Dalhousie Law Journal

Volume 47 | Issue 2 Article 6

9-2024

Public Sector Use of Private Sector Personal Data: Towards Best Practices

Teresa Scassa
University of Ottawa Faculty of Law

Recommended Citation
Teresa Scassa, "Public Sector Use of Private Sector Personal Data: Towards Best Practices" (2024) 47:2 Dal LJ.

Teresa Scassa*

Public Sector Use of Private Sector Personal Data: Towards Best Practices

Governments increasingly seek to use personal data sourced from the private sector for purposes that range from the generation of statistics to municipal planning. The data collected by companies is often high volume and rich in detail. Location and mobility data—which have many applications—are collected by multiple private sector actors, from cellular service providers to app developers and data brokers. Financial sector organizations amass rich data about the spending and borrowing habits of consumers. Even genetic data is collected by private sector companies. The range of available data is constantly growing as more and more data is harvested, and as companies seek secondary markets for their data.

This paper explores issues raised by public sector access to and use of personal data held by the private sector. The analysis is framed around two examples from Canada that involved actual or attempted access by government agencies to private sector personal data for public purposes. The first involved Statistics Canada's attempt to collect and use data held by credit monitoring companies and financial institutions to generate economic statistics. The second was the use, during the COVID-19 pandemic, of mobility data by the Public Health Agency of Canada (PHAC) to assess the effectiveness of public health policies in reducing the transmission of COVID-19 during lockdowns. Each of these instances led to public outcry and to investigation or inquiry. The paper examines the legal bases for the use of private sector data by government, the safeguards available to protect privacy, and the framing of concerns over the use of these data by different interested parties. Given that legal frameworks for data sharing may not keep pace with data needs and methods, this paper also takes a normative approach which examines whether and in what circumstances such data sharing should take place.

Les gouvernements cherchent de plus en plus à utiliser les données personnelles provenant du secteur privé à des fins qui vont de la production de statistiques à la planification municipale. Les données collectées par les entreprises sont souvent volumineuses et très détaillées. Les données de localisation et de mobilité—qui ont de nombreuses applications—sont collectées par de multiples acteurs du secteur privé, des fournisseurs de services cellulaires aux développeurs d'applications et aux courtiers en données. Les organisations du secteur financier accumulent de riches données sur les habitudes de dépense et d'emprunt des consommateurs. Même les données génétiques sont collectées par des entreprises du secteur privé. L'éventail des données disponibles ne cesse de s'élargir, car de plus en plus de données sont collectées et les entreprises cherchent des marchés secondaires pour leurs données.

Dans le présent article, nous explorons les questions soulevées par l'accès du secteur public aux données personnelles détenues par le secteur privé et par leur utilisation. L'analyse s'articule autour de deux exemples canadiens qui impliquent l'accès réel ou la tentative d'accès par des agences gouvernementales à des données personnelles du secteur privé à des fins publiques. Le premier concerne la tentative de Statistique Canada de collecter et d'utiliser des données détenues par des sociétés de surveillance du crédit et des institutions financières pour produire des statistiques économiques. La seconde est l'utilisation, pendant la pandémie de COVID-19, de données sur la mobilité par l'Agence de santé publique du Canada (ASPC) pour évaluer l'efficacité des politiques de santé publique dans la réduction de la transmission de la COVID-19 pendant les fermetures d'établissements. Chacun de ces cas a suscité des protestations publiques et a donné lieu à une enquête ou à une investigation. Dans l'article, nous examinons les bases juridiques de l'utilisation des données du secteur privé par le gouvernement, les garanties disponibles pour protéger la vie privée et la formulation des préoccupations relatives à l'utilisation de ces données par les différentes parties intéressées. Étant donné que les cadres juridiques pour le partage des données peuvent ne pas suivre l'évolution des besoins et des méthodes en matière de données, nous adoptons également une approche normative qui examine si et dans quelles circonstances un tel partage de données devrait avoir lieu.

[*] Canada Research Chair in Information Law and Policy, University of Ottawa. This paper was written during the course of a visitorship at the University of Macerata, Faculty of Law in Italy. A version of this paper was presented at the Data for Policy Conference held in Brussels in December 2022.

Introduction

Governments seeking to make data-driven decisions require sufficient data to do so. Although they may already hold large stores of administrative data, their ability to collect new or different data is limited both by law and by practicality. In our networked, Internet-of-Things-connected society, the private sector has become a source of abundant data about almost anything—but particularly about people and their activities. Private sector companies continuously collect a wide variety of personal data, often in high volumes, and rich in detail. For example, many different actors collect location and mobility data from cellular service providers to app developers. Financial sector organizations amass rich data about the spending and borrowing habits of consumers. Even genetic data is collected by private sector companies. The range of available data is constantly broadening as ever more is harvested, and as companies seek secondary markets for the data they collect.

Public sector use of private sector data is, however, fraught with important legal and public policy considerations. Chief among these is privacy; government access to such data raises concerns about undue government intrusion into private lives and habits.[1] Data protection issues in this context implicate both public and private sector actors, and include notice and consent, as well as data security. Where private sector data is used to shape government policies and actions, important questions about ethics, data quality, the potential for discrimination, and broader human rights questions also arise. Alongside these issues are interwoven concerns about transparency, as well as necessity and proportionality when the public sector conscripts privately collected data.

This paper explores issues raised by public sector access to and use of personal data held by the private sector. It considers how such data sharing is legally enabled and within what parameters. Given that laws governing data sharing may not always keep pace with data needs and public concerns, this paper also takes a normative approach which examines whether and in what circumstances such data sharing should take place. To provide a factual context for discussion of the issues, the analysis in this paper is framed around two recent examples from Canada that involved actual or attempted access by government agencies to private sector personal data for public purposes.

The two case studies chosen are different in nature and scope. The first was the attempted acquisition and use by Canada's national statistics organization, Statistics Canada (StatCan), of data held by credit monitoring companies and financial institutions to generate economic statistics. The second was the use, during the COVID-19 pandemic, of mobility data by the Public Health Agency of Canada (PHAC) to assess the effectiveness of public health policies in reducing the transmission of COVID-19 during lockdowns. The StatCan example involves the compelled sharing of personal data by private sector actors; while the PHAC example involves a government agency that contracted for the use of anonymized data and analytics supplied by private sector companies. Each of these instances generated significant public outcry. This negative publicity no doubt exceeded what either agency anticipated. Both believed that they had a legal basis to gather and use the data or analytics, and both believed that their actions served the public good. Yet the outcry is indicative of underlying concerns that had not properly been addressed.

[1] In *R v Bykovets*, 2024 SCC 6 at para 78, the majority of the Supreme Court of Canada comments on the informational power of the private sector, noting: "By concentrating this mass of information with private third parties and granting them the tools to aggregate and dissect that data, the Internet has essentially altered the topography of privacy under the *Charter*. It has added a third party to the constitutional ecosystem, making the horizontal relationship between the individual and state tripartite."

Using these two quite different cases as illustrations, the paper examines the issues raised by the use of private sector data by government. Recognizing that such practices are likely to multiply, it also makes recommendations for best practices. Although the examples considered are Canadian and are shaped by the Canadian legal context, most of the issues they raise are of broader relevance. Part I of this paper sets out the two case studies that are used to tease out and illustrate the issues raised by public sector use of private sector data. Part II discusses the different issues and makes some recommendations for future policy in the area.

I. *The data sharing examples*

1. *Private financial data and national statistics: the case of StatCan*

In 2018, StatCan initiated two projects that involved the collection of financial data about Canadians from the private sector. The goal of each was to collect more detailed, accurate and lower-cost economic statistics for Canada. One project obtained information from TransUnion, a credit reporting agency; the other sought to collect a range of personal data from financial institutions such as banks and credit card companies.

Both instances involved the collection of what is known as "administrative data." This label is applied to data that are already in existence, having been collected for other purposes. The practice of national statistics organizations (NSOs) using public sector administrative data to generate statistics is well-established.[2] The use of private sector "administrative data" is much less so. Nevertheless, StatCan is not alone among NSOs in exploring the potential uses of such data.[3] StatCan maintains

[2] Office of the Privacy Commissioner of Canada (OPCC), *Statistics Canada: Invasive data initiatives should be redesigned with privacy in mind*, complaint under the *Privacy Act* (Ottawa: OPCC, 2019), online: <www.priv.gc.ca> [perma.cc/V6EC-WU2P] [OPCC 2019]. StatCan has collected and used administrative data for statistics since 1921, although these have primarily been public sector administrative data. Currently, 40 per cent of the data used by StatCan are administrative data (*ibid* at para 3).

