l'information et à la protection de la vie privée de l'Ontario

CANON/Access Privacy Webinar

19 November 2025

Presentation Outline

- Why and how we updated the IPC's de-ID guidance
- Lightning-speed guidance walkthrough
  - Terminology
  - 12 step process
  - Case studies
  - Abundant (but applicable) appendices



Key takeaways for today

- De-identification is a **use** of personal information, and the new IPC guidance can assist in ensuring that use follows best practice and is effective to mitigate privacy risks.
- **Properly executed**, de-identification can allow data to be used to draw important insights from data **without compromising privacy**.
- IPC guidance can help researchers and organisations plan, execute, and document their de-identification process(es) to demonstrate their commitment to privacy, transparency and accountability in their use of data, and support and sustain public trust.

APRIL 2021

IPC Strategic Priorities 2021–2025

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

Privacy and Transparency in a Modern Government

Advance Ontarians' privacy and access rights by working with public institutions to develop bedrock principles and comprehensive governance frameworks for the responsible and accountable deployment of digital technologies.

Trust in Digital Health

Promote confidence in the digital health care system by guiding custodians to respect the privacy and access rights of Ontarians, and supporting the pioneering use of personal health information for research and analytics to the extent it serves the public good.

Children and Youth in a Digital World

Champion the access and privacy rights of Ontario's children and youth by promoting their digital literacy and the expansion of their digital rights while holding institutions accountable for protecting the children and youth they serve.

