# De-identification of personal information and the new IPC Ontario guidelines

**October 21, 2025**

Under Canadian privacy law, “de-identified” data can be used for research, product development and other such “secondary purposes” without individual consent and without the normal protections that apply to personal information. On October 15, 2025, the Information and Privacy Commissioner/Ontario updated its guidance on the de-identification of personal information. The new guidance is set out in a 96-page document entitled *De-Identification Guidelines for Structured Data; Updated and Expanded*. This new guidance updates and expands guidance first issued by the IPC in 2016 in setting out how to transform personal information into data that neither identifies individuals nor unacceptably risks their re-identification.

De-identification is a technical subject, though we believe the basic concepts are within reach of privacy and data security professionals. In this article, we address the topic of de-identification generally, with reference to the “New Guidelines” and the “2016 Guidelines.” Both documents are technical and focus on de-identification risk analysis without explaining the broader legal context. Our aim is to help our clients understand the de-identification process, how it relates to legal obligations in privacy legislation, and how to approach developing in-house capacity for de-identification.

## Why is de-identification important?

Today, there is no privacy related topic more important than de-identification. This has been the premise of the IPC’s leadership on de-identification, which began with a 2014 research paper that stated, “Information is the new currency of the economy.”[1] With increased computing power, data-driven content is generated at an ever-increasing rate both through direct analysis of data and through data modeling. For example, “training” artificial intelligence models relies heavily on the use of data initially collected for other purposes.

Our growing reliance on data clashes with the need for privacy protection. To address this need, Canadian laws give individuals a right to control information about themselves and place accountability obligations on custodians of data containing personal information. Most significantly, custodians have a duty to keep personal information secure.

De-identification of data breaks the link between information and the identifiable individual, thus tempering custodian obligations and unlocking the data's value to drive innovation. If predictions about artificial intelligence are accurate, using de-identified data to train artificial intelligence models will boost automation and enable scientific discovery.

# De-identification and the law

Our understanding of de-identification starts with statutory text, and specifically, the definition of "personal information." In most statutes, the definition includes the word "identifiable," which invites consideration of the probability an individual is likely to be identified in data sets that do not include direct identifiers such as names, full addresses, and identification numbers.

The governing test under Canada's federal privacy legislation was articulated in *Gordon v Canada*, in which the Federal Court agreed with the Office of the Privacy Commissioner of Canada that information "will be about an identifiable individual where there is a serious possibility that an individual could be identified through the use of that information, alone or in combination with other available information."[2]

It is questionable whether the federal "serious possibility" test from *Gordon* differs significantly from the test formulated by the Information Privacy Commissioner/Ontario and approved by the Court of Appeal for Ontario in *Ontario v Pascoe*: is there a reasonable expectation that, when information in the records at issue is combined with information from sources otherwise available, the individual can be identified?[3]

There are now also statutes that establish an express test for de-identification. The Ontario *Personal Health Information Protection Act, 2004*, for example, was expressly amended in 2019 to prohibit re-identifying individuals from "de-identified" data sets.[4] PHIPA now defines "de-identify" as the removal of "any information that identifies the individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify the individual."[5]

# Three important de-identification concepts

There are three important concepts relating to de-identification that privacy professionals must understand.

### 1. The de-identification test is probabilistic, ( *i.e.*, it measures, or is about the degree of re-identification risk).

The ultimate goal of de-identification is to create a data set where the risk of re-identification is "very low." However, the meaning of "very low" – and the tolerable risk of re-identification – depends on several factors, including the nature of the personal information being de-identified, the number of individuals who would be affected by a re-identification attack, and the nature of the harms that would result from a re-identification attack.[6]

To assess whether the risk of re-identification of a data set is very low, the New Guidelines, like the old ones, propose that organizations conduct an "invasion of privacy assessment," in which one must ask: "If this data set were re-identified, to what extent would the release of the data set invade an individual's privacy?"[7] If the privacy impact of data re-identification would be high, then the re-identification probability must be below 5%.[8] Conversely, if the privacy impact would be low, then the re-identification probability can be as high as 9%.[9]

In no case does the IPC expect organizations to create de-identified data sets with zero re-identification risk.

## 2. The de-identification test contemplates re-identification through matching with available information.

De-identification analysis requires one to consider what personal information is available to an individual who is motivated to re-identify and that is unique enough to reveal identities in the de-identified data set.[10] Information, such as names, are "direct identifiers," while other information, such as street addresses, that can be linked to identifiable individuals and that is likely available to an attacker are "indirect" or "quasi-identifiers."[11] To de-identify data, direct identifiers **must** be removed or modified and indirect identifiers **may** be removed or modified to reduce the likelihood of re-identification.