[3] See e.g. Panel on Improving Federal Statistics for Policy and Social Science Research Using Multiple Data Sources and State-of-the-Art Estimation Methods, "Using Private-Sector Data for Federal Statistics," in Robert M Groves & Brian A Harris-Kojetin, eds, *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy, Report of the National Academies of Sciences, Engineering and Medicine* (Washington DC, National Academic Press, 2017) online: <nap.nationalacademies.org/download/24652> [perma.cc/W9W9-4ZR7]; United Nations Statistics Division, "UN Committee of Experts on Big Data and Data Science for Official Statistics: Mandate and Terms of Reference of the UN-CEBD," online: <unstats.un.org/bigdata/about/mandate.cshtml> [perma.cc/H5PG-DCAW]: The United Nations has also established a working group on Big Data and although its mandate is broad it does include the use of big data (both personal and non-personal) in national statistics (*ibid* at para e); Statistics Canada, *Data Strategy*, 2019–2022, Catalogue No 89-26-00032020001, online: <www.statcan.gc.ca/en/about/datastrategy> [perma.cc/XU7T-R8S7] [StatCan Data Strategy]. StatCan describes the greater use of private sector administrative data as part of its process of modernization and its agenda as "creating strategic partnerships with other

that the use of private sector administrative data can improve the quality of statistical outputs (because of its detail and comprehensiveness) and it can reduce the costs of collecting data (because it already exists).[4] StatCan also notes that the use of administrative data enables the collection of data from those who would otherwise be unwilling to complete surveys.[5]

In the first case, in 2018, StatCan quietly obtained access to personal credit history information from TransUnion, a national credit-reporting company, for a project "to measure household debt on a periodic basis by collecting credit information of individuals directly from credit bureaus."[6] The data was collected pursuant to a formal data sharing-agreement and ranged from as far back as 2002 to the present day. The collected data "included approximately 600 elements of data containing personal information including name; age; data of birth; social insurance number; address and credit information."[7] The "credit information" included data about the types of credit that had been obtained by the consumer, as well as amounts, activity, and balances. StatCan's goal was to capture the data of roughly 80 per cent of the Canadian population; at the time the project was suspended following pushback from the public, it had collected data related to 24 million Canadians.[8]

Because StatCan sought to link the data obtained from TransUnion with other data obtained through its broader statistical data collection activities, the data received was personal data associated with identifiable

organizations and researching and discovering data inputs that can be used by statistical programs, such as administrative data, open data, found data, commercial data, crowdsourced data and web-scraped data, while respecting privacy and maintaining public trust." (*ibid* at 18).

[4] OPCC 2019, *supra* note 2 at para 2; The OPCC's first reflections on the StatCan plan to use more administrative data are found in its Annual Report of 2017–2018 (see Office of the Privacy Commissioner of Canada, *Trust by Verify: Rebuilding trust in the digital economy through effective, independent oversight*, (2017–18 Annual Report to Parliament on the *Personal Information Protection and Electronic Documents Act and the Privacy Act*) (Ottawa: OPPC, 2018), online: <priv.gc.ca/en/opc-actions-and-decisions/ar_index/201718/ar_201718> [perma.cc/6T39-RCEW] [OPCC 2017–2018]); A general justification for the use of administrative data is found at StatCan's Trust Centre, where it is noted: "traditional statistics-gathering methods are no longer sufficient to accurately measure Canada's economic and societal changes. That is why Statistics Canada's focus has shifted toward leveraging administrative data, using advanced technologies and developing new, cost-effective methods to link and integrate data from a variety of sources" (see Statistics Canada, "Modernization Projects" (26 July 2022) online: <www.statcan.gc.ca/en/trust/modernization> [perma.cc/787W-USEP]).

[5] See StatCan *Data Strategy*, *supra* note 3 at 5.

[6] OPCC 2019, *supra* note 2 at para 10.

[7] *Ibid*, at para 11.

[8] *Ibid*, at para 12.

individuals. It was subsequently pseudonymized[9] according to StatCan procedures.[10]

Interestingly, prior to any complaints, the Office of the Privacy Commissioner of Canada (OPCC), in its 2017–2018 Annual Report, indicated that it had received expressions of concern from private sector companies about StatCan's growing interest in access to private sector administrative data. Private sector actors were no doubt concerned about how such practices might impact both their obligations to secure customer data and their relationships with customers who might balk at having their personal financial data shared with a government agency. The report stated:

> We have recommended the agency consider whether it could achieve the same objectives by collecting customer information that has been de-identified before it is disclosed to the agency. We also suggested it limit collection of administrative data to what is needed for the specified purposes, and that it evaluate the necessity and effectiveness of this work on an ongoing basis. To ensure transparency, we recommended StatCan let the Canadian public know how and why it is increasing its collection of data from administrative and other non-traditional sources.[11]

It is unclear whether any steps were taken following this recommendation. In terms of transparency, under Canadian provincial credit reporting legislation, TransUnion is required to notify those about whom it maintains reports of any request for data about them. Its practice is to place a note on the affected individual's file. In the case of the StatCan data sharing, TransUnion placed notes regarding a "non-credit related inquiry" on files where data was shared. This note included a telephone number at StatCan where individuals could obtain further information about the data sharing project. Nevertheless, individuals would be unaware of these notes unless they requested access to their files with TransUnion. For its part, StatCan

[9] The pseudonymization of personal information typically involves the stripping of direct identifiers (such as names, government identification numbers, and so on) from the data. See e.g. EU, *Regulations, 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)* [2016] OJ, L 119/1, art 4(5), online (pdf): <eur-lex.europa.eu> [perma.cc/9UP6-3TRB] [GDPR]: The GDPR defines pseudonymization as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person"; for a discussion of pseudonymized data, see generally: Khaled El Emam et al, "The Seven States of Data: When is Pseudonymous Data Not Personal Information?" (2016) Future of Privacy Forum, online: <www.fpf.org> [perma.cc/22XE-7R6K].

[10] OPCC 2019, *supra* note 2 at para 13.

[11] OPCC 2017–2018, *supra* note 4 at heading Statistics Canada: Increased use of administrative data sources.

generated a short supplemental privacy impact assessment (PIA) relating to this project and added it to its general PIA. It shared this with the OPCC and later published it on its website.

The second administrative data project to attract media attention also began in 2018. StatCan sought access to the banking records of 500,000 Canadian households in an effort to generate more accurate—and lower cost—economic statistical data.[12] Financial institutions resisted this attempt, expressing concerns about, among other things, their potential liability under Canada's private-sector data protection laws.[13] The tension between the parties led to a leak of the project to the media; the coverage generated a public outcry and complaints to the federal Privacy Commissioner.[14] This led to an investigation by the OPCC, with findings critical of StatCan.[15] Nevertheless, the controversy did not end StatCan's desire to use private sector administrative data. Instead, it led to changes in how StatCan planned to approach the gathering of this kind of data and its development, in conjunction with the OPCC, of a normative framework for the use of private sector administrative data.[16]

Under the federal *Privacy Act*, a government entity can only collect personal data where it has legal authority to do so.[17] StatCan maintained that its legal authority to collect private sector administrative data came from section 13 of the *Statistics Act*,[18] which permitted it to compel access to any "documents" or "records" in the hands of a broad range of organizations (including corporations) when those records contained information useful for the generation of statistics. TransUnion was of the view that the provision obliged it to comply with StatCan's request for information, and it supplied a large volume of information to the statistical agency. The financial institutions, in contrast, challenged StatCan's legal

[12] OPCC 2019, *supra* note 2 at para 20.

[13] See Canadian Bankers' Association, News Release, "Statement from the CBA on the Statistics Canada Data Acquisition Project" (2 November 2018), online: <cba.ca/statement-by-the-cba-re-statistics-canada-data-acquisition> [perma.cc/9U8A-WDEE].

[14] Jean-Guy Prévost, "Big Brother, Big Data and Statistics Canada," *Policy Options* (16 January 2020), online: <policyoptions.irpp.org> [perma.cc/2385-US66].

[15] OPCC 2019, *supra* note 2 at para 168.

[16] *Ibid* at para 111. The necessity and proportionality framework is discussed at StatCan's online Trust Centre, although no documentation on the actual framework is provided (see Statistics Canada, "Principles of Necessity and Proportionality," online: <statcan.gc.ca/en/trust/address> [perma.cc/8WB5-9U4C] [Statistics Canada, "Necessity and Proportionality"]. See also Statistics Canada, "The Chief Statistician of Canada moderated a panel discussion at the 51st annual United Nations Statistical Commission" (4 March 2020), online: <statcan.gc.ca/en/trust/un-statistical-commission> [perma.cc/HU4L-NZJT].

[17] *Privacy Act*, RSC 1985, c P-21, ("No personal information shall be collected by a government institution unless it relates directly to an operating program or activity of the institution" at s 4).

[18] *Statistics Act*, RSC 1985, c S-19.

authority to compel disclosure of financial administrative data in their databases.