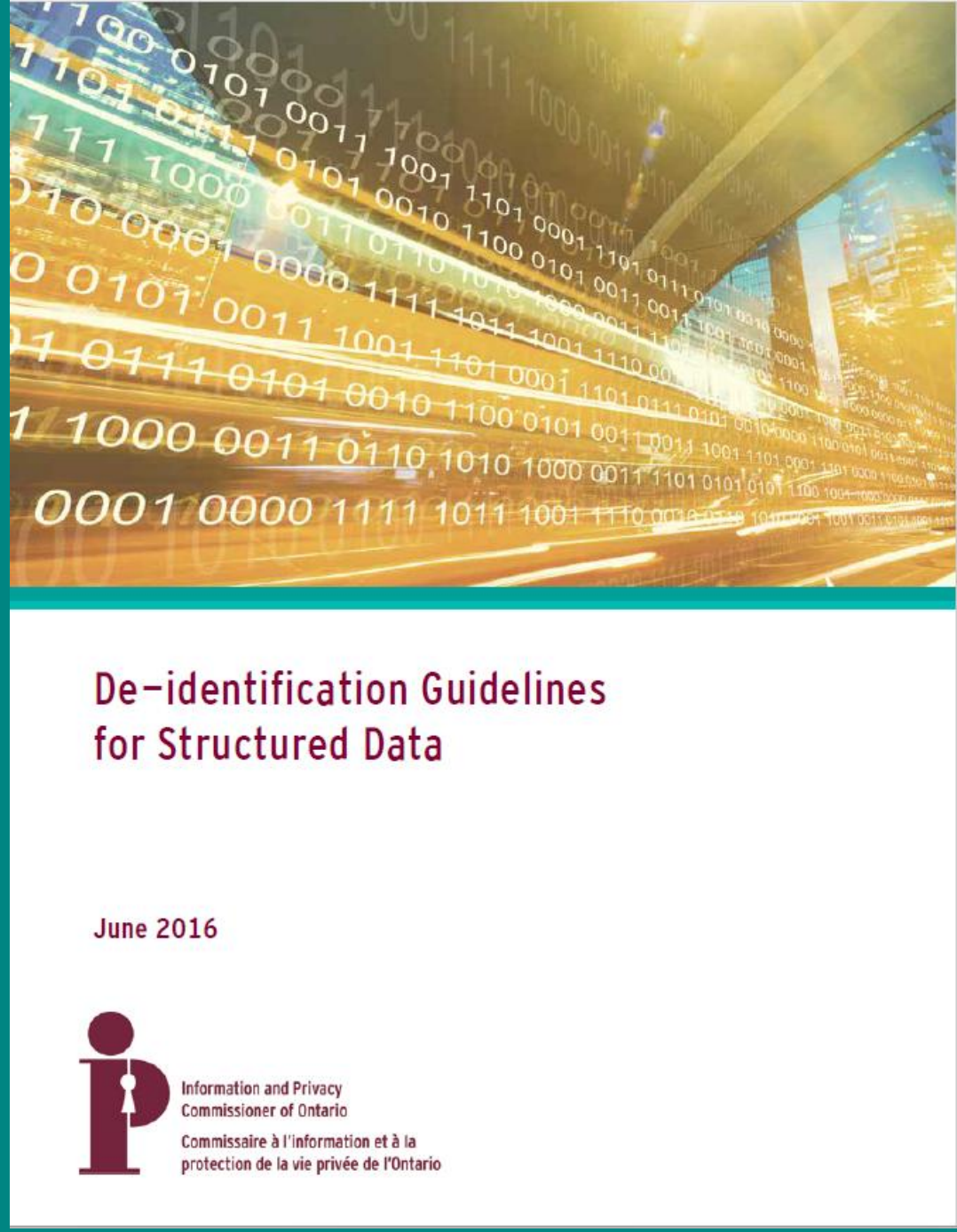
Next-Generation Law Enforcement

Contribute to building public trust in law enforcement by working with relevant partners to develop the necessary guardrails for the adoption of new technologies and community based approaches that protect both public safety and Ontarians' access and privacy rights.

Accessibility & Equity
Collaboration & Consultation
Vision & Pragmatism
Knowledge Development

New builds upon old

- The IPC published *De-identification Guidelines for Structured Data* in 2016
- Methods are still relevant today
- However:
  - Regulated organisations asked for additional operational guidance
  - Concept of identifiability has evolved
  - Scope needed to expand to cover multiple types of data release



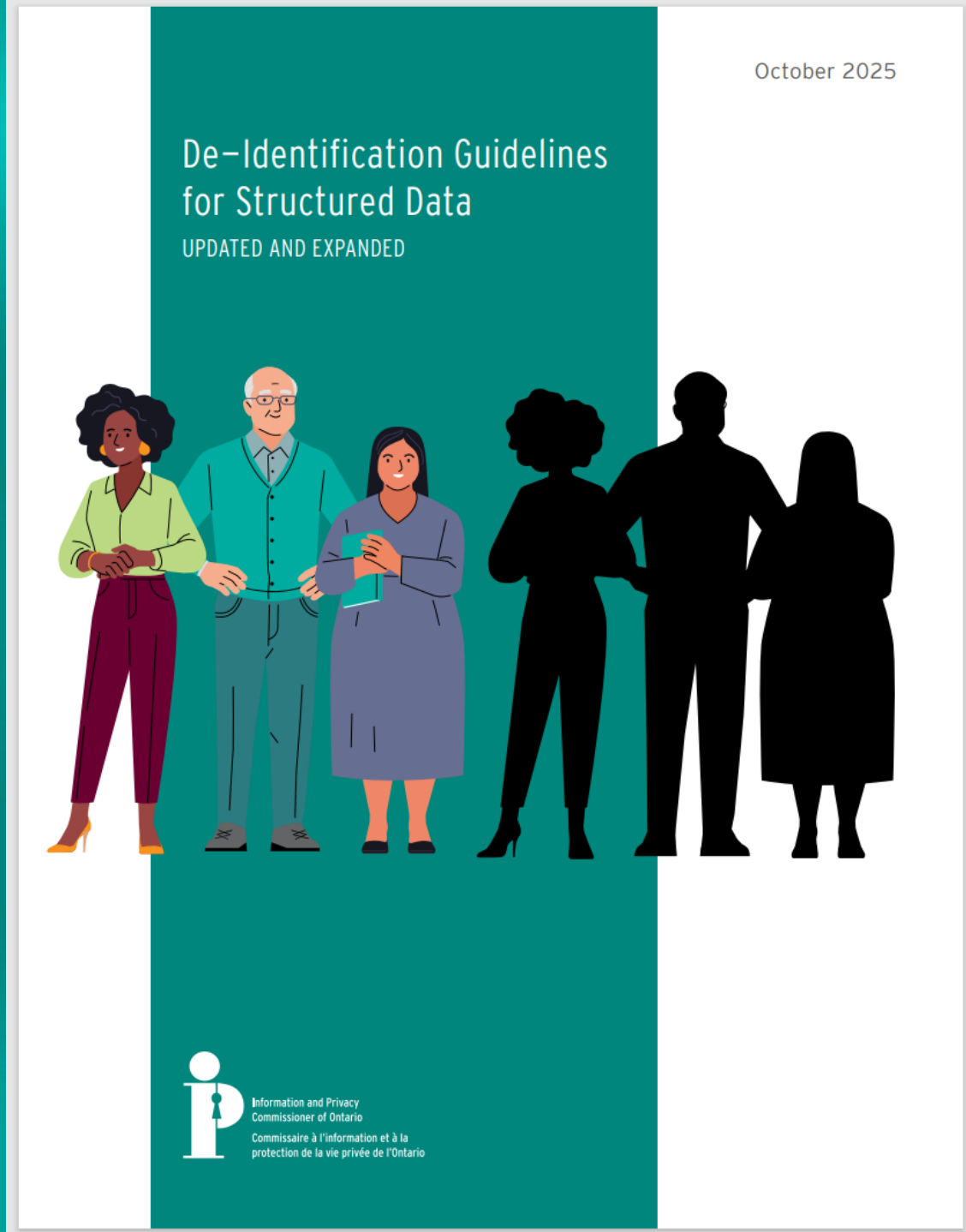
The new guide aims to:

