

# IMPLEMENTING THE PERSONAL HEALTH INFORMATION PROTECTION ACT

TORONTO MARRIOTT BLOOR YORKVILLE  
MAY 8/9 2006

## PRIORIZING YOUR PHIPA IMPLEMENTATION PLAN BY DETERMINING WHERE PHIPA APPLIES

Carole Lucock, LL.B, LL.M  
LL.D Candidate, Faculty of Law  
University of Ottawa

### I Introduction

This presentation considers the application of Ontario's *Personal Health Information Protection Act* (PHIPA) to anonymized health information from the perspectives of:

- (a) the exclusion of anonymized information from the definition of personal health information;
- (b) the standard of anonymization;
- (c) whether the act of anonymization constitutes a use of personal health information.

It should be noted that there are no definitive answers to the questions that arise in this area and this presentation should not be construed as providing legal advice on any of the matters considered.

### II The Exclusion of De-Identification Information from the Definition of Personal Health Information

PHIPA applies to 'personal health information', which is defined to mean "*identifying* information about an individual in oral or recorded form" if the information meets the criteria established in s.4(1) of the Act.<sup>1</sup> PHIPA further defines 'identifying information' to mean:

Information that identifies an individual or for which it is reasonably foreseeable in the circumstances that it could be utilized, either alone or with other information, to identify an individual.

---

<sup>1</sup> These criteria are if the information: relates to the physical or mental health of the individual, including information that consists of the health history of the individual's family, 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, is a plan of service within the meaning of the *Long-Term Care Act, 1994* for the individual, relates to payments or eligibility for health care in respect of the individual, relates to the donation by the individual of any body part of bodily substance of the individual or is derived from the testing or examination of any such body part of bodily substance, identifies an individual's substitute decision-maker. (s.4(1) a-g)

Thus, by definition, information that does not identify an individual or from which it is not reasonably foreseeable in the circumstances that it could be used, either alone or in combination with other information, to identify an individual would not constitute personal health information for the purposes of PHIPA.

PHIPA also contains a section that provides for the Minister to direct that personal health information be disclosed to a health data institute for purposes related to the analysis of the health system.<sup>2</sup> The information disclosed to the health data institute may also be linked with other information if so directed by the Minister. This process is done with the oversight of the Privacy Commissioner. For the most part, information that is provided to the Minister (or others) as a result of the data analysis must be in de-identified form. In this section (and for its purposes), PHIPA defines de-identify to mean:

to remove 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...

As can be seen, this definition flows from the definition of personal health information and the requirement of identifiability that is contained in that definition.

A final, relevant section in PHIPA relates to permitted uses to which health information can be put. Section 32 relates to the uses of health information permitted by PHIPA and these include for the purpose of “modifying the information in order to *conceal* the identity of the individual.” It should be noted that this section does not include, as a permitted use, the de-identification or anonymization of information (the only section that expressly permits this is the section identified above, which relates to de-identification for an express purpose). Moreover, the section that does relate to de-identification, which clearly contemplates the health data institute holding the key to re-identifying de-identified information provided to the Minister, does not use the term ‘conceal’ when it refers to de-identification. Which is to say that the terms conceal and de-identification are likely to mean different things under PHIPA.

### **III Standard of De-Identification**

In order to understand the complexities of determining whether or not information does or could be expected to identify an individual it is important

---

<sup>2</sup> S.47. Note that a health information institution is an institution approved by the Minister that must meet specified conditions concerning its corporate objects and privacy and confidentiality procedures and practices.

to understand something about the 'identifiers' contained in the information. Information may contain what is called, direct identifiers, "these are variables such as name and address, health insurance number, etc., that provide an explicit link"<sup>3</sup> to the person who is the subject of the information. If these direct identifiers are present on a piece of information then the information will probably be squarely within the ambit of the legislation.<sup>4</sup> However, even though information does not contain direct identifiers, it may contain indirect identifiers, which "are variables such as date of birth, sex, marital status, area of residence, occupation, type of business, etc. that, in combination, could be used to identify an individual."<sup>5</sup> It is important to note that there are gradations of identifiability, ranging from clearly identifiable to impossible to identify through any combination with other data that could re-link back to an identifiable individual.<sup>6</sup> Between these two poles is information which, on its face, does not directly link to an identifiable individual but *could be* linked to an identifiable individual either because somebody holds the 'key' to make this re-linking possible or because the information contains data elements that, in combination with other datasets, could be used to re-link to an identifiable individual.<sup>7</sup>

