Canadian Journal of Law and Technology

Volume 14 | Number 2 Article 1

6-1-2016

Strategies for Protecting Privacy in Open Data and Proactive Disclosure

Teresa Scassa

Amy Conroy

Strategies for Protecting Privacy in Open Data and Proactive Disclosure

Teresa Scassa* and Amy Conroy**

Abstract

In this paper, the authors explore strategies for balancing privacy with transparency in the release of government data and information as part of the growing global open government movement. The issue is important because government data or information may take many forms, may contain many different types of personal information, and may be released in a range of contexts. The legal framework is complex: personal information is typically not released as open data or under access to information regimes; nevertheless, in some cases transparency requirements take precedence over the protection of personal information. The open courts principle, for example, places a strong emphasis on transparency over privacy. The situation is complicated by the availability of technologies that facilitate widespread dissemination of information and that allow for the searching, combining and mining of information in ways that may permit the reidentification of individuals even within anonymized data sets. This paper identifies a number of strategies designed to assist in identifying whether government data sets or information contain personal information, whether it should be released notwithstanding the presence of the personal information, and what techniques might be used to minimize any possible adverse privacy impacts.

---

* Canada Research Chair in Information Law, University of Ottawa, Faculty of Law. We gratefully acknowledge the support of the Social Sciences and Humanities Research Council of Canada through the GEOTHINK partnership grant. Many thanks to Linda Low for her helpful comments on an earlier draft of this paper.

** Doctoral Candidate, University of Ottawa, Faculty of Law.

INTRODUCTION

Over the past several years the municipal,[1] provincial,[2] and federal[3] governments of Canada have become increasingly involved in open government initiatives.[4] In general, the open government movement is

[1] Over thirty Canadian municipalities have adopted open data and open government strategies: Datalibre.ca lists 44 cities with open data initiatives. Most are government-led while several are citizen-led portals: Datalibre, online: Open Data and Open Governance in Canada: A Critical Examination of New Opportunities and Old Tensions" (2014) 6:3 Future Internet 414 at 415. Recently, the City of Edmonton won an award at the Open Data Summit. Discussing this and other developments in open data in Canada, see Teresa Scassa, "Open Data Current Events" (26 May 2015), *Teresa Scassa* (blog), online: < www.teresascassa.ca/index.php?option = com_k2&view = item&id = 187:open-data-current-events&Itemid = 81 > .

[2] Provinces actively participating in the open government movement include British Columbia, Alberta, Ontario, Quebec, and Newfoundland and Labrador: Alberta, "Open Government," online: < https://open.alberta.ca > ; British Columbia, "DataBC," online: < https://data.gov.bc.ca > ; Ontario, "Sharing Government Data," online: < https://www.ontario.ca/page/sharing-government-data > [Ontario, "Sharing Data"]; Newfoundland and Labrador, "Open Government," online: < www.open.gov.nl.ca > ; Quebec, ''An Open and Transparent Government," online: < www.gouv.qc.ca/EN/VotreGouvernement/Pages/Gouvernement-ouvert.aspx > .

[3] A list of federal government departments proactively disclosing information is provided at: Treasury Board of Canada Secretariat, "Proactive Disclosure by Department or Agency," (TBS, 12 May 2016), online: < www.tbs-sct.gc.ca/pd-dp/gr-rg/index-eng.asp > [TBS, "Proactive Disclosure"]. Since re-launching its open government portal in June 2013, the various departments of the Government of Canada have collectively released a total of 701 new data sets: Canada, "Open Government Analytics," (15 May 2016), online: < open.canada.ca/en/content/open-government-analytics#top10 > .

[4] The Government of Canada has released a map depicting the move towards greater

motivated by multiple inter-related objectives. These include increased government transparency,[5] open engagement,[6] greater citizen participation in government affairs, general economic development, cost-savings in research, improved efficiency, and support for innovative uses of publicly-held information.[7]

---

openness in the different levels of government in Canada: Canada, "Open Government Across Canada," (12 July 2016), online: < open.canada.ca/en/maps/open-data-canada > . See also Parliamentary Information and Research Service, "Government 2.0 and Access to Information: 1. Recent Developments in Proactive Disclosure and Open Data in Canada," by Alysia Davies & Dara Lithwick, Publication No. 2010-14-E (Ottawa: Library of Parliament, 15 April 2010), online: < www.lop.parl.gc.ca/content/lop/ResearchPublications/2010-14-e.pdf > [Davies & Lithwick], noting proactive disclosure at municipal, provincial, and federal levels in Canada.

5 The concept of "transparency" has been defined in various ways, including a view of transparency as an act of "making relevant, timely and useful information available to the public in easy-to-access formats": Health Canada, "Notice (Revised): Posting information in the Drug Product Database Online," (Ottawa: Health Canada, 18 June 2015), online: < www.hc-sc.gc.ca/dhp-mps/prodpharma/activit/announce-annonce/notice_dpd-avis_bdpp-eng.php > [Health Canada, "Posting Information"]. More complex definitions of transparency characterize the concept as "a measure of the degree to which the existence, content, or meaning of a law, regulation, action, process, or condition is ascertainable or understandable by a party with reason to be interested in that law, regulation, action, process, or condition": Professor William Mock, "On the Centrality of Information Law: A Rational Choice Discussion of Information Law and Transparency" (1999) 17:4 John Marshall J. Computer & Info. L. 1069 at 1082 [emphasis omitted]. Transparency and access to information rights are considered vital to the public's ability to participate in the democratic process and hold government officials accountable for their actions: Information and Privacy Commissioner of Ontario, "Transparency, Privacy and the Internet: Municipal Balancing Acts" (Toronto: IPC, 2015) at 1, online: < https://www.ipc.on.ca/images/Resources/2015-municipal%20-guide-public%20discl-access.pdf > [IPC, "Transparency, Privacy and the Internet"]. A clear understanding of what is meant by transparency is crucial, as the concept is central to the developing open government and open data frameworks (see John Gaventa & Rosemary McGee, "The Impact of Transparency and Accountability Initiatives" (2013) 31:S1 Development Policy Rev. S3 at S4, noting that the increasingly common "social," "citizen-led," or "demand side" accountability initiatives regularly invoke the concepts of accountability and transparency as underlying values to the point that these concepts are at risk of becoming "buzzwords"). For a detailed discussion of the concept of transparency, see Amy Conroy & Teresa Scassa, "Promoting Transparency while Protecting Privacy in Open Government in Canada" (2015) 53:1 Alta. L. Rev. 175.

6 "Open engagement" may involve citizen monitoring of government activities in order to enforce government accountability, citizen participation in government activities and decision-making through mechanisms such as social media or online reporting of community issues that relate to government responsibilities: Teresa Scassa, "Privacy and Open Government" (2014) 6:2 Future Internet 397 at 399, 400 [Scassa, "Privacy and Open Government"].

7 *Ibid* at 397; Anneke Zuiderwijk & Marijn Janssen, "Open Data Policies, Their Implementation and Impact: A Framework for Comparison" (2014) 31:1 Government Information Q.17 at 17; Marijn Janssen, Yannis Charalabidis & Anneke Zuiderwijk, "Benefits, Adoption Barriers and Myths of Open Data and Open Government" (2012)

Both open access and open data are aspects of open government that relate to the disclosure of information or data[8] by governments. Open access is a familiar concept: freedom of information laws already provide access to information in the hands of government. This right of access serves the objectives of transparency and accountability. Government commitments to open access may include reform and improvement of existing legislation and processes. They also frequently include new commitments to proactive disclosure. Proactive disclosure of government information recognizes and acknowledges that the burden should not always be on citizens to make (and to pay for) access to information requests for some categories of government information.[9] Government commitments to proactive disclosure encourage agencies and departments to identify information that is frequently sought and released under access requests, or information that could and should be made available to the public as a matter of course.[10] This information is then made

---

29:4 Information Systems Management 258 at 261; Canada, "G8 Open Data Charter — Canada's Action Plan," (8 March 2015), online: < data.gc.ca/eng/g8-open-data-charter-canadas-action-plan > [Canada, "G8 Action Plan"]; Open Government Partnership, "Open Government Declaration," (OGP, 2015), online: < www.opengovpartnership.org/about/open-government-declaration > . Creative "mashups" of data are also sometimes specifically encouraged when data are released to the public. "Mashing up" of data involves computer programming that combines different data sets with computer applications to derive customized digital tools. See discussion in Davies & Lithwick, *supra* note 4 at 3, and definition of "mashing up" at 7, n 16. See also discussion in Elizabeth F. Judge, "Enabling Access and Reuse of Public Sector Information in Canada: Crown Commons Licenses, Copyright, and Public Sector Information" in Michael Geist, ed., *From "Radical Extremism" to "Balanced Copyright": Canadian Copyright and the Digital Agenda* (Toronto: Irwin Law, 2010) 598 at 631 [Judge, "Enabling Access and Reuse"].

8 The difference between "data" and "information" has been explained in Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences* (London: SAGE Publications, 2014) at 1, where Kitchin states that:

> Data are commonly understood to be the raw material produced by abstracting the world into categories, measures and other representational forms — numbers, characters, symbols, images, sounds, electromagnetic waves, bits — that constitute the building blocks from which information and knowledge are created.

He goes on to explain at 3 that data is different from "facts, evidence, information and knowledge" in that it is "pre-analytical and pre-factual." In this paper, "data" is similarly used to refer to raw facts while information refers to contextualized or interpreted data.

9 See comments in Office of the Information & Privacy Commissioner for British Columbia, "Investigation into the Simultaneous Disclosure Practice of BC Ferries," by Elizabeth Denham, Investigation Report F11-02 (OIPCBC, 16 May 2011) at 2, online: < https://www.oipc.bc.ca/investigation-reports/1243 > [Denham, "BC Ferries"].

10 In some cases, government actors have a duty to make certain information available, including personal information. Where the law requires such disclosures, the relevant legislation may not specify or restrict the ways in which the information may be made

available to the public—often through the website of the particular agency or department. Proactive disclosure may be of reports, internal procedures, background documents, commissioned research, study results, inspection reports, or other such materials. It is often the case that these materials are made available in the format in which they exist (e.g., portable document format, or PDF).

Open data is a somewhat different concept than open access and proactive disclosure, although there can be overlaps. Open data serves somewhat different objectives, although, as with open access discussed above, transparency and accountability can be among the goals served by open data.[11] The open data movement recognizes that governments collect and compile large amounts of data that are often extremely useful to the private sector. This can include statistical and geospatial data, but also data on a broad range of subjects. Open data programs involve the release of government data sets, in reusable formats, ideally according to generally accepted standards, and under an open licence that permits the royalty-free reuse of the data with few or no restrictions.[12] While, as noted earlier, these data sets can be used for the purposes of increasing transparency and accountability in government, the explicit goals of the open data movement are to foster economic development by encouraging innovative reuse of data, and to encourage civic engagement.[13] To this end, "hackathons" have become part of open data programs.[14]

Both proactive disclosure and open data make information and data available not only for the purpose of serving access rights but also to enable reuse

---

available and government actors may choose to release the information online in an effort to promote transparency. We argue further below in this guide that in some instances, the privacy concerns may weigh in favour of releasing the information but withholding it from online release since this will make the data permanently available. The benefits of releasing information online that may currently influence the method of disclosure include that online release may be more cost-effective, may increase accessibility, and may reduce administrative burdens relating to the maintenance of the information or the need to respond to access to information requests. See discussion in IPC, "Transparency, Privacy and the Internet," *supra* note 5 at 5-6.

11 See discussion in Pew Research Center: Internet, Science & Tech, "Americans' Views on Open Government Data," by John B. Horrigan & Lee Rainie, online: < www.pewinternet.org/2015/04/21/open-government-data > [Horrigan & Rainie].

12 Scassa, "Privacy and Open Government," *supra* note 6 at 399; Canadian Internet Policy and Public Interest Clinic, "Creative Commons Licenses: Options for Canadian Open Data Providers," by Kent Mewhort (Ottawa: CIPPIC, 1 June 2012) at 6, online: < https://cippic.ca/sites/default/files/Creative%20Commons%20Licenses%20-%20Options%20for%20Canadian%20Open%20Data%20Providers.pdf > [Mewhort]; Judge, "Enabling Access and Reuse," *supra* note 7 at 618.

13 Scassa, "Privacy and Open Government," *supra* note 6 at 399.

14 "Hackathons" and "appathons" involve competitions in which participants work to create the best consumer-friendly applications. See e.g., Peter Johnson & Pamela Robinson, "Civic Hackathons: Innovation, Procurement, or Civic Engagement?" (2014) 31:4 Rev. Policy Research 349.

and redistribution of the information.[15] In this way, open government policies will play a central role in making more government-held information available for reuse by members of the public and by other public, private, and not-for-profit organizations.[16] The type of information being released by the public sector—either through proactive disclosure or as open data—is diverse and tends to reflect the different roles played by the municipal, provincial, and federal governments. For instance, the City of Toronto's open data website contains much information about municipal issues and services, including data about bicycle paths, building permits, social housing, shelters, garbage and recycling programs, public transit, and public parks.[17] At the provincial level, the Government of Ontario has surveyed its public to determine areas of priority in releasing open data and has found that areas of high demand include public sector salary disclosures, labour force statistics, provincial highway traffic statistics, and information on the budget and expenditures across the different ministries.[18] At the federal level, the government has made available open data sets relating to a range of issues; the overwhelming majority of these are geospatial and statistical data sets.[19]

---

[15] Scassa, "Privacy and Open Government," *supra* note 6 at 399 citing Open Knowledge — Source Code, "The Open Definition," online: < opendefinition.org > . See also Treasury Board of Canada Secretariat, "Canada's Action Plan on Open Government 2014—16," (TBS, 10 July 2016), online: < data.gc.ca/eng/canadas-action-plan-open-government > [TBS, "Action Plan"].

[16] TBS, "Action Plan," *supra* note 15.

[17] City of Toronto, "Open Data: Data Catalogue," (City of Toronto: 7 July 2016), online: < www1.toronto.ca/wps/portal/contentonly?vgnextoid = 1a66e03bb8-d1e310VgnVCM10000071d60f89RCRD > [Toronto, "Data Catalogue"]. See also Mewhort, *supra* note 12 at 6, noting that "[m]unicipal open data web portals release data such as maps of streets and parks, transit routes, transit schedules, electoral boundaries and city budget information."

[18] Ontario, "Sharing Data," *supra* note 2.

[19] These along with many other types of data sets can be accessed at Canada, "Open Government," (12 July 2016), online: < open.canada.ca/en > [Canada, "Open Government"]. In addition to those already released, the website shows requests for new types of data that are currently being considered, including requests for information on divorce rates and maintenance payments and for the most commonly dispensed drugs in Canada (both of which have been forwarded to the appropriate department for consideration). Other data sets currently under review include databases showing all Canadian trademarks and patents. Some previous requests for the release of data on the open data portal have been denied, such as the request for information about the nuclear workforce in Canada (not released because it constituted private sector data). New data sets are being developed, including information on hospital emergency room wait times, which is listed as "in progress." In addition to the federal government's portal, the Treasury Board of Canada Secretariat operates a portal for proactive disclosure of information. This features information about travel and hospitality expenses, contracts, position reclassifications, grant and contribution awards and wrongdoing in the workplace and is proactively disclosed by individual departments: TBS, "Proactive Disclosure," *supra* note 3. Note that Canada's open data portal, which plays a crucial role in open

To achieve the above-noted goals, the open government policy process emphasizes a push for release of information;[20] however, the emphasis on making more information available to the public introduces important questions about protecting the personal information that may be contained within the data sets being released as open data, as well as the information made available through proactive disclosure.[21] Existing access to information legislation at the federal and provincial levels already sets parameters for balancing privacy rights with transparency. In this paper, we examine how this balance may be struck in the particular context of proactive disclosure and open data. This context is novel in that it requires governments to release both data and information proactively to the public at large, and, in the case of open data at least, in readily reusable formats.

