# Personalization, Analytics, and Sponsored Services: The Challenges of Applying PIPEDA to Online Tracking and Profiling Activities

*Eloise Gratton*[*]

## INTRODUCTION

In 2008, the online advertising industry was found to be worth 27 billion dollars, a figure that was projected to double over the subsequent four years.[1] The reason for this extraordinary market growth can be explained by two factors. To begin with, current technology now makes it possible to gather a great variety of information associated with a particular device or individual, including browsing history, which can be used to create a profile specific to that device or individual. This practice facilitates more personalized advertising, tailored to the interests and tastes of the consumer. Secondly, many online services, in the form of information or entertainment, are offered for free to consumers as long as they accept the presence of advertising and the eventuality that their online behaviour will be tracked to a certain degree.[2]

Internet business models are increasingly being based on the notion of greater customization of services and products. This entails that there are huge amounts of data that need to be collected about online users. Moreover, online profiles present new types of concerns. For instance, although isolated pieces of profile information may not be sensitive, their context, especially in light of profiling or behavioural analysis practices, may become extremely sensitive. With the convergence between different technologies and the growing demand for applications that include location tracking capabilities, privacy concerns pertaining to tracking and profiling activities need to be properly addressed.

---

* Eloise Gratton is Counsel to McMillan for information technology issues and also privacy officer for the Montreal office while pursuing a PhD degree at University of Montreal and Université de Panthéon-Assas (Paris II). She was previously partner at McMillan (Corporate Law group) which she joined in 2002. Prior to joining the firm, she acted as Director of Corporate & Legal Affairs for a wireless technology company. Éloïse serves as Vice-Chair for the Canadian IT Law Association's ad hoc Privacy Committee. She also teaches e-commerce law (DRT6903A and CEL6001) at University of Montreal and HEC.

1 European Parliament (Committee on Civil Liberties, Justice and Home Affairs), PROGRAMME, *Public Seminar: Data protection on the Internet (Google-DoubleClick and other case studies)*(21 January 2008), Brussels, Room PHS 3C50, at 2.

2 Peter Fleischer (Global Privacy Counsel, Google), "Response to the Article 29 Working Party Opinion On Data Protection Issues Related to Search Engines" (8 September 2008) [Fleischer, "Response"] at 3: "to support this free service, Google primarily relies on being able to serve relevant advertising to its users.".

Many authors have already outlined that there is definitely an issue with the fact that online users may not always be aware that the online profiling and tracking activities are happening in the first place (even if the website privacy policy is open about its practices, many studies have shown that consumers don't read privacy policies). While this (lack of) consent issue is a serious one, this analysis will instead be focused on other issues: firstly, whether profile data is covered under data protection laws; secondly, whether tracking and profiling activities are legal in accordance with data protection laws such as the *Personal Information Protection and Electronic Documents Act* (PIPEDA)[3]; and finally, issues pertaining to the management of profile data, more specifically as they relate to granting access to profile data to individuals and what constitutes a reasonable retention period of the profile data.

## I. DATA PROTECTION LAWS GOVERNING PROFILES

Canadian data protection laws regulate the collection of *personal information*, a notion which is defined very broadly. PIPEDA defines it as: "information about an identifiable individual, but does not include the name, title or business address or telephone number of an employee of an organization" (hereinafter referred to as: the "Definition"). Substantially similar provincial laws have similar definitions.[4] In the context of tracking and profiling activities, the Definition raises certain uncertainties.

### (a) Profiles as "Personal Information"

The current Definition can be challenged when attempting to qualify profiles that have emerged on the Internet. Profile data collected through a cookie or an IP address[5] is linked to a device connected to the Internet (instead of a physical person). This device may be used by one or more individuals and only in certain cases and circumstances (sometimes with the assistance of ISP's log files or through data aggregation or correlation across services) can it correctly identify an individual. Furthermore, although the definition from PIPEDA has been interpreted broadly to include a computer's NETBIOS information or information collected from website

3 *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5.

4 The Quebec *Act Respecting the Protection of Personal Information in the Private Sector*, R.S.Q., c. P-39.1, s. 2 [the Quebec Privacy Law] defines personal information as "any information which relates to a natural person and allows that person to be identified.". Alberta's *Personal Information Protection Act*, S.A. 2003, c. P-6.5, s. 1(k) defines personal information as "information about an identifiable individual"; and British Columbia's *Personal Information Protection Act*, S.B.C. 2003, c. 63, s. 1 defines *personal information* as "information about an identifiable individual and includes employee personal information but does not include (a) contact information, or (b) work product information."

5 An IP address is a numerical identification assigned to a device each time that it connects to the Internet in order to enable that device to communicate with other connected devices.

cookies in one instance,[6] it is not always clear at what point an online profile may actually be associated with an *identifiable individual*.

For example, a 2006 case illustrates how the content of search queries can sometimes identify an individual.[7] On August 4, 2006, AOL Research had published a compressed text file on one of its websites containing twenty million search keywords which had been punched into AOL's search engine for over 650,000 anonymous AOL users over a three-month period. This text file was intended to be used for research purposes. According to reports in the press, it was possible to identify individual users on the basis of the content of their combined search queries.[8]

A similar privacy concern can arise in the mobile space. Some time ago, several US companies such as Intelligent Transportation Society of America[9] had requested the FCC to allow them to anonymously track the location of mobile users over time without having to disclose this tracking to the users.[10] The companies claimed that there may be great value in knowing, for instance, that a particular individual, associated with a particular profile (let's call it profile ABC) lives in a certain area, works in another one and uses a certain road at a specific time of the

---

6 See Office of the Privacy Commissioner of Canada, *PIPEDA case summary #2001-25: A Broadcaster accused of collecting personal information via Website* (20 November 2001); *PIPEDA case summary #2003-162: Customer complaints about airline's use of cookies on its Website* (16 April 2003); and *PIPEDA case summary # 2005-297: Unsolicited e-mail for marketing purposes* (31 March 2005), online: Office of the Privacy Commissioner <http://www.priv.gc.ca/>.

7 "Resolution on Privacy Protection and Search Engines" (28th International Data Protection and Privacy Commissioners' Conference, London, United Kingdom2 and 3 November 2006), [Resolution] online: <http://www.privacyconference2009.org/dpas_space/space_reserved/documentos_adoptados/common/2006_London/LONDON-EN4.pdf>.