Not all information in a data set is identifying data. Consider, for example, answers in a diversity, equity and inclusion survey about one's exposure to discrimination and harassment in the workplace. Such data (which is unlikely to be used to identify an individual) can normally remain in a de-identified data set in unmodified form. The IPC refers to this type of data simply as "sensitive data."[12]

## 3. The risk of re-identification must be assessed in the context.

Although the New Guidelines state plainly, "Context is important,"[13] this point must be underscored given the prevalence of non-public data releases. In a non-public release, the recipient is bound not to disclose de-identified data to others and is restricted from re-identification. When data are disclosed in a non-public release the IPC model significantly discounts the overall risk of de-identification based on the probability of an attack.[14] To derive an overall risk value the (contextual) probability of an attack is multiplied with the probability of the data being re-identified in an attack:

Overall risk = probability of attack in the context x data vulnerability probability[15]

In the New Guidelines, the IPC says that the probability of attack is a function of "context and controls." In a non-public data set release to a third-party, the probability of an attack relates to applicable security controls, privacy controls, contractual terms, and what can be reasonably gleaned about attacker motives and capacity.[16] Using the same model, the risk associated with releasing data publicly is higher, and is directly related to the ease with which the data can be linked to other available information both today and into the future.

The other way the IPC recognizes context is by stipulating risk thresholds based on privacy impact. According to the IPC, one who is de-identifying data must consider the following factors when rating the impact of a successful attack as low, medium or high:

- the sensitivity of the information
- the scope and/or level of detail of the information
- the number of individuals potentially affected
- the potential harms or injuries to individuals in the event of a breach or inappropriate use.[17]

# The thrust of the IPC guidance

The legal test for de-identification is set out in statute, not the IPC guidance. The New Guidelines and the 2016 Guidelines, however, contribute in two ways: one, they set out a process to soundly analyze re-identification risk; and two, they set out objective thresholds for re-identification risk – numerical probability values to aim for when de-identifying data based on the context and the properties of the data set. An organization that follows a sound analytical process and derives a risk of re-identification below the applicable threshold will be very likely to meet the IPC's expectations. Although statutory differences will prevail, the process set out is one that can be reasonably relied upon by organizations across sectors and across Canada.

The New Guidelines, like the 2016 Guidelines, set out a step-by-step de-identification process. We simplify and re-articulate this process as follows:

- ***Select a release model*** . Re-identification risk will be significantly lower for a private data release than a public data release. When data is released privately the recipient must be contractually restricted from disclosing the de-identified data to third parties and the contract must govern the de-identified data.[18] We provide further input on this requirement below.
- ***Identify and classify identifiers in the data set*** . Determine which variables in the dataset are "direct identifiers" and which are "quasi-identifiers."[19]
- ***Select an appropriate risk threshold value*** . As explained above, the appropriate risk threshold value depends on the privacy impact of a successful re-identification attack, and according to the New Guidelines (and ISO/IEC 27559), can range from 5% to 9%.[20]
- ***Measure the context or attack vulnerability*** . This is a percentage figure – one representing the highest probability of a re-identification attack for all foreseeable threat scenarios. To derive this figure, one must identify foreseeable threat scenarios (*e.g.*, deliberate insider attack, data breach, inadvertent recognition of a data subject) and evaluate the probability of an attack for each scenario using defensible logic or available models. The New Guidelines, for example, incorporate a table that sets out probabilities for deliberate insider attacks based on the level of security (low, medium, or high) and the level of attacker motives and capacity (low, medium, or high).[21]
- ***De-identify and measure the resulting data vulnerability*** . Data vulnerability is a second percentage figure – one representing the risk of re-identification based on the properties of the de-identified data set.[22] Once direct identifiers are removed or securely masked then, data vulnerability is a function of the number of records in the disclosed data set that have unique quasi-identifier values. For example, "Sage" may be a male aged 45 who lives in Toronto. If Sage is the only individual

in a de-identified data set with these attributes, he is more identifiable than if he was one of three individuals in the de-identified data set with identical attributes. If Sage's data values are unique, a data analyst could remove the age variable altogether to protect Sage from being identified, although removing the age variable could significantly reduce the utility of the data set. Translating the age variable so it identifies individuals in five-year age bands (e.g., 45 to 49) is another way to reduce uniqueness and the risk of re-identification. This de-identification technique is called "generalization."[23]

