Hidden Surveillance by Consumer Health Websites

Behavioural Tracking Practices and Disclosure

Final report to the Office of the Privacy Commissioner of Canada
Contributions Program 2013-2014

Jacquelyn Burkell
Alexandre Fortier

Faculty of Information and Media Studies
The University of Western Ontario
London, Ontario

March 31, 2014

Table of Contents

List of Tables

Executive Summary

Behavioural tracking presents a significant privacy risk to Canadians, particularly when their online behaviours reveal sensitive information that could be used to discriminate against them. This concern is particularly relevant in the context of online health information seeking, since searches can reveal details about health conditions and concerns that the individual may wish to keep private. The privacy threats are exacerbated because behavioural tracking mechanisms are large invisible to users, and many are unaware of the strategies and mechanisms available to track online behaviour. In this project, we seek to document the behavioural tracking practices of consumer health websites, and to examine the privacy policy disclosures of these same practices. The results of our research demonstrate that tracking is widespread on consumer health information websites; furthermore, sites recommended by Information Professionals are similar to sites returned in Google searches in terms of overall tracking, though they show lower levels of third-party advertiser presence. Privacy policy disclosure of tracking practices is largely ineffective, and website visitors cannot easily determine tracking practices from a review of the website privacy policies. Taken together, these results suggest that alternative mechanisms are required to detect and/or mitigate or neutralize the behavioural tracking measures used on many consumer health information websites.

Our goal is to raise awareness of behavioural tracking and potential responses by communicating these results, and information about the risks of and responses to behavioural tracking, to three different groups: the academic community, Library and Information Science professionals, and the general public. This communication is carried out using a variety of mechanisms including presentations, publications, public lectures, and an educational video. In addition, we will provide education regarding behavioural tracking and associated risks to an important group of professional intermediaries: librarians. Armed with this education, librarians will be better able to select privacy-respecting information resources for their clients, and they will also be better prepared to address behavioural tracking as part of the information literacy education for the general public that they undertake as part of their professional practice.

Sommaire

Le pistage comportemental présente un risque important pour la protection de la vie privée des Canadiens, particulièrement lorsqu'il se produit dans des domaines où l'information récoltée pourrait être utilisée comme outil de discrimination entre utilisateurs. La nature privée de l'information en matière de santé amène un risque particulièrement élevé à cet égard pour la vie privée des internautes, puisque les recherches d'information à ce sujet peuvent révéler des détails qu'on désirerait garder pour soi. L'invisibilité des mécanismes utilisés pour le pistage comportemental exacerbe de plus les risques pour la protection de la vie privée et bien des Canadiens ne sont pas au courant des stratégies qu'ils peuvent employer pour contrer — ou, du moins, atténuer — les effets du pistage comportemental.

Les résultats de ce projet de recherche indiquent que le pistage comportemental est présent sur la majorité des sites web offrant de l'information relative à la santé. Les sites web recommandés par les professionnels de l'information, en outre, ne font pas meilleure figure que ceux trouvés à l'aide de recherches sur Google, bien qu'ils présent un taux un peu moins élevé de pistage par des annonceurs. La divulgation des pratiques de pistage comportemental dans les politiques de confidentialité est par ailleurs peu efficace et, à leur lecture, il est difficile pour un internaute de déterminer quelles sont les pratiques de pistage comportemental en cours sur le site web qu'il visite. Ces résultats suggèrent que des mécanismes alternatifs sont requis pour détecter le pistage comportemental ou pour, du moins, en amoindrir les effets.

Notre but est de sensibiliser à la fois la communauté universitaire, celle des professionnels de l'information et le public en général sur les effets du pistage comportemental et des mécanismes qui peuvent être employés pour en atténuer les effets. Divisé en plusieurs volets, cet exercice de dissémination prend plusieurs formes destinées à ces diverses communautés : présentations, publications et production d'une vidéo. Nous procéderons de plus à un atelier de formation continue auprès de bibliothécaires, qui pourront ainsi proposer à leurs usagers des sites web plus respectueux de la vie privée et également disséminer ce savoir auprès de leurs usagers.

Introduction

> *Most Canadians consider health information to be extremely sensitive. It is inappropriate for this type of information to be used in online behavioural advertising.*
>
> (Chantal Bernier, Interim Privacy Commissioner, January 15, 2014)

Health information is indeed considered among the most sensitive of personal information (see, for example, Nass et al, 2009), and most discussions of health information privacy focus on one particular issue: the protection of medical information, usually in the form of electronic health records (e.g., Pritts, 2008; Whetton, 2013; Norgren, 2013). In the quote above, however, Ms. Bernier is referring to a different kind of health information: details about health-related online searches, collected by Google Adsense and used to target health-related advertisements across multiple websites and over time[1]. She made her remark in a 2014 press release reporting the results of an investigation of a consumer complaint regarding this activity that concluded that the practice contravenes both online behavioural advertising guidelines issued by the Office of the Privacy Commissioner of Canada (OPC) and Google's own privacy policy. The investigation, triggered by a consumer complaint of being "followed" by advertisements for CPAP devices after searching for information online using Google, documented the use of sensitive personal information to deliver interest-based advertisements by one online advertising service (Google Adsense). This raises an important question: how common is such tracking across consumer health information websites?

Other research and the results of investigative reporting (see Soltani et al., 2009; Angwin, 2010) demonstrate that online behavioural tracking is a common and perhaps even ubiquitous practice. These investigations examine a cross-section of commonly visited websites, and the results are thus broadly representative of the behavioural tracking that Internet users are likely to encounter on a regular basis. Among the websites examined in these studies, however, only a small proportion are consumer health information sites: Soltani et al., 2009, for example, reports the results of an analysis of the top 50 sites identified by Quantcast in 2009, and among these

[1] https://www.priv.gc.ca/media/nr-c/2014/nr-c_140115_e.asp

only one (WebMD) is a site that delivers consumer health information. Thus, these results tell us relatively little about behavioural tracking practices on sites where users are likely to reveal, by their activities, detailed information about their health status, health-related concerns, and health-related activities. The current project seeks to address this gap by examining behavioural tracking practices on English and French language consumer health information websites. In addition, we examine the disclosure practices regarding tracking on these sites. One important aspect of our project is to contrast tracking activities on sites recommended by Library and Information Science professionals with those sites, not recommended by these professionals, that consumers would find on their own through internet searches for common health conditions. This contrast provides insight into the degree to which information professionals are successfully protecting the privacy-related interests of their patrons in their online health information recommendations.

Background

Behavioural Tracking and Privacy

It has long been recognized that Internet users face privacy risks as they navigate online spaces. Historically, these privacy concerns have focused on the collection, use, and retention of personally identifying information (PII) that is explicitly provided by users in the course of online activities (e.g., registration information that includes name, email, etc.). More recently, however, websites and associated advertisers have increased their use of behavioural tracking measures that collect non-personally identifying information (NPII) that cannot be associated with a specific identifiable individual, including IP address, browser configuration information, and details of browsing behaviour (Soltani et al., 2009; McDonald & Cranor, 2010; Ayenson et al., 2011; Chester, 2012).

Behavioural tracking is often justified as a tool that supports positive outcomes such as website personalization and targeted advertising that delivers information on products and services that are of interest to the user. The information gathered through this tracking, however, can also be used to discriminate against consumers through activities such as price discrimination or even denial of service (e.g., insurance applications; Center for Digital Democracy et al., 2009; Castelluccia & Narayanan, 2012). The detailed personal profile that can be developed on the basis of behavioural tracking, especially when that information is integrated across multiple visits and/or multiple websites, is of potential interest to employers, insurers, and providers of financial

services — in fact, to anyone who would derive value from the segmentation of Internet users according to their online behaviour and characteristics inferred on the basis of that behaviour (Kosinski, Stillwell, & Graepel, 2013). Privacy threats associated with this profiling are particularly acute in the context of health information, since the searches that individuals conduct can reveal sensitive and potentially damaging information regarding health-related concerns and interests (Anderson-Inman & Horney, 1998; Berger, Wagner & Baker, 2005; Cline & Haynes 2001).

Various privacy guidelines have been proposed for the collection, retention and use of personal information in the online environment (e.g., the *Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data*, developed by the Council of Europe, and the Organisation for Economic Co-operation and Development's *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*). Arguably foremost among these is the set of Fair Information Practice Principles (FIPs) proposed in 1973 by The US Secretary's Advisory Committee on Automated Personal Data Systems. FIPs and other guidelines are not themselves enforceable, but these principles form the basis of legally enforceable regulatory frameworks including the European Commission Data Protection Directive (Directive 95/46/EC) and Canada's *Personal Information Protection and Electronic Documents Act* (PIPEDA).

It is important to recognize that regulatory frameworks did not originally contemplate the collection of NPII, and were focused solely on the regulation of the collection and use of personally identifiable information. Regulatory bodies have noted this problem, and are beginning to respond. The *European Data Protection Directive*, for example, has recently been extended to cover any information that a website causes to be stored in a users' browser (thereby covering some if not all forms of NPII; 2009 EU directive 2009/136/EC). The OPC has recently developed guidelines[2] and a policy position[3] on online behavioural advertising that address the application of PIPEDA to the collection and use of NPII in the context of online behavioural advertising. The guidelines and policy position extend PIPEDA coverage to at least some NPII through the argument that this information can be personally identifying, requiring opt-in consent for collection and use of sensitive information, and opt-out or implied consent for information that is less sensitive. These advances achieve the positive outcome of increasing the reach of

[2] https://www.priv.gc.ca/information/guide/2011/gl_ba_1112_e.pdf
[3] http://www.priv.gc.ca/information/guide/2012/bg_ba_1206_e.asp

legislative regulatory frameworks with respect to online behavioural tracking. At the same time, enforcement is challenging and thus uneven, given that is typically reliant on consumer complaint to identify breaches. Moreover, unless the regulations require opt-in consent, users may remain unaware of behavioural tracking, since many will fail to read privacy policies that provide disclosure for opt-out consent (Milne et al., 2004; Vila et al., 2003).

Increasingly, consumers are seeking health information online (Fox, 2011; Statistics Canada, 2011), revealing in the process potentially sensitive information about their health status, health concerns, and health-related activities. It is important, therefore, that we understand the behavioural tracking practices of the consumer health websites used by consumers, in order to develop educational interventions and other strategies that will help to ensure consumer privacy with respect to this type of health information.

Consumer Health Information

Consumer interest in health information has been increasing in response to a movement toward patient participation in health care decisions (Entwistle, 2000, Holmes et al., 2005, Coulter, 1997, Gafni et al., 1999) and sociocultural changes including the general consumer movement and women's health movements of the mid to late 1990s (Marshall, 1992). Health information is and has been available through a variety of sources, including health care professionals, family and friends, and various forms of media (Hesse et al, 2005). Increasingly, however, consumers are seeking this information online, and searches for health information are among the most common of online activities. The latest iteration of the Canadian Internet Use Survey, for example, indicates that 69.9% of Canadians who use Internet at home search for medical or health related information, making it one of the most widespread online activities, ranking above paying bills online (Statistics Canada, 2011).