Section 13 of the *Statistics Act* reads:

> A person having the custody or charge of *any documents or records* that are maintained in any department or in any municipal office, corporation, business or organization, from which information sought in respect of the objects of this Act can be obtained or that would aid in the completion or correction of that information, shall grant access thereto for those purposes to a person authorized by the Chief Statistician to obtain that information or aid in the completion or correction of that information.[19]

A key interpretive issue was, therefore, whether what was being compelled was a "record" or "document" within the meaning of section 13. The financial institutions had argued that the data sought by StatCan was detailed and complex and was found throughout a series of databases that were maintained separately for internal and security purposes. In their view, the desired data thus did not exist as a record or document and could not be compelled under section 13. Although TransUnion's data was also contained in a database, it had taken the position that since the data sought by StatCan could be extracted from the database using the same technical process used to generate credit reports, the requested data fell within the definition of a "record." The Privacy Commissioner accepted TransUnion's interpretation of the application of section 13 in its particular case and assumed that the data collection from TransUnion was authorized under the *Statistics Act*. As a result, its collection complied with the legal authority requirement of the federal *Privacy Act*. However, the OPCC also agreed with the financial institutions that in their case, the data sought was not part of an existing "record." The OPCC therefore concluded that any collection of this data (had the program not been paused) would have been unauthorized.

Disagreement over the legal authority to obtain the data was an important part of the conflict between StatCan and the financial institutions. The legislative provision that enabled StatCan to compel access to records had been drafted in an era in which most records were compiled or produced in paper format. Amendments in 2017 to modernize the *Statistics Act* had provided for the Chief Statistician to make mandatory requests for information under a different provision but this would require a level of

[19] *Ibid*, s 13 [emphasis added].

transparency and pre-clearance that had not occurred in this case.[20] As a result, StatCan could not rely on this new provision for its authority.[21]

Quite apart from the legal authority of StatCan to compel disclosure of the data, the financial sector organizations were also concerned about their own potential liability under private sector data protection laws if they provided large volumes of sensitive customer data to StatCan. Such laws typically contain exceptions that allow for disclosure of personal data without knowledge or consent—and some of these exceptions may be specifically tailored to the statistical context. For example, article 89 of the EU's *General Data Protection Regulation* (GDPR) allows states to derogate from other GDPR requirements in the case of processing personal data for statistical purposes.[22] Nevertheless, any derogations are only permissible "in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes."[23] One of the rights that can be overridden in this manner is the data subject's right to object to processing of that data, although in the case of statistical purposes, that right to object can be limited only if "the processing is necessary for the performance of a task carried out for reasons of public interest."[24]

In Canada, the applicable law for the private sector was the *Personal Information Protection and Electronic Documents Act* (PIPEDA).[25] PIPEDA allows private sector organizations to disclose information without the knowledge or consent of those to whom it pertains where the disclosure is "for statistical, or scholarly study or research, purposes that cannot be achieved without disclosing the information, it is impracticable to obtain consent and the organization informs the Commissioner of the disclosure before the information is disclosed."[26] However, because the public and private sectors are subject to different data protection laws in Canada, this exception did not obviate the need for StatCan to have legal authority to collect the data under the *Privacy Act*. Further, sharing for statistical purposes under PIPEDA is not mandatory—section 7(3)(f) merely *permits* an organization to share data without knowledge or consent in the specified circumstances.

[20] *Ibid,* ss 8(2), 8(3). Mandatory requests for information have to be published before the request is made, and the minister of industry would have to be notified of the request 30 days before its publication.

[21] OPCC 2019, *supra* note 2 at para 62.

[22] GDPR, *supra* note 9, art 89.

[23] *Ibid*, art 89(2).

[24] *Ibid*, art 21(6).

[25] *Personal Information Protection and Electronic Documents Act*, SC 2000, c 5 [PIPEDA].

[26] *Ibid*, s 7(3)(*f*).

2. *Mobility data and public health: the case of PHAC*

In 2022, controversy erupted following media reports that the Public Health Agency of Canada (PHAC) had acquired and used the mobility data of Canadians during the COVID-19 pandemic. PHAC is a government entity charged with "[p]romoting and protecting the health of Canadians through leadership, partnership, innovation and action in public health."[27] According to the Minister of Health, "anonymized, aggregated mobility data was used to monitor the trajectory of the pandemic and how best to respond to it."[28] The goal was to help the government "determine how the public is responding to public health directives so that they can tailor their approach and communications."[29] It transpired that PHAC had two different contracts for mobility data: one with Telus, a Canadian cellular service provider, and one with BlueDot, a US-based data analytics company that focuses on infectious diseases.[30] The controversy led to hearings before the Canadian Parliament's Standing Committee on Access to Information, Privacy and Ethics (ETHI), which issued a report and recommendations. It also led to complaints filed with the OPCC, and a subsequent investigation and report of findings.[31]

In the case of its contract with Telus, PHAC accessed data and analytics via the company's Data for Good[32] platform. The available data was described as "aggregated indicators derived from cell-tower/operator location data, or in some cases aggregated and de-identified GPS location data."[33] The data source was Telus' roughly nine million subscribers. No actual mobility data was supplied to PHAC. Instead, the Data for Good program allows data scientists from contracting organizations such as PHAC to run queries on Telus' data. According to Telus, any such queries are reviewed to ensure that they are consistent with the purpose for which access was contracted. Further, Telus only releases results after they

[27] Government of Canada, "Public Health Agency of Canada mandate" (25 June 2018), online: <www.canada.ca/en/public-health/corporate/mandate.htm> [perma.cc/9LSH-YEP7l.

[28] House of Commons Standing Committee on Access to Information, Ethics and Privacy, *Collection and Use of Mobility Data by the Government of Canada and Related Issues*, 44-1 (22 May 2022) at 15, online (pdf): <ourcommons.ca/DocumentViewer/en/44-1/ETHI/report-4> [perma.cc/HD53-4M9Q] [ETHI Report].

[29] *Ibid*.

[30] BlueDot, online: <www.bluedot.global> [perma.cc/KZF7-VPVZ].

[31] Office of the Privacy Commissioner of Canada, *Investigation into the collection and use of de-identified mobility data in the course of the COVID-19 pandemic*, complaints under the *Privacy Act*, (Ottawa: OPCC, 2023), online: <priv.gc.ca> [perma.cc/4HR4-NQB9] [PHAC Investigation].

[32] Telus, "Data for Good: Benefiting Canadians through data insights" (2022), online: <telus.com/en/about/privacy/data-for-good> [perma.cc/CUS6-4DP4]. Telus' secure data analytics platform is Privacy by Design certified, which is the international standard.

[33] ETHI Report, *supra* note 28 at 7; for a breakdown of the data, see PHAC Investigation, *supra* note 31 at para 5.

are assessed to ensure that they met reidentification risk metrics.[34] In its report on its investigation into the complaints, the OPCC found that a combination of de-identification techniques and measures put in place to protect against reidentification meant that "there is not a serious possibility that the information collected by PHAC could be linked to an identifiable individual."[35]

Under the contract with BlueDot, the data relied upon came from approximately five million mobile devices and was sourced by BlueDot from Pelmorex Corp[36] and Veraset LLC.[37] Like Telus, BlueDot did not supply raw data to PHAC; rather, it performed the analytics required by PHAC on its own data holdings. According to BlueDot, the data that it receives from its suppliers is de-identified and pre-aggregated, and the company has procedures in place to manage and secure the data, thus providing assurances that reidentification is not possible.[38] It was this type of anonymized mobility data that was of particular interest to public health authorities. According to testimony before ETHI, BlueDot did not provide PHAC with any identifiable personal information. Instead, it provided

[34] ETHI Report, *supra* note 28 at 13.

[35] PHAC Investigation, *supra* note 31 at para 67.

[36] Pelmorex Corp, "Fueled by Data, Driven by Creativity, Powered By Technology:Developing Insights to Move Businesses Forward" (2022), online: <www.pelmorex.com/en/data> [perma.cc/DW8C-7LFX]: This is a Canadian company which provides, among other services, data analytics. According to its website, "Pelmorex is deeply committed to data privacy and security. We are a gold standard of privacy in Canada in consumer data. We only use data derived with clear and transparent consent from consumers and only use aggregated and anonymous insights on all our platforms that are built by us. We do not sell or share any individual user data."

[37] See "About Veraset" (2022), online: <www.veraset.com/about> [perma.cc/AC26-7768]: Veraset is a US based data analytics company that specializes in "raw and pre-processed population movement datasets."