Dr. Latanya Sweeney has demonstrated that by linking three shared variables (e.g. data of birth, a portion of a ZIP code, and gender) from two sets of data (voter list and medical data), apparently anonymous medical data could be re-identified.<sup>8</sup> This study relied on the general availability of a matching data source, which in this case was a voter list that could be purchased for \$20. Dr. Sweeney uses the term *quasi-identifiers* for those

---

<sup>3</sup> Canadian Institutes of Health Research, *Best Practices for Protecting Privacy in Health Research (September 2005)* (Ottawa: Public Works and Government Services, September 2005) ["CIHR, *Best Practices*"] at p. 111.

<sup>4</sup> Unless the information is excluded by other provisions in the legislation; for example, private sector legislation in Canada will sometimes exclude business contact information such as might be found on a business card if the information is used for a purpose that is consistent with the reason for its collection (i.e. a business purpose).

<sup>5</sup> *Ibid.*

<sup>6</sup> For example, the CIHR Best Practices document, *supra* note 3 at p. 33, describes the following gradation: **(1) Directly identifiable:** The data contains direct identifiers of an individual (e.g. name, address, health number). **2) Coded: Single coded:** A participant's data are assigned a random code. Direct identifiers are removed from the dataset and held separately. The key linking the code back to direct identifiers is available only to a limited number (e.g. senior members) of the research team.

ii) **Double or multiple coded:** Two or more codes are assigned to the same participant's data held in different datasets (e.g. health administrative data, clinical data, genetic samples and data). The key connecting the codes back to participants' direct identifiers is held by a third party (such as the data holder) and is not available to the researchers.

**3) Not directly identifiable and not coded:** Direct identifiers were never collected or have been deleted, and there is no code linking the data back to the individual's identity.

**4) Non-identifiable:** Any element or combination of elements that allows direct or indirect identification of an individual was never collected or has been removed, although some elements may indirectly identify a group or region. There is no code linking the data back to the individual's identity.

<sup>7</sup> See, L. Sweeney, *Uniqueness of Simple Demographics in the U.S. Population*, LIDAPWP4. Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA: 2000, which describes the re-linking of apparently anonymous data through matching data elements such as date of birth and postal code with the same elements contained in (readily available) voters' lists.

<sup>8</sup> L. Sweeney, *Uniqueness of Simple Demographics in the U.S. Population*, LIDAPWP4. Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA: 2000.

variables that, while not explicit like a name or address, can nevertheless, in combination with an external data source, be used to re-identify data. In her work on *k*-anonymity,<sup>9</sup> Dr. Sweeney notes that data-holders who wish to release data anonymously often do not know what data sources are available to the data recipient and are therefore unaware of which quasi-identifiers in their data set are risky. Consequently, release of data could be re-identified through the use of quasi-identifiers. Although Dr. Sweeney focused on externally (publicly) available data-sources, her insight with respect to the ability to re-identify through the use of quasi-identifiers would be equally applicable to the combination of two (or more) private data sets.

In Canada, Dr. Khaled El Emam has tried to replicate Dr. Sweeney's research in Ontario using her three variables, data of birth, postal code and gender. He has found that there is no comparable data set that is externally available to enable the same re-linking in the case of medical data. His study did find, however, that readily available information for doctors and lawyers (at a cost) did permit replication of Dr. Sweeney's work. Dr. El Emam has also conducted a qualitative study on how persons engaged in clinical research perceive privacy risks. Through interviews with 20 persons – investigators, study coordinators, Research Ethics Board (REB) members and IT personnel – Dr. El Emam found that while REBs may require anonymization there is no systematic or evidence-based approach concerning how this will be achieved. For example, although data limitation (data with some variables eliminated) was the method used for anonymization, knowledge of which variables to remove or which variables were high-risk was lacking and there was wide variation among practices. In general, decisions were made on the basis of intuition and hearsay rather than justified according to evidence. He also found that no one used statistical methods extensively. Dr. El Emam's findings are supported by the study of REB chairs and coordinators by Don Willison and others, which found considerable variation in the ability to recognize the potential for re-identification through the combination of variables.<sup>10</sup>

It is not clear from PHIPA what standard of de-identification should be employed to be assured that information no longer meets the definition of personal health information. In particular, if direct identifiers are not contained on the face of the information, but the information could be re-identified through combination with other data or by obtaining the 'key' held by another institution is the information still considered personal health information for the purposes of the Act?