The proliferation of health information on the Internet has certainly brought many positive outcomes. Even though consumers report that their preferred source of health information is a health professional (Hesse et al., 2005), they often consider that the information they receive from these professionals is insufficient (Coulter, Entwistle and Gilbert, 1999; Chen and Siu, 2001; Saver et al., 2007). In this context, access to online health information increases consumers' health knowledge, such as their understanding of the risks and benefits of treatment, and enables them to make more informed decisions (Coulter, 2006; Coulter et al., 2006). For

consumers, the Internet also improves anonymous access to information on stigmatized conditions or issues such as mental illness or sexual health (Anderson-Inman and Horney, 1998; Cline and Haynes, 2001; Berger, Wagner and Baker, 2005).

At the same time, online health information presents risks to consumers. There are the obvious risks associated with inaccurate or misleading information, particularly acute because consumers use health information to make important decisions (Fox and Rainie, 2002). Of particular relevance in the current context is that fact that consumers seeking health information online are also subject to the privacy risks associated with the collection of personal information and behavioural tracking data (Soltani et al., 2009; McDonald and Cranor, 2010; Ayenson et al., 2011; Chester, 2012), and the often sensitive nature of health information seeking serves only to heighten the potential negative consequences. Many Internet users are unaware of behavioural tracking, and even those who are familiar with the issue may not fully understand the range and power of behavioural tracking mechanisms (McDonald and Cranor, 2010). Given this background, online health information seeking creates a 'perfect storm' of privacy concern for Canadians.

Digital Literacy and the Role of Information Professionals

Librarians are and have long been important intermediaries in the search for health information (Lunin, 1987; Marshall et al., 1991; Murray, 2008; Rees, 1991; Rubenstein, 2012), and health related questions are common at library reference desks (Marshall et al., 1991; Wood et al., 2000). In the past, information professionals have directed patrons to print sources for health information. Increasingly, however, the best and most up-to-date health information is available on the Internet, and this shift has brought about attendant changes in the role of information professionals with respect to this important aspect of information practice. In particular, information professionals are now directing patrons to online health information resources, and at the same time taking up the new responsibility of assisting their patrons to identify on their own the best possible online resources. As a result, these professionals are faced on two fronts with fundamental issues of digital literacy: first, information professionals must themselves have the skills and expertise necessary to select the best possible online resources; second, they must support their patrons in developing those same skills.

One digital literacy issue that is increasingly significant for information professionals is that of user privacy, not only with respect to library practices regarding the collection and use of personal information (Burkell and Carey, 2011), but also with respect to the privacy challenges associated with the resources that consumers access online. Protecting users' privacy is very important to librarians, and the issue is identified in the code of ethics of many library associations (e.g., Canadian Library Association, 1976). Librarians, from the earliest days of internet-based health resources, have worked to overcome barriers preventing consumers to make effective use of online health information (Cline and Haynes, 2001). They advise consumers who look online for health information to evaluate it with respect to factors such as intended audience, sponsorship, and information accuracy, currency and completeness (Fox and Rainie, 2002; Medical Library Association, n.d.). In helping users find good health information online, librarians have long advised them to pay attention to whether a health website includes advertisements (Medical Library Association, n.d.). These concerns about advertising are typically tied to the balance, coverage and objectivity of the information delivered on the site. More rarely, librarians advise users to read privacy policies for disclosure of collection and use of PII (e.g. MedlinePlus, n.d.). To date, however, information literacy guidelines have not addressed the privacy risks associated with behavioural tracking (Office of the Privacy Commissioner of Canada, 2011; 2012).

As information professionals responsible for promoting the information literacy of the public, librarians should be aware that advertisers, using powerful data collection and profiling apparatus, can form a rich and nuanced portrait of Internet users — without the collection of any identifying information. This requires an understanding of behavioural tracking mechanisms and strategies, including "data optimization, 'self-tuning' algorithms, 'intent' data and 'immersive' multimedia" (Chester, 2012) that leverage information collected through cookies and beacons. In particular, information professionals should understand that user activity can be tracked across multiple websites, linking key aspects of Internet users' digital identities into a comprehensive user profile. This understanding will support identification of privacy-respecting internet resources for recommendation to patrons, and also support digital literacy initiatives that will assist patrons in identifying privacy threats in the information they access online.

Behavioural Tracking Mechanisms

Mechanisms for tracking a user's activities online—and worries about potential privacy breaches they can allow—are not new. At their inception, HTTP cookies (cookies), which were originally introduced by Netscape in its Navigator 1.1 in the mid 1990s, were generating controversy about the potential invasion of privacy (Randall, 1997). HTTP cookies, however, are relatively easy for users to manage using accessible browser settings that limit or even entirely disallow the practice of setting cookies. As a result, websites, advertisers and others who benefit from web audience segmentation and behaviour analytics now deploy these mechanisms along with newer and more obscure tracking technologies including 'supercookies' and web beacons (Sipior, Ward and Mendoza, 2011). Cookies can be set by directly by the website (first-party cookies) or by advertising companies through ads embedded in first-party sites (third-party cookies). 'Supercookies' and web beacons, similarly, can be first-party or third-party mechanisms. First-party tracking mechanisms collect information about a site visit and visitor and deliver that information to the site itself. Using first-party tracking, web sites can provide personalized interaction, integrating visit and visitor information both within a single visit and across multiple visits. This information, however, is only available to the web site itself, and thus neither includes information about visits to other sites nor is accessible by other websites. Third-party tracking mechanisms, by contrast, deliver information about a site visit and visitor to a third party, often an advertising company. Third-party tracking represents a greater menace to privacy, since third parties have a presence on multiple sites, and are able to collect information about users and their activities on all those sites and integrate that information across sites and across visits into a single detailed user profile (McDonald and Cranor ,2010).

HTTP Cookies

HTTP cookies (also known as browser cookies) were originally meant to help web developers gather information about users in order to personalize and optimize user experience (Randall, 1997). These cookies are simply a few lines of text shared in an HTTP transaction, and a typical cookie might include a user ID, the time of a visit, and the IP address of the computer. Cookies do not usually include identifying information such as name or address, and they are able to do so if an only if the user has explicitly provided this information to the website. When users want to access a web page, their browser sends a request to the server for the specific

website and the server searches the hard drive for a cookie file from this site. If there is no cookie, a unique identifier code is assigned to the browser and a cookie file is saved on the hard drive. If there is a cookie, the browser transfers the cookie file contents back to that site using the previously recorded identifier code (Harding, Reed and Gray, 2007).

HTTP cookies can record visited pages, a user's chosen values and all mouse-clicking choices. They also provide the server with information such as a user's IP address, service provider, operating system and browser type (Harding, Reed and Gray, 2007). They enhance the experience of users in many ways and make the browsing experience more efficient. HTTP cookies can record the preferences of users on a web page and enable them to resume interaction with a website at the point where they were on the previous visit, which is the basis of site personalization. Using the IP address of a user, for instance, HTTP cookies can allow a website to display information relevant to the geographic area where a user is located. HTTP cookies also allow a website to remember registration details and the content users have put in their shopping basket (Harding, Reed and Gray, 2007). HTTP cookies inform webmasters of users' movements on their websites: what pages are visited, how often they are visited, and in what order. They can also indicate the common entry and exit points for a specific website, information that can be used to increase user satisfaction and traffic. This information has obvious value for website optimization and personalization. At the same time, however, the detailed profile of user activities, potentially aggregated over multiple visits, presents potential privacy risks. The information stored in HTTP cookies can allow a website to know what topics or products are of particular interest for a user, and identify browsing and information access habits.

Some HTTP cookies, called session or transient cookies, automatically expire at the end of a session. They are mainly used to keep track of what a consumer has added to a shopping cart or to allow users to navigate on a website without having to log in repeatedly. Other HTTP cookies, called permanent, persistent or stored cookies, are configured to keep track of users until the cookie reaches its expiration date, which be set many years after creation. Permanent HTTP cookies can be easily deleted using browser management tools (Sipior, Ward and Mendoza, 2011). Studies have shown that over 30 percent of users delete cookies once a month (e.g Marshal, 2005). Such behaviour, however, displeases advertisers, as it leads to an overestimation of the number of true unique visitors on a website and impede user tracking (Abraham, Meierhoefer and Lipsman, 2007).

Local Shared Objects

To palliate this 'attack' on HTTP cookies, an online advertising company, United Virtualities, developed a backup system for cookies: local shared objects (also known as Flash cookies), now a feature of Adobe's Flash Player plug-in (Soltani et al., 2009). Unlike HTTP cookies, Flash cookies do not have an expiration date. They are also not handled by a browser, but are stored in a location accessible to different browsers and Flash widgets, which are thus all able to access the same cookie. Flash cookies represent a more resilient technology for tracking than HTTP cookies, and erasing traditional cookies within a browser does not affect Flash cookies (Soltani et al., 2009). Moreover, Flash cookies have the ability to 'respawn' (or recreate) deleted HTTP cookies, and a website using Flash cookies can therefore track users across sessions even if the user has taken reasonable steps to avoid this type of online profiling (Ayenson et al., 2011).

Web Beacons

Users' online behaviour can also be monitored by beacons (also called web beacons or web bugs), which tiny are image tags embedded within the coding of a document placed on a website or an e-mail (Martin, Wu and Alsaid, 2003). The image tag creates a holding space for a referenced image residing on the Web, and beacons transmit information to a remote computer when the page is viewed. As with cookies, beacons can also be first- or third party (McDonald and Cranor, 2010). Unlike cookies, beacons are not tied to a specific server and, in the case of third-party beacons, can track users over multiple web sites (Schoen, 2009). User interaction on the web page, such as typed entries and mouse movement, can be tracked directly using beacons, and web beacons can also be retrieve information from a previously set cookie (Angwin, 2010). Such capacity means, according to the Privacy Foundation (2000), that beacons could potentially transfer to a third-party demographic data and personally identifiable information (name, address, phone number, email address, etc.) that a user has typed on a page.

Objectives

This research project seeks to document the tracking practices of consumer health websites that Canadians are likely to encounter. It also explores the disclosure of the collection of NPII in the privacy policies of these websites that engage in behavioural tracking.

Plan of the Report

The first chapter analyzes the presence of behavioural tracking mechanisms on both English language and French language consumer health websites, and the disclosure of these practices on a subset of consumer health websites are documented in Chapter 2. Chapter 3 discusses strategies for detecting, mitigating, and neutralizing behavioural tracking. Chapter 4, finally, describes the activities that we have undertaken to disseminate these results to the various communities and presents future dissemination plans.

Chapter 1: Behavioural Tracking Practices

Introduction

In this chapter, we analyse the presence of behavioural tracking mechanisms on both English language and French language consumer health websites. The analysis of the English language websites contrasts websites recommended by health librarians to those retrieved using Google searches for the ten most commonly searched conditions on the Internet. In the absence of a recommended list of consumer health websites in French, the analysis of French language websites does not include this contrast. Instead, for the French language websites we report the tracking detected on relevant websites retrieved using Google searches for the ten most commonly searched conditions on the Internet.