- ***Measure the overall risk*** . The percentage figure for overall risk is derived by multiplying the attack vulnerability by the data vulnerability.[24]
- ***Confirm overall risk is below the risk threshold*** . If the overall risk value is below the selected risk threshold (*i.e.*, the risk of re-identification is sufficiently low), the data is deemed to be appropriately de-identified. Conversely, if the overall risk value exceeds the risk threshold, further data manipulation is required to reduce the overall risk below the selected risk threshold (see Figure below). Data manipulation should be conducted in a manner sensitive to requirements relating to data utility given data utility is generally reduced by de-identification.[25]

Given that the risk environment can change over time, the IPC says that entities who de-identify and release data should periodically re-analyze the risk of re-identification. In the New Guidelines, the IPC says, "In practice, the typical interval between re-identification risk assessments is every two years or three years."[26]

The New Guidelines expand the process set out in the 2016 Guidelines in two important ways:

- **New planning guidance** . The New Guidelines prescribe initial planning steps and include guidance about deciding whether to outsource de-identification, assembling a team of individuals with appropriate expertise, and determining whether to conduct a privacy impact assessment.[27] The New Guidelines also recommend that organizations that engage in de-identification provide advanced notice to individuals that de-identification is being conducted and describe its "general purposes."[28] This guidance aligns with PHIPA Decision 175, in which the IPC held that Ontario health information custodians must give notification of routine de-identification practices in their written public statements if they routinely de-identify personal health information.[29]
- **More emphasis on post-release oversight** . Also consistent with Decision 175, the New Guidelines emphasize that one who releases data in de-identified form has ongoing obligations, whether the release is done privately to another organization under a data sharing agreement or publicly. These obligations include monitoring the risk environment for changes that increase the risk of re-identification and auditing compliance with data sharing agreements. We discuss governance in more detail and revisit Decision 175 in the section immediately below.

To help data custodians manage these additional responsibilities, the IPC provides several resources in its appendices.

# Governance of released de-identified data

“Private release” of de-identified data is premised on sharing with a trusted third party bound to contractual obligations that protect against onward disclosure. In this release model, a disclosing entity continues to be accountable for the handling of the data it de-identifies.

“Public release” of de-identified data is different. Given the releasing entity cannot impose contractual protections or effectively recall released data sets, the risks of public release are much higher. One who releases data publicly must engage in greater data manipulation to address this risk, rendering the need for ongoing oversight debatable.

Differences across Canada in the legal duties that apply post release are developing. For example, Alberta recently passed a new public sector privacy statute that governs “non-personal information” – “data, including data derived from personal information, that has been generated, modified or anonymized so that it does not identify any individual, and includes synthetic data and any other type of non-personal data identified in the regulation.”[30] The statute permits disclosure of non-personal data to another public body without any restriction, but disclosure of non-personal data to others (including commercial entities) will be limited to specified purposes and subject to requirements that render downstream users accountable to the disclosing public body.[31] Specifically, the new legislation only permits the disclosure of non-personal information if:

> (ii) the head of the public body has approved conditions relating to the following: (A) security and confidentiality; (B) the prohibition of any actual or attempted re-identification of the non-personal data; (C) the prohibition of any subsequent use or disclosure of the non-personal data without the express authorization of the public body; (D) the destruction of the non-personal data at the earliest reasonable time after it has served its purpose under subclause (i), unless the public body has given the express authorization referred to in paragraph (C),
>
> and
>
> (iii) the person has signed an agreement to comply with the approved conditions, this Act, the regulations and any of the public body’s policies and procedures relating to non-personal data.[32]

In Ontario, under the *Personal Health Information Protection Act, 2004* (PHIPA), the IPC held that governance obligations arise out of the duty to safeguard personal health information. In PHIPA Decision 175 the IPC said:

> In this circumstance, ‘taking reasonable steps’ includes determining whether appropriate measures were taken to ensure that the information sold was properly de-identified and that it is sufficiently unlikely that the information can be re-identified. These measures will generally include both the masking and de-identification strategy applied to the personal health information, and the safeguards established to protect against re-identification of the de-identified information.[33]

The IPC held that the disclosing health information custodian met its governance related obligations by contractually prohibiting the recipient from using information determined to be identifiable and by disclosing to an entity known to have reasonable privacy and security controls.[34]

Neither the Alberta statute nor PHIPA Decision 175 explicitly contemplate the need to notify affected individuals of "re-identification incidents" or other events that raise a re-identification risk. The New Guidelines, however, call for data custodians to take action to address claimed or confirmed re-identification, including by retrieving datasets (if appropriate) and notifying affected individuals.