---

<sup>9</sup> L. Sweeney. *k*-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 2002; 557-570, online: <<http://privacy.cs.cmu.edu/dataprivacy/projects/kanonymity/kanonymity.pdf>>.

<sup>10</sup> The findings of Drs El Emam and Willison were discussed at a workshop on the Electronic Health Record (<http://www.ocri.ca/ehip2005/workshop.html>) and should be published shortly.

Different jurisdictions have taken different approaches to these issues. For example, in the United States the *Health Insurance Portability and Accountability Act of 1996*<sup>11</sup> and specifically privacy rule for research,<sup>12</sup> lists eighteen data elements and if any of these data elements are contained in the information then it is not considered de-identified. These elements are:

1. Names
2. All geographic subdivisions smaller than a State, including:
  - street address
  - city
  - county
  - precinct
  - zip codes and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly-available data from the Bureau of the Census: (1) the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people, and (2) the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.
3. Telephone numbers
4. Fax numbers
5. E-mail addresses
6. Social Security numbers
7. Medical record numbers
8. Health plan beneficiary numbers
9. Account numbers
10. All elements of dates (except year) for dates related to an individual, including:
  - birth date
  - admission date
  - discharge date
  - date of death
  - all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying numbers, characteristics, or codes

In addition, the Privacy Rule permits information to be de-identified by

---

<sup>11</sup> Public Law 104-191, 104<sup>th</sup> Congress. Online: <http://aspe.hhs.gov/admsimp/pl104191.htm>.

<sup>12</sup> [http://privacyruleandresearch.nih.gov/pr\\_08.asp#8a](http://privacyruleandresearch.nih.gov/pr_08.asp#8a)

entities covered by the legislation; however, the entity can only consider the information to be de-identified if the eighteen elements have been removed AND the covered entity has “no actual knowledge that the remaining information could be used alone or in combination with other information to identify the individual who is the subject of the information.” The Privacy Rule does provide for an alternative way to certify that information has been de-identified even if some of eighteen data elements remain. This is done through the use of an expert in statistics and is described as follows:

Covered entities may also use statistical methods to establish de-identification instead of removing all 18 identifiers. The covered entity may obtain certification by “a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable” that there is a “very small” risk that the information could be used by the recipient to identify the individual who is the subject of the information, alone or in combination with other reasonably available information. The person certifying statistical de-identification must document the methods used as well as the result of the analysis that justifies the determination. A covered entity is required to keep such certification, in written or electronic format, for at least 6 years from the date of its creation or the date when it was last in effect, whichever is later.

In the United Kingdom, the Data Commissioner has taken a fairly strong position with respect to de-identification for the purpose of excluding information from the ambit of the U.K. legislation.<sup>13</sup> In the Legal Guidance the Data Commissioner states:

The Commissioner recognises that the aim of anonymisation is to provide better data protection. However, true anonymisation may be difficult to achieve in practice. Nevertheless, the Commissioner would encourage that, where possible, information relating to a data subject, which is not necessary for the particular processing being undertaken, should be stripped from the personal data being processed. This may not amount to anonymisation but is in line with the requirements of the Data Protection Principles.

The Commissioner considers anonymisation of personal data difficult to achieve because the data controller may retain the original data set from which the personal identifiers have been stripped to create the “anonymised” data. The fact that the data controller is in possession of this data set which, if linked to the data which have been stripped of all personal identifiers, will enable a living individual to be identified, means that all the data, including the data stripped of personal

---

<sup>13</sup> U.K. Information Commissioner *Data Protection Act 1998: Legal Guidance* (London: Information Commissioner, 2002).

identifiers, remain personal data in the hands of the data controller and cannot be said to have been anonymised. The fact that the data controller may have no intention of linking these two data sets is immaterial. A data controller who destroys the original data set retaining only the information which has been stripped of all personal identifiers and who assesses that it is not likely that information will come into his possession to enable him to reconstitute the data, ceases to be a data controller in respect of the retained data. Whether or not data which have been stripped of all personal identifiers are personal data in the hands of a person to whom they are disclosed, will depend upon that person being in possession of, or likely to come into the possession of, other information which would enable that person to identify a living individual.