1.1 Methodology

1.1.1 Sampling of the Websites

For the English language websites, two sets of consumer health websites were identified:

1. The consolidated set of websites recommended by the Consumer and Patient Health Information Section (CAPHIS) of the American Medical Library Association (2010) and the Consumer Health Information Providers Interest Group (CHIPIG) of the Canadian Health Libraries Association (2010; see Appendix I for a full list);
2. The consolidated set of websites returned on the first two pages of Google searches of the ten most commonly searched conditions, as identified by the Pew Research Center's Internet & American Life Project (Fox, 2011; see Table 1 for a list of these conditions).

Obviously irrelevant results (e.g., sites for roofing companies that were returned for the 'shingles' search) were eliminated from the Google results, as were any sites that were included on the consolidated 'recommended' list. The Google searches were repeated three times using a different computer each time. The complete list of 'Google Only' websites used for the analysis can be found in Appendix II.

For the French language websites, in the absence of a recommended list of consumer health websites, the websites examined included the consolidated set of websites returned on the first two pages of Google searches of the ten most commonly searched conditions as identified by the Pew Research Center's Internet & American Life Project (Fox, 2011). The ten conditions identified in that report were translated into French by the second author for the purposes of searching (see Table 1). Obviously irrelevant results (e.g., kijiji.ca was returned in a search for 'vésicule biliaire') were eliminated from the results. The Google searches were repeated three times using a different computer each time, and all relevant returned websites were recorded. The complete consolidated list of websites used for the analysis can be found in Appendix III.

Table 1
The Ten Most Commonly Searched Conditions on the Internet (Fox 2011)

Original conditions	French translation
Shingles	Zona
Gallbladder	Vésicule biliaire
Gout	Goutte
Hemorrhoids	Hémorroïdes
Lupus	Lupus
Skin problems	Problèmes de peau
Allergies	Allergies
Heart disease	Maladie cardiaque
Diabetes	Diabète
Sleep disorders	Problèmes de sommeil

1.1.2 Data Collection and Analysis

The data for each website (Recommended, Google Only, and French language) was collected separately, following a protocol that was developed to avoid any contamination of tracking results between the websites. Each website was visited in an independent session. Each session began with the browser at an about:blank page, with clean data directories (no HTTP and Flash cookies, and an empty cache). The website was then accessed directly by entering the domain name into the browser's navigation bar. A typical user interaction with the website was mimicked by visiting approximately 10 pages on the site. Search functions on the site were used and any surveys that did not ask for personal information were completed (e.g., 'Question of the day' surveys). We did not click through on any ads or follow any external links; thus, user interaction was confined to the website in question.

At the end of the session, HTTP cookies in the browser cookie file were recorded along with any flash cookies stored in Adobe's Website Storage Settings panel. These results were augmented by those returned by Ghostery[4], a browser extension that records web beacons, and Charles[5], an application that captures and analyzes data being sent between the browser and the visited website, and between the browser and third-party sites. Using these sources, we created a comprehensive list of the tracking mechanisms present on the site and the domains from which these trackers originated. After these data were recorded, the browser cache was cleared, all HTTP cookies were removed and the flash cookie folder was emptied using Adobe's Website Storage Settings panel, in preparation for a new data collection session. In a separate session, each website was visited to identify a privacy policy; if a privacy policy was located, it was saved for later analysis.

Once all data collection was complete, we created a consolidated list of third party tracking domains identified on all websites. Using a combination of results from Ghostery and PrivacyChoice[6], we assigned each domain to one of two categories: third-party analytics, or third-party advertisers[7]. The results provide an overview of third-party tracking presence on these consumer health websites, with additional separate focus on third-party advertising, since it is these trackers that create the greatest privacy risk. For the purpose of the analysis, we also divided websites into three categories: Government, Not-for-Profit (e.g., Alzheimer's society) and Other (usually commercial sites).

1.2 Results for English Language Websites

The recommended lists from CAPHIS and CHIPIG yielded a total of 83 distinct consumer health information sites. The first two pages of the Google searches for the ten most commonly searched conditions yielded a total of 81 relevant websites that did not also appear on the recommended list.

[4] https://www.ghostery.com
[5] http://www.charlesproxy.com
[6] www.privacychoice.org/trackerlist
[7] We recognize that this determination is not always clearcut. For example, the distinction between third-party analytics companies and third-party advertisers is not always absolute (e.g., PrivacyChoice identifies Dataium as both an ad network and an analytics company). In general, for those cases where a domain undertook both types of activities, we identified it as an analytics rather than advertising domain, in order not to over-estimate the prevalence of third party advertising.

Third party behavioural tracking mechanisms were identified in the large majority of websites. Overall, 87% of the English language consumer health information websites included in the two samples had third party trackers, representing an average of 10 different domains for each site. Over half of the sites had trackers from third party advertisers, on average from 2 different domains. Across all sites we identified trackers from a total of 230 different tracking domains, of which 23 were identified as advertisers (see Appendix IV for a list of all tracking domains identified in the website scans).

1.2.1 Recommended vs. Google Only Sites

Table 2 identifies the tracking domains that appear on more than 25% of all websites, indicating separately the presence of trackers from these domains on Recommended and Google Only sites. There was no difference between Recommended and Google Only sites with respect to the presence of third party tracking ($X^2_{(1)}$=0.3, n.s.). Among the 83 Recommended websites, 86.7% (n=72) were associated with at least one tracking domain, compared to 87.7% (n=71) of the Google Only sites. Recommended sites, however, were significantly less likely to have third party advertisers: 41% (n=34) of Recommended sites had at least one third-party advertiser, compared to 61.7% (n=50) of Google Only sites ($X^2_{(1)}$=7.13, p<.01). We also examined the number of different domains represented on sites that had trackers present. Among those sites with third-party trackers, Recommended sites had trackers from an average of 13.1 domains, and Google Only sites had trackers from an average of 17.2 domains. The difference, however, only approached but did not reach significance ($t_{(141)}$=1.82, p=.071). Among those sites with third-party advertisers, Recommended sites had trackers from an average of 4.8 different domains, while Google Only sites had trackers from an average of 5.3 different domains. The difference was not significant ($t_{(82)}$=4.89, n.s.). These results demonstrate that third party tracking on consumer health websites is widespread, and that Recommended websites show similar levels of tracking to Google Only sites, although the presence of third-party advertisers (as compared to all third party trackers) is lower on Recommended sites.

Table 2 Common Tracking Domains (on more than 25% of sites)		
Cookie / Beacon	Number in Recommended sites (% of sites)	Number in 'Google Only' sites (% of sites)
Third Party Trackers (All)		
Google Analytics	60 (72.3%)	48 (59.3%)
Facebook	27 (32.5%)	30 (37.0%)
AddThis	29 (34.9%)	21 (25.9%)
ScorecardReesarch	23 (27.7%)	26 (32.1%)
Third Party Advertising		
DoubleClick	21 (25.3%)	38 (46.9%)
Microsoft Atlas Solutions	20 (24.1%)	24 (29.6%)
Google Adsense	11 (13.3%)	31 (38.3%)

1.2.3 Government, Not-for-Profit, and Other Sites

Considering all of the sites (Recommended and Google Only) in the sample, 18.3% (n=30) are Government sites, 32.3% (n=53) are Not-for-Profit sites, and 49.4% (n=81) are Other sites. When Recommended and Google Only sites are considered separately, significant differences emerge between the two groups ($X^2_{(2)}$=29.9, p<.001): among Recommended sites, 30.1% (n=25) are Government sites, 39.8% (n=33) are Not-for-Profit sites, and 30.1% (n=25) are Other sites; among Google Only sites, 6.2% (n=5) are Government sites, 24.7% (n=20) are Not-for-Profit sites, and 69.1% (n=56) are Other sites. Thus, other (including commercial) sites are far less likely to be included among those recommended by Library and Information science professionals compared to those returned by a Google search.

Among the Government sites, 83.3% (n=25) have at least one tracker, compared to 86.8% (n=46) of the Not-for-Profit sites, and 88.9% (n=72) of the Other sites. Across the three groups, there is no significant difference in the presence of trackers ($X^2_{(2)}$=0.173, n.s.). The picture is quite different, however, when only advertising trackers are considered. Among Government sites, only 3.3% (n=1) had this type of tracker, compared to 49.1% (n=26) of Not-for-Profit sites, and 70.4% (n=57) of Other sites. The difference across types was significant ($X^2_{(2)}$=8.543, p<.001). The number of trackers present was compared across Government (average of 3.6), Not-for-Profit (average of 6.4) and Other sites (average of 18.0); revealing a significant effect ($F_{(2,140)}$=14.6, p<.001). Post-hoc tests (Tukey's) revealed that Other sites have significantly more third-party trackers than either Government or Not-for-Profit sites, which do not differ from each other. We also compared the number of different advertising trackers present on Not-for-Profit

and Other sites (Government sites were eliminated from this analysis because only one had any advertising trackers). Among those sites with at least one advertising tracker, Not-for-Profit sites had an average of 2.8 advertising trackers, compared to 6.2 for Other sites; the difference is significant ($t_{(81)}$=3,41, p<.01). Thus, Government sites have fewer third-party trackers than do Not-for-Profit and Other sites, which do not differ significantly from each other. Other sites also have significantly more third-party advertisers present on their sites than do websites Not-for-Profit agencies.

1.3 Results for French Language Websites

The first two pages of the Google searches for the ten most commonly searched conditions yielded a total of 195 relevant French language websites. The presence of behavioural tracking was identified in the large majority of websites. Overall, 91% (n=179) of the consumer health websites included in the sample had third party trackers, representing an average of 13 different domains for each site. Forty percent (n=78) of the sites had trackers from third party advertisers, on average from 2.4 different domains. Across all sites we identified trackers from a total of 251 different tracking domains, of which 25 were advertisers (see Appendix IV for a list of all tracking domains identified in the website scans).

Among the French language websites, 17.1% (n=33) were Government sites, 20.2% (n=39) were Not-for-Profit sites, and 62.7% (n=121) were Other sites. When looking at the presence of third-party trackers, significant differences were present across the three categories ($X^2_{(2)}$=16.909, p<.001). Results indicate that Other websites (including commercial) are more likely to contain third party trackers (75.8% (n=25) of Government sites have third-party trackers compared to 84.6% (n=33) of Not-for-Profit sites and 97.5% (n=118) of other sites). The number of trackers present was compared across Government (average of 2.3), Not-for-Profit (average of 6.0) and Other sites (average of 18.2). The difference was significant ($F_{(2,173)}$=11.396, p<.001), and post-hoc tests (Tukey's) indicate that other sites have significantly more third-party tracking domains than do either Government or Not-for-Profit sites, which do not differ from each other. Significant differences were also be found when looking at the presence of advertisers ($X^2_{(2)}$=43.251 p<.001): among Government sites 3% (n=1) have advertising trackers, compared to 17.9% (n=7) of Not-for-Profit sites and 53.7% (n=65) Other sites. We contrasted the number of advertisers present on Not-for-Profit and Other sites (government sites were eliminated because

only one had any tracking by third-party advertisers). The results indicate that Other sites have significantly more third-party advertisers (7.0) than do Not-for-Profit websites (3.57; $t_{(9.905)}$=2.858, p<.05[8]). Table 3 identifies the tracking domains that appear on more than 25% of all websites, and identifies the proportion of French language websites that have trackers from each of these domains. Overall, the results for French language sites mirror those for English language sites: third party tracking is widespread, occurring on the vast majority of sites. Tracking by third-party advertisers, while less prevalent, is still common. Commercial sites show the highest levels of tracking by third-party advertisers.