To avoid ongoing accountability, disclosing entities can de-identify data so it is suitable for public release. The Alberta public sector privacy statute stipulates that public bodies are free to publish non-personal data in "aggregate" or "statistical" form.[35] Neither "aggregate" nor "statistical" is defined, but these terms likely refer to data for which the risk of re-identification is so low that public release is appropriate. Most surprisingly, the New Guidelines suggest that one who releases de-identified publicly does bear ongoing accountability for it, though the New Guidelines also recognize that publicly released data may not be amenable to recall.[36]

## Ethics, purpose limitations, and transparency

The legitimacy of de-identifying personal information for secondary purposes is a fundamental question that future Canadian privacy statutes and regulatory interpretations will address. Research is considered to be a legitimate purpose, but how broad should research be defined? Should it encompass product development by commercial entities? And what about other commercial purposes? Even if consent to de-identify is not required, how does an individual's interest in their personal information, and their interest in the receipt of health or other services that generate such personal information, restrict the uses for de-identified data?

The sitting Ontario Privacy Commissioner, Patricia Kosseim, has taken a particular interest in this question, writing a 2008 paper in which she commented, "Quite apart from the legal complexity inherent in ascertaining what constitutes [de-identified] data for the purposes of interpreting a given statute, it is not clear that individuals' legal and moral interests in their personal information dissipate simply because it is de-identified and falls outside the scope of data protection regimes."[37] She further commented on the ethical implications of this question in a blog post on Decision 175, stating:

> While the focus of PHIPA Decision 175 is on compliance with PHIPA, as it existed at the time of the investigation, the broader ethical questions inherent in selling or disclosing even properly de-identified personal health information are ripe for public debate. Inferences made or derived from de-identified personal health information can have significant impacts on groups that share similar characteristics, exposing individual members of those groups to potential harms, such as stigmatization and discrimination, unfair distribution of services or benefits, loss of jobs, or denial of insurance coverage. Even in good hands, and for appropriate purposes, the sale or disclosure of de-identified data without clear and meaningful transparency can seriously undermine public trust.
>
> In today's digital world where health data is an increasingly valuable commodity, the stakes have never been higher. In a digital economy that's becoming increasingly opaque, the need for transparency has never been greater. And in a context where the differences between research and commercialization, and public and private goods, get progressively murkier, the time for public debate has never been more pressing.

> Regardless of how the broader ethical discussion shapes up, there must at a minimum be greater transparency around the sale or disclosure of de-identified health data and greater accountability for what happens to that data after its release. This will serve to protect both individuals and health information custodians, by supporting trust between them, and upholding general confidence in the health care system.[38]

These ethical questions are becoming a matter of law. As discussed, the Alberta public sector privacy statute now includes a purpose-based restriction on the use of de-identified data by persons other than public bodies – restricting such use to research and analysis, planning, administering, delivering, managing, monitoring or evaluating a program or service, or other purposes that are prescribed. The Québec private sector privacy statute has a broader purpose-based restriction, arguably restricting the use of "anonymized" personal information to "serious and legitimate purposes."[39]

Neither of these novel requirements impose an obligation to address the potential discriminatory impact of secondary use that Commissioner Kosseim has raised, nor do they address the matter of transparency. Transparency may be required already by existing provisions. In PHIPA Decision 175, for example, the IPC held that a health information custodian has a duty to disclose routine de-identification in its written public statement.[40] A recent study on the perspectives of Canadian privacy regulators indicated a uniform view among regulators that those who de-identify personal information for secondary use should be transparent about it.[41]

# Building your de-identification capacity

When establishing an internal de-identification function, organizations should begin by developing a comprehensive framework that clearly defines roles, responsibilities, and decision-making authority. In developing this framework, consider:

- stipulating the purposes for which de-identification is conducted and for which de-identified data is released;
- establishing minimum contractual requirements for private data release;
- assigning responsibility for risk assessment and private data release oversight;
- committing to re-identification risk assessment and transparency in accordance with the requirements of applicable law and guidance (including the IPC guidance).

The IPC guidance is intended to make de-identification risk assessment accessible, though some expertise is required, and there is a unique set of knowledge and skills that a de-identification analyst must draw upon. In other words, de-identification is arguably a competency in and of itself. Organizations can adopt a de-identification competency as core to their privacy or security analyst positions. De-identification courses are available, there are ample writings available for self-study, and the basic concepts will readily be grasped by those who have prior experience in applying statistical methods.