8 While none of the records on the file were personally identifiable *per se*, certain keywords contain personally identifiable information by means of the user typing in their own name (ego-searching), as well as their address, social security number or by other means. The *New York Times* was able to locate individuals from the released and anonymized search records by cross referencing them with phonebooks or other public records. This list, although quickly withdrawn by AOL recognising that it was an error, had by the time of the withdrawal reportedly been downloaded and re-posted many times, and made available in searchable form on a number of websites. See Nate Anderson, "AOL releases search data on 500,000 users (updated)" *ars technica* (7 August 2006), online: ars technica <http://arstechnica.com/news.ars/post/20060807-7433.html>.

9 Public/private partnership serving as a utilized Federal Advisory Committee to the U.S. Department of Transportation, Educational and scientific research organization created in 1991 for the purpose of fostering the development and deployment of intelligent transportation systems.

10 Before the Federal Communications Commission, Washington D.C., In the Matter of the Petition of the Cellular Telecommunications and Internet Association Regarding Proposed Location Information Privacy Principles, WT Docket No. 01-72, INTELLIGENT TRANSPORTATION SOCIETY OF AMERICA, Reply Comments (April 24, 2001), at 7.

day. This seemingly mundane information would, according to the companies, have the potential to provide traffic engineers and planners with rich data feeds necessary to promote optimal traffic flows. Anonymous tracking would also allow them to efficiently allocate transportation resources and to properly reroute traffic in emergency situations. An issue emerges from the fact that it can be a challenge to determine at what point the mobile profile can establish a clear correlation with a specific individual, since it is not always evident to what degree the location data is in fact anonymized. In this example, the location data collected may be anonymized in the sense that the phone number relating to a specific profile may have been removed and instead replaced by a profile number (for example profile ABC). Still, if the mobile data collected is very accurate and collected over a long period of time, then one may at some point be able to determine that the individual "behind" profile ABC, who spends every night at a specific location (his residence?) and spends his days at another one (work place?) can be identified.

Identification tools on the Internet enable the correlation of different types of data made available on the web or through online services.[11] Internet technologies also allow for the grouping of widespread information of various types that pertain to a single person, which can then lead to identification.[12] Data correlation across services raises additional privacy concerns. For example, many search engine providers offer users the option of personalising their use of services through a personal account. With Web 2.0 and online social networks and the new trend towards increased cross-site profile linkage, perhaps certain types of data which could not previously be used to identify an Internet user may now be used to identify such a user as recently suggested by technical experts.[13]

---

11 Many services providers may also, using IP addresses and correlating it with other data that they have collected, identify an individual behind an IP address. For example, a search engine provider may be able to link an IP address to an individual by linking different requests and search sessions originating from a single IP address to track and correlate all the web searches originating from a single IP address if these searches are logged. See Resolution, *supra* note 7.

12 <www.123people.fr> is an example of a service provider that groups and aggregates all kinds of information (such as pictures, email addresses, links, etc.) pertaining to the name of an individual searched and displays the data available, which would otherwise be more difficult to obtain, in a logical and comprehensive manner. Article 29 Working Group has raised their concerns with regards to the retrieving and grouping capabilities of search engines. Article 29 Data Protection Working Party, *Opinion 1/2008 on data protection issues related to search engines* (Adopted on 4 April 2008), 00737/EN WP 148 [Article 29, *Opinion 1/2008*], at 5 and 14.

13 Dan Brickley, "YouAndYouAndYouTube: Viacom, Privacy and the Social Graph API" (3 July 2008), online: danbri's foaf stories <http://danbri.org/words/2008/07/03/359#comment-15692>:

> YouTube users who have linked their YouTube account URLs from other social Web sites (something sites like FriendFeed and MyBlogLog actively encourage), are no longer anonymous on YouTube. (. . .) It can give them a mechanism for sharing "favourited" videos with a wide circle of friends, without those friends needing logins on YouTube or other Google services. This clearly has business

Many service providers on the Internet mention in their privacy policy that they enrich data collected from users with data from third parties.[14] New algorithms are being developed that allow extraction of information from a veritable sea of collected data.[15] Also, more and more technologies and more of the new types of data will make it possible to collect data that are more intrusive and that are of a far more personal nature.[16] Some have raised the issue that the changes to the core architecture of the Internet and its protocols (to Internet Protocol version 6) will permit many more physical objects to have an Internet address, paving the way for a wide range of devices to be connected, and that combining these technologies with RFID could affect privacy in many ways.[17] All of these examples illustrate

---

> value for YouTube and similar "social video" services, as well as for users and Social Web aggregators. Given such a trend towards increased cross-site profile linkage, it is unfortunate to read that YouTube identifiers are being presented as essentially anonymous IDs: this is clearly not the case. If you know my YouTube ID "modanbri" you can quite easily find out a lot more about me, and certainly enough to find out with strong probability my real world identity. (. . .) To understand YouTube IDs as being anonymous accounts is to radically misunderstand the nature of the modern Web.

14 See Microsoft privacy policy, online:<http://privacy.microsoft.com/en-ca/fullnotice.mspx#EAB>, which states:

> We may also use technologies, such as cookies and web beacons to collect information about the pages you view, the links you click and other actions you take on our sites and services. Additionally, we collect certain standard information that your browser sends to every website you visit, such as your IP address, browser type and language, access times and referring Web site addresses. We also deliver advertisements and provide Web site analytics tools on non-Microsoft sites and services, and we may collect information about page views on these third party sites as well.

Google privacy policy, online:<http://www.google.com/privacypolicy.html>, which states "We may combine the information you submit under your account with information from other Google services or third parties in order to provide you with a better experience and to improve the quality of our services." and Yahoo! privacy policy, online: <http://info.yahoo.com/privacy/us/yahoo/details.html>, which states "Yahoo! may combine information about you that we have with information we obtain from business partners or other companies."

15 James Waldo, Herbert S. Lin, and Lynette I. Millett, eds., *Engaging Privacy and Information Technology in a Digital Age* (Washington, D.C.: The National Academies Press, 2007).