[38] See PHAC Investigation, *supra* note 31 at paras 55-58. Note, however, the transparency and ethical concerns raised by Christopher Parsons regarding the source of the data, "More information is needed to know exactly how this location information is collected. If it is derived from the data brokerage economy—which largely operates unknown to the individuals who have their information is collected, and where that information is regularly and routinely re-identified—then it would be troubling to see the Government of Canada participate in this arguably unethical, if ostensibly legal, brokerage economy" (see Miles Kenyon, "Christopher Parsons Delivers Testimony to the Standing Committee on Access to Information, Privacy and Ethics" (14 February 2022) at para 18, online: <citizenlab.ca/2022/02/christopher-parsons-delivers-testimony-to-the-standing-committee-on-access-to-information-privacy-and-ethics/> [perma.cc/R54G-ESB8]). PHAC Investigation, *supra* note 31 at para 52 notes that BlueDot data is collected both through third party apps and from the use of Software Development Kits, as well as from data brokers. According to the ETHI report, although BlueDot does receive some data from individual devices, these data only have an approximate location and time stamp, and are without identifying information, although the data may be accompanied by a reference to a "home." This designates the primary location of the device in order to distinguish between devices that stay close to their primary location and those that move about. See ETHI Report, *supra* note 28 at 15.

only reports containing "statistical summaries, numbers of devices, and proportions and percentages."[39]

The federal minister of health testified before ETHI that mobility data analytics were also being used in other countries and were helpful to governments in determining "how the public is responding to public health directives so that they can tailor their approach and communications."[40] This is supported by a number of studies and reports exploring such practices,[41] and was confirmed in the OPCC's investigation.[42]

Unlike StatCan, which was responsible for its own data collection and management practices and was subject to the federal public sector *Privacy Act* because it collected personal data, PHAC did not collect any *identifiable* personal data.[43] The OPCC's decision that the complaints were unfounded turned on the conclusion that there was no "serious possibility" that reidentification could take place.[44] It was the companies that provided data analytics services that had to conform to the requirements of the applicable private sector data protection law—in this case, PIPEDA. Before ETHI, both Telus and BlueDot maintained that their collection and use of personal data complied with the law. Telus' data came from its customers and was presumably collected with their consent. No customer data—de-identified or otherwise—was shared with PHAC. Instead, analytics were performed on behalf of PHAC on the available data. What PHAC ultimately received were analytics reflecting aggregate

[39] ETHI Report, *supra* note 28 at 15.

[40] *Ibid*: The Minister of Health testified that mobility data were in use in at least 22 countries, including the US, the UK, and countries in South American and Europe. See also Satchit Balsari et al, *The Use of Mobility Data in Public Health Emergencies* (Cambridge, MA: Crisis Ready, 2022) Crisis Ready, online: <crisisready.io/wp-content/uploads/2022/06/The-Use-of-Human-Mobility-Data-in-Public-Health-Emergencies.pdf > [perma.cc/P95E-N998].

[41] See e.g. European Commission eHealth Network, "Towards a common approach for the use of anonymised and aggregated mobility data for modelling the diffusion of COVID-19, and optimising the effectiveness of response measures: Version 4.3" (30 June 2020), online (pdf): <www.health.ec.europa.eu> [perma.cc/DC2V-YNHC]; see also Balsari et al, *supra* note 40; see further Nishant Kishore, "Mobility data as a proxy for epidemic measures" (2021) 1 Nature Computational Science 567, DOI: <10.1038/s43588-021-00127-7>; Kristofer Ågren, Pär Bjelkmar & Elin Allison, "The use of anonymized and aggregated telecom mobility data by a public health agency during the COVID-19 pandemic: Learnings from both the operator and agency perspective" (2021) Data & Pol'y e17-1, DOI: <10.1017/dap.2021.11>; Pascale Chambreuil, Ju Y Jeon & Thierry Barba, "The value of network data confirmed by the Covid-19 epidemic and its expanded usages" (2022) 4:e4 Data & Pol'y e4-1, DOI:<10.1017/dap.2021.31>.

[42] PHAC Investigation, *supra* note 31 at paras 73, 79-80.

[43] *Privacy Act*, *supra* note 17 at s 3: personal data is defined as "information about an *identifiable individual* that is recorded in any form" [emphasis added]. See also Amanda Cutinha & Christopher Parsons, "Minding Your Business: A Critical Analysis of the Collection of De-identified Mobility Data and Its Use Under the Socially Beneficial and Legitimate Interest Exemptions in Canadian Privacy Law" (22 November 2022), online: <www.citizenlab.ca/2022/11> [perma.cc/8KRU-WB7B].

[44] PHAC Investigation, *supra* note 31 at para 67.

and anonymized data about the mobility of Canadians. The situation with BlueDot was similar. The company did not provide raw data to PHAC; instead, it performed analytics on behalf of PHAC on its stores of data which it also claimed to have been provided with consent.

Although the collection of location by Telus and Bluedot appears superficially compliant with data protection law, critics nonetheless raised concerns. PIPEDA, which is sorely outdated,[45] makes consent the primary basis for legitimizing the collection, use or disclosure of personal information. As some critics pointed out in the ETHI hearings,[46] most individuals would not read Telus' privacy policy in detail and, even if they did, might not understand that their data would be shared in this way or for these purposes. BlueDot maintained that data subjects had consented to the use of all mobility data that they sourced, and that data subjects had the ability to opt out of further sharing of this data.[47] However, the actual sources of location data used by the company were unclear, and critics questioned the legitimacy or traceability of any consents that may or may not have been provided, as well as the challenges for individuals to opt out in such circumstances.[48] Concerns were also raised that some of the data may have been sourced through data brokers.[49] These concerns coloured perceptions of PHAC's use of the data.

The OPCC report of findings on the complaints it received centred on PHAC since the complaints were made against it, under the *Privacy Act*, as opposed to being against the private sector companies involved, under the PIPEDA. It therefore did not consider the privacy policies or practices of the companies involved. This approach left unexamined the

[45] Bill C-27, *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts*, *House of Commons Debates*, 44-1, No 91102 (second reading 24 April 2022) <www.parl.ca/legisinfo/en/bill/44-1/c-27> [perma.cc/V2PE-S98P]: At the time of writing, Bill C-27 which would substantially reform and modernize PIPEDA is at the committee stage. The antiquated nature of PIPEDA has been the subject of considerable commentary. See e.g. Office of the Privacy Commissioner of Canada, "Key recommendations for a new federal private sector privacy law" (4 May 2022) online: <www.priv.gc.ca> [perma.cc/BR8H-W8KK]; Standing Committee on Access to Information, Privacy and Ethics, "Towards Privacy by Design: Review of the Personal Information Protection and Electronic Documents Act," 42-1, No 12 (28 February 2018) online: <www.ourcommons.ca> [perma.cc/WP5P-FGLX]; see also ETHI Report, *supra* note 28: Several of the recommendations proposed amendments to PIPEDA.

[46] ETHI Report, *supra* note 28 at 26: The Privacy Commissioner of Canada also observed that "that privacy policies that mention that mobility data may be used for the public good are not a good way of informing Canadians about how their data is used, as these policies are often long, complicated and difficult to understand." See also Kenyon, *supra* note 38 at para 23.

[47] ETHI Report, *supra* note 28, at 27-28.

[48] See e.g. Kenyon, *supra* note 38 at paras 7, 18: at the time, BlueDot did not even indicate in its privacy policy that it collected mobility data; see also Cutinha & Parsons, *supra* note 43 at 29, 32.

[49] See e.g. Kenyon, *supra* note 38 at para 18.

issue as to the legitimacy of the consent to the collection of data from mobile apps (including from Software Development Kits[50]) and the potential sourcing of these data from data brokers. It is important to note that these issues remain unresolved. They raise interesting questions for public sector organizations, particularly since the OPCC has elsewhere taken the position that government actors should ensure that third parties from whom they obtain services involving personal data have complied with applicable laws in collecting that data.[51]

II. *Discussion*

Although the actual or proposed use of data in the StatCan and PHAC cases was different, both examples have common elements. The discussion below considers a series of normative issues raised by these examples, and pertinent to the public sector use of personal data from the private sector. These are organized under the headings of: necessity and proportionality; privacy, data protection and data ethics; transparency, trust and public engagement; and data quality and access.

1. *Necessity and proportionality*

Because StatCan put its financial data project on hold when the OPCC began its investigation, the OPCC found that StatCan had not breached the *Privacy Act*, since it had not yet actually collected any financial data. However, as noted above, the Privacy Commissioner had expressed concerns that the financial data collection was not authorized by section 13 of the *Statistics Act*. In its report, the OPCC also opined that had the project been authorized by law, it might still have been non-compliant because it failed to meet a necessity and proportionality approach to data collection. This latter point was interesting since the *Privacy Act*, an outdated statute in its own right,[52] does not make "necessity" a condition of data

[50] Software Development Kits are covert trackers that are part of a kit for the development of mobile apps that facilitates the serving of ads to app users. The kit is free—in exchange for the tracker feeding data back to the provider. See e.g. Sara Morrison, "The hidden trackers in your phone, explained," *Vox* (8 July 2020), online: <vox.com/recode/2020/7/8/21311533/sdks-tracking-data-location> [perma.cc/3Q6G-BTQA]; Charli Warzel, "The Loophole That Turns Your Apps Into Spies" *New York Times* (14 September 2019), online: <nytimes.com/2019/09/24/opinion/facebook-google-apps-data.html> [perma.cc/7R9P-7Q74].