In this paper we outline strategies and techniques for those involved in decision-making processes about whether to proactively release information or to release data sets as open data. Public sector freedom of information legislation mandates a balance between transparency and other competing considerations, including privacy.[22] However, both proactive disclosure and open data raise

---

government, is considered to be "[p]articularly strong in geospatial data": Mewhort, *supra* note 12 at 6.

[20] This push for release is seen most clearly in "open by default" principles reflected in Canada, "G8 Action Plan," *supra* note 7.

[21] This discussion has also taken place more specifically in the health research context, where the benefits of making health research data available for reuse must also be pitted against the risks of disclosure. The Committee on Strategies for Responsible Sharing of Clinical Trial Data has summarized the four main concerns as:

i. the potential risk to individual privacy;
ii. use of data for "unfair commercial practices" (a particularly relevant consideration in the area of clinical drug trials where unfair commercial practices may reduce the incentive for companies to assume the costs of researching new treatments);
iii. the dissemination of invalid information produced through secondary analysis of the information without adequate consideration of the relevant issues of data quality; and
iv. ensuring proper professional credit for those who conduct the research involving the original data collection in subsequent publications that rely on the information.

Committee on Strategies for Responsible Sharing of Clinical Trial Data, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks* (Washington, D.C.: National Academies Press, 2015) at 113.

The second and final considerations are more relevant to the private sector than the public sector context considered in this paper. The primary focus of this paper is how to balance personal privacy with the benefits of making more information available to secondary users. The issue of data quality also arises in the current context and is discussed further below.

[22] These obligations are set out in a complex web of laws that includes the federal *Privacy Act,* R.S.C. 1985, c. P-21 [*Privacy Act*], which regulates information-handling practices for the public sector and numerous access to information statutes. The first Canadian

distinct challenges in striking that balance.[23] The paper uses principles and concepts that are common across all levels of government in Canada with a view to helping decision-makers determine when data sets or information targeted for release raise privacy issues that may need to be balanced against transparency considerations. In addition, open data and proactive disclosure raise new issues of scale in the release of data/information: its release in formats and according to standards that facilitate reuse in the form of data mashups and a broad range of analytics raise additional privacy concerns.

The strategies suggested below can be applied individually or in combination depending on the particular data set under consideration. They include data

---

access to information statute was the federal *Access to Information Act,* R.S.C. 1985, c. A-1 [*Access to Information Act*]. Every province now has at least one statute that governs access to information in the hands of public organizations: Alberta: *Freedom of Information and Protection of Privacy Act,* R.S.A. 2000, c. F-25 [*FIPPA-AB*]; British Columbia: *Freedom of Information and Protection of Privacy Act,* R.S.B.C. 1996, c. 165; Manitoba: *The Freedom of Information and Protection of Privacy Act,* C.C.S.M. c. F175; New Brunswick: *Right to Information and Protection of Privacy Act,* S.N.B. 2009, c. R-10.6; Newfoundland and Labrador: *Access to Information and Protection of Privacy Act, 2015,* S.N.L. 2015, c. A-1.2; Northwest Territories: *Access to Information and Protection of Privacy Act,* S.N.W.T. 1994, c. 20; Nunavut: *Access to Information and Protection of Privacy Act*, S.N.W.T. (Nu) 1994, c. 20, as duplicated for Nunavut by s. 29 of the *Nunavut Act*, S.C. 1993, c. 28; Nova Scotia: *Freedom of Information and Protection of Privacy Act,* S.N.S. 1993, c. 5; Ontario: *Freedom of Information and Protection of Privacy Act,* R.S.O. 1990, c. F.31 [*FIPPA-ON*]; Prince Edward Island: *Freedom of Information and Protection of Privacy Act,* R.S.P.E.I. 1988, c. F-15.01; Quebec: *An Act Respecting Access to Documents Held by Public Bodies and the Protection of Personal Information,* C.Q.L.R. c. A-2.1; Saskatchewan: *The Freedom of Information and Protection of Privacy Act,* S.S. 1990-91, c. F-22.01; Yukon: *Access to Information and Protection of Privacy Act,* R.S.Y. 2002, c. 1. In some provinces, there are statutes outlining access to information regimes for municipal and local public bodies. See e.g., Ontario's *Municipal Freedom of Information and Protection of Privacy Act,* R.S.O. 1990, c. M. 56 and Saskatchewan's *The Local Authority Freedom of Information and Protection of Privacy Act,* S.S. 1990-91, c. L-27.1.

[23] Peter Conradie & Sunil Choenni, "On the Barriers for Local Government Releasing Open Data" (2014) 31:S1 Government Information Q S10 at S12-S14 [Conradie & Choenni, "Municipal Barriers"]; Peter Conradie & Sunil Choenni, "Exploring Process Barriers to Release Public Sector Information in Local Government" in J. Ramon Gil-Garcia, Natalie Helbig & Adegboyega Ojo, eds., *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance, Albany, 2012* (2012, Albany New York) 5 [Conradie & Choenni, "Process Barriers"]. It is important to note that practices can emerge by default when such issues are not managed from an early stage in the process, and it is therefore crucial that this discussion be held as part of Canada's already developing open government movement. See comments in Zuiderwijk & Janssen, *supra* note 7 at 19. See similar comments about the need to ensure clear policies for the release of court information online in: Canadian Judicial Council, *Court Information Management: Policy Framework to Accommodate the Digital Environment*, by Jo Sherman (Ottawa: CJC, 2013) at 15, online: < www.cjc-ccm.gc.ca/cmslib/general/AJC/Policy%20Framework%20to%20Accommodate%20the%20Digital%20Environment%202013-03.pdf > [Sherman].

minimization, cooperation and communication across the many different public sector actors involved in the open government initiative, individualized assessment of data sets to clarify the risks and benefits of disclosure, anonymization, licence requirements, and the use of technological barriers to the reuse of information disclosed online. The first two strategies—data minimization and cooperation and communication between government actors—represent general strategies to guide the development of open government in a broad sense. The remaining considerations are more specifically aimed at evaluating individual data sets or information in other formats that may be disclosed as part of the open government movement.

The decision-making process around the release of government data or information must be informed by an understanding of the meaning of personal information. We therefore begin our discussion with a review of the law defining personal information and reidentification risk as well as a discussion of the particular category of personal information known as "publicly available personal information."

I. THE DEFINITION OF PERSONAL INFORMATION

In the move towards greater openness in government, the need to protect individual privacy plays a central role.[24] While the definition of "personal information" is crucial to this balancing act, a specific legislative framework or body of case law to regulate the developing open government movement does not yet exist.[25] As much of the discussion on the definition of personal information

---

[24] The Article 29 Working Party, established under Directive 95/46/EC of the European Parliament ("Article 29 Working Party"), emphasizes that in releasing information to the public, government organizations must strike a balance between the need to respect the individual right to privacy and the need to support public use of government-held information: EC, Article 29 Data Protection Working Party, *Opinion 06/2013 on Open Data and Public Sector Information ('PSI') Reuse* (Brussels: EC, 2013) at 3, online: <ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp207_en.pdf> [Article 29 Data Protection Working Party].

[25] As open government develops, it will be important to emphasize the differences across the various information regimes and the need for individualized policies for open data and the broader open government movement. For instance, there appears to be continuing confusion in terms of the meaning of identifiability for health information released for the purpose of facilitating secondary research. The area is governed by federal and provincial privacy legislation and several statutes geared specifically to the regulation of health information. Discussion of the appropriate threshold to attach to the concept of "identifiability" in this particular context may involve unique concerns relevant to the health research context, such as the need to facilitate medical advancements. See general discussion in Niko Yiannakoulias, "Understanding Identifiability in Secondary Health Data" (2011) 102:4 Can. J. Public Health 291. Despite such differences, access to information schemes will always remain closely tied with the movement towards more proactive disclosure of government-held information. While the proactive disclosure of information as part of the open government initiative may lead to a reduction in access to information requests, there will be a continued role for

held by government institutions has taken place in the access to information context, this area of the law provides a useful starting point for understanding the concept within the emerging open government arena.[26]

The statutory definitions of personal information vary somewhat across the different federal and provincial access to information frameworks, but at their core they typically characterize personal information as "information about an identifiable individual."[27] In Ontario, the leading test on identifiability was explained in *Ontario (Attorney General) v. Pascoe,* and was presented as a question of whether there is a *reasonable expectation* that an individual may be identified upon disclosure of the information.[28] The courts in British Columbia,

---

access to information schemes in order to serve requests for information that is not proactively disclosed or for information that constitutes personal information belonging to the requester: Scassa, "Privacy and Open Government," *supra* note 6 at 399.

[26] Data protection legislation informs open government policies, given that it is motivated by the aim of balancing the information needs of different actors with the risks of disclosing government-held information. Data protection statutes in Canada include the federal *Privacy Act, supra* note 22 and the *Personal Information Protection and Electronic Documents Act,* S.C. 2000, c. 5 [*PIPEDA*], which applies to the private sector. There are also provincial data protection statutes such as Alberta's *Personal Information Protection Act,* S.A. 2003, c. P-6.5 [*PIPA-AB*]; British Columbia's *Personal Information Protection Act,* S.B.C. 2003, c. 63 [*PIPA-BC*]; and Quebec's *An Act Respecting the Protection of Personal Information in the Private Sector,* C.Q.L.R. c. P-39.1 [*PPIPS-QC*], all of which have been deemed substantially similar to *PIPEDA* and therefore replace the federal scheme in certain situations. The provincial legislation further includes statutes that replace *PIPEDA* for certain situations relating to the handling of health information. In Ontario, see the *Personal Health Information Protection Act, 2004*, S.O. 2004, c. 3, Schedule A; in New Brunswick, the *Personal Health Information Privacy and Access Act,* S.N.B. 2009, c. P-7.05; and in Newfoundland and Labrador the *Personal Health Information Act,* S.N.L. 2009, c. P-7.01. See also the federal, provincial, and municipal access to information statutes, *supra* note 22.

[27] Under *PIPEDA*, *supra* note 26, s. 2, "personal information" is defined as "information about an identifiable individual." The federal *Privacy Act, supra* note 22, s. 3, defines "personal information" as "information about an identifiable individual that is recorded in any form" and provides examples of information that qualifies as well as exceptions to the rule at section 3. See further discussion in Teresa Scassa, "Geographic Information as 'Personal Information'" (2010) 10:2 O.U.C.L.J. 185 [Scassa, "Geographic Information]; Teresa Scassa, Jennifer A. Chandler & Elizabeth F. Judge, "Privacy by the Wayside: The New Information Superhighway, Data Privacy, and the Deployment of Intelligent Transport Systems" (2011) 74:1 Sask. L. Rev. 117 at 138; and IPC, "Transparency, Privacy and the Internet," *supra* note 5 at 3, explaining that information is considered personal information if there is a "reasonable expectation that an individual can be identified from the information itself or in combination with other information, whether or not it is publicly available."

[28] *Ontario (Attorney General) v. Pascoe,* 2002 CarswellOnt 3825, [2002] O.J. No. 4300, 166 O.A.C. 88 (Ont. C.A.) at para. 1, additional reasons 2002 CarswellOnt 4072 (Ont. C.A.) [*Pascoe*]. The *Pascoe* test has been applied in numerous tribunals and court decisions in Ontario. See e.g., *Ontario (Ministry of Correctional Services) v. Goodis*, 2008 CarswellOnt 365, 89 O.R. (3d) 457, 290 D.L.R. (4th) 102, [2008] O.J. No. 289 (Ont. Div. Ct.), additional reasons 2008 CarswellOnt 3853 (Ont. Div. Ct.); *Canada*

Newfoundland and Labrador, Prince Edward Island, and Saskatchewan appear to have adopted the Ontario test.[29]

At the federal level, the concept of "personal information" is also broadly defined.[30] The result is that personal information may include not only information that is viewed as highly personal or that may disclose details relating to "an individual's identity, intimacy, dignity and integrity," but also less-sensitive information like a person's name, phone number, or professional title.[31] Much like the provincial threshold set out above, the courts have explained that to constitute personal information, information must be "about" an individual and must permit or lead to the possible identification of that person.[32] In 2008, after *Pascoe* had provided an interpretation of "personal information" as this concept relates to the Ontario legislation, the Federal Court of Canada took a slightly different approach and set out an alternative test for identifiability in relation to the federal access to information regime in *Gordon v. Canada (Health)*.[33] Considering the meaning of "personal information" under Canada's *Access to Information Act,*[34] the court (without referring to the *Pascoe*

(Information Commissioner) v. Canadian Transportation Accident Investigation & Safety Board, 2006 CAF 157, 2006 FCA 157, 2006 CarswellNat 2738, 2006 CarswellNat 1277, [2007] 1 F.C.R. 203, 267 D.L.R. (4th) 451, [2006] F.C.J. No. 704 (F.C.A.), leave to appeal refused 2007 CarswellNat 800, 2007 CarswellNat 801 (S.C.C.) [*Information v. Accident Investigation and Safety*]; *Ontario (Energy Board)* (December 27, 2006), Doc. PO-2536, PA-060066-1 (Ont. Information & Privacy Comm.); *Ontario (Community Safety and Correctional Services) (*March 2, *2006),* Doc. PO-2456, PA-040268-2 (Information & Privacy Comm.); *Ontario (Community Safety and Correctional Services)* (October 31, 2006), Doc. PO-2518, PA-030365-2, PA-040280-1, PA-030407-3 (Information &d Privacy Comm.); *Ontario (Community Safety and Correctional Services) (*August 7, *2009),* Doc. PO-2811, PA08-213-2 (Information & Privacy Comm.) [*Community Safety and Corrections* (IPC-Ont.)].

[29] *British Columbia (Health Services)* (December 17, 2003), Doc. 03-42 (B.C. Information & Privacy Comm.); *Eastern Regional Integrated Health Authority* (June 26, 2007), Doc. 2007-008 (N.L. Information & Privacy Comm.); *Workers Compensation Appeal Tribunal* (April 19, 2005), Doc. 05-005 (P.E.I. Information & Privacy Comm.); *Regina Qu'Appelle Regional Health Authority* (January 9, 2013), Doc. LA-2013-001 (Sask. Information & Privacy Comm.).

[30] See discussion in Office of the Privacy Commissioner of Canada, "Legal Information Related to PIPEDA: Interpretation Bulletin," (11 December 2015), online: <https://www.priv.gc.ca/leg_c/interpretations_02_e.asp>, citing *Dagg v. Canada (Minister of Finance),* 1997 CarswellNat870, 1997 CarswellNat 869, [1997] 2 S.C.R. 403, 148 D.L.R. (4th) 385, [1997] S.C.J. No. 63 (S.C.C.); *Information v. Accident Investigation and Safety*, *supra* note 28; *Canada (Information Commissioner) v. Royal Canadian Mounted Police Commissioner,* 2003 CSC 8, 2003 SCC 8, 2003 CarswellNat 448, 2003 CarswellNat 449, REJB 2003-38212, [2003] 1 S.C.R. 66, [2003] S.C.J. No. 7 (S.C.C.) [*Information v. RCMP*]. See also *Canada (Information Commissioner) v. Canada (Minister of Natural Resources)*, 2014 CF 917, 2014 FC 917, 2014 CarswellNat 3963, 2014 CarswellNat 4596, 464 F.T.R. 308 (F.C.) at para. 41 [*Information v. Natural Resources*].