- Provide more clarity on terminology
- Make some of the basic principles of de-identification and non-personal data explicit and more precise
- Provide further considerations of other types of disclosures around inferences
- Clear separation of public vs non-public data sharing for secondary purposes, and internal versus external reuse of data
- Be more precise about the quantitative aspects of risk assessment and management to support implementation
- Provide more information about the necessary documentation of de-identification
- Cover some of the more modern methods for de-identification
- Reflect practical experiences with de-identification gained over the last decade or so

De-identification Guidelines for Structured Data

Contents:

- Terminology & key concepts
- Scope
- De-identification use cases
- Principles of de-identification
- Process for de-identifying structured data (12 steps)
- Conclusions
- Appendices



Definitions

- **Pseudonymization** is the process of transforming direct identifiers that exist within a dataset.
  - Eg. Names & Addresses
  - Almost always, pseudonymized data is still considered personal information
  - Other terms used include masking, obfuscation, key-coding, redacting, tokenization and (confusingly) de-identification.

- **De-identification** is the process of performing pseudonymization, **plus** transforming indirect identifiers that remain in the dataset following pseudonymization.
  - A properly de-identified dataset no longer contains information that identifies an individual or information that could be used, either alone or with other information, to identify an individual based on what is reasonably foreseeable in the circumstance.

Scope of Guidance

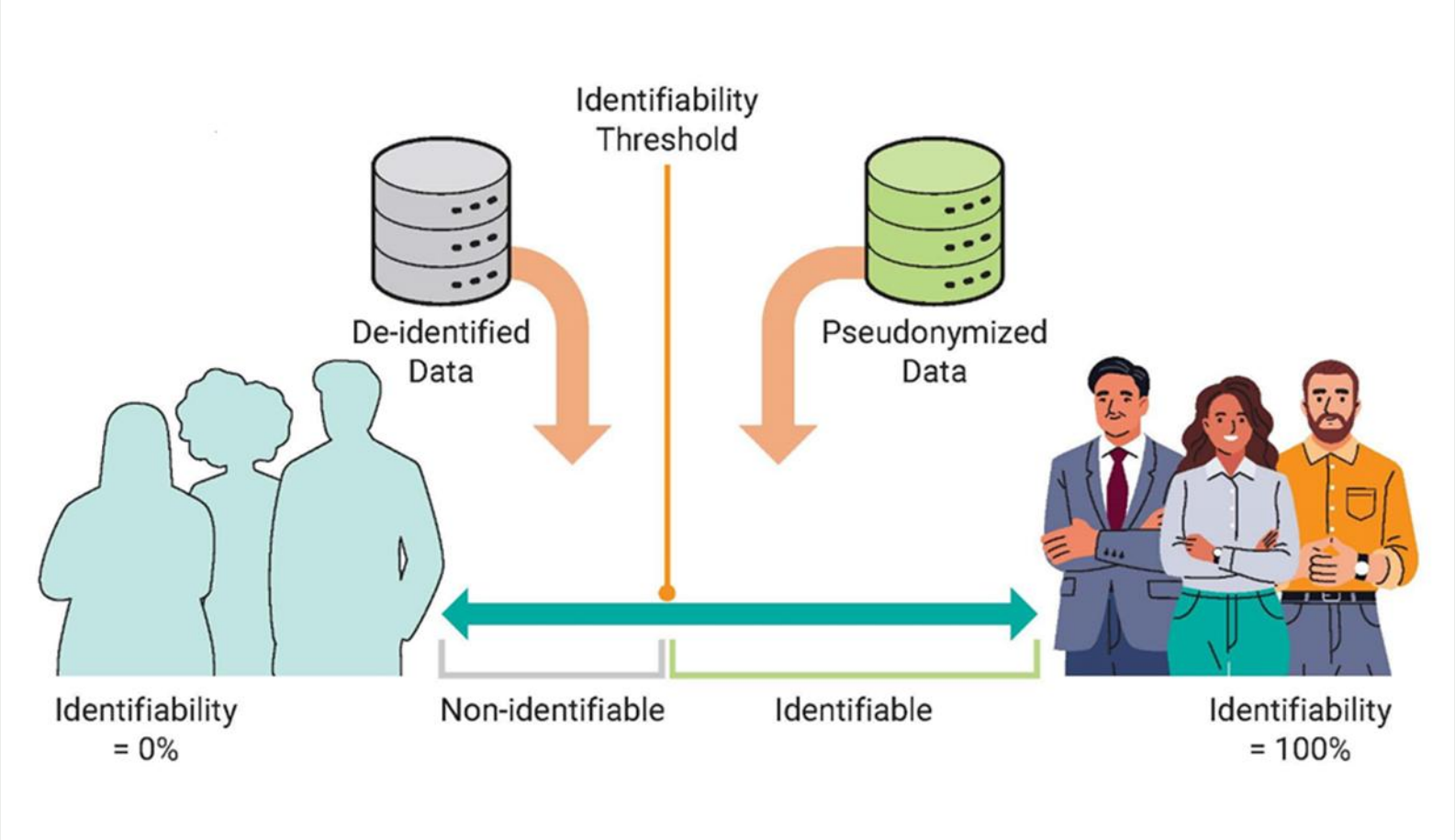
- Ontario privacy laws
- Structured data
- Identity disclosure focus
- Incorrect re-identification is out of scope
- Focus on model-based re-identification risk assessment