[51] Office of the Privacy Commissioner of Canada, *Police use of Facial Recognition Technology in Canada and the way forward: Special report to Parliament on the OPC's investigation into the RCMP's use of Clearview AI and draft joint guidance for law enforcement agencies considering the use of facial recognition technology* (Ottawa: OPCC, 2021), online: <priv.gc.ca/en/opc-actions-and-decisions/ar_index/202021/sr_rcmp/> [perma.cc/QKV6-8HL5] [OPCC 2021]. For example, in its report on its investigation into the RCMP's reliance on Clearview AI's facial recognition service, the OPCC stated: "[T]he RCMP is obligated to inform itself of the lawfulness of the collection practices of partners from whom it collects personal information" (*ibid* at para 31).

[52] Reform of the *Privacy Act* is long overdue. See Canada, Department of Justice, *Respect,*

collection, a fact which was admitted by the OPCC.[53] Nevertheless, the OPCC considers necessity and proportionality to be a generally accepted approach to public sector data protection.[54]

The OPCC's necessity and proportionality approach askes four questions:

a. Is the measure demonstrably necessary to meet a specific need?
b. Is it likely to be effective in meeting that need?
c. Is the loss of privacy proportional to the need?
d. Is there a less privacy-invasive way of achieving the same end?[55]

The first question requires that the data collection be rationally connected to a pressing and substantial goal, and that there is an evidentiary basis for this connection.[56] Each of StatCan's two projects involved the collection of high quality economic statistical data and the filling of data gaps in order to enable the government to develop appropriate economic and social policies. The OPCC found that these objectives could be considered "pressing and substantial" but it expressed concerns that StatCan had not clearly demonstrated that the volume and detail of data sought were necessary to meet these objectives.[57] StatCan had explained the challenges it faced collecting statistical data and that the collection of administrative data presented an "effective alternative" to traditional methods of collection. Acknowledging StatCan's statistical expertise, the OPCC nonetheless expressed concerns that the projects had not been defined "at a level of detail sufficient to fully assess their effectiveness."[58] It noted that one hundred per cent accuracy or confidence in statistics is not required in order for them to be valid and reliable. However, StatCan should be able to define the level of accuracy needed in order to meet their objectives. In other words, it might be possible that a sufficient level of accuracy could be obtained with less data; the onus was on StatCan to

Accountability, Adaptability: A discussion paper on the modernization of the Privacy Act, (Ottawa: DOJC, 2021) online: <justice.gc.ca/eng/csj-sjc/pa-lprp/dp-dd/raa-rar.html> [perma.cc/P8DP-DT6N].

[53] Nonetheless, the OPCC considered that the federal government required that the collection of personal information by government entities be "demonstrably necessary" (see OPCC 2019, *supra* note 2 at para 79 citing Canada, Treasury Board, *Directive on Privacy Practices*, (directive issued pursuant to paragraph 71(1)(d) of the *Privacy Act*), effective 26 October 2022 (Ottawa: Treasury Board, 2020) s 4.2.9, online: <tbs-sct.canada.ca/pol/doc-eng.aspx?id=18309> [perma.cc/A9B7-7LEY]).

[54] OPCC 2019, *supra* note 2 at para 80. See also: Daniel J Therrien, "Incorporating Privacy into Statistical Methods—Necessity and Proportionality" (2 March 2020), online: <priv.gc.ca/en/opc-news/speeches/2020/sp-d_20200303> [perma.cc/8RRY-F336].

[55] OPCC 2019, *supra* note 1 at para 82.

[56] *Ibid*.

[57] *Ibid* at para 88.

[58] *Ibid*.

demonstrate that the volume and detail of data collection was necessary to meet its objectives.

On the issue of proportionality, the OPCC was concerned that the administrative data programs were overly intrusive. It disagreed with StatCan that the agency's obligations to protect the confidentiality of data and its production of statistics in aggregate format were sufficient to address privacy concerns. Although important, these factors did not satisfy the proportionality requirement. The OPCC commented on the extent of the mandate and interests involved:

> Otherwise there would be seemingly no limit to what personal information Statistics Canada could collect pursuant to its mandate… financial transaction information can paint an intrusively detailed portrait of an individual's lifestyle, consumer choices and private interests, including lawful choices individuals would not want the government to know about. We consider a complete record of financial transactions to be extremely sensitive personal information.[59]

A proportionality analysis would require an assessment not just of the importance of the *demand* for the economic statistics, but also the importance of the underlying public policy need. The OPCC found that "Statistics Canada has not demonstrated here that this depth of surveillance is proportional to the objectives of the Financial Transactions Project."[60] In other words, just because something can be done and is useful, it does not mean that it should be done, or that the benefits of doing it automatically outweigh any possible harms.

The OPCC also found other considerations to be relevant to the necessity and proportionality analysis, including whether the information in pseudonymized format would be retained by StatCan and used for other possible linkages. StatCan indicated to the OPCC that it "intended to link the credit information and financial transaction information with census information, and to other source(s) that have yet to be determined, for statistical purposes."[61] For the OPCC, this open-ended retention of pseudonymized information for future purposes weighed against a finding of proportionality.

The OPCC was also unpersuaded by arguments that the sample sizes in each project were relatively small and thus were a factor in favour of a finding of proportionality. StatCan had argued that they sought only 600 data elements from 44 million files, representing only four per cent of

[59] *Ibid* at paras 91, 93.

[60] *Ibid* at para 94.

[61] *Ibid* at para 95.

data available from credit bureaus.[62] In the case of the data from financial institutions, StatCan argued that its sample size of half a million households was only three per cent of all households.[63] Nevertheless, the OPCC found that the volume and detail of information sought in both cases was too high to be considered "limited" in nature.

StatCan argued that it had considered alternatives to the scale of collection of private sector administrative data, but neither the use of anonymized data nor increasing the sample size of their Survey of Household Spending would meet the demand for quality statistics. The OPCC observed that there were other innovative methods available, and that some of these were being considered by NSOs in other countries. In looking to experience elsewhere, the report found widespread use of public sector administrative data for statistics, but much more limited use of private sector data. They found that other NSOs had made tentative and experimental uses of this kind of data. They noted that: "The conservative and careful approach is in part due to concerns with respect to privacy and complying with associated data-protection legislation."[64] They also noted that some NSOs had adopted less privacy invasive approaches to the use of such data, giving the example of Portugal's use of "somewhat aggregated" credit and debit card data.[65] France had also considered the use of more aggregated mobile phone data for statistical purposes because of the sensitivity of that information. The OPCC noted that in New Zealand, a consideration of the use of private sector administrative data had raised privacy concerns that required the adoption of a necessity and proportionality approach. They concluded that:

> Overall, we were informed that while there is general statistical interest by NSOs (National Statistical Offices) internationally in gaining access to individual-level financial information, including credit and banking information, many barriers and countervailing considerations exist to collecting such information, including both legal barriers and concerns relating to privacy intrusion and safeguard concerns.[66]

Given the approaches taken by other NSOs, the OPCC concluded that alternative measures existed and had not properly been considered.

The upshot of the OPCC's report was the development by StatCan of a necessity and proportionality framework for the collection of

[62] *Ibid* at para 97.

[63] *Ibid*.

[64] *Ibid* at para 31.

[65] *Ibid* at para 32.

[66] *Ibid* at para 40.

private sector administrative data for use in statistical projects,[67] and the establishment of an Advisory Council on Ethics and Modernization of Microdata Access.[68] The OPCC expressed the view that this was "a positive change in direction," but that it remained too early to tell how effective such an approach would be. It would not, however, resolve the issue of legal authority for the collection of financial data.

The complaint to the OPCC about PHAC's use of mobility data was also brought under the *Privacy Act*. Nevertheless, the OPCC did not apply a necessity and proportionality analysis because it found that the information relied upon by PHAC was sufficiently irreversibly deidentified to not constitute "personal information." A necessity and proportionality analysis would apply only once it is determined that the *Privacy Act* applies because personal information is at issue. This highlights another issue that is becoming increasingly challenging in the data society—not just the siloing of issues regarding ethical data use, but the tendency to address most of these through the vehicle of privacy or data protection legislation. Privacy and data protection are important, but they are not the only issues, and more holistic approaches are needed. A necessity and proportionality approach offers one way to do this—particularly where it is approached from more than just a privacy perspective.[69]

The concept of necessity and proportionality emerges in the StatCan example as an important means of assessing the balance between data collection for public purposes on the one hand and privacy/ethical concerns on the other. Just because government entities can compel the production of data does not mean that they should do so, or that they should do so in an unfettered way. Although a necessity and proportionality approach was not specifically recommended by the ETHI Committee as a framework for determining what data at what granularity were required for public sector programs, such an approach could have application more broadly.[70] However, ideally a necessity and proportionality framework in

[67] Statistics Canada, "Necessity and Proportionality," *supra* note 16: Note again that although the framework is described in general terms in the Trust Centre, no actual published explanation or flow chart is provided.