[31] *Information v. Natural Resources*, *supra* note 30 at paras. 29, 42.

[32] *Information v. Accident Investigation and Safety*, *supra* note 28 at para. 43.

standard) stated that the proper test for determining whether information was "about an identifiable individual" asks "[whether] there is a *serious possibility* that an individual could be identified through use of [the] information, alone or in combination with other available information."[35] As of yet there is little case law exploring the *Gordon* standard, though it has been considered in relation to provincial legislation in Alberta.[36] It remains somewhat unclear whether the *Gordon* standard is substantively different than that set out in *Pascoe*.[37] It is worth noting that on the facts of *Gordon,* the information released had already led to the identification of at least one individual, so the ministry easily convinced the judge that the standard of a "serious possibility" of identification had existed. As more cases apply *Gordon*, it may be possible to determine whether and to what extent the standard of proof required under each test differs.

It is clear that the concept of personal information relates to the risk of identifiability attached to the information being considered for disclosure to the public. The issue is that it is often quite difficult to determine reidentification risk in relation to a specific data set.[38] The world of "big data"[39] continues to see growth in the amount of data available[40] and the development of information

---

[33] *Gordon v. Canada (Minister of Health)*, 2008 CF 258, 2008 FC 258, 2008 CarswellNat 522, 2008 CarswellNat 6510, 324 F.T.R. 94(Eng.), [2008] F.C.J. No. 331 (F.C.) [*Gordon*].

[34] *Access to Information Act*, *supra* note 22.

[35] *Gordon, supra* note 33 at para. 34 [emphasis added].

[36] This was in relation to the province's *FIPPA-AB*, *supra* note 22. See *Alberta (Employment and Immigration)* (January 12, 2011), Doc. F2010-018 (Alta. Information & Privacy Comm.); *Out-Of-Country Health Services Committee* (January 26, 2012), Doc. F2012-04, H2012-01 (Alta. Information & Privacy Comm.). Recently, however, Alberta's Court of Appeal reviewed an adjudicator's decision that driver's licence and vehicle licence numbers constituted personal information, determining that the decision had been reasonable. The court cited both *Pascoe*, *supra* note 28 and *Gordon*, *supra* note 33 in its review of the authorities on the meaning of "personal information," but did not clearly rely on either standard to come to its final conclusion, and instead was satisfied that the information constituted "personal information" because it was possible to identify a particular person through the information collected with access to the proper database: *Leon's Furniture Ltd. v. Alberta (Information & Privacy Commissioner)*, 2011 ABCA 94, 2011 CarswellAlta 453, [2011] 9 W.W.R. 668, [2011] A.J. No. 338 (Alta. C.A.) at para. 49, leave to appeal refused 2011 CarswellAlta 1938, 2011 CarswellAlta 1939 (S.C.C.). It is therefore unclear whether Alberta will adopt the *Gordon* standard in relation to provincial legislation. In terms of federal legislation, see PIPEDA Case summary #2010-004 where a manager's remark revealing an employee's salary was found to have infringed the complainant's privacy, despite the fact that the complainant had consented to public disclosure of his salary in audited financial statements.

[37] See discussion in Scassa, "Geographic Information," *supra* note 27. In El Emam & Fineberg, *infra* note 85 at 15, the authors argue that the serious possibility threshold is more stringent than the reasonable expectation test.

[38] See UK, Information Commissioner's Office, *Anonymisation: Managing Data Protection Risk Code of Practice* (Cheshire, UK: ICO, 2012) at 16, online: < https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf > . [ICO, *Anonymisation*], noting that data custodians may determine that little or no risk of reidentification

technologies that enable cross-referencing of information from different sources.[41] This includes information proactively disclosed by governments, data released as part of the open data initiative, and an unknown amount of information in the hands of third parties.[42] Within this "data revolution,"[43] it may be almost impossible to determine whether reidentification can be achieved, or how easy it will be. This places public servants making decisions about the release of government-held data in a difficult position and raises questions about the extent to which reidentification risk must be considered in light of unknown sources of data and emerging information technologies.

The extent to which these future realities are relevant to reidentification risk is a question that arose indirectly in the recent Supreme Court of Canada decision in *Ontario (Ministry of Community Safety & Correctional Services) v. Ontario (Information & Privacy Commissioner)*.[44] In that case, a journalist

---

attaches to the release of information under their control, but may be unaware of the other data available to facilitate reidentification.

[39] "'Big data'" refers to "the massive and ever-expanding volume of digital data": Scassa, "Privacy and Open Government," *supra* note 6 at 407. It has similarly been defined as "data linked together, to create a digital picture that is bigger than the sum of the parts": Andy Williamson, "Big Data and the Implications for Government" (2014) 14:4 Leg. Info. Mgmt. 253 at 253. Rob Kitchin notes that the concept is often explained in relation to three main concepts: volume, velocity, and variety. It is huge in volume in that it exists in terabytes or petabytes; it is high in velocity in that it is created in real or near-real time; and it is available in a variety of formats. Other characteristics that may separate big data from small data are that it is "exhaustive in scope," "fine-grained in resolution," "relational in nature" and "flexible": Kitchin, *supra* note 8 at 68 [emphasis omitted].

[40] Recent estimates indicate that approximately 90 per cent of the world's data has been created in the last two years: Jodi LeBlanc, "Understanding Open Data: Don't Get Left Behind" (2014 February 18), *Canadian Government Executive* (blog), online: < https://cgexecblog.wordpress.com/2014/02/18/understanding-open-data-dont-get-left-behind > .

[41] Examples of reidentifications that have occurred from what was believed to have been anonymized data are discussed below in relation to the continued use of anonymization techniques.

[42] See Kitchin, *supra* note 8 at Chapter 4, discussing the sources of big data, which include numerous privately-held devices that enable private production and collection of information. See also Cynthia M. Gayton, "Beyond Terrorism: Data Collection and Responsibility for Privacy" (2006) 36:4 J. Information & Knowledge Management Systems 377 at 377, noting that "[l]ittle by little, seemingly insignificant pieces of data are being collected by not only the government entities and companies with whom consumers conduct business, but third party data brokers."

[43] As Kitchin, *supra* note 8 at Preface, xv explains,

> This revolution is founded on the latest wave of information and communication technologies. . .[which] are leading to ever more aspects of everyday life. . .and the worlds we inhabit to be captured as data and mediated through data-driven technologies. Moreover, they are materially and discursively reconfiguring the production, circulation and interpretation of data [and] producing "big data."

[44] *Ontario (Ministry of Community Safety & Correctional Services) v. Ontario (Information*

requested information about the whereabouts of sex offenders registered on Ontario's Sex Offender Registry, asking for the information to be disclosed by reference to the first three letters of the offenders' postal codes.[45] The adjudicator who first heard the dispute ordered disclosure of the information, rejecting the ministry's claims that the information could be withheld based on the personal privacy exemption or due to risks presented for law enforcement objectives.[46] The case was appealed to the Supreme Court of Canada, where the order for disclosure was upheld.[47]

In determining whether the information was personal information, the courts applied the *Pascoe* test, which, as previously noted, asks whether there is a *reasonable expectation* that an individual may be identified upon disclosure of the information. The ministry argued that there was a reasonable expectation that an individual could be identified from the information because the information could be combined with other publicly available information sources to identify the home addresses of such individuals.[48] The ministry cited numerous examples of publicly available sources that might lead to this outcome, including information available through the Internet; newspapers; voter registration lists; occupational licensing registries; property records; crime/court records; corporate proxy statements; stock holding reports; city directories; and birth, death and marriage records.[49] The adjudicator disagreed and found that the standard of identifiability had not been established and that the requested information was therefore not "personal information" within the meaning of the personal privacy exemption in section 2(1) of Ontario's *Freedom of Information and Protection of Privacy Act* (*FIPPA*).[50] The Court of Appeal and the Supreme Court of Canada upheld this decision on judicial review in oral reasons.[51]

& Privacy Commissioner), 2014 CSC 31, 2014 SCC 31, 2014 CarswellOnt 5105, 2014 CarswellOnt 5106, [2014] 1 S.C.R. 674, [2014] S.C.J. No. 31 (S.C.C.) [*Community Safety and Corrections (S.C.C.)*]. For a detailed discussion of this case and its implications for privacy and transparency, see Conroy & Scassa, *supra* note 5.

45 *Community Safety and Corrections* (IPC-Ont.), *supra* note 28 at 1.

46 *Ibid* at 17.

47 *Community Safety and Corrections (S.C.C.)*, *supra* note 44.

48 *Community Safety and Corrections (IPC-Ont.)*, *supra* note 28 at 7.

49 *Ibid* at 7.

50 *Ibid* at 11. Note that the adjudicator also rejected the ministry's attempt to establish an exemption under section 14(1)(e) and 14(1)(l), which both relate to risks to law enforcement objectives. The adjudicator's determination was based on the fact that both sections depended on the information being personal information: *Ibid* at 14-16.

51 *Ontario (Ministry of Community Safety & Correctional Services) v. Ontario (Information & Privacy Commissioner)*, 2012 ONCA 393, 2012 CarswellOnt 7088, 292 O.A.C. 335, [201] O.J. No. 2575 (Ont. C.A.), affirmed 2014 CarswellOnt 5105, 2014 CarswellOnt 5106 (S.C.C.), *Community Safety and Corrections (S.C.C.)*, *supra* note 44 at paras. 22-23.

The Supreme Court of Canada's decision provided new guidance on the standard for identifiability in the access to information context.[52] The court's opinion suggests that to show a risk of identifiability, the government must provide evidence that is directly related to the information under consideration. This interpretation is largely based on the court's rejection of what it characterized as "unconvincing and generic scholarly research" as evidence of a risk of identifiabilty in the case.[53] This supports the view that going forward, risk of identifiability will need to be established by reference to specific information posing an immediately observable risk in the circumstances at hand. In other words, a general and developing understanding of privacy risks in the access to information, open data, and proactive disclosure contexts will not be viewed as adequate reason to withhold information.

To the extent that the decision is relevant to the open government context, it places public servants deciding on the release of data in a difficult position. Some data sets may appear entirely free of personal information—but their use in combination with other data might lead to the identification of specific

---

[52] Note that by the time the case reached the Supreme Court of Canada, the issues had been reformulated into a question about the standard of review to be applied to the adjudicator's decision, a question of whether the adjudicator had made a reviewable error by ordering disclosure for purposes inconsistent with *FIPPA* or with *Christopher's Law* (which governs the Ontario Sex Offender Registry), and whether the Commissioner had erred by interpreting the scope of *FIPPA*'s law enforcement exceptions by applying an elevated evidentiary standard to its provisions: *Community Safety and Corrections (S.C.C.)*, *supra* note 44. The personal information issue remained relevant, however, because the court agreed with the adjudicator that the law enforcement exemptions depended on the information being personal information within the meaning of *FIPPA*.

[53] *Ibid* at para. 60. The Supreme Court of Canada did not specify the literature advanced by the ministry in its reasons. In the Privacy Commissioner's initial order, the only literature reviewed was a piece co-authored by then Ontario Information and Privacy Commissioner Ann Cavoukian: Information and Privacy Commissioner of Ontario, "The New Federated Privacy Impact Assessment (F-PIA): Building Privacy and Trust-Enabled Federation," by Ann Cavoukian (Toronto: IPC, 2009). The decision specifically referred to the following passage, which had been quoted by the ministry:

> Personal information is any information, identifying or otherwise, relating to an identifiable individual. Specific [personal information] may include one's name, address, telephone number, date of birth, age, marital or family status, e-mail address, etc. For example, credit cards, debit cards, social insurance/security numbers, driver's licenses and health cards contain a great deal of sensitive personal information. Moreover, it is also important to point out that almost any information, once linked to an identifiable individual, becomes personal information, be it biographical, biological, genealogical, historical, transactional, locational, relational, computational, vocational or reputational.

(at section 2.1 in the original paper and discussed in *Community Safety and Corrections* (IPC-Ont.), *supra* note 28 at 7 [emphasis omitted]). The ministry argued that the description of personal information in the passage "'broadens'" the definition of personal information in the context of the *FIPPA-ON*, *supra* note 22. The Commissioner disagreed and explained that while personal information may be viewed as an expanding category, its "hallmark" is the requirement of identifiability.

individuals. In other cases, anonymization techniques may be applied to data that are known to contain identifying information. Yet as discussed below, while information may sometimes appear to have been anonymized, reidentification may occur when such information is combined with publicly available information or with the unknown amount of privately-held data.[54] The extent to which outside data sources will become available online is difficult to predict and will depend on the exercise of discretionary judgment by data custodians in possession of those data sets. This discretionary authority was emphasized in a recent Federal Court of Canada judgment reviewing the ministry of natural resources' decision not to disclose the names, phone numbers, and business titles of individuals in the private sector following an access to information request for those details.[55] The court agreed with the minister that the information constituted personal information under the federal *Privacy Act* and determined that the decision to withhold the information was reasonable.[56] By the time the decision had reached the court, the personal information at issue was publicly accessible through online searches, and the applicant argued that the government was therefore required to disclose the data.[57] The court, however, determined that the fact that the information was publicly accessible did not change the fact that the minister had the discretionary authority, not an obligation, to disclose the personal information.[58] The court refused to usurp that authority and indicated that it was "a matter for the parties to address and not the court."[59]

Given the uncertainties about the amount of outside information that may be relevant to reidentification risk, the challenge for public servants is to know how significant the risk is, and how to weigh this level of risk with the competing transparency values. Because the risks and benefits of disclosure depend on the nature of the information itself, decisions relating to the release of individual data sets require a case-by-case assessment. Considering the information that may be contemplated for open data or proactive disclosure, it may sometimes be clear that the data do not contain personal information about any individual. For example, this may be the case with data about the location of bicycle station facilities in a given municipality, boundaries for urban green spaces, or the number of drinking fountains in public parks.[60] In such cases the goals of open

---

[54] Based on this and other experiences in which reidentifications have taken place through use of what was previously believed to have been anonymized information, emerging academic discussion warns of the vulnerabilities of anonymization techniques. See discussion in Paul Ohm, "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization" (2010) 57:6 U.C.L.A. L. Rev. 1701; and Jon P. Daries et al, "Privacy, Anonymity, and Big Data in the Social Sciences" (2014) 57:9 Communications ACM 56 at 57.

[55] *Information v. Natural Resources, supra* note 30.

[56] *Ibid* at paras. 52-59.

[57] *Ibid* at paras. 54-56.

[58] *Ibid* at para. 61.

[59] *Ibid.*

government support the view that such information should be released.[61] Sometimes data may contain information that would be highly unlikely to lead to the reidentification of specific individuals—e.g., general statistics relating to the prevalence of pollen allergies.[62] When this type of information cannot assist in reidentification, it may be viewed as non-identifying information and released.[63] In other cases however, the information may be potentially useful

---

[60] This information is provided on the City of Toronto's open data portal: Toronto, "Data Catalogue," *supra* note 17.

> The significant interest in harnessing public transit data illustrates how open data (in either the public or private sector) can support innovation. The release of public transit data has so far resulted in numerous useful applications and new commercial products. The situation introduced complex questions about the copyright in information relating to public transit routes and schedules. This was partly motivated by concerns over the quality of information being released to the public; however, as open data evolves, data custodians in both the public and private sectors are adjusting to a cultural shift in which concerns about how information will be used do not provide a compelling reason for refusing to disclose data. See discussion in Teresa Scassa, "Public Transit Data through an Intellectual Property Lens: Lessons about Open Data" (2014) 41:5 Fordham Urb. L.J. 1759 at 1777 [Scassa, "Public Transit Data"].