16 Some, for instance, raise the fact that search engine providers may now be able to use more sophisticated technology such as facial recognition technology in the context of image processing and image search. See Article 29, *Opinion 1/2008, supra* note 12 at 14.

17 Neil Robinson et al., *Review of the European Data Protection Directive*, (RAND Corporation, Europe) [Robinson], online: (2009) <http://www.rand.org/pubs/ technical_reports/2009/RAND_TR710.pdf> at 17: "Communications networks and changes to the core architecture of the Internet and its protocols (e.g. Internet Protocol version

how the notion of "identifying" can be challenged on the web and as a result, it is not always clear whether profiles generated on the Internet should be included in the Definition and therefore covered by data protection laws such as PIPEDA.

A huge legal uncertainty is thus created for online business models that may need this data. This uncertainty is creating a situation in which industry players now have no choice but to decide for themselves what is or should be covered under the Definition and what isn't or shouldn't.[18]

A strict interpretation of the Definition may trigger a situation where certain profiles may not be covered under the Definition. In this case, the privacy of online users may be adversely affected. In the event that profile data can identify an individual, this data may reveal intimate details of the lifestyle and personal choices and interests of the user based on his browsing patterns and other online activities. At the same time, a broad interpretation under which all of these new types of data and profiles should be considered as *personal information* may bring about the governance of profile data by data protection laws; implying certain obligations for the data controller, which may be problematic in certain cases. For instance, it may be difficult for an organization collecting the data to grant access if this data has not even been processed.[19] It may also be a challenge to provide disclosure and obtain consent from an individual without actually identifying the individual.

In Europe, the Article 29 Working Group, having noticed that European jurisdictions differ in their interpretations of the definition of "personal data" has issued an opinion in 2007 in which they propose a more relative interpretation of the definition.[20] While the interpretation proposed by the Article 29 Working Group is

---

6, IPv6) will permit many more physical objects to have an Internet address, paving the way for a wide range of devices to be connected, such as vehicles, white goods and clothing. Combining these technologies with Radio Frequency Identification (RFID) could affect privacy in many ways, both good and bad."

18 Many website privacy policies take the position that certain data that they collect is Non-PII (Personally Identifiable Information) which may imply that this type of data is not covered by the Definition. See For example Tumri privacy policy, online: <http://www.tumri.com/privacy.html>:

> We collect Non-Personally Identifiable Information ("Non-PII") from visitors to this Website. Non-PII is information that cannot by itself be used to identify a particular person or entity, and may include your IP host address, pages viewed, browser type, Internet browsing and usage habits, Internet Service Provider, domain name, the time/date of your visit to this Website, the referring URL and your computer's operating system. We use this information to understand our Website traffic and for maintenance of the Website.

19 Yves Poullet et Jean-Marc Dinant, *Rapport sur l'application des principes de protection des données aux réseaux mondiaux de télécommunications, L'autodétermination informationnelle à l'ère de l'Internet, Éléments sur la réflexion sur la Convention no 108 destinés au travail futur du Comité consultatif* Centre de recherches informatique et droit (Strasbourg, 18 November 2004) [Poullet] at 34.

20 In order to find that data *relates* to an individual, either a *content* element, a *purpose* element or a *result* element should be present. This means that data is personal data when it contains information about a specific person (content), when it is used or likely

more flexible, it does not address the situation pertaining to profile data. More analysis may be necessary in order to determine at what point profile data should be considered as "personal information" and therefore, governed by PIPEDA. A guide on the notion of "personal information" which would focus on profile data or new types of data which can be collected on the Internet would probably be useful and welcome.[21]

## (b) Re-evaluating the Notion of "Identifying"

The Definition focuses on information that relates to an individual that is "identifiable". It is debatable whether the notion of identity is still relevant in the context of the Internet. For instance, certain authors have raised the fact that profiles of individuals, although they may be anonymous and not covered under the Definition in all cases, may still be used, for instance, to take decisions about an individual (or a profile), such as providing certain specific advertisement messages, granting a loan, etc.[22] One example could be Amazon which was accused of practising *adaptive pricing* using cookies that would identify the profile of a specific client in order to readjust and raise the price of certain items in accordance with the profile of the potential purchaser. For this reason, certain authors take the position that profiles should always be covered by the Definition.[23]

European privacy expert Chris Pounder suggests that "identifying" an individual does not necessarily involve correlating certain data (such as an IP address) to someone's name. He suggests that identifiability can involve something where there is a focus on a particular characteristic.[24] For example, this could mean that the user from a certain IP address is likely to be attracted to advertisement related to a certain area of interest because he/she has visited certain websites that pertain

---

to be used to determine the treatment of a specific person (purpose), or when it is likely to have an impact on a specific person (result). See Article 29 Data Protection Working Party, *Opinion 4/2007 on the concept of personal data*, (Adopted on 20 June 2007), 01248/07/EN WP 136.

21 The author of this paper is pursuing a PhD degree at University of Montreal under the supervision of Professor Vincent Gautrais on the topic of "Redefining personal information in the Context of the Internet" and is currently working on these issues.

22 See Roger A. Clarke, "Profiling: A hidden Challenge to the Regulation of Data Surveillance", (1993) 4 J.L.& Info. Sci. 403; and, "IP addresses and the *Data Protection Act*" (March 2008), online: out-law.com <http://www.out-law.com/page-8060> ["IP addresses"]: "An IP address in isolation is not personal data under the *Data Protection Act*, according to the Information Commissioner. But an IP address can become personal data when combined with other information or when used to build a profile of an individual, even if that individual's name is unknown." See also Poullet, *supra* note 19, at 25.

23 See Poullet, *supra* note 19, at 29.

24 Comment from Chris Pounder in answer to blog. Alma Whitten (Software Engineer), "*Are IP addresses personal*?" Official Google Blog (22 February 2008), online: <http://googlepublicpolicy.blogspot.com/2008/02/are-ip-addresses-personal.html>: "Identifiablity does not need a name — it can involve something where there is a focus on a particular characteristic (e.g., the user from the IP address 330.09.08.07 is likely to be interested in XYZ because he/she has visited web-sites P, Q and R)."

to that same area of interest. It is also worthy of note that in Sweden, the *Personal Data Act 1998* defines personal data as *"all kinds of information that directly or indirectly may be referable to a natural person who is alive"*[25] This definition does not refer to the fact that the data needs to "identify" an individual. For example, a website that would propose life insurance policies online, could conclude, rightfully or not, that the visitor that he is dealing with is homosexual and is stricken with AIDS based on the profile information collected by the cookies.[26] The Swedish law would therefore apply if the description "homosexual person who probably has AIDS" relates, at the time of connection, to a living physical person, even if such person is not identifiable by name.