Table 3
Common Tracking Domains (on more than 25% of sites)

Cookie / Beacon	Number (% of sites)
Third Party Trackers (All)	
AddThis	56 (28.7%)
Facebook	93 (47.7%)
Google +1	61 (31.3 %)
Media6Degrees	55 (28.2%)
Twitter	58 (29.7%)
Third Party Advertising	
DoubleClick	69 (35.4%)
Google Adsense	57 (29.2%)

Conclusion

The results of this research demonstrate that third-party behavioural tracking is present on the large majority (at least 4 out of 5) of English and French language consumer health websites, and almost half of consumer health websites have trackers from third-party advertisers (50% of English sites, 40% of French sites). Furthermore, the English language websites recommended by library associations are not significantly better with respect to third party tracking, although these sites do show lower levels of tracking by third-party advertisers. Government sites (both French and English) show high levels of third party tracking (4 out of 5 English government sites and 3 out of 4 French government sites), but they are much less likely to include trackers by third-party advertisers (less than 5% of both French and English language government sites have one or more advertising trackers). 'Other' (primarily commercial) sites show the highest levels of third-

[8] Levene's test for equality of variances was significant (F=4.084, p<.05, so equal variances were not assumed).

party trackers (well over 4 out of 5 French language and English language sites in this category have at least one third-party tracker) and they are most likely to include trackers from third-party advertisers (over half French language and English language sites in this category have trackers from at least one third-party advertiser). Thus, government consumer health websites are relatively free from tracking by third-party advertisers, but 'other' sites show relatively high levels of this type of tracking, with not-for-profit sites falling in between these two categories.

It is evident from these results that consumers seeking health information on the Internet are very likely to be subject to third-party tracking of their health information seeking behaviours. It is important, therefore, to examine privacy policies to determine whether a consumer could effectively learn, from reading those policies, about the tracking activities present on the websites they visit.

Chapter 2: Disclosure of Behavioural Tracking in Privacy Policies

Introduction

There exist, as discussed in the introduction, a variety of privacy guidelines have regulating the collection, retention and use of personal information in the online environment. One important aspect of these regulatory frameworks is the requirement for notice: users should be given notice of website practices with respect to the collection and use of personal information. This notice is typically provided in privacy policies that identify what information is collected, how it is used, and with whom it is shared.

In general, regulatory frameworks did not originally contemplate the collection of NPII, and instead were focused on the regulation of the collection and use of personally identifiable information. Although there is no explicit and universal requirement that users be apprised of the collection and use of NPII, such a provision would seem to be consistent with FIPs and other guidelines, and in the US new *Self-Regulatory Guidelines for Online Behavioral Advertising* identify the need to provide notice to users when behavioural data is collected that allows the tracking of users across websites and over time (United States Federal Trade Commission, 2009). Indeed, within the Self-Regulatory Guidelines it is noted that with changes in technology and increasingly powerful data analytic techniques the distinction between PII and NPII becomes "less and less meaningful and should not, by itself, determine the protections provided for consumer data." (United States Federal Trade Commission, 2009, p. 21-22). The Office of the Privacy Commissioner of Canada has also determined that PIPEDA protections extend in at least some circumstances to NPII, therefore requiring notice of these data collection practices in at least some circumstances. Thus, there seems to be general agreement that users should be informed of behavioural tracking measures active on the websites they visit.

Website privacy policies are often difficult to understand (Micheti, Burkell, & Steeves, 2010), apparently written with the goal of protecting a website owner against lawsuits rather than informing users (Earp et al. 2005; Pollach, 2005). Pollach (2005), for example, details a variety of linguistic strategies that serve to undermine user understanding of website practices, including mitigation and enhancement, obfuscation of reality, relationship building, and persuasive appeals.

Thus, it is legitimate and indeed important to examine whether the privacy policies of websites engaged in behavioural tracking effectively disclose these practices, particularly in the case of websites recommended by library and information science professionals.

In this chapter, we analyse disclosure of behavioural tracking practices on a subset of consumer health information websites. For the English language websites, these sites represent a purposive sample of the recommended sites: third party trackers were observed on all selected sites, and the set includes government and commercial sources, encompassing sites with relatively low levels of tracking (e.g., Mayo Clinic), as well as those with much higher levels (e.g., What to Expect; see Burkell and Fortier, 2013). For the French language sites, these sites represent the French language versions of the two sites that were recommended in the English language recommended lists and also appeared in the Google results (Cancer.ca and PasseportSante.net) plus four other websites purposively selected for the presence of third party trackers. These include not-for-profit and commercial sources and encompass sites with relatively low levels of tracking (e.g., FmCoeur.qc.ca), as well as sites with much higher levels of tracking (e.g., Vulgaris-Medical.com).

2.1 Methodology

Seven English language websites and six French language websites were selected (see Tables 4 and 5). Their privacy policies were examined qualitatively for disclosure of first- and third-party tracking mechanisms and NPII data collection. The analysis draws on the critical linguistics approach used by Pollach (2005), particularly focusing on linguistic strategies used for mitigation and enhancement and obfuscation of reality. These include the use of:

- Qualitative adjectives that emphasize or de-emphasize specific qualities;
- Temporal adverbs (e.g., ‘occasionally’ or ‘occasionnellement’) that downplay frequency;
- Conditional verbs (e.g., ‘may’ or ‘pourrait’) or structures that introduce uncertainty;
- Nominalizations (e.g., ‘the collection of data’ or ‘la collecte de données’) and the passive voice (e.g., ‘data are collected’ or ‘les données sont collectées’) that obscure agency.

Tables 4 and 5 present the websites selected for the analysis along with a summary of tracking mechanisms found on these websites.

Table 4
English Language Websites Used the Analysis of Privacy Policies

Website	Number of first party cookies	Number of third party cookies	Number of beacons
WhatToExpect.com	17 (6 sessional, 11 persistent, valid for up to 2 years)	119 (most persistent, valid for up to 33 years)	40
MedicineNet.com	18 (13 sessional, 6 persistent, valid for up to 17 years, 1 flash cookie)	118 (most persistent, valid for up to 3 years)	37
HealthyWoman.org	7 (2 sessional, 5 persistent, valid for up to 2 years)	42 (all persistent, valid for up to 2 years)	12
MayoClinic.com	11 (4 sessional, 9 persistent, valid for up to 30 years)	40 (all persistent, valid for up to 2 years)	9
FamilyDoctor.org	18 (9 sessional, 9 persistent, valid up to 5 years)	36 (all persistent, valid for up to 2 years)	15
MedHelp.org	8 (6 sessional, 2 persistent, valid for up to 15 years)	14 (all persistent, valid for up to 2 years)	10
Feminist.com	4 (1 sessional, 3 persistent, valid for up to 2 years)	13 (all persistent, valid for up to 6 months)	5

Table 5
French Language Websites Used in the Analysis of Privacy Policies

Website	Number of first party cookies	Number of third party cookies	Number of beacons
PasseportSante.net	8 (4 sessional, 4 persistent, valid for up to 2 years)	115 (most persistent, valid for up to 10 years)	51
Vulgaris-Medical.com	11 (2 sessional, 9 persistent, valid for up to 10 years)	262 (most persistent, valid for up to 10 years)	96
TopSante.com	15 (3 sessional, 12 persistent, valid for up to 2 years)	91 (most persistent, valid for up to 10 years)	66
InfoBebes.com	10 (1 sessional, 9 persistent, valid for up to 2 years)	218 (most persistent, valid for up to 30 years)	75
FmCoeur.qc.ca	21 (16 sessional, 5 persistent, valid up to 6 years)	46 (most persistent, valid for up to 14 years)	24
Cancer.ca	11 (5 sessional, 6 persistent, valid for up to 2 years)	55 (most persistent, valid for up to 5 years)	22

2.2 Results for English language websites

Each of the selected websites sets first-party cookies, both sessional (i.e. deleted when the browser is closed) and persistent (stored on a user's hard drive until its expiration date). The minimum number of first-party cookies observed was 4 (Feminist.com), and the maximum number was 18 (medicine.net and FamilyDoctor.org). Each website had at least one persistent first party cookie that lasted for 2 years or longer, and one site (MayoClinic.com) set a persistent first party cookie that lasted for 30 years. In addition, the seven selected health information sites set between 119 (WhatToExpect.com) and 13 (Feminist.com) third party cookies, and included between 40 (WhatToExpect.com) and 5 (Feminist.com) beacons. At least one advertiser (DoubleClick, AddThis, etc.) was included among the third parties present on each of the sites. Thus, we know that at every one of the selected sites users are subject to first party behavioural tracking as well as third party tracking by various entities including advertising agencies using both cookies and web beacons to monitor user behaviour.

2.2.1 Analysis of Privacy Policies

We analyzed the privacy policies of each of the seven websites for disclosure regarding behavioural tracking practices. One site (Feminist.com) had a very short 'privacy' policy (less than one page) that did not actually address any privacy issues. In our testing, this website showed the lowest level of behavioural tracking among the seven selected sites (see Table 1). Nonetheless, there was plenty to disclose, since this website does participate in third party behavioural tracking, setting first party cookies that persist for up to 2 years, and third party cookies that persist for up to six months. It is therefore notable that they make no attempt to disclose this behavioural tracking in their privacy policy. The remainder of this analysis addresses the six policies that included some discussion of privacy issues.

2.2.2 Notice of NPII Collection

We first examined each of the remaining privacy policies for explicit discussion of the collection of NPII, using the keywords 'collect', 'gather', or 'log'. Each of the six policies had some direct mention of first party collection of NPII, while five of the six policies explicitly discussed third party collection of this type of information.

In some cases, disclosures about first-party collection of NPII are explicit and easy to follow. FamilyDoctor.org (AAFP) has a particularly clear disclosure, identifying that they collect NPII, and telling the user what information this entails. This list appears early in the privacy policy under the heading 'What information does the AAFP collect?':

> The following information is collected from all visitors to AAFP Web sites and is recorded in a log file:
> - Time and date of the visit
> - The Internet address of the computer
> - The browser and operating system used
> - The page that is viewed
> - The previous page that was visited

Note the use of the third person ('What information does the AAFP collect?') and the passive voice; these linguistic strategies serve to reduce the perceived agency of the website with respect to NPII collection.

Among the remaining sites, WhatToExpect.com provides the clearest disclosure, but the indication of what information is collected is scattered throughout the privacy policy and is cast in conditional language. Under the heading 'Information we collect through your use of the Site', the privacy policy includes the following:

> As you use the Site and Services, certain information may also be passively collected. Through cookies, pixels, beacons, log files and other technologies, we may collect information about how you use the Site and the Services. For example we may determine through an IP address that a particular computer or device is located in New York City and we may use this information to deliver advertisements promoting New York City-based businesses.

The user is then directed to another part of the policy ('Cookies and Targeted Advertising') for further information. Under that heading, this text appears:

> We may … gather information regarding the date and time of your visit, the features and information for which you searched and viewed, the email you opened, or on which advertisements you clicked.