[61] In some instances, non-personal information may be subject to a specific statutory exception. For instance, the federal *Access to Information Act* provides exemptions for information relating to the management of the economy, national security and defence, international affairs, law enforcement, Cabinet confidences, and information that is subject to solicitor-client privilege: *Access to Information Act, supra* note 22, ss. 15-26. See discussion of exemptions in the federal access to information scheme in Office of the Privacy Commissioner of Canada, *Access to Information and Privacy: Process and Compliance Manual* (OPC, 2008) ch. 5 at 59ff, online: <https://www.priv.gc.ca/au-ans/atip-aiprp/manual_e.pdf>. It is worth noting that the Canadian government has identified four high priority categories of information, which may provide further guidance in making the decision to proactively disclose information in this category. The high priority areas represent information relating to national statistics, election results, government budget information, and national maps. Secondary priorities, identified as areas of "high value data" include information relating to companies, crime and justice, earth observation, education, energy and environment, finance and contracts, geospatial data, global development, government accountability and democracy, health, science and research, social mobility and welfare, statistics, and transportation and infrastructure. The government has expressed a commitment to engaging with the public to further assess user needs and specific interests for open government data: Canada, "G8 Action Plan," *supra* note 7. Priority goals are also emerging in other jurisdictions. In the U.S., these appear to be primarily driven by the Obama administration's focus on increasing transparency, public engagement, and collaboration, while the European Commission emphasizes the economic gains to be derived from the open data movement: Zuiderwijk & Janssen, *supra* note 7 at 17.

[62] Health Canada, "Overview of Factors Affecting the Risk of Re-identification in Canada," by Khaled El Emam (Ottawa: Health Canada, 2006) at 5, online: <www.ehealthinformation.ca/wp-content/uploads/2014/07/2006-Overview-of-Factors.pdf>. [El Emam].

[63] *Ibid* at 5.

to secondary users while also presenting a reidentification risk. This concern may arise, for example, with restaurant inspection reports;[64] health and safety reports on personal services offered within a city;[65] or geodemographic data such as information relating to family structure, marital status, and language spoken within households in different regions across Canada.[66] The extent to which this type of data presents a risk of reidentification when combined with other information will depend on the exact nature and level of specificity within the information. One relevant information source that may be combined with individual data sets to reveal personal information is publicly available personal information, which is defined in the next section.

II. THE DEFINITION OF PUBLICLY AVAILABLE PERSONAL INFORMATION

"Publicly available personal information" can still be personal information, and has been specifically included in private sector data protection legislation.[67] Some publicly available personal information is collected or disclosed by governments.[68] Examples include the types of data found in public registries (e.g., court files or land titles registries).[69] Because the public can already access this type of personal information, albeit with effort required to conduct the

---

[64] A finding that employees at a specific restaurant were not properly washing their hands could, for example, be combined with other available information to identify the delinquent hand-washers.

[65] See City of Ottawa, "Public Health Inspection Data — Restaurant" (Ottawa: City of Ottawa, 2014), online: <data.ottawa.ca/dataset/public-health-inspection-data/resource/2301656b-3695-4bee-9048-e0312441f637>, which provides inspection data for food establishments and for personal care services such as those provided in hair salons and barbershops, nail salons, and tattoo shops.

[66] See e.g., Statistics Canada, "GeoSearch," (Statistics Canada, 11 April 2016), online: <https://www12.statcan.gc.ca/census-recensement/2011/geo/ref/geosearch-georesearche-eng.cfm>. New research is exploring individual levels of comfort among Americans with respect to government disclosures as part of the growing open government movement. 82 per cent of adults surveyed expressed that they were comfortable with government sharing of health and safety records from restaurants, while 62 per cent were okay with government sharing of information about criminal records of individual citizens. In contrast, only 22 per cent were comfortable with government sharing of information about mortgages for individual homeowners. See these and additional findings at: Horrigan & Rainie, *supra* note 11.

[67] See, for example, *PIPEDA*, *supra* note 26 and the accompanying *Regulations Specifying Publicly Available Information*, S.O.R./2001-7, s.1 [*Regulations Specifying Publicly Available Information*].

[68] See discussion of publicly available personal information in the related context of PIPEDA in Teresa Scassa, "Privacy and Publicly Available Personal Information" (2013) 11:1 C.J.L.T. 1 at 10 [Scassa, "Privacy and Publicly Available Personal Information"].

[69] Scassa, "Privacy and Open Government," *supra* note 6 at 403; Elizabeth F. Judge, "Copyright, Access, and Integrity of Public Information" (2008) 1 J.P.P.L. 427 at 430.

physical or online search for the records, it might be viewed as unobjectionable to release the data as open data.[70] However, the broad disclosure of publicly available personal information in digital formats introduces certain privacy concerns that must be addressed as part of the open government policy process. Moreover, the release of this type of personal information as open data may conflict with certain areas of the law as it applies to secondary uses of this information.

With respect to the first issue, the online disclosure of publicly available personal information may raise privacy concerns that are not as relevant when the information is only available in hardcopy.[71] For example, information stored exclusively in a physical registry (and therefore that requires a certain degree of effort to retrieve) becomes easy to access and disseminate once it is released online. An example from the U.S. illustrates some of the issues.[72] During the campaign leading up to the vote in California on a proposal to amend the state constitution to ban same-sex marriage (also known as Proposition 8), opponents of the amendment obtained information about the names, street addresses, and contribution amounts of donors who had supported the ban with contributions of $100 or more.[73] The information was already publicly accessible as election contribution information,[74] but the opponents published an interactive map broadly disseminated online that gave viewers quick visual representations of

---

[70] The court in *Gombu v. Ontario (Assistant Information & Privacy Commissioner)*, 2002 CarswellOnt 1599, 59 O.R. (3d) 773, 214 D.L.R. (4th) 163, [2002] O.J. No. 1776 (Ont. Div. Ct.) at para. 28, additional reasons 2002 CarswellOnt 2132 (Ont. Div. Ct.), leave to appeal allowed 2002 CarswellOnt 2874 (Ont. C.A.) took this view and held that the disclosure of personal information in electronic format where that is already available in hardcopy represents only a "minimal further intrusion upon. . .personal privacy." See discussion of the case in IPC, "Transparency, Privacy and the Internet," *supra* note 5 at 4.

[71] Canadian Judicial Council, "Synthesis of the Comments on JTAC's Discussion Paper on Open Courts, Electronic Access to Court Records, and Privacy," by Lisa M. Austin & Fréderic Pelletier (CJC, 2005) at 17, online: < http://www.publications.gc.ca/site/eng/319625/publication.html > [Austin & Pelletier]

[72] While the situation described would need to be considered under the applicable American legislation, the example illustrates the privacy considerations that arise when publicly available information that was once only available in hardcopy is published online. As such, it is relevant to consider this situation in the Canadian environment, where certain types of publicly available information have also traditionally been held in hardcopy and where the interest in making such information available online will likely continue to grow.

[73] Scassa, "Privacy and Open Government," *supra* note 6 at 404; David Lourie, "Rethinking Donor Disclosure after the Proposition 8 Campaign" (2009) 83:1 S. Cal. L. Rev. 133 at 134.

[74] The rules requiring disclosure of donation contributions have become a point of debate. On the one hand, it is argued that disclosure is required in order to ensure transparency in relation to the role played by money in campaigns; on the other hand, due to the potential for negative consequences such as the type of publicity experienced by donors who supported Proposition 8, it is argued that identifying information relating to donors should not be made public at all. See discussion in Lourie, *supra* note 73 at 136.

donor name and amount matched to specific plotted addresses on the map.[75] The publication of the map made the information much more accessible to the public and raised privacy concerns that were not as relevant when the information was simply held within a public registry.[76] It is important to consider that the digital availability of publicly available personal information was unlikely to have been thoroughly contemplated when decisions were being made to make different types of personal information publicly available.[77] In addition to the information becoming more visible, a crucial consideration in deciding whether to make this type of personal information available online is that the move would make the data available for cross-referencing with other non-identified data, thereby increasing the risk of reidentification of purportedly anonymized data.[78]

In addition to the above, the release of publicly available personal information as open data may conflict with other areas of the law regulating the collection, use, and disclosure of personal information. While open data involves the release of information in reusable formats and encourages reuse under open licences,[79] some secondary users are restricted in terms of how they can use publicly available personal information. The issue is particularly relevant in the case of commercial organizations whose activities fall under the scope of the federal *Personal Information Protection and Electronic Documents Act* (*PIPEDA*) or its provincial equivalents, which set out rules for the collection, use, and disclosure of personal information by private sector companies engaged in commercial activities.[80] While these statutes would generally not apply to uses

---

75 Scassa, "Privacy and Open Government," *supra* note 6 at 404.

76 *Ibid*. See also Lourie, *supra* note 73 at 133-135, explaining that the impact on individuals who were revealed as having supported Proposition 8 included targeted protests and boycotts of businesses.

77 See Lourie, *supra* note 73 at 138, making the same argument in relation to campaign disclosure laws that led to the situation with Proposition 8 donor information, discussed above. See also Austin & Pelletier, *supra* note 71 at 6, discussing the shift in the balance of the open court principle and privacy considerations as information technology continues to advance. Finally, see comments in Teresa Scassa, "Balancing Privacy with Online Access to Court and Tribunal Decisions: Lessons for Open Government" (16 March 2015), *Teresa Scassa* (blog), online: <www.teresascassa.ca/index.php?option=com_k2&view=item&id=182:balancing-privacy-with-online-access-to-court-and-tribunal-decisions-lessons-for-open-government&Itemid=80> [Scassa, "Balancing Privacy"].

78 See again discussion of reidentification risk and the definition of personal information (above).

79 Office of the Chief Information Officer & Ministry of Labour, Citizens' Services and Open Government, "Open Information and Open Data Policy" (Office of the CIO, 2011) at 7, online: <www.cio.gov.bc.ca/local/cio/kis/pdfs/open_data.pdf>. See also Canada, "G8 Action Plan," *supra* note 7, emphasizing the aim of releasing "as much data in as many open formats as possible."

80 *PIPEDA*, *supra* note 26. The provincial legislation includes *PIPA-AB*, *supra* note 26; *PIPA-BC*, *supra* note 26; and *PPIPS-QC*, *supra* note 26. With respect to the public sector legislation, nothing in the *Privacy Act, supra* note 22 specifically prohibits the online

for journalistic or private purposes,[81] they would limit the use of the information by private sector companies engaged in commercial activities. Specifically, for publicly available personal information released to the public, private sector actors are limited by the rule that secondary use of the information must be directly related to the purpose for which the information was collected.[82] These restrictions as set out in *PIPEDA* (or in B.C. or Alberta's *PIPA* statutes) limit how private sector actors may use information released by government. It is therefore important to consider that by making such data available under an open licence that permits its unrestricted use, governments would likely facilitate unauthorized uses of the information. As such, we consider options for dealing with publicly available personal information as well as other personal information in the open government context further below.

---

disclosure of publicly available personal information. Under s. 69(2), publicly available personal information is specifically excluded from the rules in ss. 7 and 8, which set out a consent-based regime for government use of personal information beyond that for which the information was collected (including specific and limited exemptions to the requirement that consent be obtained).

[81] Austin & Pelletier, *supra* note 71 at 18. One of the important issues to consider is that private individuals (such as nosy neighbours or co-workers) may be motivated to access the information once it is made accessible in digital form; however, this activity would not be covered by *PIPEDA*. See Scassa, "Privacy and Publicly Available Personal Information," *supra* note 68 at 16, discussing the release of records of judicial and quasi-judicial bodies as a category of publicly available personal information and noting that "some of the privacy concerns that relate to administrative tribunal decisions stem from the potential for nosy neighbours, co-workers, ex-spouses or even malefactors to browse electronic decisions for information about individuals." The issue as it relates to court documents is considered in further detail below.

[82] *PIPEDA*, *supra* note 26, s. 7(1)(d) allows organizations to collect personal information without the knowledge or consent of the individual if the information is publicly available and is specified by the regulations. The regulations set out a list of categories of "information and classes of information" that represent publicly available personal information for the purposes of the *Act*: telephone directory information, professional or business directories, registry information, records of judicial or quasi-judicial bodies, and published information: *Regulations Specifying Publicly Available Information*, *supra* note 67, s.1. For instance, a private sector organization may only use information that appears in a government-run registry for purposes directly related to that for which the information appears in the registry. In the case of registry information, this purpose may be set out in the authorizing legislation; in other cases, permissible collection, use, or disclosure of the information will be judged against the standard of what a reasonable person would consider reasonable in the circumstances. See further discussion, including a discussion of the comparable provisions in the B.C. and Alberta legislation, in Scassa, "Privacy and Publicly Available Personal Information," *supra* note 68 at 14.

III. STRATEGIES FOR MANAGING PRIVACY IN OPEN DATA AND PROACTIVE DISCLOSURE

(a) Data Minimization

The new information environment into which more and more government data and information is being released introduces a need to reassess approaches to collecting and disclosing personal information within the public sector.[83] To avoid creating unnecessary privacy risks, Ontario's former Information and Privacy Commissioner recommended that public sector actors adhere to a policy of "data minimization" according to which interactions with the public "begin with non-identifiable interactions and transactions as the default."[84] By collecting only what personal information is required to operate government programs, public sector actors can alleviate some of the privacy risks that arise when information relating to those same programs is released to the public. The policy of data minimization does not need to be limited to new collections by government departments; data minimization principles can also apply to the disclosure of information to be used for secondary purposes. In such cases, the amount of personal information disclosed should be limited to what is required to fulfill the intended purposes, whether it be health research, innovation, or increased transparency in relation to government operations.[85]

In terms of information collected by government, however, it is important to note that a policy of data minimization does not address all of the concerns that may exist with respect to information already in the hands of government. To the extent that this information may contribute to the goals of the open government movement, it should be considered for proactive release in accordance with the guidelines set out in this paper. Moreover, it remains the case that the

---

[83] See discussion in Lourie, *supra* note 73 at 139, specifically arguing for a new framework for disclosure of publicly available personal information because of the emerging information environment and resultant privacy concerns. The issue of publicly available personal information is discussed in more detail below.

[84] Information and Privacy Commissioner of Ontario, "Privacy and Government 2.0: The Implications of an Open World," by Ann Cavoukian (Toronto: IPC, 2009) at 10, online: <https://www.ipc.on.ca/images/Resources/priv-gov-2.0.pdf> [Cavoukian, "Privacy and Government 2.0"]. See also IPC, "Transparency, Privacy and the Internet," *supra* note 5 at 8.

[85] The need to minimize disclosures of personal health data intended for secondary use to what is required to fulfill the intended purposes is discussed in Khaled El Emam & Anita Fineberg, "An Overview of Techniques for De-Identifying Personal Health Information" (Ottawa: CHEO Research Institute, 2009) at 6. The authors note at 7 that even where the law permits disclosure of personal information, data custodians should consider releasing information in anonymized form if this satisfies the legitimate needs of the secondary user. See also Sherman, *supra* note 23 at 50, emphasizing a similar principle of "[c]ollection [l]imitation" as a way to address privacy concerns relating to the online dissemination of information about court proceedings, an issue that is discussed in further detail below.

government must collect a great deal of personal information in order to operate its various programs effectively.[86] Due to these considerations, data minimization as a best practice only represents a starting point to addressing the privacy concerns relating to the release of government information.