Use Cases

5 Use cases:

- Open data
- Data sharing among related organisations
- Data sharing with an external third party
- Custodian-controlled access to data by a third party
- Internal data reuse

De-identification principles: the identifiability spectrum



Process for de-identifying structured data

- 1. Prepare
- 2. Determine the release model
- 3. Classify variables
- 4. Pseudonymize the data
- 5. Determine an acceptable re-identification risk threshold
- 6. Measure the data vulnerability
- 7. Measure the probability of attack
- 8. Calculate the overall re-identification risk
- 9. Transform the data and include controls to get risk below the threshold
- 10. Assess data utility
- 11. Document the process & results
- 12. Monitor the environment periodically/ongoing governance

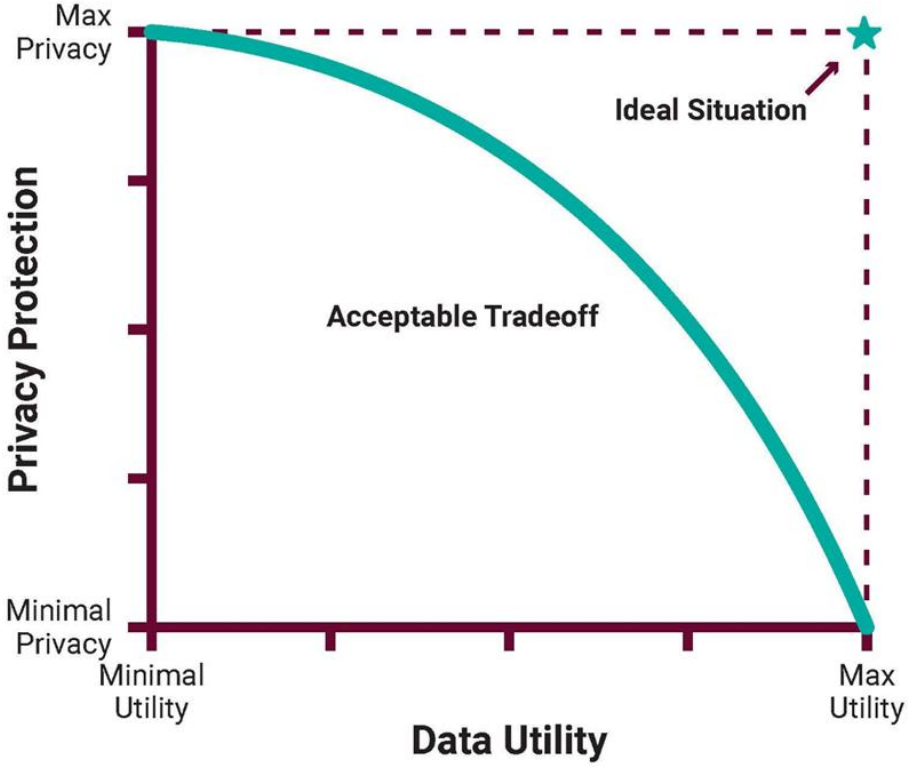
Re-identification risk thresholds

Invasion of Privacy Values	Re-identification Risk Threshold (very low)	Cell Size Equivalent
Low	0.09	11
Medium	0.075	15
High	0.05	20

Appendices

- A) **Checklist for assessing recipient controls**
- B) Separations required for secondary use
- C) **Data sharing agreement checklist**
- D) Measuring identity disclosure data vulnerability for public data release
- E) Measuring identity disclosure data vulnerability for non-public data sharing & re-use
- F) Pseudonymization techniques
- G) Methods for transforming indirect identifiers
- H) Synthetic Data Generation
- I) **Checklist for documenting pseudonymization**
- J) **Checklist for documenting de-identification**

Conclusion



- De-identification, done well, minimizes the risk of exposing sensitive information while allowing data analysis and use.
- De-identification, done badly, erodes public trust and social license for data collection and use.
- IPC guidance provides data custodians with a solid base from which to begin their de-identification processes in keeping with privacy best practice.

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