The other sites provide less detailed disclosure about this type of first party data collection. The Mayo clinic, for example, acknowledges that they 'collect and log the Internet Protocol address (IP) of all visitors to MayoClinic.com', following later with the information that they use cookies to 'provide us with information relating to the sources of our site traffic'. MedHelp.org similarly indicates that they collect non-personal information 'about your use of our website and your use of the Web sites of selected sponsors and advertisers'. MedicineNet.com

indicates that they 'may collect "Non-Personal Information"'— information that cannot be used to identify you', and later in the policy they state that they 'collect Non-Personal Information about your use of the WebMD Web Sites and your use of other web sites…'. HealthyWoman.org also explicitly acknowledges the collection of IP addresses, and they later acknowledge that they 'may collect' information about 'your use of the Website'.

Not surprisingly, explicit disclosures of third party NPII collection were less frequent and more limited, at least in part because (correctly) the sites indicate that they do not control the practices of the third parties that are active (with permission, obviously) on their websites. At the same time, it is critical to recall that the tracking measures we documented occurred in the process of regular browsing on the sites: in particular, we did not 'click through' on any advertisements or link to any outside sites. Thus, while the sites do not control the behaviour of the third parties with respect to the NPII that is collected, they certainly control the presence of those third parties on the website, and thus the ability of those third parties to collect personal information.

Five of the six policies make at least some mention of third party NPII collection. Two of these disclosures (MayoClinic.com and MedHelp.org) were quite detailed, providing the reader with a list of the NPII collected by third parties, including browser type, operating system, Web pages visited, time of visits, content viewed, ads viewed, and 'other clickstream data'; while MayoClinic.com indicates that third party advertisers collect this information, MedHelp.org notes only that they may collect it. In another section, however, their policy indicates that MedHelp.org 'receives' (from where or whom is not indicated) NPII, including

> your IP address, the URLs of sites from which you link to or leave our website, your type of browser and ISP.

Under the heading 'Information Collected by Third Party Advertisers', WhatToExpect.com includes the following:

> 'Advertisers or other third parties on the Sites may also engage in Behavioral Advertising and use cookies and web beacons in the manner described above.'

One must infer that the 'text above' refers to this passage, appearing earlier in the document under the heading 'targeted advertising':

> ‘These third party vendors may connect information about pages you visit on our Sites with information about pages you visit on other websites and show you advertising based on this combined information.’

Note, however, that the verb used here is ‘connect’, leaving open the question of what data are collected and by whom; moreover, conditional language is used once again to describe third party collection (‘may also engage’).

In the privacy policies of two sites (FamilyDoctor.org, MedicineNet.com) disclosure about third party NPII collection is limited to the assertion that collection is limited to NPII, or that there is no collection of PII by third parties (without explicit acknowledgement that NPII is collected by these third parties). No further details are provided in these cases.

We also examined the privacy policies for oblique disclosure of the collection of NPII (first or third-party). In this case, we were looking for text in the privacy policy that provided to the user an indication of the NPII that the website or third party had or used, without explicit discussion of the actual collection of that information. Thus, in some cases a user could infer first or third-party NPII collection through careful reading of the policy for these oblique references. Although MayoClinic.com explicitly identifies only the collection of IP address (see above), they acknowledge elsewhere in their policy the use (and therefore, necessarily, prior collection) of additional NPII including ‘traffic patterns’, ‘site usage’, and ‘length of stay’. MedicineNet.com acknowledges that they ‘statistically analyze user behaviour and activity including how frequently areas of the site are visited’: from this, the user can surmise that MedicineNet.com retains information about user visits that includes both the page(s) visited and the date of any visit. Similarly, HealthyWoman.org indicates that cookies enable them ‘to track site navigation, such as what sections users are visiting and how long they stay there’, while they explicitly acknowledge only the collection of IP address.

There were relatively few oblique references to NPII collection by third parties. MedHelp.org notes in their policy that “third-party advertisers can see the content of any page on the MedHelp website, with the exception of Personal Health Records”, indicating that advertisers ‘target ads based on the content of those pages but do not store any personally identifiable information.’ It is unclear from this passage whether NPII (including the page visited) is stored by the advertisers, but it is evident that this information is at the very least used for contextual advertising that is selected on the basis of the page the user is currently visiting.

WhatToExpect.com acknowledges that third parties 'may' use cookies to understand 'web usage patterns', but they leave it up to the user to infer the type of information that would be required to support this understanding.

2.2.3 Notice of Behavioural Tracking Mechanisms

Each of the website privacy policies provides a definition of the term 'cookie'; none, however, includes an explicit discussion of web beacons. FamilyDoctor.org offers the most comprehensive discussion:

> Cookies are a technology used by the AAFP to identify a user as the user moves through the AAFP Web sites. The user's browser allows the AAFP to place some information on the user's hard drive that identifies the computer utilized. Two types of cookies are commonly used. A session cookie is a temporary file stored in memory on the user's computer drive whenever a Web site is accessed and is terminated by closing the browser. A persistent cookie is a file stored on the user's hard drive that may be deleted manually by the user or expired by the Web server.

Three of the websites offer only a very brief definition, identifying cookies as 'small data files' or 'small pieces of information' that are stored or placed on the user's computer. In each of these three cases, a minimizing adjective is used to describe the information collected, suggesting that this information (and therefore any privacy risk it entails) is negligible.

HealthyWoman.org provides a more detailed description, one that is inconsistent with the description provided by the other sites:

> When you logon to the Website, a cookie is generated on the server, or the machine that hosts the site. The cookie is a randomly generated number that does not include any of your Personal Information. This randomly generated number or cookie remains on the server machine, not on your computer, until you leave the site. When you visit the Website again, a different, unique randomly generated number or cookie is assigned.

This description includes some misleading or even factually inaccurate statements (e.g., the 'cookie remains on the server machine'). Moreover, the emphasis on the 'random' nature of the cookie, paired with the assertion that the cookie 'does not include any of your Personal Information' suggests to the user that cookies have little if anything to do with them, yet nothing could be further from the truth.

All six of the privacy policies discuss the use of first party cookies for behavioural tracking, most identifying the use of this information as a basis for improving user experience on the website. This disclosure, on FamilyDoctor.org is typical (if a little more detailed than some):

> The AAFP uses cookies on areas of its Web sites to personalize a member's visit, to offer greater functionality, and to track visitor practices. The information generated from these cookies is used to help determine which services are most important and guide editorial decisions.

MayoClinic.com notes that their practice is 'like many websites', while MedHelp.org as well as HealthyWomen.org and WhatToExpect.com emphasize the benefits that users experience as a result of the use of cookies. MedicineNet.com is the only website that conditionalizes their disclosure regarding the use of first party cookies, noting that they 'may collect non-personal information… via cookies'; in another part of the policy, however, they indicate that they do collect information (about the use of the website) through cookies. Interestingly, WhatToExpect.com indicates that:

> As you use the Site and Services, certain information may also be passively collected.

This is a surprising (or perhaps inaccurate) use of language, since it is the provision of the information, and not the collection, that is passive.

Four of the six policies acknowledge the use of third party cookies and web beacons on their sites, although they use conditional language to describe these practices. These disclosures run from the minimalist (WhatToExpect.com):

> Advertisers or other third parties on the Sites may also engage in Behavioral Advertising and use cookies and web beacons in the manner described above.

to the relatively comprehensive (MayoClinic.com):

> .. third party network advertisers, along with other advertisers and sponsors on the website, may use cookies, Web beacons (also called single pixel GIFs or action tags) or similar technologies (and, in the case of cookies, access them on your computer if you choose to have cookies enabled in your browser) to serve you advertisements tailored to interests you have shown by browsing on this and other sites you have visited, to determine whether you have seen a particular advertisement before, to avoid sending you duplicated advertisements and to serve you advertisements on other sites. In doing so, the provider collects non-personal data such as your browser type, your operating system, Web

> pages visited, time of visits, content viewed, ads viewed and other clickstream data.

All four of these policies correctly identify that the collection and use of NPII by third parties is controlled by the privacy policy of the third party site. This statement, in the MayoClinic.com privacy policy, is typical:

> The use of third party cookies, Web beacons and similar technologies by these ad network providers is governed by each third party's specific privacy policy, not this one.

None of the sites, however, explicitly indicates to the user that this third party NPII collection occurs during simple browsing on the website, and does not require clickthrough on an ad or hyperlink to another website. Given that the text of the policies explicitly indicates that the third parties control the NPII that is collected, users might be forgiven for assuming that the data collection itself is activated if and only if the user interacts directly with that third party. It is also worth noting the use of conditional language to describe deployment of third party cookies and web beacons. Every policy that mentioned these techniques used the term 'may' to describe their use, even though the site itself allows the web beacons and third party cookies to operate, and indeed must have included the relevant code in their own web page. Thus, the conditional language serves only to obfuscate the actual practice on the sites.

Five of the six policies (the policy for MedHelp.com was the sole exception) offer information about opting out of first party cookies. In every case, there is an accompanying warning that opting out could reduce the website functionality and compromise browsing experience. Thus, for example, MayoClinic.com indicates "If you reject cookies from our site, some parts of the site may not work properly for you'. Three of these policies provide some (limited) information to users about how to reject cookies, directing users to the 'help' section in their browser toolbar. It is worth noting that none of the website privacy policies acknowledge that opt-outs limit only collection of NPII through traditional cookies, and thus do not affect web beacons or other newer mechanisms (e.g., flash cookies). Without this information, the policies could lead users to incorrectly assume that by refusing cookies they are stopping all NPII collection.

Four of the sites provide information on how to opt out of third party cookies, directing users to the privacy policy of the third party in question (e.g., Google, or DoubleClick) or to the

Network Advertising Initiative. Thus, the privacy policies typically send users to outside sites for additional information about data collection practices and information about user choices; only after receiving this information from an outside source can the user choose to opt out of third party tracking.

2.3 Results for French Language Websites

Each of the selected French language websites also sets first-party cookies, both sessional and persistent. The minimum number of first-party cookies observed was 8 (PasseportSante.net) and the maximum number was 21 (FmCoeur.qc.ca). Each website had at least one persistent first party cookie that lasted for 2 years or longer, and one site (Vulgaris-Medical.com) set a persistent first party cookie that lasted for 10 years. In addition, the six selected health information sites set between 46 (FmCoeur.qc.ca) and 262 (Vulgaris-Medical.com) third party cookies, and included between 22 (Cancer.ca) and 96 (Vulgaris-Medical.com) beacons. At least one advertiser (DoubleClick, AddThis, etc.) was included among the third parties present on each of the sites. Thus, we know that at every one of the selected sites users are subject to first party behavioural tracking as well as third party tracking by advertising agencies using both cookies and web beacons to monitor user behaviour.

2.3.1. Notice of NPII Collection

We first examined each of the privacy policies for explicit discussion of the collection of NPII. Only three of the six policies (FmCoeur.qc.ca, Vulgaris-Medical.com and TopSante.com) had some direct mention of first party collection of NPII and none explicitly discussed third party collection of this type of information. Somewhat surprisingly, no disclosure of third party NPII collection is made in any of the policies.

