David Young Law
Compliance Bulletin

July 2023

OPC provides rules for collecting and using non-personal data

In the Report of its investigation into the collection and use by the Public Health Agency of Canada (PHAC) of mobile location data during the COVID-19 pandemic, released at the end of May,[1] the federal Office of the Privacy Commissioner provides important guidance for organizations seeking to utilize information about individuals anonymized so that it is no longer considered personal information, and is therefore outside application of existing privacy laws.

Usefully, the Report also will serve to inform the review at Committee of Bill C-27, the government's *Digital Charter Implementation Act, 2022,* which is scheduled for this fall. It will be recalled that the Bill now includes provision for the categories of "de-identified information" and "anonymized information", the former being information that continues to be subject to the privacy law and the latter, outside the law.

The Report, entitled *Investigation into the collection and use of de-identified mobility data in the course of the COVID-19 pandemic,* examined whether mobile device data collected and used by PHAC in its response to the pandemic contained personal information as defined under the *Privacy Act*. Specifically, whether PHAC and its data providers implemented de-identification techniques and safeguards against re-identification deemed sufficient to reduce the risk of an individual being identified below the threshold of a "serious possibility". The "serious possibility" threshold is the rule articulated in relevant judicial guidance for determining whether data should be considered personal information or, conversely, whether it can be considered sufficiently de-identified to no longer be considered personal information.[2]

Following publication by PHAC of a Request for Proposal for providers to assist it to continue to acquire mobility data, media articles raised privacy concerns and the OPC received certain complaints, claiming that PHAC secretly had collected and used mobility data without authorization, in contravention of the *Privacy Act*.

Based on the information provided by PHAC and its data providers, and a review of the concepts of identification and de-identification, the OPC determined that the combination of the de-identification measures used by PHAC's data providers and the safeguards against re-identification implemented by those providers and by PHAC reduced the risk of identifying individuals below the "serious possibility" threshold, sufficient to conclude that

---

[1] *Investigation into the collection and use of de-identified mobility data in the course of the COVID-19 pandemic*, Office of the Privacy Commissioner of Canada, May 29, 2023.

2 *Gordon v. Canada (Health)*, 2008 FC 258. Information is personally identifiable if 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.

the data in question did not constitute personal information under the *Privacy Act*. Consequently, the OPC concluded, the complaints were not well-founded.

Specifically, the OPC evaluated the degree to which data collected by PHAC could be linked to identifiable individuals either directly or indirectly, through inference and/or in association with other data sources, and concluded that due to the measures taken to de-identify the data, and the protections used to reduce the risk of re-identification, there was no serious possibility that the information collected by PHAC and its providers could identify an individual.

However, consistent with the parliamentary ETHI Committee's report in May 2022 regarding the same set of facts,[3] the OPC recommended that public organizations such as PHAC, should be transparent with regards to their use of de-identified information and make every effort to publicize such uses and to inform concerned individuals of its purposes, the sources of data, and the safeguards implemented to protect it and maintain its anonymity.

The OPC also took a cue from the ETHI Committee in recommending that the federal privacy laws be amended to include a clear legal framework that defines the different types of de-identified data and that specifies the rules that govern the collection, use and disclosure, of such data.

De-identified mobility data and the risks of re-identification

PHAC had determined that gaining "mobility insights" on population movements, interactions and gatherings would assist in understanding how the coronavirus spread. It also determined that such mobility insights could be useful in planning, monitoring, and refining the effectiveness of key measures implemented by health authorities such as stay at home, quarantine and lockdowns.

Mobility insights collected by PHAC were derived from data about the movements of individuals over time (mobility data) that PHAC indicated was de-identified and aggregated information. This information was deduced from location data that is continuously produced by devices typically in the same location as their users. The most common examples of these devices are mobile phones.