(b) Cross-Government Communication

A second general best practice involves cooperation and communication between different levels of government as well as between different agencies and departments within the same government, in order to facilitate knowledge sharing about the decision-making process relating to the release of government-held information. Effective communication between government actors could also spread awareness about the kind of information being considered for release as part of the broader open government initiative.[87] As discussed below, such information may be relevant to the assessment of reidentification risk for new data sets.[88] This effort towards communication and cooperation may require a

---

[86] It is worth emphasizing that even when a policy of data minimization is in place, errors in judgment and over-collection of personal information may still occur. This was seen, for instance, with the controversial Longitudinal Labour Force File (LLFF), which was operated by Human Resources Development Canada (HRDC). A debate arose surrounding the database in the late 1990s after Canada's Privacy Commissioner at the time, Bruce Phillips, performed an audit of the database. Commissioner Phillips acknowledged HRDC's need to collect information required to fulfill its mandate of monitoring health and safety issues in the workforce, administering income security programs, and assisting individuals to secure paid employment, but challenged HRDC's extensive collection and retention of information beyond what was required to perform its duties. Prior to the discussion instigated by the LLFF, it was believed that government departments were only collecting and storing personal information to the extent required by the mandates of the individual departments, and that privacy was being protected by keeping information organized into "silos" within each department. The audit revealed, however, that HRDC was compiling information from different departments to create individual profiles that were attached to personal identifiers. The LLFF was eventually scrapped due to the privacy concerns relating to the creation of these extensive personal profiles: Office of the Privacy Commissioner of Canada, *Annual Report to Parliament 1999-2000* (Ottawa: Public Works and Government Services, 2000) at 64, online: <www.priv.gc.ca/information/ar/02_04_08_e.asp>; Office of the Privacy Commissioner of Canada, News Release, "Privacy Commissioner Applauds Dismantling of Database" (29 May 2000), online: <https://www.priv.gc.ca/media/nr-c/archive/02_05_b_000529_e.asp>.

[87] Such practices might increase consistency in the type of information released across all municipalities or all provincial governments, which in turn might enhance certain uses of such data for comparative analyses or for the design of applications that could be used in multiple cities or regions.

[88] ICO, *Anonymisation*, *supra* note 38 at 40, arguing that:

> Organisations should seek to share information about planned disclosures with other organisations, to assess risks of jigsaw identification. For example it would be helpful for public authority A to know that public authority B is also planning an anonymised disclosure at the same time, one on health and one on welfare,

restructuring of relationships between departments as well as between decision-makers at the ground level and higher-level politicians communicating information about the open government movement to the public.[89]

As with data minimization, communication and cooperation between government actors only serve as a starting point in addressing the risks involved in making more and more government-held information available to the public. For the case-by-case decisions that must be made for individual data sets and other types of information, public servants must consider the particular privacy issues that relate to the specific type of information being contemplated for release. This paper proposes a series of questions to use in evaluating individual data sets being contemplated for release, particularly where the information may be made available as open data.

(c) Assessing Privacy Risks

When privacy concerns arise in relation to the release of information as part of government programs and services, Privacy Impact Assessments (PIAs) may be used to gain a better understanding of the risks and benefits of disclosing the information at issue. The Canada Revenue Agency defines a PIA as "a process used to determine how a program or service could affect the privacy of an individual. . .[which may] also help to avoid or lessen possible negative effects on privacy that might result from a program or service."[90] PIAs have become standard practice in evaluating the privacy risks of various programs and services in many parts of Canada's public sector.[91] When a government department plans

---

> both using similar geographical units. They can then assess the risks collectively and agree [on] mitigation for both datasets.

89 See Zuiderwijk & Janssen, *supra* note 7 at 26, noting that "[a] large gap exists between the key objectives of the open data policies which reflect the ambitions of politicians on the one hand, and, on the other hand, the realities of public servants working for government organizations" and arguing for "systematic collaboration" to address such issues. See also Roy, *supra* note 1 at 426, arguing for a national (as opposed to federal) strategy for the coordination of open data initiatives.

90 Canada Revenue Agency, "Privacy Impact Assessment," (Ottawa: CRA, 2016), online: < http://www.cra-arc.gc.ca/gncy/prvcy/pia-efvp/menu-eng.html > [CRA, "PIAs"]. Another definition offered by the Treasury Board of Canada Secretariat describes a PIA as:

> [A] process that helps departments and agencies determine whether new technologies, information systems and initiatives or proposed programs and policies meet basic privacy requirements. It also assists government organizations to anticipate the public's reaction to any privacy implications of a proposal and as a result, could prevent costly program, service, or process redesign.

Treasury Board of Canada Secretariat, "Privacy Impact Assessment Guidelines: A Framework to Manage Privacy Risks Guidelines," (TBS, 2006), online: < www.tbs-sct.gc.ca/pol/doc-eng.aspx?id = 12451&section = HTML > [TBS, "PIA Guidelines"].

91 The Treasury Board of Canada Secretariat has listed a number of situations in which a PIA should be used, including in the design of a new program or service as well as where information previously collected for a government program is to be used for research or

to undertake an initiative that requires the disclosure of personal information, a PIA may help illustrate the benefits and risks of disclosure in the specific circumstances at hand and may facilitate an approach that avoids unnecessary consequences for personal privacy.[92]

To evaluate records being considered for proactive disclosure or data sets being considered for release as open data, public servants may similarly benefit from a set of predetermined questions designed to underscore the privacy risks at stake as well as the potential benefits of disclosure. Based on the standard PIA

---

statistical purposes. See "A checklist to determine when to do a PIA" in TBS, "PIA Guidelines," *supra* note 90. The Government of Canada's "Directive on Privacy Impact Assessment" (Canada, "Directive on Privacy Impact Assessment," (2010), online: < https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id = 18308 >) requires that most federal government departments use PIAs to evaluate their programs and services. See discussion of the federal policy in Office of the Privacy Commissioner of Canada, "Expectations: A Guide for Submitting Privacy Impact Assessments to the Office of the Privacy Commissioner of Canada," (Ottawa: OPC, 2011) at 1, online: < https://www.priv.gc.ca/information/pub/gd_exp_201103_e.pdf > [OPC, "PIA Guide"]. PIAs have played a role in the developing open government movement: see Office of the Information and Privacy Commissioner for British Columbia, "Early Notice and Privacy Impact Assessments to the OIPC under the *Freedom of Information and Protection of Privacy Act*" (OIPCBC, 2012), online: < https://www.oipc.bc.ca/guidance-documents/1434 > , promoting the use of PIAs in the access to information context. On its open data website, the City of Calgary similarly notes the use of PIAs in responding to access to information requests: City of Calgary, "Open Data Catalogue," online: < https://data.calgary.ca/OpenData/Pages/DatasetListingAlphabetical.aspx > . Finally, the City of Toronto supports the use of PIAs in a privacy policy designed to support the city's efforts to move towards open government: City of Toronto, "Protection of Privacy Policy," (Toronto: City of Toronto, 2014) online: < http://www1.toronto.ca/City%20Of%20Toronto/City%20Clerks/Corporate%20Information%20Management%20Services/Files/pdf/P/ProtectionOfPrivacyFinalAODA.pdf > . The use of PIAs for open government has been considered as a possible best practice or as a mandatory statutory requirement: Article 29 Data Protection Working Party, *supra* note 24 at 6.

92 For instance, the Edmonton Police Service was criticized for actions taken as part of Project Operation Warrant Execution (Project OWE). Project OWE involved the online release of pictures and personal details of individuals who had outstanding warrants and was intended to publicly shame such individuals into coming forward to resolve the outstanding matters. The Office of the Privacy Commissioner of Alberta issued an investigation report on the events, which criticized the police for having disclosed more personal information than what was necessary to carry out its responsibilities in a reasonable manner. The Commissioner argued that a PIA could have helped to identify the privacy impact of the operation (which would have facilitated police action to reduce that impact). See Office of the Information and Privacy Commissioner of Alberta, "Investigation Report F2014-IR-01: *Investigation into 'Project Operation Warrant Execution' for Compliance with the Freedom of Information and Protection of Privacy Act*" (OIPCAB, 2014) online: < https://www.oipc.ab.ca/media/127989/F2014-001IR.pdf > ; and discussion of the events in Chris Berzins, "Public Shaming by Public Bodies: The Investigation of Project OWE" (8 May 2015), *Privacy Scan* (blog), online: < www.privacyscan.ca/issues/2015/may-8-2015-public-shaming-by-public-bodies-the-investigation-of-project-owe > .

process[93] used in the evaluation of programs and services as well as relevant questions about the disclosure of records or data sets as identified in the surrounding literature, we offer a series of questions and factors to be considered as part of the decision-making process as it relates to records or data sets being contemplated for disclosure to the public.

(i) Table 1: Questions Relating to Purpose of Disclosure/Privacy Risk

Question	**Explanation/Examples**
1. What is the intended purpose of disclosure of the information/release of data?	E.g., support for government transparency, open engagement, economic development, research, or innovation.[94]

---

[93] The Office of the Privacy Commissioner of Canada has identified six key steps of a PIA:

i. Identify all personal information within the program or service being evaluated and determine how it will be used;
ii. Apply a four-part test (developed in *R. v. Oakes*, 1986 CarswellOnt 95, 1986 CarswellOnt 1001, EYB 1986-67556, [1986] 1 S.C.R. 103, 53 O.R. (2d) 719, [1986] S.C.J. No. 7 (S.C.C.)) that asks:
   a. whether the measure is demonstrably necessary to meet a specific need;
   b. whether the measure is likely to be effective in meeting that need;
   c. whether the loss of privacy is proportional to the need; and
   d. whether there exists a less privacy-invasive method of achieving the same end

to measure necessity and proportionality for highly invasive initiatives or information technologies;
iii. Apply the ten privacy principles (accountability; identifying purposes; consent; limiting collection; limiting use; disclosure, and retention; accuracy; safeguards; openness; individual access; and challenging compliance);
iv. Map out where personal information will be sent after collection;
v. Identify the existence and level of privacy risks; and
vi. Determine ways to eliminate or reduce privacy risks to an acceptable level.

Office of the Privacy Commissioner of Canada, "Fact Sheets: Privacy Impact Assessments," (OPC, 2011), online: <https://www.priv.gc.ca/resource/fs-fi/02_05_d_33_e.asp>; OPC, "PIA Guide," *supra* note 91. The Treasury Board of Canada Secretariat (TBS) describes a similar process and has issued questionnaires to assist in the application of PIAs. The TBS divides the process into the following four key steps:

(i) project initiation, which involves defining the scope of the PIA process and designating resources to accomplish the PIA;
(ii) data analysis of the personal information contained in the relevant data set;
(iii) privacy analysis, which involves completion of privacy analysis questionnaires (which reflect the ten privacy principles outlined above in the guide to PIAs provided by the OPC); and
(iv) a privacy impact assessment report.

The final step is considered crucial as a privacy impact assessment report could be used for future situations in which the privacy concerns are similar: TBS, "PIA Guidelines," *supra* note 90.

[94] See above discussion on the goals of open government.

Question	Explanation/Examples
2. Does the data set contain any identifying variables? If so, are the identifying variables relevant to the purpose of disclosure identified under question 1?	Includes any information that relates directly to an individual—e.g., name, full address, telephone number, email address, or social insurance number.[95]
3. Does the data set contain any quasi-identifiers? If so, are these quasi-identifiers relevant to the purpose of disclosure identified under question 1?	Variable(s) that may indirectly identify a given individual—e.g., postal code or other location information, sex, country of birth, profession, criminal history.[96]
4. Is the information being disclosed for a transitory purpose?	If there is only a temporary need for the public to have access to the information, this may weigh against online disclosure (which makes the information permanently available).[97]
5. If the data set contains quasi-identifiers, are any individuals uniquely identifiable by reference to those variables?	May result from the inclusion of any quasi-identifier or combination of quasi-identifiers that relate only to one individual within the sample; the more quasi-identifiers included the greater the probability that the data include a unique individual as described by those variables.[98]

---

[95] El Emam & Fineberg, *supra* note 85 at 7. If a data set includes identifying variables, it is clearly not de-identified or anonymized: El Emam, *supra* note 62 at 4.

[96] El Emam & Fineberg, *supra* note 85 at 9.

[97] See IPC, "Transparency, Privacy and the Internet," *supra* note 5 at 7, where the Commissioner argues that

> [I]n the case of a minor variance application, the application and supporting materials are created for purposes related to the application process. Once the application process is complete and the records have reached the end of the applicable retention period, there may be no ongoing need for those records to be made publicly available. In such cases, online disclosure, which makes the information permanently available online, may not be required in order to fulfill the government's goal of supporting transparency.

[98] See discussion in El Emam, *supra* note 62 at 7, emphasizing uniqueness within data containing quasi-identifiers as a factor contributing to the risk of reidentification. The author notes at 13 that the inclusion of postal codes and other geographical information as quasi-identifiers within data that reflects personal information may increase the likelihood that an individual will be unique and therefore identifiable. As noted above, the Government of Canada's open data portal is viewed as "[p]articularly strong in geospatial data": see Mewhort, *supra* note 12 at 6.

Question	Explanation/Examples
6. In its identified form, is the information highly personal?	E.g., data relating to a person's physical or mental health condition, racial or ethnic origin, political opinions, religious beliefs or other beliefs of a similar nature, sexual life, criminal history, allegations of criminal conduct or involvement in ongoing criminal proceedings.[99]
7. Are any of the variables in the data easy to identify in specific individuals?	E.g., characteristics that are visible such as an illness that can be outwardly observed, information about a person's physical appearance.[100]
8. Is it reasonably likely that any one individual could be reidentified from the data set?	Requires consideration of outside sources of information that may be combined with the data, other factors explored above in relation to the *Pascoe* standard of reidentification.
9. Could reidentification be expected to have serious negative consequences for the individual(s) concerned?	E.g., potential for damage, distress, financial loss.[101]

The first question in the table above is meant to engage the balance between privacy and the benefits of disclosure. If the second and third questions reveal that the information at issue includes personal or quasi-personal identifiers that are not relevant to the purpose for disclosure, the identifiers should be removed

---

[99] See Information and Privacy Commissioner of Ontario, "Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy," by Ann Cavoukian & Khaled El Emam (Toronto: OPC, 2011) at 4, online: <https://www.ipc.on.ca/images/Resources/anonymization.pdf> [Cavoukian & El Emam], noting that personal health information is a particularly sensitive type of personal information. See also the description of "[s]ensitive personal data" in: ICO, *Anonymisation*, *supra* note 38 at 49-50. It is important to emphasize, however that the definition of personal information remains broad enough to encompass information that may be viewed as relatively less sensitive than health data or the other types of data noted above. For instance, in *Information v. Natural Resources*, *supra* note 30, the court held that the name, phone number, and professional title of individuals working in the private sector constituted personal information under the federal *Privacy Act*.

[100] El Emam, *supra* note 62 at 20.

[101] ICO, *Anonymisation*, *supra* note 38 at 20. See also Sherman, *supra* note 23 at 9, discussing these and other more specific risks relevant to the online distribution of information about court cases (such as risks to the fairness of a trial or to vulnerable people involved in criminal or family proceedings).

before the information is released. If, however, both parts of the second or third question are answered in the affirmative and the personal information has not been made publicly available,[102] the challenge will be to balance the goal of releasing the data in a form that serves the purpose for disclosure and the need to protect individual privacy. In this type of situation, if the data are required for a transitory purpose, the risk to personal privacy may support the decision not to release the information online where it will be made permanently available. In cases where it is determined that the online release of the information would be appropriate, anonymization techniques may provide an option for balancing the public's right to information with the need to protect individual privacy. In addition, the privacy risk may be addressed through the use of licence restrictions or technological barriers to reuse of information. We review these options in turn before summarizing the best practices in a decision-making tree below.