It should be noted that the disclosure of personal data by a data controller amounts to processing under the Act. For example: The obtaining of clinical information linked to a National Health Service number by a person having access to the National Health Service Central Register will amount to processing of personal data by that person because that person will have access to information enabling him to identify the individuals concerned. It will be incumbent upon anyone processing data to take such technical and organisational measures as are necessary to ensure that the data cannot be reconstituted to become personal data and to be prepared to justify any decision they make with regard to the processing of the data. For example: In the case of data collected by the Office of National Statistics, where there is a disclosure of samples of anonymised data, it is conceivable that a combination of information in a particular geographic area may be unique to an individual or family who could therefore be identifiable from that information. In recognition of this fact, disclosures of information are done in such a way that any obvious identifiers are removed and the data presented so as to avoid particular individuals being distinguished. If data have been stripped of all personal identifiers such that the data controller is no longer able to single out an individual and treat that individual differently, the data cease to be personal data. Whether this has been achieved may be open to challenge. Data controllers may therefore be required to justify the grounds for their view that the data are no longer personal data.

If there is a known link back to the individual (which may be held by the entity de-identifying the information) then the Data Commissioner does not consider the information to be de-identified. Moreover, the Data Commissioner places the onus of demonstrating that the information is appropriately de-identified on the entity responsible for its de-identification.

## **IV Distinguishing Good Privacy Practices from Removal from PHIPA's Provisions**

As noted in the excerpt from the U.K. Data Commissioner above, in many instances it may be a good privacy practice to conceal the identity of an individual from those who have no need to know the identity in order to perform a specific function. Concealing identity for this purpose (to greater protect privacy by limiting the disclosure of identity on a need to know basis) must be distinguished from de-identifying information for the purposes of removing it from the ambit of PHIPA entirely. In the former case, the provisions of PHIPA will continue to apply and must be followed. In the latter case, once de-identified, the information is no longer subject to PHIPA's provisions.

Although PHIPA does not provide clear guidance as to when information will be sufficiently de-identified for the purposes of removing it from the Act's ambit, there is sufficient guidance from other jurisdictions to indicate that caution should be deployed before concluding that information no longer meets the test of identifying an individual and is therefore excluded from the provisions of PHIPA.

## **V The Authority to De-Identify**

To some extent there is an assumption that the act of de-identifying or anonymizing information can occur without constraint and in particular without the consent of the person to whom the information pertains (prior to de-identification). This view enjoys some support from one of the few cases that has considered the issue directly. When this question was raised before the U.K. Court of Appeal, the act of anonymizing data was seen to be unproblematic.<sup>14</sup> It should be noted that the Legal Guidance of the U.K. Data Commissioner, noted above, occurred after this decision. There are also decisions by the federal and Alberta privacy commissioners that could be seen to imply that the act of anonymisation is unproblematic with respect to patient prescription data that has been de-identified.<sup>15</sup> In both cases, while not the issue under consideration, neither Commissioner found the act of removing patient identifiers from prescriptions to be problematic.

Whether PHIPA supports the position that the de-identification of information for the purposes of removing it from the ambit of the Act is unproblematic is unclear. There are certainly reasons to assert that the de-identification of personal health information would be a use of the information and therefore

---

<sup>14</sup> *Re. Source Informatics Ltd.* [1999] E.W.J. No 6880 (C.A.)

<sup>15</sup> See, Privacy Commissioner of Canada, Case Summary #15, online: [http://www.privcom.gc.ca/media/an/wn\\_011002\\_e.asp](http://www.privcom.gc.ca/media/an/wn_011002_e.asp) and Office of the Information and Privacy Commissioner of Alberta, Order H2002-003 (note this order has been stayed pending judicial review).

would be subject to the Act's provisions (particularly with respect to consent). Support for this assertion can be found in at least two places.

First, the provisions that relate to the Minister's requirement that information be provided to a data institute for the purposes of health system analysis (noted above) make specific reference to de-identification and clearly direct the data institute to de-identify information. This direction can also be characterized as an authorization to de-identify the information and the issue of consent becomes irrelevant.

Second, the provisions that relate to permitted uses of health information recognize that the modification of information for the purposes of concealing identity is a use of information and a use that is expressly permitted. It would seem reasonable to assert that de-identifying information is also a modification of the information and also subject to the PHIPA's provisions. Since de-identification is not listed as a generally permitted use under PHIPA then it also seems reasonable to assert that it is an act that would require consent before being conducted.<sup>16</sup>

---

<sup>16</sup> S.29 prohibits a health custodian from using health information unless an individual has consented to the use or the use is permitted or required under the Act.