Through its service providers, PHAC collected mobility data from two streams: aggregated mobile cell-tower data and individual mobile device geolocation data. PHAC procured the first data stream from TELUS and the Communications Research Centre Canada (CRC), a branch of Innovation, Science and Economic Development Canada (ISED) which was processing the TELUS Data to generate reports of aggregated data and statistics. PHAC acquired the second stream, geolocation data, through commercial providers that collect and supply this data.

The OPC set forth a description of the processes that can be applied to personal data in order to make it "non-personal" so as to meet the threshold of no serious possibility of re-identification. It then reviewed the procedures applied to the mobility information collected by PHAC to determine whether they were sufficient.

[3] *Collection And Use of Mobility Data by the Government of Canada and Related Issues,* Report of the Standing Committee on Access to Information, Privacy and Ethics, May 2022.

As articulated by the OPC, "de-identification" is the process by which any personal identifiers, such as names, phone numbers or mobile device IDs, are stripped from the data about a specific individual and typically replaced with a randomly assigned identifier. In the OPC's view, de-identification alone is insufficient to render mobility data outside the scope of the law. Because such data provides the geographic location of where a person or device has been over time, it may be used to infer information about the device user, such as their place of home or work which in turn can be compared to other readily available information to link the de-identified data to the user. The OPC cited several studies that demonstrated the ready ability of de-identified mobile location data to be re-identified using a small number of other readily available "spatio-temporal points".[4]

The OPC posited that de-identified data retains a residual risk of re-identification that can be attributed to both factors *intrinsic to the data* itself and the de-identification techniques – meaning processes such as removing personal identifiers, applying encryption and aggregating – as well as factors *external to the data.* External factors include the availability of additional data that can be linked to the de-identified data, such as who has access to the dataset and for what purposes, their motivation to re-identify data and their knowledge that a specific individual's information is included in the dataset, as well as the expertise and the resources available for application to a re-identification process.

Protections against re-identification

In the context of collecting mobility data, the OPC focussed on whether sufficient additional protections against re-identification were in place – addressing both intrinsic and external factors, over and above the primary intrinsic de-identification processes.

In this regard, the OPC considered firstly whether the data as collected was sufficiently *aggregated* to render it non-identifiable per se. Aggregation reduces the risk of re-identification by combining data about multiple individuals together so that in each data element or "cell" any one individual's own data is obscured. Where an organization has access to aggregated mobility data, a sufficient number of individuals would be aggregated in each cell to reasonably reduce the risk of extrapolating the data of a single individual, in accordance with current statistical guidelines or expert advice. [5]

Secondly, the OPC considered whether the *contractual and physical protections* on access and use reduced the risk of re-identification by limiting the number of individuals (or organizations) that could have the opportunity to attempt re-identification, and the likelihood those individuals will attempt re-identification.

---

[4] *Unique in the Crowd: The privacy bounds of human mobility*, a study conducted on 1.5 million users of a mobile phone operator in a western country concluded that four spatio-temporal points are enough to uniquely identify 95% of the individuals because mobility traces are highly unique and consistent.

[5] Regarding recommended cell sizes, the OPC noted that the Treasury Board Secretariat's Privacy Implementation Notice 2020-03 (Protecting privacy when releasing information about a small number of individuals), states "there is no minimum cell size that is appropriate for all data releases, and Treasury Board of Canada Secretariat policies do not specify a mandatory minimum cell size. However, the following best practices may serve as a starting point for a case-by-case analysis: A minimum cell size of 10 is often cited as a best practice for public data releases of data that is less sensitive, while a minimum cell size of 20 is cited for more sensitive data".

Robust contractual and physical protections for properly de-identified mobility data need to be in place to: (a) limit access to that data to a specified number of individuals; and (b) limit the purposes for which individuals are permitted to use the data (i.e. not re-identification attempts). Protection against re-identification attempts should include, at a minimum, a contractual prohibition, and safeguard controls such as audit capability and monitoring of data access and use to guard against unauthorized attempts.

In the OPC's view, in order for PHAC's information to be considered 'non-personal' and therefore outside of the scope of the *Privacy Act,* both of these requirements would need to be met.