FmCoeur.qc.ca had the clearest disclosure of first-party NPII collection, indicating that the IP address, the internet provider, the time of the visit, the webpage from which the user accessed the website, the operating system and all the content seen on the website might be automatically recorded:

> Il est possible que les sites Web recueillent automatiquement certains renseignements non identificatoires au sujet de leurs utilisateurs, notamment l'adresse IP de leur ordinateur, l'adresse IP de leur fournisseur d'accès Internet, la date et l'heure à laquelle ils ont accédé aux sites Web, l'adresse URL du site à partir duquel ils se sont rendus directement aux

> sites Web, le système d'exploitation qu'ils utilisent, les sections du site Web qu'ils consultent, les pages des sites Web qu'ils ont lues et les images qu'ils ont vues, ainsi que les documents qu'ils affichent sur les sites Web ou qu'ils téléchargent à partir de ceux-ci. Les renseignements non identificatoires sont utilisés pour l'exploitation de nos services Web, le maintien de la qualité des services et la compilation de statistiques générales au sujet de l'utilisation de nos services Web.

Note the use of the third person ("les sites Web") and a conditional structure ("Il est possible que"); these linguistic strategies serve to reduce the perceived agency of the website with respect to first party NPII collection.

Vulgaris-Medical.com and TopSante.com are much more succinct in their disclosure of the collection of NPII. Vulgaris-Medical.com indicates only that the webpage from which the user accessed the website, the internet provider and the IP address are recorded (we cannot determine exactly what NPII the site collects):

> A l'occasion de l'utilisation du site www.vulgaris-medical.com, sont notamment recueillies les informations suivantes qui ne sont pas considérées comme personnelles (les « Informations Non Personnelles »):
> - l'adresse Internet URL des liens par l'intermédiaire desquels l'Utilisateur a accédé au site www.vulgaris-medical.com
> - le fournisseur d'accès de l'Utilisateur.
>
> A l'occasion de l'utilisation du site www.vulgaris-medical.com, est recueillie l'adresse de protocole Internet (IP) de l'Utilisateur l'information qui est considérée comme une Information Personnelle par une partie de la jurisprudence et par la CNIL.

Top-Sante.com indicates that "one cookie" is placed on the user computer, recording information that will be used in the future such as the pages visited and time of the visit, but they do not provide an explicit list of recorded information:

> topsante.com vous informe qu'un cookie est placé dans votre ordinateur lorsque vous naviguez sur son site. Un cookie ne nous permet pas de vous identifier. De manière générale, il enregistre des informations relatives à la navigation de votre ordinateur sur notre site (les pages que vous avez consultées, la date et l'heure de la consultation, etc.) que nous pourrons lire lors de vos visites ultérieures.

Although these sites do not use linguistic strategies to reduce the perceived agency of the website with respect to NPII collection, both sites provide only a partial list of the NPII that is collected (the lists are qualified by 'notamment' and etc.).

The three other websites provide less detailed disclosure about this type of first party data collection. PasseportSante.net, for instance, indicates that they may record "audience measures", such as number of pages visited and user activities, through cookies:

> Afin de vous assurer le meilleur service possible, nous pourrions être amenés à procéder à des mesures d'audience de notre site (nombre de pages vues, activité des visiteurs etc.) via la technologie des cookies.

The other websites, Cancer.ca and InfoBebes.com, only mention that they automatically collect "data" to improver their website without mentioning what information is collected.

No disclosure of third party NPII collection is made in any of the policies. Only two websites, PasseportSante.net and TopSante.com, mention that NPII could be linked to advertisement. PasseportSante.net indicates that cookies are used for ends linked to advertisement:

> Les cookies sont alors utilisés notamment pour des finalités liées à la publicité […]

Top-Sante.com indicates that NPII can be used to limit the number of times a user sees an advertisement (clearly a form of targeting):

> […] de limiter éventuellement le nombre de délivrance d'une même bannière publicitaire à un même utilisateur.

PasseportSante.net also mentions that it is possible to block third-party cookies, without mentioning them elsewhere in the policy, and thus not disclosing that these cookies are being set, how they are being used, etc.

2.3.2 Notice of Behavioural Tracking Mechanisms

Five of the privacy policies provide a definition of the term 'cookie'; none, however, mentions web beacons. Moreover, none of them provides a comprehensive discussion about cookies. Cancer.ca provides the most thorough definition:

> Un témoin est un petit fichier de données que le serveur d'une page Web transfère dans votre navigateur et qui ne peut être lu que par le serveur qui vous l'a transmis. Il s'agit en fait d'une « carte d'identité »; ce fichier n'est pas un programme et ne peut servir à exécuter un code informatique ou être porteur d'un virus. La Société n'analyse pas et ne rapporte pas sur les sessions personnelles d'un utilisateur, et ne redirige jamais les témoins vers d'autres serveurs.

While accurate, this definition does not mention the different types of cookies (session or permanent, first- and third-party, etc.). Passeport-Sante-net, Vulgaris-Medical.com and

FmCoeur.qc.ca offer only a very brief definition, identifying cookies as ‘small data files’ or ‘small pieces of information’ that are stored or placed on the user’s computer. In each of these three cases, a minimizing adjective is used to describe the information collected, suggesting that this information (and therefore any privacy risk it entails) is negligible. Interestingly, Top-sante.com indicates — inaccurately according to our results — that “one cookie” is set when navigating its website, while at the same time failing to define the term:

> topsante.com vous informe qu'un cookie est placé dans votre ordinateur lorsque vous naviguez sur son site.

Another inaccuracy can be found the policy of Passeport-Sante.net, where it is mentioned that cookies will only last for one year (data indicate at least one first-party cookie on this website was set to last for two years). Finally, InfoBebes.com does not mention cookies in its privacy policy.

Three of the six policies (Vulgaris-Merical.com, Top-Sante.com and Passeport-sante.net) offer information about rejecting cookies. Passeport-sante.net provides the most complete information, indicating that users can refuse all cookies or third-party cookies:

> Nous informons les internautes que les versions récentes des principaux navigateurs permettent non seulement de s'opposer à l'enregistrement de cookies mais également d'effectuer des sessions de navigation à l'issue desquelles tous les cookies installés lors de cette session sont automatiquement effacés indépendamment de leur durée de vie prévue, offrant ainsi une meilleur protection des traces.
>
> Ces derniers disposent encore d'outils permettant de gérer les cookies et notamment:
>
> - de bloquer les cookies issus de sites "tiers", c'est à dire ceux qui sont affichés par un autre site que celui qui affiche le contenu principal,
> - de créer des "listes noires" de sites pour lesquels il faut bloquer les cookies.

The two other websites only mention that it is possible to use the browser setting to manage cookies. It is worth noting that none of the website privacy policies acknowledge that opt-outs limit only collection of NPII through traditional cookies, and thus do not affect web beacons or other newer mechanisms (e.g., flash cookies). Without this information, the policies could lead users to incorrectly assume that by refusing cookies they are stopping all NPII collection.

Conclusion

In this chapter, we analyzed the disclosure, in two sets of website privacy policies, of the collection of non-personally identifiable information (NPII). The websites selected all engage in first and third party behavioural tracking using cookies and web beacons. For the English language websites, these sites represent a purposive sample of the recommended sites and they include government and commercial sources, encompassing sites with relatively low levels of tracking as well as those with much higher levels. For the French language sites, these sites represent the French language versions of two sites that appeared in the English recommended lists as well as the Google results plus four other websites purposively selected where third party trackers were observed. These include not-for-profit and commercial sources and encompass sites with relatively low levels of tracking as well as sites with much higher levels of tracking.

Our focus in this chapter was an analysis of the disclosure of these practices. Such disclosure is not generally required under regulatory frameworks stemming from *Fair Information Practice Principles* (such as PIPEDA) except insofar as this information is deemed to be personally identifiable. Nonetheless, disclosure of NPII collection falls within the spirit of the underlying guidelines (e.g., *Fair Information Practice Principles*), which are designed to protect user privacy in online spaces. Moreover, disclosure is required under self-regulatory principles (e.g., United States Federal Trade Commission, 2009), newer privacy regulations (e.g., the 2009 *EU directive 2009/136/EC*), and newer interpretations of existing guidelines (e.g., the Office of the Privacy Commissioner of Canada Policy Position on Online Behavioural Advertising). It is relevant to ask, therefore, whether privacy policies effectively disclose behavioural tracking practices.

The majority of the French and English language privacy policies we analysed (10 of 13) include at least some disclosure of first party behavioural tracking, but less than half of them (5 of 13) acknowledge behavioural tracking on their websites by third parties. The effectiveness of this disclosure furthermore is limited by the use of complex language, and passive and sometimes conditional grammatical constructions (see Pollach, 2005). The majority of the policies (11 of 13) also provide some information about behavioural tracking mechanisms, including the fact that users can opt-out of cookies (8 of 13). While opt-out information is important and in fact required in emerging regulatory frameworks (e.g., *EU Directive 2009/136/EC*), the direct pairing of opt-

out information with discussion of the negative consequences for user experience is likely to deter people from using this option to limit behavioural tracking.

Within the privacy policies we examined, disclosure of documented third-party tracking practices was limited, and the language used in the disclosures that did appear tended to be difficult to interpret. Thus, reading a privacy policy might not provide users with a full understanding of the behavioural tracking practices of the websites they visit, and as a result we need to consider additional mechanisms to identify and respond to behavioural tracking.

Chapter 3: Detecting, Mitigating, and Neutralizing Behavioural Tracking

The results of this research indicate that behavioural tracking on consumer health information websites is widespread, with the large majority of such sites including some form of third-party tracking, and approximately half of the sites participating in tracking by third-party advertisers, a practice that raises the most significant privacy concerns. Our analysis of website privacy policies with respect to disclosure of these tracking practices reveals that even those policies that acknowledge tracking do so in ways that make it difficult to determine the tracking practices that are in place. From the perspective of consumers, these issues are exacerbated by the fact that sites recommended by librarians and other information science professionals are also likely to engage in behavioural tracking. Government sites, and to a lesser extent not-for-profit sites, show a lower level of behavioural tracking compared to commercial sites. This difference, however, is restricted to third-party advertising trackers, and does not apply to trackers in general.

Increasingly, regulatory frameworks are being developed, or extended, to cover the collection and use of NPII, and internet users who are concerned about the collection and use of their personal information through behavioural tracking measures can launch complaints to relevant bodies, including the Office of the Privacy Commissioner of Canada and the Information Commissioner's Office in the United Kingdom. However, in the absence of direct investigation, triggered by such a complaint or launched by the organization responsible for regulatory enforcement, consumers cannot be assured that websites comply with applicable regulations. The situation is complicated by the fact that regulatory interpretation can be challenging. In Canada, for example, the Privacy Commissioner of Canada recently reported findings on the use of sensitive health information for targeting of Google ads[9], determining that this use contravened both PIPEDA and the organization's privacy policy. The complaint was determined to be well-founded, and in response Google undertook remedial measures that included period searches for the use of terms such as 'CPAP' or 'sleep apnea' within Google advertising products. This response, while effective for the particular complaint in question, does not provide general

[9] [1] https://www.priv.gc.ca/cf-dc/2014/2014_001_0114_e.asp

protection with respect to Google use of other information that users might consider sensitive: the use of behavioural tracking information on searches for 'epilepsy' or 'weight loss', for example, is not being monitored by Google. Regulation of the collection and use of NPII is important, and the actions of regulatory bodies in this respect are valuable and effective. At the same time, however, users cannot rely on these regulatory responses for privacy protection, and instead must become active participants in the preservation of their privacy online.