# Conclusion

De-identification can realize the value of datasets by making them available for public benefit, health research, and other critical innovation. Organizations can reduce the risk

of reidentification to levels that enable disclosures that would otherwise violate privacy legislation by implementing techniques set out in the New Guidelines. Nevertheless, even when deidentification is carried out to the standard of the New Guidelines, institutional responsibility for de-identified data sets can continue to exist.

If you have questions about de-identification, compliance with the New Guidelines, or building your organization's internal capacity, BLG's national information and privacy team is here to help. We regularly assist clients with drafting and implementing internal policies tailored to privacy and data governance needs, and we provide opinions on de-identification risk to support decision-making and regulatory compliance. Please contact us.

## Footnotes

[1] Ann Cavoukian and Khaled El Emam, "De-identification Protocols : Essential for Protecting Privacy," June 25, 2014.

[2] *Gordon v. Canada (Health)*, 2008 FC 258 (CanLII), at para 34.

[3] *Ontario (Attorney General) v. Pascoe*, 2001 CanLII 32755 (ON SCDC), at para 15.

[4] 2019, c. 15, Sched. 30, s. 3.

[5] PHIPA, s. 2.

[6] New Guidelines, p. 39.

[7] Ibid.

[8] New Guidelines, p. 40.

[9] Ibid.

[10] New Guidelines, p. 94.

[11] New Guidelines, p. 7.

[12] New Guidelines, p. 41.

[13] New Guidelines, p. 29.

[14] New Guidelines, p. 30.

[15] Ibid.

[16] Ibid.

[17] New Guidelines, p. 39.

[18] 2016 Guidelines, p. 7.

[19] 2016 Guidelines, p. 8.

[20] 2016 Guidelines, p. 9.

[21] New Guidelines, p. 75.

[22] See New Guidelines, p. 40.

[23] See New Guidelines, pp. 82, 83.

[24] New Guidelines, p. 42.

[25] See New Guidelines, p. 27.

[26] New Guidelines, p. 31.

[27] New Guidelines, pp. 35 to 37.

[28] New Guidelines at p. 37.

[29] PHIPA Decision 175 (25 March 2025), para 64.

[30] *Protection of Privacy Act*, SA 2024, c P-28.5, (PPA), s 1.Notably, the statute does not account for the degree of de-identification, imposing obligations in respect of data that is highly de-identified but sill "derived from personal information."

[31] PPA, s 23.

[32] Ibid.

[33] PHIPA Decision 175 (25 March 2025), para 98

[34] Ibid., para 89.

[35] PPA, s. 23(4).

[36] New Guidelines, p. 45 ("where possible").

[37] P. Kosseim, Policy by Procrastination: Secondary Use of Electronic Health Records for Health Research Purposes, 2008 CanLIIDocs 5.

[38] P. Kosseim, Ripe for public debate: Legal and ethical issues around de-identified data (17 May 2022), <ipc.on.ca>.

[39] Act respecting the protection of personal information in the private sector, CQLR c P-39.1, s 23.

[40] PHIPA Decision 175, paras 63 and 64.

[41] K. El Emam et al, Perspectives of Canadian privacy regulators on anonymization practices and anonymized information: a qualitative study in International Data Privacy Law, 2024, Vol. 14, No. 4.

By

Daniel J. Michaluk, Marc Vani, Avital Sternin

Expertise

Cybersecurity, Privacy & Data Protection

---

## BLG | Canada's Law Firm

As the largest, truly full-service Canadian law firm, Borden Ladner Gervais LLP (BLG) delivers practical legal advice for domestic and international clients across more practices and industries than any Canadian firm. With over 725 lawyers, intellectual property agents and other professionals, BLG serves the legal needs of businesses and institutions across Canada and beyond – from M&A and capital markets, to disputes, financing, and trademark & patent registration.

**blg.com**

## BLG Offices


**Calgary**

Centennial Place, East Tower
520 3rd Avenue S.W.
Calgary, AB, Canada
T2P 0R3

T 403.232.9500
F 403.266.1395

**Ottawa**

World Exchange Plaza
100 Queen Street
Ottawa, ON, Canada
K1P 1J9

T 613.237.5160
F 613.230.8842

**Vancouver**

1200 Waterfront Centre
200 Burrard Street
Vancouver, BC, Canada
V7X 1T2

T 604.687.5744
F 604.687.1415

**Montréal**

1000 De La Gauchetière Street West
Suite 900
Montréal, QC, Canada
H3B 5H4

T 514.954.2555
F 514.879.9015

**Toronto**

Bay Adelaide Centre, East Tower
22 Adelaide Street West
Toronto, ON, Canada
M5H 4E3

T 416.367.6000
F 416.367.6749