[68] Statistics Canada, "Advisory Council on Ethics and Modernization of Microdata Access" (22 January 2024), online: <statcan.gc.ca/en/about/relevant/acemma> [perma.cc/ZE2T-ETCH]. The inaugural meeting of this Council was held on 18 July 2019.

[69] See e.g. Teresa Scassa "The Surveillant University: Remote Proctoring, AI and Human Rights" (2022) 8:1 Can J Comp & Contemporary L 271, online: <cjccl.ca/wp-content/uploads/2022/10-Scassa.pdf> [https://perma.cc/KDX8-TXLJ]: Note the discussion of the use of this approach to analyzing the use of artificial intelligence-enabled remote proctoring tools during the pandemic.

[70] See Office of the Privacy Commissioner of Canada, "Principles for responsible, trustworthy and privacy-protective generative AI technologies" (7 December 2023), online: <priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/> [perma.cc/49SR-AJAJ].

these contexts should consider more than just individual privacy rights; it should consider the impact of the proposed collection/use on groups and communities and should take into account other human rights considerations such as equality and non-discrimination. If the issues are approached solely under existing privacy legislation, this broader scope is not possible.

2. *Privacy, data protection, and data ethics*

In addition to any necessity and proportionality analysis, the StatCan projects raised data protection issues, including concerns over cybersecurity.[71] The OPCC reviewed StatCan's de-identification and encryption practices, as well as its approaches to logging and monitoring of database access. It also considered whether information might be disclosed by StatCan for secondary purposes. For the most part, the OPCC found that StatCan's handling of personal data met the necessary standards, although it found that measures to monitor internal access to data were lacking and posed a security risk.[72] In some respects, the internal access issue may speak to the (lack of) currency in StatCan's traditional practices given the novel collection of sensitive private sector administrative data. The more such data is collected by StatCan, the greater the risk might become of improper internal access.

One important issue that was not explored in the OPCC report relates to the growing use of statistical data, in combination with other available data, by private sector companies in profiling individuals. The production of high-quality statistical data based on consumer financial data could indirectly lead to new forms of profiling that can have significant impacts on individuals and groups.[73] In addition, there are risks that high quality micro-data can contribute to overall re-identification risk with anonymized data. Groves and Harris-Kojetin observe that: "The proliferation of publicly accessible data, outside of the statistical agencies, has dramatically increased the risks inherent in releasing micro-data because these other data sources can be used to re-identify putatively anonymized data."[74] Although high-resolution economic statistics may have public benefits, they may also present novel privacy harms that should be properly assessed.

[71] Groves & Harris-Kojetin, *supra* note 3 at 75: Growing cybersecurity threats can increase the risk to the public resulting from data in the hands of statistical agencies.

[72] OPCC 2019, *supra* note 2 at para 137.

[73] See e.g. Alessandro Mantelero, "From Group Privacy to Collective Privacy: Towards a New Dimension of Privacy and Data Protection in the Big Data Era," in Bart van der Sloot, Luciano Floridi & Linnet Taylor, eds, *Group Privacy* (Dordrecht: Springer, 2017) 139.

[74] *Ibid* at 77.

In terms of the mobility data program, even if PHAC did not actually collect personal information, it did acquire services from companies that did. Both companies with which PHAC contracted maintained that their data was collected and used in compliance with the law. Yet, in the case of Telus, there were clearly some concerns that individual consent, although formally obtained, was not substantively adequate, and that the possibility of opting out of usage of data for the Data for Good platform was not readily accessible.[75] Further, there were also interesting questions regarding the "pedigree" of the mobility data relied upon by BlueDot. This situation reflects one of the contemporary challenges of data protection law: technical legal compliance by an organization may still not accord with the knowledge or understanding of data subjects. Concerns of this nature highlight the distinction between what is strictly legal and what is ethical when it comes to data use and suggest that the public sector use of private sector data may need to conform to ethical standards that are superior to bare legality.[76]

The ability of individuals to "opt out" of having one's data used for secondary purposes was clearly seen as an important means of control for individuals. Although PHAC's 2021 call for tenders required bidders to demonstrate that individuals could "opt out" of having their data used as part of the company's mobility analytics program, critics noted that information about opting out was far from evident or accessible.[77] One of the recommendations in the ETHI Report was that government entities that use data in this way provide clear information on their own websites about how individuals can opt out of the use of their data by the parties with whom they have contracted for data-related services.[78] Since opting out is a far less realistic option where data are acquired from data brokers, this raises the further issue of whether government actors should pay closer attention to the sources of the data on which they rely, exhibiting a preference in procurement for data from original sources where opt out options can be more easily managed and supervised.

The issue of the sources of data relied on by public sector actors is clearly important. In its investigation into the RCMP's reliance on

[75] ETHI Report, *supra* note 28, at 26 (recommendation 4 suggests that information on opting out of data collection should be made readily available).

[76] Cutinha & Parsons, *supra* note 42 at 33-35 (the authors discuss how Canadian law fails to adequately govern the use of de-identified data, not just in terms of ethics but also in terms of reidentification risk).

[77] ETHI Report, *supra* note 28 at 23. Note that Telus confirmed at the hearings that individuals could opt out of the Data for Good program (*ibid* at 26). See also Cutinha & Parsons, *supra* note 43 at 18-19.

[78] ETHI Report, *supra* note 28 at 26.

Clearview AI's facial recognition services, the OPCC took the position that public entities had an obligation to ensure that data they relied upon had been legally acquired.[79] There may be many grey areas for legal data collection. For example, vague terms of use may be insufficient to obtain consent for the kind of third-party collection of sensitive location data via Software Development Kits. Widespread scraping of data from the global internet may also be acceptable in some jurisdictions but not in others depending on the approach to publicly accessible data.[80] Issues of the legitimacy of consent may turn on their own facts. This places an important onus on public sector bodies to exercise care and due diligence when it comes to contracting for access to data and analytics provided by private sector entities.[81] Procurement, requests for proposals, or other tendering processes are an area where government entities should pay attention to what the ethical boundaries are for the use of private sector data and analytics.

If privacy remains the dominant framework for assessing the appropriateness of the use of private sector data by public sector entities, then there may be important legitimacy gaps when it comes to such uses. This suggests that it might be desirable to develop an independent framework—perhaps in the form of a ministerial directive or formal guidance—for the use by the public sector of data and analytics furnished by the private sector; such a framework should direct attention to both the legitimacy of consent and the ethics underlying the data collection by the private sector actors.

3. *Transparency, trust and public engagement*

Transparency, trust and public engagement were perhaps the central issues in the context of both the PHAC and the StatCan controversies. The three concepts are not identical, but they are closely related. Transparency involves sharing information publicly about one's data collection and processing practices, and it can certainly build trust. Public engagement also involves transparency, but it has a further element of consultation

[79] OPCC 2021, *supra* note 51.

[80] *Ibid* (web scraping of data was at issue in the investigation into the RCMP's use of Clearview AI); see also *AT v Globe24h.com*, 2017 FC 114 regarding the scraping of court judgements by a Romanian web service.

[81] OPCC 2021, *supra* note 51. Note that in a special report on the use of Clearview AI's facial recognition database by the Royal Canadian Mounted Police, the Privacy Commissioner stated that "a government institution cannot collect personal information from a third party agent if that third party agent collected the information unlawfully" (*ibid*, at "Overview of investigation into RCMP's use of Clearview AI"). See also Ågren et al, *supra* note 41 at e17-9. The Swedish public health authority attended to privacy issues because it was concerned that its reputation not be harmed by the use of private sector mobility data during the COVID-19 pandemic.

and responsiveness that is also essential to trust. In very simple and direct terms, trust has been defined as "the foundation upon which the legitimacy of democratic institutions rest."[82]

The lack of trust is perhaps best demonstrated by the media furore around each of these projects. The media broke the story about StatCan's plans regarding the use of financial data. This led to significant public backlash,[83] although the government defended the programs.[84] The framing of the programs by the news media no doubt contributed to the controversy. More proactive agency messaging might have emphasized the nature and extent of the benefits to the public of the programs as well as the privacy and security measures in place. By contrast, one news report referred to "attempts by Statistics Canada to obtain sensitive banking details of more than 500,000 Canadians without their consent."[85] Another media report suggested that the StatCan data program could threaten Canada's trade relations with Europe because of its inconsistency with the GDPR.[86] The media were also quick to report on PHAC's use of mobility data. In the PHAC case, some of the news stories were highly inflammatory. One report in the *National Post* stated: "The Public Health Agency of Canada accessed location data from 33 million mobile devices to monitor people's movement during lockdown, the agency revealed this week."[87] Not only were the numbers greatly exaggerated, the same article linked the PHAC data mobility program with a general increase in use of surveillance technology during the pandemic, creating the impression that individual Canadians were tracked during the pandemic in order to ensure they were complying with quarantine and lockdown requirements. Reporting of this kind clearly played upon trust concerns.