New online tracking technologies and traffic on the web makes it possible to identify the behaviour of a machine (device, computer) and the behaviour of the individual behind the machine as well. It may therefore be possible to detect the personality of an individual in order to apply certain decisions to a profile, without any actual need for the identity (such as name and contact information) of this individual. Further analysis is needed in order to determine if the notion of "identity" is still relevant in the context of the web or if the interpretation of the Definition should be re-evaluated in light of the above.

In the event that the profile data is found to be included under the Definition and therefore governed by data protection laws such as PIPEDA, issues pertaining to the legality of the collection of the profile data and relating to the management of the profile data should be addressed.

## II. LEGALITY OF PROFILE DATA COLLECTION

Organizations active in cyberspace are collecting new types of data using new types of collection tools. The data collected may be used for various purposes. For example, *clickstream* data[27] can be collected through online tracking tools such as *cookies*, which can collect basic information from a web user (such as type of computer and Internet browser used) and more private information (web pages visited, how long the individual has looked at any given page, as well as geographical location and any transactions or comments made). They may also be collecting information through IP addresses which may also be used to disclose the physical loca-

25 Swedish *Personal Data Act* (1998:204), Section 3.

26 See Poullet, *supra* note 19, at 33-34.

27 Mouse clicks translate into an electronic signal which is then sent by the user's computer to other computers on the Internet, sending or requesting certain information from them.

tion of a device[28] although not very accurate at the present time.[29] Search engines may collect and process a variety of data over and above mere IP addresses, clickstream data and information collected through cookies. This may include the content of the queries made and the preferences of the user. Other service providers may request that users create an account in order to use their online services (participate in blogs, view or post videos, participate in an online social network), and will therefore collect the username and potentially additional profile information of the user participating in the service.

Many websites use cookie-based technology in order to enable them to deliver user-specific solutions for each device that accesses their web pages. It is also common for websites to keep a record of IP addresses of their online visitors, for demographic purposes such as counting visitors, their countries of origin and their choice of ISP, sometimes even as a security measure.[30] A majority of web portals and Internet companies would be severely limited, if not rendered useless, in the absence of clickstream data.[31] Search engines collect and also process vast amounts of data generated from the browsing of online users, including log files belonging to specific individuals' which record their use of search engine services using technical means, such as cookies. These log files may include the content of the search queries, histories of search queries, the date and time, source (IP address and cookie), the preferences of the user, and data relating to the user's computer); data on the content offered (links and advertisements as a result of each query); data on the subsequent user navigation (clicks).[32] The search engines claim to collect some of this data to improve the quality of their services, particularly their search ser-

---

28 Damien Cave, "Do They Know Where You Live?" (28 February 2000), online: salon.com <http://www.salon.com/tech/feature/2000/02/28/geographic/index.html>: "Ad-serving companies like Double Click offer services that they say can target ads to users by location. And Digital Island introduced technology last year called TraceWare, which can identify the location of Web site visitors with 96 percent accuracy. TraceWare works by scanning worldwide traffic as it passes through ISPs, then matching users' IP addresses with a database of IP address locations that Digital Island has built."

29 See Article 29, *Opinion 1/2008*, *supra* note 12, at 6.

30 For example, if a customer regularly accesses his account from an IP address in London, access to that customer's account from an IP address in Moscow might indicate fraud. See "IP Addresses", *supra* note 22.

31 Rebecca Wong and Daniel B. Garrie, "Demystifying Clickstream Data: A European and U.S. Perspective" (2006) 20 Emory Int'l L. Rev 563 at note 24: "Elimination of clickstream data or cookies would impact such websites as: www.yahoo.com; www.google.com; www.wamu.com; www.schwab.com; www.ibm.com. Adjoining these web sites are a slew of Internet and web applications that utilize cookies and clickstream data for authentication. Elimination would impact not only businesses but also a large number of government enabled web applications."

32 See Article 29, *Opinion 1/2008*, *supra* note 12, at 6.

vices,[33] to keep their services secure[34] and their users safe from *malware* or *phishing* attacks,[35] to detect and prevent advertising "click fraud," and for accounting requirements.[36]

Most data protection laws provide that an organization may only collect and store data which is relevant for the product or service to be provided. More specifically, under PIPEDA, principle 4.3.3 states that an organization shall not, as a condition of the supply of a product or service, require an individual to consent to the collection, use, or disclosure of information beyond that *required* to fulfil the explicitly specified, and legitimate purposes. In Quebec, a similar principle can be found under article 9 of the *Act Respecting the Protection of Personal Information in the Private Sector*[37] (the "Quebec Privacy Law"). An organization may not "refuse to respond to a request for goods or services (. . .) by reason of the applicant's refusal to disclose personal information except where collection of that information is *necessary* for the conclusion or performance of a contract, collection of that information is authorized by law; or there are reasonable grounds to believe that the request is not lawful." In case of doubt, personal information is deemed to be non-necessary. The *Civil Code of Quebec* also provides that an organization establishing a file on an individual shall have a serious and legitimate reason for doing so and may only gather information which is *relevant* to the stated objective of the file.[38]

Online tracking tools enable websites and other online service providers to gather information to track the online conduct of individuals and profile them in order to (i) send out personalized advertising more focused on consumer behaviour and offer different sponsored products and services; and (ii) better understand the behaviour of online users in order to improve or personalize their services and products. It is not always clear whether personal information collected in exchange for a service or a product or that is intended to improve the organization's products and services is "required", "necessary" or "relevant" in accordance with Canadian data protection laws.

---

33 Hal Varian, "Why data matters" (3 April 2008) online: Official Google Blog <http://googleblog.blogspot.com/2008/03/why-data-matters.html>.