(d) Anonymization of Data

Data that have been anonymized have been rid of all personal identifiers that would allow for reidentification to occur; various tools have been developed to assist in the process.[103] The approach used may depend on what type of data are being anonymized.[104] For instance, statistical data can be aggregated,[105] randomized or coded,[106] or pseudonymized.[107] Personal information contained in qualitative data may be redacted.[108] Where the data contain quasi-identifiers,

---

[102] See further discussion on "publicly available personal information" above.

[103] Khaled El Emam, *Guide to the De-Identification of Personal Health Information* (Boca Raton: CRC Press, 2013); Latanya Sweeney, "k-Anonymity: A Model for Protecting Privacy" (2002) 10:5 Intl. J. Uncertainty, Fuzziness & Knowledge-Based Systems 557; and Ross Fraser & Don Willison, "Tools for De-Identification of Personal Health Information," Prepared for the Pan Canadian Health Information Privacy (HIP) Group (September 2009), online: < https://www.infoway-inforoute.ca/en/component/edocman/supporting-documents/500-tools-for-de-identification-of-personal-health-information >, discussing the following: aggregation of data, k-anonymity, data reduction, data modification, data suppression, and pseudonymization. See also El Emam & Fineberg, *supra* note 85 at 17-25, discussing the techniques of randomization, irreversible coding, reversible coding, heuristics, and analytics.

[104] It is worth noting that the use of anonymization techniques to facilitate the disclosure of information held by the government relates to an additional issue of transparency in that government actors should clearly communicate policies about how personal information collected for government programs may later be used. Where information may be anonymized for disclosure to the public, this should be communicated to individuals whose data are being collected along with information about how the risks inherent in such disclosures will be managed. See ICO, *Anonymisation*, *supra* note 38 at 40, recommending that this information be disclosed within privacy policies.

[105] See *ibid* at 36, noting that depending on the level of granularity, aggregation of data can be considered relatively low risk in terms of reidentification, but may make it difficult to carry out the individual analysis required by some forms of research.

[106] El Emam & Fineberg, *supra* note 85 at 10.

[107] Pseudonymization involves replacing "real world" identities with a pseudonym that

other techniques such as heuristics or analytics may be used.[109] Because no two documents or data sets are alike, both the approach used and the nature of the data itself will affect the degree of reidentification risk that remains once the data are released.[110] Options for anonymizing data (which are not mutually exclusive) are outlined in Table 2, below.

(i) Table 2: Anonymization Techniques[111]

Technique	Description	Notes
Aggregation	Displays statistical data as totals, averages or in ranges, with personal identifiers removed.	May be particularly useful for population statistics and demographic information; may not adequately protect against reidentification for data with small number of cell counts; may impede some forms of research that require specific information about individuals whose information is contained within the data.
Randomization	Involves the "scrambling" of direct and indirect identifiers in data to create a credible data set.	May provide a useful way to retain identifiers such as postal codes and telephone numbers where the purpose for disclosure requires reference to such variables.

users can refer to in order to understand the data without accessing the actual identity of the individual: ICO, *Anonymisation*, *supra* note 38 at 49, 51.

[108] *Ibid* at 22.

[109] El Emam & Fineberg, *supra* note 85 at 10.

[110] See Cavoukian & El Emam, *supra* note 99 at 13, noting that:

> The risk threshold should reflect the amount of re-identification risk that the health information custodian is willing to take. For example, one approach is to ensure that for each record contained in the data set that describes characteristics of a data subject, there are at least four other individuals also represented by records in the data set who share these same characteristics.

[111] Discussion of these various options is found in El Emam & Fineberg, *supra* note 85; El Emam, *supra* note 62; Cavoukian & El Emam, *supra* note 99; and Information and Privacy Commissioner of Ontario, "The Unintended Consequences of Privacy Paternalism," by Ann Cavoukian, Alexander Dix & Khaled El Emam (Toronto: IPC, 2014) at 17, online: < https://www.ipc.on.ca/images/Resources/pbd-privacy_paternalism.pdf > .

Technique	Description	Notes
Coding/ Pseudonymization	Replaces "real world" identities with pseudonyms that allow users to associate data with the pseudonym but not the real personal identity.	May be particularly useful for clinical health data as it allows for reidentification of individuals by the data custodian (which may be useful where research reveals information that should be communicated to the individual about whom it relates).
Heuristics	Uses threshold rules to determine the risk that (i) an individual may be uniquely identifiable through any combination of quasi-identifiers; (ii) outside sources of information may be combined with the data to identify one or more individuals.	May be particularly useful for geodemographic data; usually accounts for a limited number of variables; unlikely to sufficiently protect against complex data (e.g., epidemiologic analysis that includes many variables).
Redaction	Conceals personal identifiers (e.g., by blacking out text).	Useful for qualitative data sets but lacks consistent methodological approach.
Data Suppression	Eliminates details that would lead to a high risk of reidentification.	May significantly distort data (especially where suppression is not random).
Sampling	Involves the release of a sample of records included in a broader data set.	Normally used in combination with other de-identification techniques.

Using the above options alone or in combination, data custodians can attempt to remove personal information from government-held data and make more information available to the public. Yet, while many advocate for the continued use of anonymization techniques in order to enable public release of information that would otherwise be withheld,[112] recent experiences with data

[112] Support for the use of anonymization techniques is often accompanied by an acknowledgment that anonymization can mitigate but not erase the privacy risks. See discussion in Information and Privacy Commissioner of Ontario, "Big Data and Innovation, Setting the Record Straight: De-identification *Does* Work," by Ann Cavoukian & Daniel Castro (Toronto: IPC, 2014), online: <www.ipc.on.ca/images/Resources/pbd-de-identification_ITIF.pdf>; Cavoukian & El Emam, *supra* note 99; El

anonymization have called the value of these techniques into question. There have been cases in which one or more individuals have been reidentified from information released on the assumption that it had been cleared of personal details.[113] A much-cited example followed the disclosure of twenty million search queries by media corporation America Online (AOL); in this case the disclosure of the information was meant to support independent research on Internet behaviours.[114] The belief that the data had been anonymized was invalidated when two reporters used clues in the information along with outside sources to identify a woman and link her to her individual search queries.[115]

The type of reidentification that occurred in the AOL case can be expected to occur more frequently as the amount of information available to the public grows and as the technologies that facilitate mass and indefinite storage[116] and

---

Emam, *supra* note 62 at 22. See also comments in Khaled El Emam et al, "A Systematic Review of Re-Identification Attacks on Health Data" (2011) 6:12 PLoS One e28071 at 1, noting the particular gains that have been derived from the secondary use of de-identified data in the areas of population health research, health services research, and public health. The authors also note at 2 that an exception may apply for genomic data due to evidence that anonymization techniques may not sufficiently protect against reidentification for this type of information.

[113] The Article 29 Data Protection Working Party notes that there are a number of reasons for which an individual might be motivated to attempt to reidentify persons whose information is released in the open government framework or elsewhere, including to reveal identities for commercial or law enforcement purposes, to reveal personal information that may be newsworthy or relevant in an adversarial political setting, or simply to satisfy individual curiosity: Article 29 Data Protection Working Party, *supra* note 24 at 16. See also ICO, *Anonymisation*, *supra* note 38 at 23.

[114] See a description of these events in Ohm, *supra* note 54 at 1717.

[115] *Ibid* at 1718. In this work, Ohm also describes two other high profile examples of reidentifications of individuals through the use of data that was believed to have been anonymized. In the first, William Weld, Governor of Massachusetts, was identified from data relating to state employee hospital visits, which was released by the Group Insurance Commission, a government organization. In the second, Netflix released records of user ratings and details on movies watched by individual customers. Researchers then illustrated the ease with which individuals could be linked with their ratings and to the movies they had watched with just a few details about the individual. See also discussion in Daries et al, *supra* note 54 at 3. See additional examples of reidentifications in El Emam & Fineberg, *supra* note 85 at 13.

[116] See Kitchin, *supra* note 8 at 70, discussing the ongoing replication of data on the internet, and at 31 explaining the move to digital storage solutions for long-term archiving of data. See also Rob Kitchin & Tracey P. Lauriault, "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work" (2014) The Programmable City Working Paper 2, online: <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112>, discussing modern storage solutions and machine data mining and analysis of data. Former Information and Privacy Commissioner of Ontario Ann Cavoukian has noted that "[w]e can no longer speak meaningfully of information destruction, as we once did with paper records, because digital bits and bytes have now attained near immortality on the Internet, thwarting efforts to successfully remove them from the public domain": Cavoukian, "Privacy and Government 2.0," *supra* note 84 at 4.

manipulation of these data become increasingly sophisticated.[117] Due to these concerns, academic study into the strength of anonymization techniques continues.[118] At present, data custodians can test the strength of anonymization techniques by attempting to reidentify individuals prior to disclosure using outside sources of information. This exercise can be approached from the perspective of a "motivated intruder" who has no prior knowledge of the personal details in the information but wishes to reidentify one or more individuals from the data set.[119] It is also good practice to review the information to determine whether someone with prior knowledge of a certain individual whose personal information is reflected in the data set would be able to reidentify that person and learn new personal information about him or her in the process.[120]

In evaluating the strength of the anonymization technique and the extent to which the anonymization is likely to protect against reidentification risk, data custodians should return to the purpose of disclosure to evaluate whether the risk to personal privacy is relevant to the objective. For instance, if the purpose of disclosure is to increase transparency with respect to government operations, the likelihood that releasing the information will genuinely support that objective should play a role in the decision-making process. Where the transparency value to be gained from disclosure is high, the situation may lean towards disclosure in anonymized form despite some degree of risk of reidentification. On the other hand, where the privacy risk remains high but the transparency value is less clear, it may be appropriate to withhold the information or take further steps to address the risk (e.g., through applying additional anonymization techniques

---

Finally, see comments in Article 29 Data Protection Working Party, *supra* note 24 at 22; and ICO, *Anonymisation*, *supra* note 38 at 29.

[117] Office of the Information and Privacy Commissioner for British Columbia, "Evaluating the Government of British Columbia's Open Government Initiative," by Elizabeth Denham, Investigation Report F13-03 (OIPCBC, 25 July 2013) at 33, online: < https://www.oipc.bc.ca/investigation-reports/1553 > [Denham, "BC Open Government"]; Article 29 Data Protection Working Party, *supra* note 24 at 13.

[118] See e.g., Ohm, *supra* note 54; Sweeney, *supra* note 103; Khaled El Emam, Ann Brown & Philip AbdelMalik, "Evaluating Predictors of Geographic Area Population Size Cut-offs to Manage Re-identification Risk" (2009) 16:2 J. American Medical Informatics Association 256.

[119] ICO, *Anonymisation*, *supra* note 38 at 22. A four-step approach to evaluating reidentification risk has been offered in Cavoukian & El Emam, *supra* note 99 at 13, which is to consider "the re-identification probability; the mitigating controls that are in place; the motives and capacity of the data recipient to re-identify the data; and the extent to which an inappropriate disclosure would be an invasion of privacy."

[120] For instance, it might be asked whether parents might be able to use health information to learn details their teenage child had not previously disclosed to his or her parents. Another example would be a family member who knows about a crime committed by one of his or her relatives and may be able to learn additional details of which he or she was not previously aware: see these and other examples in ICO, *Anonymisation*, *supra* note 38 at 25.

outlined in Table 2, above). An additional consideration is the fact that removal of all personal identifiers is likely to sacrifice some quality in the data, making it unfit for certain secondary purposes.[121] As the release of quality data is one of the underlying goals of open data,[122] this presents an important consideration and emphasizes the need for good metadata[123] as well as data literacy initiatives.[124]

---

[121] This is an important consideration in the open government context, where the issues of fitness for use and the need for metadata are particularly relevant. Depending on the level of deidentification, the process may skew the data and make them less useful for certain purposes. See Daries et al, *supra* note 54 at 57, noting that "[i]t is impossible to anonymize identifiable data without the possibility of affecting some future analysis in some way," and further discussion of the issue at 61-63. See further discussion and examples of data fitness issues in R. Devillers et al, "Towards Spatial Data Quality Information Analysis Tools for Experts Assessing the Fitness for Use of Spatial Data" (2007) 21:3 Intl. J. Geographical Information Science 261; Sylvie Servigne, Nicolas Lesage & Thérèse Libourel, "Quality Components, Standards, and Metadata" in Rodolphe Devillers & Robert Jeansoulin, eds., *Fundamental of Spatial Data Quality* (London: ISTE, 2006) 179; R. Devillers et al, "How to Improve Geospatial Data Usability: From Metadata to Quality-Aware GIS Community" (Workshop delivered at the A AGILE Conference, Aalborg, Denmark, 8 May 2007) [unpublished], online: < https://www.researchgate.-net/publication/253972014_How_to_Improve_Geospatial_Data_Usability_From_-Metadata_to_Quality-Aware_GIS_Community > .

[122] Emphasis on quality data is reflected in open data principles: Canada, "G8 Action Plan," *supra* note 7. As noted above, open data specifically supports the open government goals of innovation and economic development. See again Scassa, "Privacy and Open Government," *supra* note 6 at 399.

[123] The primary way to communicate the known limitations of data to potential users is to include metadata with each new dataset disclosed in the open government context. Metadata is defined simply as "data about the data" and can include information about the authors and editors of a document; the date that a document was created, edited, accessed, emailed, or printed; or the name, type, and size of a file. In addition to helping to clarify limitations of data for secondary use, metadata may serve to support the goals of increased government transparency and accountability in the open government movement. For instance, it could be relevant to know when a particular data set was created, as this could indicate a latest possible time at which the government knew of the information. Metadata might also reveal whether a document was backdated, which might be highly relevant to enforcing transparency and accountability in a given situation: see Peter S. Kozinets, "Access to Metadata in Public Records: Ensuring Open Government in the Information Age" (2011) 28:1 Computer & Internet Lawyer 25 at 25; Zuiderwijk & Janssen, *supra* note 7 at 19, 22; Conradie & Choenni, "Municipal Barriers," *supra* note 23 at S11-S14; Conradie & Choenni, "Process Barriers," *supra* note 23; Conroy & Scassa, *supra* note 5. Canada's "Open Government Portal" currently contains data sets with some accompanying metadata: Canada, "Open Government," *supra* note 19.

[124] Cavoukian and El Emam argue for "the use of proper de-identification techniques and re-identification risk measurement techniques. . .[which enable] a high degree of privacy, while at the same time preserving the required level of data quality necessary for the secondary purpose": Cavoukian & El Emam, *supra* note 99 at 12. While it is useful to consider the ways in which data quality might be maintained while ensuring adequate protection for privacy, this requires a certain level of data literacy. In this way, the

(e) Licence Restrictions

Because of the uncertainties relating to the use of anonymization techniques to protect individual privacy, and because some anonymized government-held data released as open data may have been derived from highly personal information, additional precautions may be required before the data can be disclosed. When analyzing the information reveals a particular privacy risk for a given data set, decision-makers may consider several options. First, data custodians may decide not to release the information as open data.[125] If, however, the transparency value or other benefit from disclosure outweighs the privacy risk and it is determined that the risk does not reach the threshold set by the governing legislation (e.g., the *serious possibility* of identification threshold set by the Federal Court of Canada in relation to Ontario legislation in *Gordon v. Canada*),[126] an option would be to release the information with warnings to potential users about their responsibilities in using the information. These could be imposed in addition to existing open government licences[127] to remind users

---

situation highlights the importance of active promotion of data literacy to accompany the move towards open government. See general comments in Denham, "BC Open Government," *supra* note 117 at 36; and Janssen, Charalabidis & Zuiderwijk, *supra* note 7 at 264. See also Conroy & Scassa, *supra* note 5.