[82] Organization for Economic Cooperation and Development, "Trust in Government," (n.d.) online: <oecd.org/governance/trust-in-government> [perma.cc/E67U-EBF6].

[83] See generally David Akin, "Privacy Commissioner of Canada launches investigation into StatCan over controversial data project" *Global News* (31 October 2018), online: <globalnews.ca> [perma.cc/9YZV-SU9R] (members of Parliament were quoted as saying their phones were "ringing of the hook" after news broke about the StatCan initiatives).

[84] See generally Geoff Zochodne, "Statistics Canada putting financial data collection project on hold after public outcry" *Financial Post* (18 November 2018), online: <financialpost.com> [perma.cc/D2UX-T7BH].

[85] CTV News, "Privacy commissioner investigating StatCan's attempt to get banking info" *CTV News* (31 October 2018), online: <ctvnews.ca/politics/privacy-commissioner-investigating-statcan-s-attempt-to-get-banking-info-1.4157136> [perma.cc/ST8T-F3VV].

[86] See Amanda Connolly, "Personal data scooping by StatCan could threaten trade with Europe under tough new privacy rules" *Global News* (4 November 2018), online: <globalnews.ca/> [perma.cc/9FP5-N8PM].

[87] Swikar Oli, "Canada's public health agency admits it tracked 33 million mobile devices during lockdown" *National Post* (24 December 2021), online: <nationalpost.com> [perma.cc/FEN9-UBP6].

In both the StatCan and PHAC examples, transparency, trust, and public engagement were found to be sorely lacking. In its report on the complaints that it received about StatCan, the OPCC stated that it "did not find evidence that Statistics Canada adequately engaged the public or affected individuals about the Projects."[88] This was despite public commitments by StatCan claiming it was practicing transparency. The OPCC considered both the nature and sheer volume of the over one hundred complaints it received to be evidence of this lack of transparency. In its view, the complaints revealed that "the public was concerned, surprised and unclear about what was being collected and for what purpose."[89] A lack of transparency in the PHAC case was also linked to the public backlash. PHAC's messaging, had it been more proactively communicated to the public, might have explained that the data used were not those of specific individuals.[90] It might also have communicated the value of such analytics in assessing the effectiveness of public health measures introduced to stop the spread of disease. As one witness noted in the ETHI hearings, poor transparency practices can result in polarizing messages regarding data usage gaining traction.[91] Another witness noted that opposition politicians were also likely to use exaggerated claims to amplify messages attacking the credibility of the current government.[92]

Because the StatCan data collection was from organizations that separately had obligations to their customers under private sector data protection laws, StatCan appeared to rely heavily on those obligations to provide notice to affected individuals. This was in spite of the fact that under the private sector laws, as noted earlier, disclosures for statistical purposes could take place without the knowledge or consent of the individual. Transparency was thus not mandatory for the financial sector companies, although TransUnion was under an obligation to place a note on individual files under credit reporting laws. In any event, the OPCC was clearly of the view that it was not appropriate for StatCan, as the party collecting and processing the personal data, to rely upon others to provide notice. It stated: "Considering the scope and breadth of the collection of sensitive information in the Projects, it is equally surprising that Statistics Canada

[88] OPCC 2019, *supra* note 2 at para 116.

[89] *Ibid* at para 117.

[90] Cutinha and Parsons, *supra* note 43 at 14 (note the "qualitative difference in how states can potentially use mobility information as compared to private organizations").

[91] See ETHI Report, *supra* note 28 at 24 (summary of testimony by Daniel Weinstock).

[92] *Ibid* at 25 (summary of testimony by David Murakami-Wood).

would see it as appropriate or effective from a transparency perspective, to rely on third parties to notify affected individuals."[93]

Timing is also relevant to transparency. In the case of the credit reporting project, TransUnion's file notations would only be seen by the relatively small subset of individuals who requested access to their credit files. The OPCC considered that this, along with StatCan's position that financial institutions could place a note about the request for information on customer online banking statements, was not timely notice, since in both cases it would only be provided after data collection had taken place.

The OPCC observed that over the course of its investigation and in response to its recommendations, StatCan agreed to take further steps to increase transparency around its collection of administrative data. Part of StatCan's response was the launch of a Trust Centre,[94] where information can be found about its administrative data programs, as well as about its privacy and security measures. At the time of the report on the investigation, the OPCC was cautious about this response, noting that the Trust Centre "lacked detail, notably with regards to the two Projects, and its usability could be improved."[95]

In the ETHI hearings on the PHAC mobility data usage, the Committee noted that transparency was lacking both with respect to PHAC's actual use of mobility data and with respect to any precautions it had taken to ensure that privacy was protected. A second level of transparency issue arises with respect to the notice provided to individuals by private sector companies that their data might be shared for various purposes, including with government entities. While information about such practices can be included in privacy policies, such policies are generally not effective as a means of communicating with the public.[96]

The OPCC also raised concerns about transparency in its report on the investigation of the PHAC complaints. It noted that because PHAC did not collect anything that fell within the definition of personal information, the *Privacy Act* did not impose any transparency obligations.[97] Nonetheless, it commented on the government's submissions that it had been transparent. The OPCC found that oblique references in a Prime Ministerial news release and on the COVIDTrends website were not sufficient notice.

[93] OPCC 2019, *supra* note 2 at para 121.

[94] Statistics Canada, "Statistics Canada's Trust Centre" (20 February 2024), online: <statcan.gc.ca/en/trust> [perma.cc/7N4P-8CP9].

[95] OPCC 2019, *supra* note 2 at para 122.

[96] See, e.g. Florian Schaub, Rebecca Balebako & Lorrie Faith Cranor, "Designing Effective Privacy Notices and Controls" (2017) 21:3 IEEE Internet Computing 70, DOI: <10.1109/MIC.2017.75>.

[97] PHAC Investigation, *supra* note 31 at para 82.

Such measures required the public to proactively seek out information. The OPCC recommended that in future "more efficient, targeted, and accessible communications channels be used in order to achieve better transparency."[98] The report cited former privacy commissioner Daniel Therrien, who stated that "greater flexibility to use personal information for public good should come with greater transparency and accountability."[99]

Transparency can be linked to trust in the sense that clear, public information can dispel misgivings at the outset and breed greater acceptance of certain practices. However, trust is complex, involving a range of factors including the perception of public institutions and the relationship that individuals have with particular types of personal data. A Nanos survey carried out shortly after the media disclosures regarding the StatCan projects reported that "Nearly two in three Canadians say protecting the privacy of financial data is more important than Statistics Canada better understanding consumer behaviour and trends."[100] The same survey found that almost three quarters of Canadians were either opposed or somewhat opposed to the sharing of financial data with StatCan without consent. A slightly greater proportion of Canadians (seven out of ten) expressed trust in banks to protect this information. The trust levels were significantly lower for credit card companies. According to Nanos, potential cyber attacks on StatCan were a major concern of almost half of the Canadians surveyed,[101] whereas just over 60 per cent of Canadians expressed some level of confidence in banks to protect their data against cyber attack.

The survey is interesting on a number of levels. In the case of financial data, it indicates that Canadians were more willing to trust the institutions to whom they had expressly provided the data, rather than StatCan, as a secondary user of that data. It is interesting to reflect on the importance of this lack of a direct relationship and what role it plays in building and maintaining a sense of public trust. From the perspective of private sector corporations, it might also highlight the risks they face by allowing others (such as StatCan) to access data entrusted to them by the public.

[98] *Ibid* at para 83.

[99] *Ibid* at para 84.

[100] "Canadians choose protecting data privacy over Statistics Canada better understanding consumer behavior and trends: Statistics Canada and Privacy Survey, Summary" (November 2018) at 2, 3, online (pdf): <nanos.co/wp-content/uploads/2021/04/2018-1326-StatsCan-and-Privacy-Populated-Report-with-Tabs.pdf> [perma.cc/6SHT-SC4J].

[101] Hessie Jones, "Canadians Up In Arms: Privacy Without Consent And The Dangerous Precedent" *Forbes* (4 November 2018), online: <forbes.com> [perma.cc/8RC8-M7LV]. This Forbes article suggested that StatCan's collection of more and more detailed and sensitive data would increase its risk of cyberattack.