34 For instance, Google claims that it needs users' data for improving security and fighting web spam. Web spam is junk that the user sees in search results when websites successfully cheat their way into higher positions in search results or otherwise violate search engine quality guidelines. See Matt Cutts, "Using data to fight webspam" (27 June 2008), online: Official Google Blog <http://googleblog.blogspot.com/2008/06/using-data-to-fight-webspam.html>.

35 Niels Provos, "Using log data to help keep you safe" (13 March 2008), online: Official Google Blog: <http://googleblog.blogspot.com/2008/03/using-log-data-to-help-keep-you-safe.html>.

36 For example, services such as clicks on sponsored links, where there is a contractual and accounting obligation to retain data, this data would be useful at least until invoices are paid and the period for legal disputes has expired. See Article 29, *Opinion 1/2008*, *supra* note 12, at 15-16.

37 See the Quebec Privacy Law, *supra* note 1.

38 Art. 37 C.C.Q.

### (a) Collecting data in exchange for a service or product

On the web, many services, much information as well as entertainment are offered for free to consumers as long as they accept the presence of advertising and submit their online behaviour to be tracked to a certain degree.[39] Some even raise the fact that among consumers, there seems to be a growing and implicit understanding that the use of their personal data is intrinsic to the provision of most online (and an increasing number of offline) services.[40]

In the recent CIPPIC complaint against Facebook, one of the issues was the fact that since users were not allowed to opt out of Facebook ads, Facebook was unnecessarily requiring users to agree to such ads as a condition of service, in violation of Principle 4.3.3 of PIPEDA (the "Facebook finding").[41] The finding of the privacy commissioner on this issue took into account the fact that the site is free to users and that since advertising is essential to the provision of the service, individuals who wish to use the service must be willing to receive a certain amount of advertising.[42]

This Facebook finding may illustrate a change in mentality as to what is acceptable from a privacy and business perspective, where a certain trade-off is necessary. It may also have an impact in the mobile space. Wireless devices are powerful communication devices with respect to immediacy, interactivity and mobility and can act as the most powerful marketing communications devices. Advertisers may wish to sponsor content alerts and location-specific services which may include traffic, navigation information, proximity and directory or information services, mobile gaming, mobile-commerce and shopping support, mobile dating ser-

---

39 See Fleischer, "Response", *supra* note 2.

40 See the 2008 Eurobarometer results, online: <http://ec.europa.eu/public_opinion/archives/flash_arch_en.htm>; commented on in Robinson, see *supra* note 17, at 4.

41 Office of the Privacy Commissioner of Canada, *PIPEDA Case Summary #2009-008, Report of Findings into the Complaint Filed by the Canadian Internet Policy and Public Interest Clinic (CIPPIC) against Facebook Inc., Under the Personal Information Protection and Electronic Documents Act* [Case Summary #2008-008], online: Office of the Privacy Commissioner <http://www.priv.gc.ca/>.

42 See Case Summary #2008-008, *ibid.*, at Section 3, Finding 131: "Facebook has a different business model from organizations we have looked at to date. The site is free to users but not to Facebook, which needs the revenues from advertising in order to provide the service. From that perspective, advertising is essential to the provision of the service, and persons who wish to use the service must be willing to receive a certain amount of advertising."

vices[43] and buddy lists.[44] Since these advertisers may play a significant sponsorship role in the financing of mobile data services, similar issues will potentially take place in the mobile space as well.

While this Facebook finding may provide for some type of guidance on the legality of the tracking and profiling activities in the event that the business model is based on providing a free or sponsored service, it is still not clear that the tracking and profiling activities would be legal if the service or product being offered was otherwise not free. While many services are sponsored in the online environment (the profitability of search engines generally relies on the effectiveness of the advertising that accompanies the search results,[45] certain email services are also supported by online advertising,[46] etc.), many online or mobile service providers that provide non-sponsored services may very well wish to, and potentially benefit from, using analytics solutions in order to better understand their customers' behaviour. It is debatable whether this type of collection would be valid in all cases.

### (b) Collecting Data in order to Improve Websites, Products, and Services

Many websites and online service providers disclose through their privacy policies that they may collect some type of information in order to improve their websites, products, or services.[47] It is unclear whether any service provider (online,

---

43 Wireless users might be interested in receiving a service that would provide them with movie schedules, locations and reviews based on their location, for example when and if they are downtown on a weekend night. Others may be interested in a dating service that would alert them if someone corresponding to the desired profile were in their area. At the same time, a content provider, like a specific coffee shop, might want to sponsor this dating service by inviting these people, through their wireless devices, to meet at the closest coffee shop for a free coffee.

44 For instance Facebook friends signed up with this service could be alerted on their mobile device when they are in close proximity, for example, within a half kilometer range.

45 See Article 29, *Opinion 1/2008*, *supra* note 12, at 6.

46 Users of Microsoft Outlook email application pay for a licence, download emails and store them on their own laptop. For users of Google Gmail services, their emails are managed through a web browser and are stored remotely with Google. The user pays for Gmail account by being exposed to the advertisements that Google places on the far right edge of the screen. See Randal C. Picker, "Competition and Privacy in Web 2.0 and the Cloud", U of Chicago Law & Economics, Olin Working Paper No. 414 (26 June 2008), p. 7. [Randall C. Picker, "Competition and Privacy in Web 2.0 and the Cloud" (2008) 103 N.W. U. L. Rev. Colloquy 1]

47 See Microsoft privacy policy which states: "Microsoft collects and uses your personal information to operate and improve its sites and services. These uses may include (. . .) performing research and analysis aimed at improving our products, services and technologies;" online: <http://privacy.microsoft.com/en-ca/fullnotice.mspx#EAB>; Google privacy policy which states: "Google only processes personal information for the purposes described in this Privacy Policy (. . .) such purposes include: (. . .) protect and improve our services; (. . .) and Developing new services" online: <http://www.google.com/privacypolicy.html>; and, Yahoo! Privacy policy which

mobile, or other) can legally collect data for the purpose of better understanding their customers' behaviour if they are not providing "free" services. They may wish to collect users' data, using data mining, analytics and similar tools or calculations, in order to capture, analyze, and correlate the data as a means to uncover hidden patterns within the otherwise raw jumble of information. The online service providers may then be able to determine or predict the future behaviour of consumers and identify different trends and patterns over time from large amounts of data, from sometimes disparate sources. This process may result in enabling online service providers to manage the wealth of customer information strategically, capitalize on the information collected and optimize the value of each customer.