In considering responses to limit or even eliminate behavioural tracking, it must first be acknowledged that these practices provide some benefit to website users and thus it may not be their desire or even in their best interests to eliminate behavioural tracking. Behavioural tracking mechanisms enhance the experience of users in many ways and make the browsing experience more efficient. They are used to personalize websites, display information relevant to the geographic area where a user is located, remember registration details and content users have put in a shopping basket. Targeted advertising can also be positive for many people. Ultimately, it should be a user's choice to decide when he or she wants to be tracked. Different levels of actions are possible for users to control, on their own, when they are being tracked. Each of them, however, comes with a downside.

The easiest step is for users to learn how to manage HTTP cookies in every web browser that they use. Users can decide to refuse third-party cookies or even all cookies. The latter, however, will make the make the browsing experience much less efficient and may impede users from accessing some websites. Users should also learn how to delete cookies and think about emptying the cookie file of each of their browsers periodically. A more advanced and more complex step, yet crucial considering the capabilities of Flash cookies, is to learn how to manage Flash Cookies through the Adobe Website Storage Settings Panel. Browser extensions, such as Ghostery and AdBlock Plus[10], can be added to most browsers. Ghostery allows users to block trackers, either on a tracker-by-tracker basis, a site-by-site basis or a mixture of the two. Also customable, Adblock Plus allows users to block either all advertisements or only the ones they do not want to see. These extensions, however, may slow down Internet browsing.

[10] https://adblockplus.org/

Users can also change their Internet use habits. It is possible for user to use search engines that do dot store any NPII, such as Ixquick[11] and DuckDuckGo[12]. Ixquick returns the top ten results from multiple search engines. It only sets one cookie that remembers a user's search preferences and that is deleted after a user does not visit Ixquick for 90 days. DuckDuckGo, which returns the same search results for a given search term to all users, aims at getting information from the best sources rather than the most sources. While these search engines do not have all the functionality of the major search engines, both of them have received praise (e.g. McCracken, 2011). The ultimate solution, one that allows a user to navigate online total anonymity, is to use the Tor[13] web browser, which impedes network surveillance or traffic analysis and which the U.S. National Security Agency has characterized as "the King of high secure, low latency Internet anonymity" (Schneier, 2013). The anonymity afforded by Tor, however, comes at the price of reduced speed and limitations to available content.

[11] https://www.ixquick.com/
[12] https://duckduckgo.com/
[13] www.torproject.org/torbrowser/

Chapter 4: Dissemination and Knowledge Mobilization

4.1 Dissemination to the Academic Community

Dissemination and knowledge mobilization activities for this project were carried out in three domains: academic, professional, and general public. With respect to the first group, the results of this research were presented to academic audiences at a number of academic conferences (I^3: Information, Interactions and Impact, June 25-28, 2013, Aberdeen, Scotland; Association for Information Science and Technology: *Beyond the Cloud: Rethinking Information Boundaries*, November 1-5, 2014, Montréal, Québec), and future presentations are planned at several other conferences (Association French language pour le savoir, May 12-16, 2014, Montréal, Québec; Graphics, Animation, and New Media, Annual Conference, May 14-16, 2014, Ottawa, Ontario). The results of this research are in preparation for publication in academic journals. The first of these papers will examine the presence of tracking on consumer health websites, contrasting the results for sites recommended by library and information professionals with the results for sites returned by Google searches. The second of these papers will examine privacy policy disclosures of tracking mechanisms on consumer health websites. In future publications, we will examine tracking mechanisms and privacy policy disclosures on French language websites.

4.2 Dissemination to the Professional Community

An important aspect of dissemination involved communication of the research to professionals in the library and information science community. Our goals in this professional outreach are twofold: first, to improve the understanding of LIS professional regarding behavioural tracking and associated privacy issues in order that they can make privacy-respecting decisions and recommendations for themselves and their patrons; second, to provide background to LIS professionals to support their digital literacy outreach initiatives. We have presented at the Ontario Library Association Superconference (January 29 – February 1, Toronto, Ontario, and a presentation on this research has been accepted at the Canadian Health Libraries Association conference (June 16-20, 2014, Montréal, Québec). We are planning a webinar on behavioural

tracking mechanisms and strategies to manage tracking for the Ontario Library Association; this presentation will take place in the summer of 2014. Further opportunities to provide direct education to information science professionals are available to each of the investigators in their teaching activities. Dr. Burkell is a faculty member at the University of Western Ontario in the Faculty of Information and Media Studies, where she regularly teaches a course on Consumer Health Information, and she will integrate training on privacy and behavioural tracking in this and other relevant courses. Mr. Fortier is completing his PhD in Library and Information Science. He regularly teaches in the LIS program at UWO, and he will be seeking a faculty appointment once he has completed his PhD. He too will have the opportunity to incorporate education regarding privacy and behavioural tracking into his course syllabi. We are promoting the results of the research and the educational initiatives including the video (see below) produced as part of the project through a 'Focus on Research' profile on the Canadian Library Association website. Finally, we are preparing a paper on behavioural tracking mechanisms and strategies to manage these mechanisms for Library Quarterly, which is a journal directed to information professionals.

4.3 Dissemination to the Public

Finally, we have had opportunity for direct outreach to the general public. In partnership with the London Public Library, we provided public lectures on behavioural tracking (January 27, 2014) and social network privacy (February 24, 2014). In addition, we have provided links to our educational video and related publications for incorporation into library materials on online privacy.

4.3.1 Educational Video

Another important aspect of dissemination involved the production and promotion of an educational video on behavioural tracking mechanisms and responses. Copies of this educational video were provided (on USB keys) to attendees at the Ontario Library Association conference, and links to this educational material will be provided to the Canadian Library Association and the Ontario Library Association for inclusion in their educational materials regarding privacy and online behavioural tracking. This video was promoted at the public lecture offered at the London Public Library, and links to the online versions (French and English) of the video were provided to the London Public Library for use on their website and in their privacy-related educational materials.

The educational video was designed and produced for dissemination in presentations and online through websites (such as public library websites). The video uses animation and testimonials to create awareness of online tracking mechanisms with the following objectives:

- To describe different kinds of mechanisms exist and how they work.
- To present advantages and disadvantages these mechanisms may have for individuals in the online environment.
- To offer potential tools and protocols (through browsers or software) for managing these mechanisms.
- To present the implications of behavioural tracking from a privacy standpoint both on general and consumer health websites.

Expert interviews were conducted with Valerie Steeves (University of Ottawa, Department of Criminology), Avner Levin (Ryerson University, Ted Rogers School of Management, Director of the Privacy Institute), Andrew Clement (University of Toronto, Faculty of Information, Identity Privacy and Security Institute), and Jacquelyn Burkell (Western University, Faculty of Information and Media Studies). These experts described the implications of behavioural tracking and the broader picture of the state of privacy on the internet, including a need for public awareness and ultimately public debate. In addition, we interviewed a number of internet users regarding their understanding of and attitudes toward behavioural tracking. These interviews were coupled with animated sequences explaining behavioural tracking mechanisms and strategies for reducing behavioural tracking. Appendix V includes a copy of the final video scripts in both English and French.

Conclusion

Our investigation focused on consumer health information websites because, as we argue in the introduction to this report, online searches for health information reveal highly sensitive information. The privacy risks associated with tracking of health information seeking are, therefore, exacerbated, and particular care must be taken to ensure that privacy is protected in the health domain.

The results of this research demonstrate that third-party behavioural tracking is present on the large majority (4 out of 5) of consumer health websites, and over half of consumer health websites have trackers from third-party advertisers. This level of tracking is observed on French and English language websites, and on those recommended by Library and Information Science associations as well as those returned by a regular Google search. Government sites are essentially free from tracking by third-party advertisers, and overall commercial sites show much higher level of tracking (by advertisers and by third parties in general including analytics companies) than do government sites or sites from not-for-profit agencies. Even government sites, however, have high levels of tracking by non-advertising third parties including analytics companies, and although the privacy implications of this type of third-party tracking are reduced relative to advertising uses, privacy considerations remain (Mayer and Mitchell, 2012). Our analysis of privacy policy disclosure of tracking practices indicates that disclosure is inconsistent, incomplete, and difficult to understand. Thus, behavioural tracking is widespread on consumer health information sites, and privacy policies are not effectively informing consumers of website tracking practices.

This research provides a snapshot of behavioural tracking practices and disclosures on consumer health information websites at a particular point in time (Spring 2013 for the English language sites and Fall 2013 for the French language sites). We recognize that behavioural tracking is a moving and indeed expanding target, and we have every reason to believe that tracking practices on these sites will have changed since we completed our data collection: new tracking companies will have emerged, new uses for behavioural tracking data will have been identified, and new technologies will have been developed. The importance of our results is less

the specific number and identity of trackers, and more the demonstration that behavioural tracking is widespread on consumer health websites, that many of these websites allow third party advertisers to collect behavioural data on their sites, and that the disclosure of these data collection practices is insufficient to fully inform users.

Online privacy management is a digital literacy issue (Park, 2013), and users of consumer health information websites need to be supported in identifying privacy-respecting resources and making informed choices regarding their privacy. We consider Library and Information Science professionals to play a critical role both in directing consumers to privacy-respecting resources and in providing digital literacy education that will assist users of consumer health information websites (and other websites as well) to make informed choices regarding their online privacy. In order to address theses issues, Library and Information Science professionals must first become familiar with these tracking mechanisms, the risks they present, and the strategies (such as setting browsers to refuse cookies) that consumers can use to combat them. Second, they must monitor the behaviour tracking practices of the websites they recommend to consumers, identifying those websites that use behavioural tracking mechanisms and therefore present to users the privacy risks associated with profiling. Finally, Library and Information Science professionals should engage in digital literacy initiatives that promote an understanding, among the general public, of online behavioural tracking, including the related privacy risks and responses to mitigate these risks (Lankshear and Noble, 2008). The research results reported here and the activities undertaken to disseminate these results will assist in achieving these goals.

While regulatory frameworks serve to limit behavioural tracking and mandate disclosure of tracking practices, it is also important that users take measures to understand the privacy risks they encounter, and make informed decisions about their own online privacy. Ultimately, the Internet is a place where users must protect themselves, and we must provide them with the tools and information they need to do so.

Acknowledgements

The authors would like to thank the Office of the Privacy Commissioner of Canada for their support of this project through the Contributions Program. The opinions expressed in this report are those of the authors and do not necessarily reflect those of the Office of the Privacy Commissioner of Canada.

This project could not have been completed without the assistance of Ruby Lavallee, whose technical expertise was critical in documenting website tracking practices. Lola Wong developed the educational video, and her skills are evident in the quality of the finished product. Cliff Lonsdale contributed provided invaluable consultation regarding the content and organization of the video, and we are grateful for his expert assistance.