Public engagement received relatively little attention in the OPCC StatCan report and in the ETHI report on the PHAC's data practices. This may be because the traditional framing of privacy issues has tended to focus on the individual and their rights regarding their personal data. Public engagement addresses a broader set of issues relating to legitimacy and trust. In cases where data are sought at a group or population-level, public engagement becomes important because these processes form a kind of public analogue for knowledge and consent. It is especially notable that the OPCC, in its report, observed that NSOs in other countries that were exploring the use of administrative data for statistics were proceeding cautiously and with active public engagement.[102]

Linnet Taylor and others have advanced the concept of group privacy as a means of shifting the focus from the individual to the group when it comes to the collection and use of data.[103] In this framing, the concept of the identifiable individual is no longer central. Instead, issues of trust require a form of group knowledge and consent to collect and use human-derived data.[104] This will remain so even if the data are deidentified. This approach to data also recognizes that the collection of group-level data can be used in ways that have impacts on groups and communities. Statistical data is a prime example of this, since such data are often used to help governments determine where resources and programs should be directed. Although group privacy should not be a barrier to public sector uses of private sector data in the public interest, addressing group privacy concerns is crucial. Some data have the potential to adversely impact certain groups more than others, and the collection of these data may require public engagement and consultation at the community or group level to avoid issues of bias and discrimination. For example, in his testimony to ETHI on the PHAC mobility data inquiry, Christopher Parsons noted that even if it did not impact individual privacy, the aggregated and anonymized data relied upon by PHAC could have population level effects:

> Some communities may be forced to travel more frequently during the pandemic to fulfil essential work, and other communities may be less represented in mobility data if not all members in a household have a mobile phone, and governments might modify how they allocate policing

[102] OPCC 2019, *supra* note 2 at para 42.

[103] Linnet Taylor, Luciano Floridi and Bart Van der Sloot, "Introduction: A New Perspective on Privacy," in Bart van der Sloot, Luciano Floridi & Linnet Taylor, eds, *Group Privacy* (Dordrecht: Springer, 2017) 1-12; Luciano Floridi, "Open Data, Data Protection, and Group Privacy" (2014) 27:1 Philosophy of Technology 1, DOI: <10.1007/s13347-014-0157-8>.

[104] For a definition and discussion of the need for governance of human derived data see Teresa Scassa, "Governing Human-Derived Data" *Platform Governance in Canada* (2023), online: <democracy.ubc.ca> [perma.cc/F2ST-J5WU].

> or service resources as a result of such mobility data. All of which is to say that even aggregated and anonymized data can have impacts on communities. It is insufficient to just consider whether an individual privacy violation has taken place—though it is possible that one may have occurred—and is imperative to also consider the community impacts of how data is collected or used in policy making or resource allocation.[105]

The Privacy Commissioner's legislative mandate does not include group privacy, and the OPCC cannot therefore be expected to address these issues in its response to complaints. Unfortunately, there is currently no oversight body or mechanism to address these issues, and public engagement appears to be largely voluntary and ad hoc within the public sector. As government entities shift to greater reliance on private sector administrative data or analytics based on such data, public engagement and some form of ethics review should become normalized either through a government policy directive or through an institution's enabling legislation.

If they demonstrate nothing else, the StatCan and PHAC examples show how transparency is crucial, not just at the individual level (the more traditional notice and consent to data collection or use) but also at the project-wide and collective level. Governments must come to terms with the group privacy dimensions of human-derived data and must understand that in some cases, legitimacy requires meaningful public engagement. Such engagement may also be required not just to provide notice and gain public acceptance, but in some cases also to define and circumscribe projects.

4. *Data quality and access*

A recent study on the availability and use of mobility data in public health crises addressed some of the challenges to access to such data. In some cases, concerns over the legality of data sharing made private sector companies reluctant to share; in others, the risks of harming relationships with customers who might object to the data sharing was also a factor in their reluctance. The report noted that such concerns could influence what was shared and with what degree of aggregation, leading to issues of data quality and interoperability.[106] Quality is also an issue for the public sector. For example, a Swedish study found that trust in data quality was a key

[105] Kenyon, *supra* note 38 at para 17. See also Cutinha & Parsons, *supra* note 43 at 35.

[106] Balsari et al, *supra* note 40; See also Groves & Harris-Kojetin, *supra* note 3 at 64-65.

consideration for public health authorities seeking to use mobility data during the COVID-19 pandemic.[107]

Data quality issues are certainly front and centre in the StatCan examples, where StatCan sought high-volume, detailed data to improve statistical output. Nevertheless, the OPCC cautioned that there must be a balance between statistical quality on the one hand and privacy on the other; data collection should be limited to what is reasonably necessary to meet the project objectives, not driven by what is required to produce the most refined statistics possible. This tension between data quality and privacy manifests itself in other situations as well. For example, different anonymization techniques may variously degrade data quality.[108] While the goals of data scientists may focus on optimal quality, the message of the OPCC is that privacy concerns must be balanced with the data's fitness for purpose.

This balancing is evident in other jurisdictions as well. Balsari et al note the relationship of data sharing practices to the legal regime in place, suggesting that privacy-forward laws such as the EU's GDPR may limit data sharing in the public interest. As a result, the ability and willingness of private sector companies to share their data with the public sector may be lower, and data sets may also be differently licenced depending on their quality and granularity.[109] The more aggregated the data, the more likely it is to be widely shared (some companies even published freely accessible mobility data analytics on their websites during the pandemic).[110] More detailed and granular data might have to be shared under restrictive data sharing agreements.

Finally, on the issue of access to data, the PHAC example demonstrates that many companies develop secondary markets for their data, providing for-fee access to both data and analytics services. In the StatCan example, rather than contracting for the data it sought, StatCan asserted its legislative

[107] Ågren et al, *supra* note 41 at e17-10.

[108] See e.g. Sam Fletcher & Md Zahidul Islam, "Measuring Information Quality for Privacy Preserving Data Mining" (2015) 7:1 Intl J Computer Theory & Engineering 21, DOI: <10.7763/IJCTE.2015.V7.924>; Felix N Wirth et al. "Privacy-preserving data sharing infrastructures for medical research: systematization and comparison" (2021) 21 BMC Medical Informatics & Decision Making 242, DOI: <10.1186/s12911-021-01602-x>.

[109] Balsari et al, *supra* note 40 at 8.

[110] See e.g. Google, "COVID-19 Community Mobility Reports" (17 October 2022) online: <google.com/covid19/mobility> [perma.cc/C6YP-M78H]; Stanford University, "COVID-19 Mobility Network Modelling" (n.d.) online: <covid-mobility.stanford.edu/> [perma.cc/GFN2-78G3]: Stanford University, "COVID-19 Mobility Network Modelling" (n.d.), online: <covid-mobility.stanford.edu/> [perma.cc/GFN2-78G3] (researchers at Stanford University also used mobility data available from the private sector for modeling purposes);Serina Chang et al, "Mobility network models of COVID-19 explain inequities and inform reopening" (2021) 589 Nature 82, DOI:<10.1038/s41586-020-2923-3>.

authority to compel the production of data by the private sector. However, the fact that private sector data is now extremely marketable for secondary purposes raises a further question about the legitimacy of compelled access. At a certain point, it could resemble a form of expropriation. Even if there is a public policy justification for permitting the compulsion of such data, the costs involved in providing large volumes of highly sensitive data may necessitate some form of cost-recovery model at the very least.

Conclusion

The two examples considered in this paper reflect quite different contexts for public sector use of private sector data. Together, they permit the exploration of a range of issues that arise where private sector data are accessed and used by government entities.

One of these issues is clearly the state of the law—and the relevant law may include not just public and private sector data protection laws, but other laws, directives and policies that enable or constrain government acquisition and use of data. The examples discussed in this paper reveal the challenges caused by outdated statutes, including both public and private sector data protection laws and StatCan's enabling legislation. In the case of StatCan, not only was the enabling legislation poorly adapted to data maintained in databases by private sector entities, it was also not suited to the volume and detail of data held by the private sector and the need for much greater attention to transparency and oversight when it came to attempts to collect such data. This type of data access raised concerns for individuals about data privacy and security; for organizations, it raised concerns about their own obligations and relationships with customers, as well as potential concerns about costs and impacts on their business. The use of data and analytics by PHAC ran up against a private sector data protection paradigm that depended heavily upon individual consent in a context in which reasonably informed consent was largely unrealistic. Governments must address deficiencies in the laws. In addition, government policies, ethical frameworks and/or directives that specifically apply to the ethically appropriate acquisition of data and analytics by government are becoming imperative. Such policies can address issues such as transparency and public engagement, as well as any assurances required by private sector partners as to the legitimacy of any collection of data from individuals.