The knowledge gained by organizations using analytics solutions and having them better understand the behaviour of their users may, in certain cases, be translated into direct on indirect benefits for consumers. Direct benefits would include personalized services, products and advertising where online businesses may be in a position to offer the right services to the right users at the right time. Personalizing certain products and services would improve the users' experience in the online and mobile worlds. Indirect benefits may include the upgrading of current products and services based on users' needs, developing and deploying new applications and services or the "repackaging" of certain products and services. These developments could eventually ensure that users only be charged for the services that they actually use instead of sponsoring other users' usage of certain services that they have no interest for. This may potentially result in reduced costs for these users.

The fact remains that more and more analytics solutions are available and service providers are looking to benefit from them. A flexible interpretation of the privacy principle under which only personal information "required," "necessary," or "relevant" to a service or a product can be collected should be adopted, in order to provide for the proper balance between the right for service providers to benefit from these tools (which may include certain potential benefits for consumers as well) while protecting users' privacy. This may be especially true on the Internet where business models are often based on personalization and sponsorship. Given that the "value" obtained by consumers is not always very clear,[48] service providers should be encouraged to disclose privacy policies to their users and educate them on the specific kinds of benefits that will result from their use of analytics solutions.

## III. CHALLENGES IN THE MANAGEMENT OF PROFILE DATA

There are additional uncertainties when attempting to translate certain privacy principles pertaining to the management of the profile data into business practices.

---

states: "Yahoo! uses information for the following general purposes: to customize the advertising and content you see, (. . .) improve our services (. . .).", online: <http://info.yahoo.com/privacy/us/yahoo/details.html>.

48 See Nicole Ferraro, "Users Will Trade Privacy for Value, Say Industry Leaders" Internet Evolution (1 April 2010), online: <http://www.internetevolution.com/author.asp?section_id=466&doc_id=190020&f _src=internetevolution_sitedefault>.

### (a) Retention Period for the Profile Data

Data protection laws usually restrict the retention of personal information for a period longer than what is necessary.[49] In the context of recent security breaches, privacy commissioners have outlined the importance of not retaining personal information which is no longer necessary.[50]

Given that in the context of tracking and profiling practices, personal information will need to be collected by organizations in order to track and analyze online users' behaviour over time, this data may need to be collected over a certain period of time in order to be useful to organizations. For example, Google collects personal data online for purposes such as improving its services and developing new ones and, as a result, has been criticized by Article 29 Working Party for retaining the data collected though its search engine for too long.[51] Google has agreed to limit its retention period from eighteen to nine months.[52] Microsoft has reduced its search data retention period to six months.[53]

Again, in the mobile space, similar concerns will arise. Some have been suggesting for quite a while that, similar to the ad networks such as DoubleClick, the more efficient companies looking to send personalized content to wireless users based on their geographical location will understand the behavioural patterns of the wireless users using location data.[54] Real-time location data may be useful to send

---

49 PIPEDA: 4.5 Principle 5 — Limiting Use, Disclosure, and Retention "Personal information shall not be used or disclosed for purposes other than those for which it was collected, except with the consent of the individual or as required by law. Personal information shall be retained only as long as necessary for the fulfilment of those purposes." 4.5.2 "Organizations should develop guidelines and implement procedures with respect to the retention of personal information. These guidelines should include minimum and maximum retention periods. Personal information that has been used to make a decision about an individual shall be retained long enough to allow the individual access to the information after the decision has been made. An organization may be subject to legislative requirements with respect to retention periods." The Quebec Privacy Law, *supra* note 4, s. 12, states "Once the object of a file has been achieved, no information contained in it may be used otherwise than with the consent of the person concerned, subject to the time limit prescribed by law or by a retention schedule established by government regulation."

50 Office of the Privacy Commissioner of Canada and Office of the Information and Privacy Commissioner of Alberta, *Report of an Investigation into the Security, Collection and Retention of Personal Information, TJX Companies Inc. /Winners Merchant International L.P.*, (25 September 2007), online: Office of the Privacy Commissioner <http://www.priv.gc.ca/>.

51 Letter from Article 29 Data Protection Working Party to Google (16 May 2007).

52 See Fleischer, "Response", *supra* note 2; and Peter Fleishcer, Jane Horvath, and Alma Whitten, "Another step to protect user privacy" The Official Google Blog, (8 September 2009) [Fleischer, "Another Step"], online: <http://googleblog.blogspot.com/2008/09/another-step-to-protect-user-privacy.html>.

53 Thomas Claburn, "Microsoft Reduces Search Data Retention to Six Months" *Information Week* (8 December 2008).

54 Goldman Sachs, Technology: Mobile Internet, MOBILE INTERNET PRIMER, U.S., (14 July 2000) at 5.

a content message to a wireless user that appears to be at the right geographic location at the right time to make a certain message relevant. Historical location data which is location data gathered and stored over a certain period of time may enable even greater personalisation in the mobile space. It will enable mobile service providers to create dynamic profiles about users' movement patterns, lifestyle, and habits though time, especially since a wireless device is time-sensitive and typically used by only one individual.

Certain Canadian jurisdictions have not adopted or established retention schedules. For instance, the Quebec Privacy Law states that the government, after obtaining the advice of the Commission, may make regulations to establish retention schedules.[55] Since this was not done, the Commission d'accès à l'information in Quebec has been refusing to take a position on what constitutes a reasonable retention period, most notably in a recent case brought before it.[56] It would be useful for Canadian industry players involved in online and mobile profiling and tracking activities to have some indication of a reasonable retention period; particularly with regards to collecting profile data for the purpose of either providing a free or sponsored product or service, improving current products and services or developing new ones or, finally, for personalizing products or services. Retention periods should take into account the various benefits resulting from the use of analytics solutions for online businesses and their users. For instance, when Google was requested to reduce the period of time that it retains the data collected, it raised concerns about the potential loss of quality and innovation that may result from having less data.[57]

Certain organizations may wish to anonymize the profile data in their possession instead of destroying it once the fulfilment of the purpose of collection is completed. Guidance as to what "anonymizing" data actually means would be useful. For example, Google and Article 29 Working Party recently did not agree on what anonymization of data meant. After Google revealed its anonymization process,[58]

---

55 See *supra* note 2, s. 90(3).

56 In *E.P. c. TransUnion*, 2009 QCCAI 139, [2009] C.A.I. 139 (RNF), the Commission d'accès à l'information mentioned that it could not take a position as to the reasonableness of the seven years retention period by a credit score agency because the government has not established a retention schedule.