Analysis of the PHAC data

With respect to the cell-tower data, the OPC analyzed the safeguards that TELUS, the CRC and PHAC had put in place to reduce the risk of re-identification. It categorized these as follows.

Prior de-identification: all device IDs were encrypted so that the information that the CRC accessed, on behalf of PHAC, did not contain any direct identifiers - a robust algorithm was applied to hash direct identifiers and de-identify data at the device-level.

Aggregation: the CRC, on behalf of PHAC, was restricted to importing only aggregated data from TELUS, and no data at the device level, even though it was de-identified, could be copied outside of the TELUS platform. The data that CRC imported was aggregated spatially, at least at the census sub-division level, temporally to span at least over a 24 hour period, and with cells that contained at least twenty devices.[6]

Release model: the relevant data (TELUS' Data for Good) is provided as a non-public release limiting availability to a select number of identified recipients. As a condition of receiving the data, recipients must agree to terms and conditions regarding the privacy and security of the data in a data sharing agreement including requiring users to not attempt to re-identify it.

Contractual clauses: Both PHAC and TELUS included in the contract governing their commercial relationship binding provisions to use only de-identified information. Specifically, in the contract's statement of work that PHAC , TELUS was required to provide PHAC with access to de-identified information that ensured data anonymization in order to generate aggregate indicators and insights on the mobility of individuals in Canada. Additionally, TELUS' data sharing terms stipulated that PHAC must not use the derived data for any other purpose except for that specified and that it may not correlate, associate, link or combine any of the derived data with other data sources, except as consented to by TELUS.

[6] This minimum cell size was compliant with Treasury Board guidance (note 4, above) and above the minimum threshold (11) that was determined in an expert report submitted to the Federal Court in the recent case of *Cain v. Canada (Minister of Health)* that dealt with risk of re-identification.

A similar analysis addressed safeguards respecting the mobile device geolocation data stream collected by PHAC's commercial data providers.

In both data streams, information under PHAC's and/or the CRC's control was aggregated according to several criteria, either temporally and/or spatially, with minimum cell sizes between five and twenty, a minimum size that is accepted and recommended by experts, which increased the degree of data anonymity of the datasets under PHAC's or the CRC's control.

The OPC noted also that only select employees from PHAC/CRC were authorized to access mobility data either at the device-level, on a 'view only' basis, or in aggregated form, and that PHAC and its epidemiologists in this specific project were looking for macro trends on the population movement and were not engaged in contact-tracing. Therefore, it concluded that PHAC had no motivation to re-identify - as was expressly reflected in its contracts and the RFP related to the matter.

Based on its analysis, including taking recognition of the accepted practices in this field, the OPC concluded that the threshold of a serious possibility of re-identifying the data was not met and therefore the data as collected and used by PHAC was non-personal and was outside the scope of the *Privacy Act*.

Takeaways

The OPC's Report addresses usefully two important techniques for generating non-personal data (meaning data that is outside of the statutory privacy laws): data aggregation and administrative measures protective against re-identification.

The Report is subject-specific to mobility location data. However it contains important insights of wider application, both in the analysis and conclusions as well as in the research referred to (as cited in the footnotes) for potential users of non-personal information under the current privacy laws. It also provides guidance for legislators who are striving to modernize those laws to embrace explicitly categories of such information that remain subject to the laws as well as categories that are outside of the laws. In this regard, both Quebec's amended private sector privacy law (Law 25) and the proposed federal amended law, the *Consumer Privacy Protection Act,* part of Bill C-27, provide for the categories of "de-identified information" being non-personal information that remains subject to the law, and "anonymized information" being a category intended to be outside application of the law.

While using different terminology, the OPC's Report should be understood to address the category of information intended to be outside the privacy law – i.e. anonymized information. Both the Quebec and the federal initiatives stipulate this category of information as required to be generated according to "generally accepted best practices". The OPC Report can be read as an articulation of an example of such practices, specific to mobility location data but with potentially broader application.

For more information please contact: David Young 416-968-6286 david@davidyounglaw.ca