[125] In such circumstances, it might still be considered appropriate to release the data if a request is made in the access to information context where specific limitations might be attached to use (e.g., a purpose limitation or restriction on the disclosure for the recipient of the information). Data recipients who intend to use information for a specific purpose may be open to mitigating controls that are not feasible when the information is released to the public at large (such as limitations on the ability to disclose the data). In some cases this may allow secondary users to access better quality data that can be used for limited purposes. See discussion in ICO, *Anonymisation*, *supra* note 38 at 37; and Cavoukian & El Emam, *supra* note 99 at 14.

[126] *Gordon*, *supra* note 33.

[127] Many governments have adopted some version of an open licence for use with open data. The development of the federal government's open government license involved consultations and redrafting in which matters such as the need for clear definitions of key terms and for clear and accessible language were addressed. See Canada, "Open Government Licence — Canada," (2015), online: < open.canada.ca/en/open-government-licence-canada > [Canada, "Open Government Licence"]; Canada, "Open Government Licence Consultation Report," (2015), online: < open.canada.ca/en/open-government-licence-consultation-report > ; Teresa Scassa, "Canada's Open Government Licence V2.0 is Released" (18 June 2013), *Teresa Scassa* (blog), online: < www.teresascassa.ca/index.php?option = com_k2&view = item&id = 131:canada's-open-government-licence-v20-is-released&Itemid = 81 > ; and Teresa Scassa, "Canada's New Draft Open Government License" (6 December 2012), *Teresa Scassa* (blog), online: < www.teresascassa.ca/index.php?option = com_k2&view = item&id = 113:canadas-new-draft-open-government-licence&Itemid = 83 > . Similar action has been taken at the provincial level by the Government of Alberta: Alberta, "Open Government Licence — Alberta," online: < data.alberta.ca/licence > [Alberta, "Open Government Licence"]. It is important to note that the existence of an open licence does not mean that there are no intellectual property rights in the information being disclosed; the licence

that their right to use data disclosed to the public by any level of government does not entitle them to use personal information as defined under the relevant federal and provincial laws.[128] For information that raises significant privacy concerns, additional licence provisions could:

i. reiterate that data has been deliberately anonymized;[129]
ii. remind users that they are prohibited from using the data to reidentify individuals;[130] and
iii. require that users alert the licensor to any reidentifications that occur.[131]

One way of communicating these conditions would be to require that users agree to the limitations before the information can be accessed or downloaded.[132] An example of a licence provision for information that has been anonymized is:

---

may instead explicitly authorize use of a protected work under specified conditions. See discussion in Scassa, "Public Transit Data," *supra* note 60 at 1780.

[128] The federal government licence refers to the definition of personal information in the *Privacy Act, supra* note 22. The Government of Alberta's licence refers users to the *FIPPA-AB*, *supra* note 22: Canada, "Open Government Licence," *supra* note 127; and Alberta, "Open Government Licence," *supra* note 127. Note that under both licences, users are also required to specifically acknowledge the source of the information as part of secondary uses. This requirement allows subsequent users to understand the "chain of custody" and makes it possible to trace information back to its source to understand the limitations and bias within the data. The issue is therefore attached to a broader discussion on the need for good metadata in open government, discussed in the notes above. Promoting emphasis on the chain of custody of government data, see Cavoukian "Privacy and Government 2.0," *supra* note 84 at 5. Note, however, that some open data advocates decry the addition of these types of conditions in open data licences because they may make reuse more challenging. See Mewhort, *supra* note 12.

[129] Note, however, that it is not recommended that the steps taken to anonymize the data be released as this information may make it easier to reidentify individuals from within the data set: El Emam, *supra* note 62 at 23.

[130] It is important to emphasize that data sharing agreements and other licence restrictions are typically only used to address the limits on use of information shared with a specific third party: *ibid* at 19. In order for licence agreements to help address the privacy concerns in the open government context, significant development in terms of resources devoted to investigation and enforcement of these rules would be required: Daries et al, *supra* note 54 at 10; Article 29 Data Protection Working Party, *supra* note 24 at 18, 25. The difficulties in enforcing licensing restrictions are also noted in ICO, *Anonymisation*, *supra* note 38 at 38.

[131] See ICO, *Anonymisation*, *supra* note 38 at 41, recommending that governance procedures include a plan of action for when reidentifications occur through secondary use of purportedly anonymized data released by the public sector. Possible actions include notifying the individuals who have been reidentified and assisting in remedial action as well as improving the anonymization process that was used to avoid future privacy breaches.

[132] This is sometimes done in the clinical trials context where users attempt to access data for secondary purposes. Another option used in that context is to prohibit analyses of the data on personal computers (i.e., users must use standard software programs that have been deemed secure and that must be used on the website run by the data custodian). See

> Personal information has been removed from this data set. The licence governing use of this information does not authorize you to reidentify individuals from within this data. If secondary use of this information reveals the identity of one or more individual(s), users must contact [department name or contact person].

For publicly available personal information, a sample provision might be:

> The licence governing use of these data does not authorize you to combine the information with outside sources of data for the purposes of reidentifying individuals from within any other anonymized data. The licence remains subject to the rules on collection, use, and disclosure of publicly available personal information in privacy and personal data protection legislation in Canada.

An important issue to consider is that licence limitations may restrict the legal interoperability of open government data licences, making it more difficult to combine the information with other data for potentially useful purposes. The concept of interoperability of data sets includes technical as well as legal dimensions, and pertains to how easily and effectively different data sets can be combined or "mashed-up." Legal interoperability means that the licences for different data sets contain essentially the same terms and conditions, making it easy to combine and reuse them without being unduly limited by the more restrictive licence provisions. For instance, by combining open data sets released by the Government of Canada, contestants in a hackathon created numerous applications to assist users with respect to their international travel needs, decisions about where to live or work in Canada, decisions about commuting to work, questions relating to vehicle fuel efficiency, and more.[133] A lack of legal interoperability may mean that certain data sets cannot be combined in these potentially useful ways, which runs counter to the growing effort to make information—especially that which is released online—more useful to secondary users. Furthermore, a lack of licence interoperability may specifically impede some of the goals that underlie open government (particularly the objective of supporting commercial development and innovation when data are released online as open data).[134] This consideration weighs in favour of using license

---

discussion in Committee on Strategies for Responsible Sharing of Clinical Trial Data, *supra* note 21 at 117.

[133] Canada, "Canada Codes! CODE 2014 — Winners Showcase," (2015), online: < open.canada.ca/en/winners-showcase > .

[134] See Mewhort, *supra* note 12, discussing ongoing efforts to ensure that open data licences allow for data interoperability and discussing the use of licences that have been created for data and other works released in the Creative Commons as a model for governments to consider. The report acknowledges the need to protect personal information that may inadvertently be disclosed and suggests that a specific provision might be included to restrict secondary users from using personal information contained within the data (as well as information attached to "third-party rights the releasor is not authorized to waive or license; and data subject to other intellectual property rights, including patents, trade-

restrictions only where there is a significant concern over the risk of reidentification or an apparent conflict with another area of privacy or data protection law rather than as a matter of routine.

(f) Technological Barriers to Reuse

The inclusion of specific licence restrictions may mitigate some of the privacy concerns relating to the disclosure of government-held information and the option may be applied alone or, if required, in combination with technological barriers to reuse. The option to include technological barriers may be particularly relevant where the information being considered for disclosure constitutes publicly available personal information that holds both a high transparency value and a significant privacy risk.[135] As noted above, a specific privacy concern related to the digital release of publicly available personal information is that this makes the data available for cross-referencing with other non-identified data, which may increase the risk of reidentification of purportedly anonymized data.

One option (provided the situation is not one in which the government is required by law to release the information in an open format) is to release information in restricted proprietary formats such as PDF (thus using proactive disclosure instead of open data). This has been done, for instance, with the release of public sector salary information in Ontario.[136] The idea behind the disclosure of information relating to salaries in the public sector is that by making the details available to taxpayers, the government will be made more accountable for the way it spends taxpayer money.[137] Because the data contain personal details of the individual public sector employees however, there are

---

marks and official marks, and design rights: *ibid* at 18). The European Commission has endorsed the use of Creative Commons Licenses to support the reuse of information released by governments: see EC, Press Release, "Commission Encourages Re-use of Public Sector Data" (17 July 2014), online: <europa.eu/rapid/press-release_IP-14-840_en.htm> and discussion at Timothy Vollmer, "European Commission Endorses CC Licenses as Best Practice for Public Sector Content and Data" (17 July 2014), *Creative Commons* (blog), online: <creativecommons.org/weblog/entry/43316>. See also discussion on open data licences and the use of the Creative Commons model in Scassa, "Public Transit Data," *supra* note 60 at 1803; and Judge, "Enabling Access and Reuse," *supra* note 7, advocating that the government use Crown Commons licensing for public sector information until copyright reforms address the restrictions imposed by Crown copyright.

[135] Austin & Pelletier, *supra* note 71 at 11, discussing the question in relation to court documents, a matter that is examined in more detail below.

[136] The disclosure of this information is required for certain salary ranges paid by the public sector under the *Public Sector Salary Disclosure Act, 1996*, S.O. 1996, c. 1 Schedule A.

[137] See comments and the disclosure of information for 2013 with links to the information for previous years at Ontario Ministry of Finance, "Public Sector Salary Disclosure 2014 (Disclosure for 2013)," (Toronto: 2016), online: <www.fin.gov.on.ca/en/publications/salarydisclosure/pssd>.

privacy concerns attached to the release of the information online. The potential for the information to be reused for purposes beyond holding the government accountable for the salaries of its employees is partially addressed by the fact that the information has so far only been released in HTML (Hyper Text Markup Language) and PDF versions, which limits the ability to cross-reference the information.[138] It is important to note, however, that the media and other users of government data continue to seek out new ways to get around the technological barriers that have been imposed in order to facilitate cross-referencing and other secondary uses of the information.[139] In fact, online applications make it possible for the average user to defeat these technological barriers and to manipulate data into machine-readable formats.[140] Thus, a high degree of interest in the information may mean that technological barriers will provide only limited protection for the privacy interests at issue.

A second possibility is to limit the searchability of the information online.[141] The Canadian Legal Information Institute (CanLII)—which operates an open online database of Canadian laws, regulations, court and tribunal decisions—uses this technique.[142] Over a million Canadian court decisions have been made available on the website on the condition that CanLII not allow the decisions themselves to be searchable through Google or other search engines.[143] While the restrictions on the searchability of court decisions have

---

[138] *Ibid.*

[139] See Stuart A. Thompson, "Unlocking Ontario Public Sector Salary Data" (2014) 16:2 Media 27, providing a tutorial on how to manipulate the data in the format in which it has so far been released.

[140] See e.g., the following application that allows users to transform a pdf document into excel format: Nitro, "PDF to Excel Converter," online: < https://www.pdftoexcelonline.com/en > .

[141] IPC, "Transparency, Privacy and the Internet," *supra* note 5 at 10.

[142] Canadian Legal Information Institute, online: < www.canlii.org > . This initiative has had clear advantages in terms of making the law more readily accessible to Canadians. See discussion in Scassa, "Balancing Privacy," *supra* note 77.

[143] The rule that the information held on CanLII would not be searchable by Google or other search engines appears to be based on an agreement between the courts and CanLII rather than a specific order prohibiting CanLII from making the information available in this way. See Austin & Pelletier, *supra* note 71 at para. 75, discussing the limits on the searchability of Canadian court records online and noting the early existence of a "general consensus that unrestricted bulk searches should not be permitted to the public generally." The rule is reflected in CanLII's privacy policy, which states that:

> CanLII adheres to the principle of openness and transparency of legislative and judicial processes, and recognizes their fundamental importance in democratic societies. In order to minimize the negative impact of such transparency on the privacy of those involved in cases leading to judicial decisions, CanLII does not permit its case law collections to be indexed by external search engines.

The policy further notes that for the sake of openness and transparency, court decisions and other documents are published "in the form in which they are received from the institutions that issue them, such as official publishers, courts and law

prompted a debate on the costs for transparency and openness when it comes to information about the law,[144] the limitation was meant to provide a degree of privacy to the individuals involved in proceedings—especially those in the often sensitive areas of family, criminal, and immigration law.[145] Technological restrictions also prevent the court databases from being downloaded in bulk. However, the limitations of such technological restrictions were highlighted when a Romanian-based website scraped decisions from CanLII and published them

societies." Finally, the policy reinforces that external search engines are prohibited from indexing the text and style of cause of court decisions published by CanLII: Canadian Legal Information Institute, "Privacy Policy," online: < https://releve.canlii.org/en/info/privacy.html > .

See also IPC, "Transparency, Privacy and the Internet," *supra* note 5 at 10-11, noting that generally speaking, the main weakness of relying on the designation of site content as "'[o]ff [l]imits'" to search engines is that it is a voluntary standard (though one that major search engines like Google, Bing, and Yahoo typically adhere to).

144 The Alberta courts previously operated a searchable database of provincial court judgments; the site now redirects users to CanLII: Alberta Justice and Solicitor General, "Court Decisions," online: < https://justice.alberta.ca/programs_services/courts/Pages/decisions.aspx?WT.svl = programs > . On the debate brought on by this change, see Shaunna Mireau, "Trusting the System" (20 January 2015), *Slaw* (blog), online: < www.slaw.ca/2015/01/20/trusting-the-system > ; and Addison Cameron-Huff, "CanLII's Licensing Terms: How Much Access Should Canadians Have to Court Decisions?" (30 January 2015), *Addison Cameron-Huff, Tech Lawyer* (blog), online: < https://www.cameronhuff.com/blog/canlii-licensing-terms/index.html > . See also Addison Cameron-Huff, "Why Google Can't Build A Case Law Search Engine in Ontario" (11 February 2014), *Addison Cameron-Huff, Tech Lawyer* (blog), online: < https://www.cameronhuff.com/blog/ontario-case-law-private/index.html > , critiquing the scheme in Ontario under which only CanLII, Quicklaw and WestLaw are provided bulk access needed to create searchable databases of court decisions.