57 See Fleischer, "Response," *supra* note 2; and Fleischer, "Another Step," *supra* note 52:

> When we began anonymizing after 18 months, we knew it meant sacrifices in future innovations in all of these areas. We believed further reducing the period before anonymizing would degrade the utility of the data too much and outweigh the incremental privacy benefit for users. (. . .) While we're glad that this will bring some additional improvement in privacy, we're also concerned about the potential loss of security, quality, and innovation that may result from having less data. As the period prior to anonymization gets shorter, the added privacy benefits are less significant and the utility lost from the data grows.

58 Letter from Google to the Article 29 Working Party in answer to their Letter dated May 16, 2007 (10 June 2007) at 5:

> We are putting significant resources into creating processes for reliably anonymizing data. Although we are still developing our precise

Article 29 Working Party issued an opinion on data protection issues related to search engines in which it took a position that an anonymization process must be "completely irreversible" for Directive 95/49 to no longer apply.[59] With the advent of more sophisticated technology comes the possibility to link an individual to certain data, thereby challenging the very notion of "anonymization" of data. Experts claim that there is always a risk of re-identification with new technologies,[60] and that as the semantic web continues to evolve and tools become more sophisticated, re-identification arguably could become easier.[61] In this context, perhaps some guidance on what kind of anonymization methods are acceptable would be useful.

---

> technical methods and approach, we can confirm that we will delete some of the bits in logged IP addresses (i.e., the final octet) to make it less likely that an IP address can be associated with a specific computer or user. And while it is difficult to guarantee complete anonymization, the network prefixes of IP addresses do not identify individual users. Logs anonymization will not be reversible. We will intentionally erase, rather than simply encrypt, logs data so that no one (not even Google) can read it once it has been anonymized. Finally, logs anonymization will apply retroactively and will encompass all of Google's search logs worldwide.

59 See Article 29, *Opinion 1/2008*, *supra* note 12, at 20:

> Even where an IP address and cookie are replaced by a unique identifier, the correlation of stored search queries may allow individuals to be identified. (. . .) Anonymisation of data should exclude any possibility of individuals to be identified, even by combining anonymised information held by the search engine company with information held by another stakeholder (for instance, an internet service provider). Currently, some search engine providers truncate IPv4 addresses by removing the final octet, thus in effect retaining information about the user's ISP or subnet, but not directly identifying the individual. The activity could then originate from any of 254 IP addresses. This may not always be enough to guarantee anonymisation.

60 Yet, as Joel Reidenberg and Paul Schwartz suggest, anonymity in a network environment is not necessarily absolute. The mapping functions that render data anonymous are not always irreversible. Joel R. Reidenberg and Paul M. Schwartz, "Data protection law and online services: regulatory responses" (Delivered to the Commission of the European Communities, December 1998), at 34.

61 See Robinson, *supra* note 17, at 27: "Anonymity in large datasets is also complicated. Healthcare research is one area that uses large sets of anonymised clinical data for statistical analysis, data mining etc. However, regardless of how rigorously the data is de-personalized, legally speaking under this absolute interpretation it remains personal data if there is a possibility of linking the data to an individual, however remote, difficult or complex that may be." Many authors including Khaled El Emam suggests that there is indeed evidence showing that it is often possible to re-identify data sets. He suggests that de-identification should be part of an overall risk management approach. See Khaled El Emam, *De-identification Risk Assessment Model*, May 30, 2009. [Khaled el Emam, "Risk-Based De-Identification of Health Data" (2010) 8:3 I.E.E.E. Security & Privacy 64]

### (b) Access to Profile Information

It is a general privacy principle that an online user should be informed of the existence, use, and disclosure of his or her personal information and be given reasonable access to that information. The user should also be able to correct or amend that information when it is inaccurate. When it comes to the Internet, many kinds of data can be collected and many different types of profiles created. Deciding on whether access should be granted and, as the case may be, the type of data that should actually be covered by the access principle can sometimes prove to be a challenge.

PIPEDA provides that in certain situations, an organization may not be able to provide access to all the personal information it holds about an online user and that in these cases access might be denied.[62] However, PIPEDA does not specify what these exceptions are. It might be useful to have certain guidelines as to the type of data or profile data that should or should not be covered by the access principle.

It is usually accepted that data which (i) contains references to other individuals; (ii) cannot be disclosed for legal or security reasons; (iii) is subject to solicitor-client or litigation privilege; or (iv) would reveal commercial proprietary or trade secrets[63] should be off-limits. Important factors such as expense and burden should be taken into account when determining whether providing access to profiles is reasonable.[64]

Whether the profile data collected is relevant and important for the online users should also be considered. For instance, if the information is used for decisions that will significantly affect the individual, then the organization should disclose that information even if it is relatively difficult or expensive to provide. If the information requested is not sensitive or not used for decisions that will significantly affect the individual (such as non-sensitive marketing data that may be used to determine whether or not to send certain advertisement), but is readily available and inexpensive to provide, an organization should be obliged to provide access to

---

62 Schedule 1, Section 5, s. 4.9.

63 Potentially, this type of inferred or derived information could be in certain circumstances the result of a proprietary model and could provide a competitive advantage to the organization, for instance when the data is an indicator of an online user's future purchase behaviour. Disclosing the assumptions or conclusions a business makes might undermine competition by inviting competitors to attempt reverse engineering to proprietary operations and allowing them to free-ride off the analytic work of rivals. See Federal Trade Commission, "FTC Advisory Committee on Online Access and Security, Final Report" (3 May 2000), [FTC on Online Access] at 8, online: <http://www.ftc.gov/acoas/papers/acoasdraft1.htm>.

64 Navigational or *clickstream* data is being processed automatically and changing over time. Some argue that perhaps there is little benefit, and much cost, in accumulating this data in a form that could be reviewed intelligibly by the individual at any moment. See Online Privacy Alliance (OPA), "Online Consumer Privacy in the U.S. Submitted with the Comments of the Online Privacy Alliance, On the Draft International Safe Harbor Principles", (19 November 1998) legal framework White Paper, online: <http://www.privacyalliance.org/news/12031998-5.shtml>; In addition, some believe that providing access to this type of data might be too costly to provide. See FTC on Online Access, *supra* note 63, at 24.

it.

The privacy issue with granting access to a website recording navigational or *clickstream* data as an online user moves from page to page on its web site is that the data collected through these devices does not necessarily belong to one single individual. This entails that providing access to an online user to this data[65] may breach the privacy of the other users of the same computer. The profile data, clickstream data and other data that could be collected might reveal significant private information. For instance, an employee at work sharing his/her workstation with other colleagues could be afflicted with a certain embarrassing disease. Should the data regarding this disease be disclosed to online users who request it, the employee in question would be terribly embarrassed and this would be in breach of his or her privacy. Access to profile information could be provided to an online user requesting it, if the web site is one of general interest and therefore the disclosure of data collected by a cookie would not be revealing anything private about the other computer's user(s). Some websites use anonymous data collected from tracking tools sometimes for statistical use. The organization could refuse access to the online user if the data collected is anonymous or aggregated.

With regards to data derived from data mining or analytic tools, the organization may not be in a position to provide this data-mined or inferred data in an intelligible form to the users or at a low or reasonable cost. In addition, it might be impractical and difficult to enable a user to ascertain whether the inferences made using certain tools or calculations are relevant. These types of data are not usually susceptible to correction. Therefore, it could be a major challenge for the user to update or amend this data (or profile data). These arguments could be sufficient to conclude that organizations should not have to provide access to this type of data for reasons of practicality and competitive advantage. Some believe that providing access to this type of data might be costly, and therefore access should be denied.[66] On the other hand, a refusal of access could be potentially harmful when the data is used to make a decision about the user that would result in an important denial of services. This may open the door for a case-by-case evaluation. If this data can be used to make decisions that will have a serious impact on individuals, it should therefore be available to online users requesting it.

Finally, it is interesting to note that certain online service providers and websites have recently allowed users to access part or all of their profile information.[67] Perhaps this illustrates a new trend in which industry players wish to demonstrate some type of transparency in their profiling activities or wish to increase the quality

---

65 The web site could, in order to authenticate the identity of the access requester, require that the identifier (the number associated with the organization cookie) be provided in an access request.

66 See FTC on Online Access, *supra* note 63, at 24.

67 See "Transparency, choice and control — now complete with a Dashboard!" (5 November 2009), online: Official Google Blog <http://googleblog.blogspot.com/2009/11/transparency-choice-and-control-now.html>. See also Amazon website which has a section entitled "Improve your recommendations" in which the user can request that certain purchases made not be included as part of its profile by rating items or adjusting the checkboxes; online: <https://www.amazon.ca/gp/yourstore/iyr?ie=UTF8&ref;_=sv_ys_3>.

of the profile data collected which, at the end of the day, will benefit both parties.

## CONCLUSION

On the Internet, it is possible to collect new types of data such as clickstream data, IP addresses, data collected by cookies or similar tools. This data relates to a machine used by one or more individuals. Interestingly, isolated pieces of profile information, which may not be in and of themselves be included under the definition of personal information, may at some point become sensitive in light of profiling or behavioural analysis practices. These isolated fragments may actually be used to identify an individual, especially with the convergence between various different technologies and correlation across services.

The definition of personal information from PIPEDA has been interpreted broadly. Still, it is not always clear at what precise point an online profile may in fact be associated with an *identifiable individual* and therefore, governed by PIPEDA. A guide on the notion of "personal information" which would focus on profile data or new types of data which can be collected on the Internet would probably be useful and welcome. Also, further analysis may be needed in order to determine if the notion of "identity" is still relevant in the context of the web. As a matter of fact, given that new online tracking technologies on the web make it possible to identify the behaviour of a machine (device, computer), it is possible to detect the presence or the personality of an individual behind the machine in order to apply certain decisions (providing certain specific advertisement messages, granting a loan, etc.), without the explicit need for the identity (such as name and contact information) of this individual.

On the Internet, online advertising is becoming increasingly prevalent, no doubt aided by the fact that technology now makes it possible to gather a lot of information to profile individuals, track their online conduct, and properly profile them in order to produce personalized advertising more tailored to actual consumer behaviour. More and more analytics solutions are available and service providers — whether or not they are offering free services — are looking benefit from them. A flexible interpretation of the privacy principle under which only personal information "required," "necessary," or "relevant" to a service or a product can be collected should be adopted, in order to provide for the proper balance between the right for service providers to benefit from these tools (which implies certain benefits for the service providers as well as their users) while protecting users' privacy. Service providers should be encouraged to disclose their information gathering practices to their users and educate them on the specific kinds of benefits for users which will result from their use of analytic solutions.

Data protection laws usually restrict the retention of personal information for a period longer than what is necessary. Given that in the context of tracking and profiling practices, personal information will need to be collected by organizations in order to track and analyze online users' behaviour over time, this data may need to be collected over a certain period of time in order to be useful to organizations. It would be useful for Canadian industry players involved in online and mobile profiling and tracking activities to establish some guidelines concerning a reasonable retention period. Retention periods should take into account the various benefits resulting from the use of analytics solutions for online businesses and their users if the data will be used to improve or develop new products or services.

Certain organizations may wish to anonymize the profile data in their possession instead of destroying it once the fulfilment of the purpose of collection is completed. Guidance as to what "anonymizing" data actually means and what kind of anonymization methods are acceptable would be useful given that with the emergence of new, highly sophisticated technologies, the possibility of linking an individual to certain data has increased and has challenged the very the notion of "anonymization" of data.

Finally, PIPEDA provides that in certain situations, an organization may not be able to provide access to all the personal information it holds about an online user and that in these cases access might be denied. However, PIPEDA does not specify what would these exceptions really are. It might be useful to have certain guidelines as to the type of data or profile data that should or should not be covered by the access principle.