145 For instance, in the Globe24h case discussed immediately below, individuals complained to the Office of the Privacy Commissioner of Canada after discovering that personal information about them appeared online when their names were searched in a search engine. This information included the fact that one woman had worked in the sex trade (which was revealed in a case in which she acted as a witness), details of a hostile child custody dispute, and financial and medical information: Office of the Privacy Commissioner of Canada, *PIPEDA Report of Findings #2015-002: Website That Generates Revenue by Republishing Canadian Court Decision and Allowing Them to be Indexed by Search Engines Contravened PIPEDA*, (OPC, 5 June 2015), online: < https://www.priv.gc.ca/cf-dc/2015/2015_002_0605_e.asp > [OPC, "Globe24h"]. See also comments in Christine Dobby, "Canadians Upset with Romanian Website that Exposes Court Case Details," *The Globe and Mail* (5 January 2015), online: < www.theglobeandmail.com/report-on-business/industry-news/the-law-page/canadians-upset-over-romanian-website-that-exposes-court-case-details/article22284367 > . See an explanation of CanLII's approach to shielding personal information from broad access through Google and other search engines at: Colin Lachance, "Google, González and Globe24h" (26 May 2014), *Slaw* (blog), online: < www.slaw.ca/2014/05/26/google-gonzalez-and-globe24h > . See also Scassa, "Balancing Privacy," *supra* note 77, discussing the personal nature of divorce cases; cases heard by workers compensation, human rights, and pension and disability tribunals; and personal injury trials, all of which are relevant to the content released on CanLII.

on a website that could be indexed by search engines, which made the information much easier to access on the Internet.[146] Thus, while limiting the searchability of online information is a possible strategy, it is not a perfect solution.[147]

It should be noted that the privacy problems resulting from the circumvention of the technological restrictions could also be attributed to the failure of some courts to follow recommendations that personal identifiers not be included in some types of decisions, or that the inclusion of personal information in decisions be limited only to what is necessary to make the decision comprehensible and transparent. This highlights the point that technological barriers on released data alone are unlikely to suffice. In addition, technological barriers may limit innovative uses of the data. This example therefore highlights the need for new approaches to information management to address the challenges of the new digital era in which those releasing information must deal with risks that did not present in the previous paper-based world.[148] In particular, the value of data minimization, discussed above, may provide a "first line of defence" in protecting individual privacy.[149]

Even where the privacy risk is high, the use of technological barriers for government-held information that is released online may be viewed as controversial given the objectives of both proactive disclosure and open data.[150] This controversy was highlighted in 2014 when Treasury Board President Tony Clement stated that some information requested under the access to information framework could not be released in its original format

---

[146] The Office of the Privacy Commissioner of Canada became involved following complaints about Globe24h's activities. In response to those complaints, Globe24h argued that it was entitled to use the information without obtaining consent from the individuals to whom it related because it was publicly available information (see *supra* note 82, explaining that under the terms of *PIPEDA*, organizations can use publicly available information without consent provided the use is consistent with that for which the information was first collected). Globe24h argued that its purpose was to make information freely available online. The Privacy Commissioner disagreed and determined that the organization's true purpose was "to use personal information contained in court and tribunal decisions for the purpose of generating revenue through its paid removal service." Globe24h had initially charged a fee to have the personal information removed from the website; the policy was later changed to requiring a written request. It appears, however, that the company continued to charge a fee for removal of documents by negotiating payment on a case-by-case basis. See OPC, "Globe24h," *supra* note 146 at paras. 10, 68, 70. See also discussion in Dobby, *supra* note 146.

[147] See discussion in IPC, "Transparency, Privacy and the Internet," *supra* note 5 at 1.

[148] Sherman, *supra* note 23 at 10.

[149] See discussion in Scassa, "Balancing Privacy," *supra* note 77.

[150] See comments in Jordan Press, "Information Czar Investigating Whether Government Refusing to Release Data in Easy-to-Read, Digital Formats" *National Post* (24 December 2014), online: < news.nationalpost.com/2014/12/24/information-czar-investigating-whether-government-refusing-to-release-data-in-easy-to-read-digital-formats > .

because of the risk that it would be manipulated and perhaps used to spread false information.[151] The discussion that arose focused mainly on the use of technological barriers to prevent falsification of data instead of to protect against reidentification.[152] Based on this focus, some took issue with the government taking steps to obstruct use of the data, arguing that the information should be released in reusable format and false information simply rebutted.[153] While this approach may balance the public's right to access information with the risk that information will be used to spread false or misleading claims, it does not address the personal privacy concerns. As noted above, however, technological barriers might do no more than merely impose a delay to misuse where there is a high motivation to manipulate the data in a way that might reidentify individuals.[154] At the same time, they may inhibit innovative and valuable uses of the data.

---

[151] Dean Beeby, "Tony Clement Concern about Electronic Information Access Queried: Bureaucrats Query Treasury Board President's Claim that Datasets Must Be Protected from Manipulation" *CBC News* (8 March 2015), online: < www.cbc.ca/news/politics/tony-clement-concern-about-electronic-information-access-queried-1.2983561 > . Though this section considers the issue that arose following the Treasury Board President's comments in terms of the privacy risks of cross-referencing different data sets, it is important to emphasize that his comments concerning the cross-referencing of information for secondary purposes in a way that produces false information presents an important question for researchers to consider. The Committee on Strategies for Responsible Sharing of Clinical Trial Data emphasizes that in order for data to be usable, it must be accompanied by specifications about the meaning of the data. Where data are being pulled from different sources (e.g., different databases), the meaning of the data must be compatible in order to avoid producing false information when combined: Committee on Strategies for Responsible Sharing of Clinical Trial Data, *supra* note 21 at 132.

[152] Beeby, *supra* note 152; Press, *supra* note 151.

[153] Beeby, *supra* note 152.

[154] This was emphasized following the comments made by Tony Clement through media comments about the fact that recipients of the data were unhappy because the move "ha[d] meant extra time for researchers or journalists making the request to transfer data into easy-to-handle forms to identify trends or issues": Press, *supra* note 151. Similarly, in the context of personal health information obtained in clinical trials, the Committee on Strategies for Responsible Sharing of Clinical Trial Data notes that:

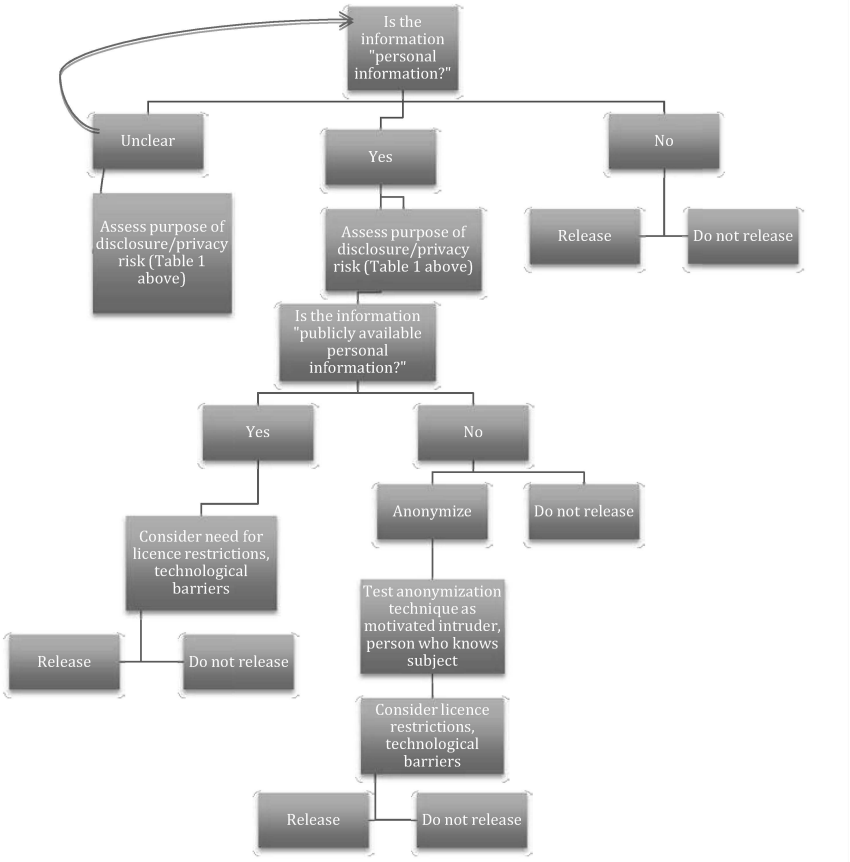
> [P]roponents of open access believe that individuals and organizations with bad intentions could easily find ways to overcome the controls instituted by sponsors, and the controls would therefore serve only to slow the rate of scientific discovery and advancement without mitigating risks.

Committee on Strategies for Responsible Sharing of Clinical Trial Data, *supra* note 21 at 141.

IV. A DECISION-MAKING MODEL FOR RELEASING OPEN GOVERNMENT DATA

The main steps involved in assessing information available for proactive release of government information are compiled below in a decision-making tree. The guide highlights the relevant questions to be asked when considering whether or not to release a given data set or other types of information as well as the options that may be considered where a particular privacy risk is identified.

(a) Decision-Making Tree for the Release of Government-Held Information



As indicated in the guide and in the discussion above, the first question to ask is whether the information being considered for proactive release is personal information. If it is clearly not, the choices are to release the information in full or to withhold the data. As previously noted, the goals of the open government

movement support disclosure in such circumstances unless a specific legislative exception applies. If the information is or may be personal information, the questions set out in Table 1 above should be asked to clarify the purpose of disclosure and the privacy risks related to the specific data set at hand. Following this, the options depend on whether the information is publicly available personal information. If the information is not publicly available personal information, it may only be released in anonymized form. If the information is anonymized for release, the strength of the anonymization technique should be tested from the perspective of a motivated intruder and an individual with some knowledge of those whose personal information is in the data set. If this process confirms that the information has been anonymized, the data may be released. Licence restrictions or technological barriers may be used if the previous assessment has highlighted a particular privacy risk. If the information is publicly available personal information, data custodians must consider the potential privacy issues relevant to the online disclosure of this type of personal information (including that the information becomes easier to access, disseminate, and cross-reference with purportedly anonymized data). These issues may be partially addressed through the use of licence restrictions and technological barriers to reuse of the information. It may also be deemed appropriate in certain cases to withhold publicly available personal information from online release where the transparency value of the information is low but the privacy risk high, or where there is only a temporary need for the public to have access to the information. Moreover, the information may not be available for release as open data due to conflict between the goals of open data and certain parts of privacy and personal data protection legislation. If publicly available personal information is released online, licence restrictions can be used to refine the continued application of this area of the law.

(b) Assessing the Proportionality of Restrictions Used

Based on the above, data custodians may decide to use anonymization techniques, licence restrictions, or technological barriers (alone or in combination), or to withhold information from online disclosure or even any form of public disclosure. As a final best practice, we offer a series of questions that can be used to evaluate the proportionality of the restrictions used in Table 3, below.

(i) Table 3: Evaluating Restrictions Used

Question[155]	**Notes**
1. Are the restrictions necessary to prevent a specific privacy risk related to the release of the information?	Refer to privacy risk identified as part of the inquiry outlined in Table 1, above.
2. Are the restrictions designed to minimally impair the benefits of disclosure?	Consider purpose of disclosure identified as part of the inquiry outlined in Table 1, above; consider whether the information is available through another means (e.g., publicly available personal information contained in a registry) that provides an alternate means to fulfill the purpose for disclosure.
3. Do the benefits of the restrictions outweigh any adverse effects attached to the modified release or withholding of the information?	

V. CONCLUSION

This paper has aimed to outline strategies for those responsible for the release of information in the open government context. These strategies are intended to assist in complying with the legislated requirement to balance transparency with privacy in the release of government information. As we have discussed, such decisions are complicated by several factors, including:

i. continuing questions about how to apply the definition of personal information;
ii. a rapidly changing information environment in which the amount of information is increasing while information technologies progress; and
iii. the fact that anonymization techniques are proving to be less reliable than previously believed.

In order to guide the decision-making process, we have suggested a series of questions to ask in evaluating the privacy risk and transparency value relating to the release of specific data sets and other types of information being contemplated for disclosure. Where this process reveals a particular privacy risk, decision-makers will need to consider options for minimizing the risks before disclosing information to the public. These options include licence restrictions as well as technological barriers to reuse of the information. The

[155] These questions have been modeled in part on the proportionality test outlined in Austin & Pelletier, *supra* note 71 at 13 (provided to assess restrictions on access to court documents).

steps to decision-making and some proposed solutions have been presented in a decision-making tree that can be used in the assessment of individual data sets. As a final step, data custodians should evaluate the proportionality of any restrictions on access to government-held information in order to ensure that the focus remains on balancing the goals of open government and open data with the risks to personal privacy that attach to the emerging open government environment.

VI. GLOSSARY

Aggregation: the displaying of statistical data as totals, averages or in ranges, with personal identifiers removed.[156]

Anonymization (or de-identification): a process through which data are freed of personal identifiers and that is unlikely to allow reidentification of any individual when combined with other information sources.[157]

Big data: the massive and increasing amount of data available in digital form.[158]

Data minimization: policy relating to the collection of information by government actors, which emphasizes non-identifiable interactions as a starting point in order to minimize the amount of personal information that is collected by the public sector.[159]

Motivated intruder test: an approach to evaluating the strength of anonymization techniques which assumes the position of a person who holds no prior knowledge of the personal details within the data and who wants to reidentify one or more individuals by combining the information with outside sources of data.[160]

Open access: a concept linked to the "right to information" movement and the availability of government-held information, especially information relating to government operations.[161]

Open data: the release of data in reusable electronic formats under an open licence for reuse.[162]

---

[156] ICO, *Anonymisation*, *supra* note 38 at 52; Fraser & Willison, *supra* note 103 at 2.

[157] ICO, *Anonymisation*, *supra* note 38 at 6; El Emam et al, *supra* note 112 at 1 defining de-identification as "the act of reducing the information content in data to decrease the probability of discovering an individual's identity."

[158] Scassa, "Privacy and Open Government." *supra* note 6 at 407.

[159] Cavoukian "Privacy and Government 2.0," *supra* note 84 at 10.

[160] ICO, *Anonymisation*, *supra* note 38 at 22.

[161] Scassa, "Privacy and Open Government," *supra* note 6 at 399; TBS, "Action Plan," *supra* note 15.

[162] Scassa, "Privacy and Open Government," *supra* note 6 at 399.

Open dialogue: a two-way conversation between Canada's different governments and their people that aims for better public engagement in the delivery of public policies and programs.[163]

Open engagement: activities involving both the government and its public, which may include citizen participation in government activities and decision-making through mechanisms such as social media or online reporting of community issues that relate to government responsibilities.[164]

Open government: movement aiming to increase transparency and accountability in the public sector and to promote the economic value of data through open access, open data, and open engagement.[165]

Personal information: information about an identifiable individual.[166]

Privacy impact assessment: a process that aims to determine how a government program or service will impact individual privacy and to highlight ways to minimize that impact.[167]

Proactive disclosure: release of information to the public without any specific requirement to disclose the information (e.g., in response to an access to information request).[168]

Pseudonymization: a de-identification technique that replaces "real world" identities with a pseudonym that allows users to associate the data with the pseudonym but not the real personal identity.[169]

Publicly available personal information: information about one or more individuals that is openly available to the public for consultation.[170]

Re-identification: the process of data-matching, cross-referencing or otherwise analyzing purportedly anonymized data to identify one or more individuals from within a data set.[171]

---

[163] TBS, "Action Plan," *supra* note 15.

[164] Scassa, "Privacy and Open Government," *supra* note 6 at 400.

[165] *Ibid* at 398.

[166] See above discussion on personal information in the access to information context in Canada.

[167] CRA, "PIAs," *supra* note 90; TBS, Privacy Impact Assessment Guidelines, *supra* note 90.

[168] Denham, "BC Ferries," *supra* note 9 at 2.

[169] ICO, *Anonymisation*, *supra* note 38 at 49, 51.

[170] Scassa, "Privacy and Publicly Available Personal Information," *supra* note 68 at 10.

[171] See a similar definition in ICO, *Anonymisation*, *supra* note 38 at 6, 49.

Secondary purpose: any use of data beyond that for which it was collected, including for analysis, research, safety and quality measurement and improvement, and marketing.[172]

Transparency: a concept given variable definitions in different contexts, including: the act of "making relevant, timely and useful information available to the public in easy to access formats,"[173] or
[A] measure of the degree to which the existence, content, or meaning of a law, regulation, action, process, or condition is ascertainable or understandable by a party with reason to be interested in that law, regulation, action, process, or condition.[174]

---

[172] El Emam & Fineberg, *supra* note 85 at 2; El Emam et al, *supra* note 112 at 1.

[173] Health Canada, "Posting Information," *supra* note 5.

[174] Mock, *supra* note 5 at 1082 [emphasis omitted].