References

Anderson-Inman L. & M.A. Horney. (1998). Transforming text for at-risk readers. In D. Reinking, M. McKenna, L. Labbo and R. Kieffer (Eds.), *Handbook of literacy and technology: Transformations in a post-typographic world* (pp. 15–44). Mahwah, NJ: Lawrence Erlbaum Associates.

Angwin, J. (2010). The web's new gold mine: Your secrets. *Wall Street Journal Online*, 30 July.

Ayenson, M., D.J. Wambach, A. Soltani, N. Good & C.J. Hoofnagle. (2011). Flash cookies and privacy II: Now with HTML5 and ETag respawning. Retrieved April 29, 2013 from http://ssrn.com/abstract=1898390

Berger M, T.H. Wagner & L.C. Baker. (2005). Internet use and stigmatised illness. *Social Science and Medicine* 61(8), 1821–1827.

Burkell, J. & Fortier, A. (2012). Consumer Health Websites and Behavioural Tracking. *Annual Conference of the Canadian Association for Information Science*. Waterloo, ON: Wilfrid Laurier University, June May 31-June 2.

Burkell, J. & Fortier, A. (2013). Hidden Surveillance by Consumer Health Websites. *Information: Interactions and Impact*. Aberdeen, U.K.: Robert Gordon University, June 25-28.

Burkell, J., & Carey, R. (2011). Personal information and the public library: Compliance with Fair Information Practice Principles. *Canadian Journal of Information and Library Science* 35(1), 1-16.

Canadian Health Libraries Association. (2010). *Top 10 Canadian consumer health websites.* Retrieved April 29, 2013 from http://www.chla-absc.ca/chipig/Events/CHLA2010poster.pdf>.

Castelluccia, C. & Narayanan, A. (2012). *Privacy considerations of online behavioural tracking*. ENISA, Retrieved April 29, 2013 from http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/privacy-considerations-of-online-behavioural-tracking

Center for Digital Democracy, Consumer Federation of America, Consumers Union et al. (2009). *Online behavioural tracking and targeting concerns and solution. Legistlative Primer*. Retrieved April 29, 2013 from https://www.eff.org/sites/default/files/OnlinePrivacyLegPrimerSEPT09.pdf

Charles C, Gafni A, Whelan T. (1999). Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. *Social Science & Medicine* 49, 651–661.

Chester, J. (2012). Cookie wars: How new data profiling and targeting techniques threaten citizens and consumers in the 'Big Data' Era. In European Data Protection: In S. Gutwirth, R. Leenes, P. De Hert & Y. Poullet (Eds.), *Good Health?,* (pp. 53-77). Dordrecht (Netherlands): Springer.

Cline, R. J. & K. M. Haynes. (2001). Consumer health information seeking on the Internet: The state of the art. *Health Education Research* 16 (6), 671–692.

Coulter A. (1997). Partnerships with patients: the pros and cons of shared clinical decision making. *Journal of Health Services Research & Policy* 2, 112–121.

Earp, J.B., A.I. Antón, L. Aiman-Smith, & W.H. Stufflebeam. (2005). Examining Internet privacy policies within the context of user values. *IEEE Transactions on Engineering and Management* 52(2), 227-237.

Entwistle V.A. (2000). Supporting and resourcing patient participation in treatment decision-making: some policy considerations. *Health Expectations* 3: 77–85.

Fox S. & L. Rainie. (2002). *Vital decisions: How Internet users decide what information to trust when they or their loved ones are sick*. Washington D.C.: Pew Internet and American Life Project. Retrieved April 29, 2013 from http://www.pewinternet.org/Reports/2002/Vital-Decisions-A-Pew-Internet-Health-Report/Summary-of-Findings.aspx

Fox, S. (2011). *Health Topics: 80% of internet users look for health information online*. Washington, D.C.: Pew Research Center's Internet & American Life Project. Retrieved April 29, 2013 from http://pewinternet.org/Reports/2011/HealthTopics.aspx.

Hesse, B.W., D.E. Nelson, G.L. Kreps, R.T. Croyle, N.K. Arora, B.K. Rimer, K. Viswanath, (2005). Trust and sources of health information. *Archives of Internal Medicine*, 165, 2618-2624.

Holmes M., H.A. Llewellyn-Thomas, G.J. Elwyn. (In press). Moving to the mainstream. In: A.G.K. Edwards & G.J. Elwyn (Eds.). *Evidence-based patient choice*. Oxford: Oxford University Press.

Kosinski, M. , D. Stillwell, D., & T. Graepel, (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America, early edition*. Retrieved April 29, 2013 from http://www.pnas.org/content/early/2013/03/06/1218772110.short.

Lankshear, C., & M. Knobel (Eds.). (2008). *Digital literacies: Concepts, policies and practices*, 30. New York: Peter Lang.

Lunin, L.F. (1987). Where does the public get its health information? *Bulletin of the New York Academy of Medicine* 63: 923–38.

Marshall J.G. (1992). A development and evaluation model for a consumer health information service. *Canadian Journal of Information Science* 17, 1–16.

Marshall, J.G., C. Sewards, and E.L. Dilworth. (1991). Health information services in Ontario public libraries. *Canadian Library Journal*, 48(1), 37-44.

Mayer, J., and J. Mitchell. (2012). Third-party web tracking: Policy and technology. *IEEE Symposium on Security and Privacy* 2012, pp 413-427.

McCracken, H. (2011). The 50 best websites of 2011. *Time*, August 16.

McDonald, A.M. & L. F. Cranor. (2010). Beliefs and behaviors: Internet users' understanding of behavioral advertising. Retrieved April 29, 2013 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1989092

Medical Library Association. (2010). *CAPHIS top 100 list health websites you can trust*. Retrieved April 29, 2013 from http://caphis.mlanet.org/consumer/top100all.pdf

Medical Library Association. (n.d.). *A user's guide to finding and evaluating health information on the web*. Retrieved April 29, 2013 from http://www.mlanet.org/resources/userguide.html

MedlinePlus. (n.d.). *Guide to Healthy Web Surfing*. Retrieved April 29, 2013 from http://www.nlm.nih.gov/medlineplus/healthywebsurfing.html

Micheti, A., J. Burkell, & V. Steeves. (2010). Fixing broken doors: Strategies for drafting privacy policies young people can understand. *Bulletin of Science, Technology, and Society*. 30(2): 130-143.

Milne, G. R., & M. J. Culnan. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing, 18*(3), 15-29.

Murray, S. (2008). Consumer health information services in public libraries in Canada and the US. *Journal of the Canadian Health Libraries Associati*on, 29, 141-143.

Nass, S.J., L.A. Levit, & L.O. Gostin. (2009). The Value and Importance of Health Information Privacy. In S.J. Nass, L.A. Levit & L.O. Gostin (Eds.) *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*.Washington (DC): National Academies Press (US).

Norgren, A., (2013). Privacy by design in personal health monitoring. *Health Care Analytics*, Epub ahead of print, published online August 27, 2013.

Park, Y. (2013). Digital literacy and privacy behavior online. *Communication Research*, 40 (2), 215-236.

Pollach, I. (2005). A typology of communicative strategies in online privacy policies: Ethics, power, and informed consent. *Journal of Business Ethics* 62 (3), 221-235.

Rees A.M, ed. (1991). *Managing Consumer Health Information Services*. Phoenix, AZ: Oryx Press.

Rubenstein, E. (2012). From social hygiene to consumer health: Libraries, health information, and the American public from the late nineteenth century to the 1980s. *Library Information History* 28 (3), 202–219

Schneier, B. (2013). Attacking Tor: How the NSA targets users' online anonymity. *The Guardian*. 4 October.

Soltani, A., S. Canty, Q. Mayo, L. Thomas & C.J. Hoofnagle. (2009). *Flash cookies and privacy*. Available: Retrieved April 29, 2013 from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862

Statistics Canada (2011). *2010 Canadian Internet Use Survey*. Retrieved April 29, 2013 from http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4432&lang=en&db=imdb&adm=8&dis=2

United States Federal Trade Commission. (2009). F*TC Staff Report: Self-Regulatory Principles for Online Behavioral Advertising*. Retrieved April 29, 2013 from http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf, accessed April 26, 2012.

Vila, T., rR. Greenstadt & D. Molnar. (2003, September). Why we can't be bothered to read privacy policies models of privacy economics as a lemons market. In *Proceedings of the 5th international conference on Electronic commerce* (pp. 403-407). ACM.

Whetton, S. (2013). Personal health information, privacy and surveillance: Do we need a critical voice? *Studies in Health Technology and Informatics*, 192, 234-238.

Wood, F.B., B. Lyon, M.B. Schell, P. Kitendaugh, V.H. Cid, E.R. Siegel. (2000). Public library consumer health information pilot project: Results of a National Library of Medicine evaluation. *Bulletin of the Medical Library Association*, 88(4), 14-22.

Appendix I: Recommended Websites

http://www.intelihealth.com/

http://my.clevelandclinic.org/

http://familydoctor.org/

http://hardinmd.lib.uiowa.edu/

http://www.healthfinder.gov/

http://www.healthlinkplus.org/

http://www.mayoclinic.com/

http://www.medhelp.org/

http://www.medicinenet.com/

http://www.medlineplus.gov

http://www.netwellness.org/

http://www.noah-health.org

http://www.feminist.com/

http://www.hormone.org/

http://www.ihr.com/

http://www.mypelvichealth.org/

http://www.healthywomen.org

http://www.menopause.org

http://www.ourbodiesourselves.org

http://www.womenshealth.gov/

http://www.ahrq.gov/

http://www.cdc.gov/

http://www.urologyhealth.org/

http://www.aap.org

http://www.aacap.org/

http://www.drgreene.com

http://www.whattoexpect.com

http://www.kidshealth.org

http://www.nichd.nih.gov/health

http://www.virtualpediatrichospital.org

http://aarp.org/health/

http://www.aoa.gov/

http://www.agingcare.com/

http://www.healthinaging.org/

http://www.medicare.gov/

http://www.caregiver.org

http://www.firstgov.gov/Topics/Seniors.shtml/

http://www.gmhfonline.org/gmhf/

http://nihseniorhealth.gov/

http://www.alz.org/

http://www.aad.org/public

http://www.ada.org/

http://www.diabetes.org/

http://www.americanheart.org/

http://www.lungusa.org/

http://www.aafa.org/

http://www.cancer.gov/

http://digestive.niddk.nih.gov/index.htm/

http://www.nei.nih.gov/health/

http://www.nhlbi.nih.gov/

http://www.nichd.nih.gov/

http://www.nia.nih.gov/

http://www.niams.nih.gov/

http://www.nimh.nih.gov/health/topics/index.shtml/

http://www.ninds.nih.gov

http://www.stroke.org/

http://orthoinfo.aaos.org/

http://www.aidsinfo.nih.gov/DrugsNew/Default.aspx?MenuItem=Drugs&Search=On

http://www.centerwatch.com/patient/drugs/drugdirectories.html

http://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?id=2115

http://www.fda.gov

http://www.longwoodherbal.org

http://www.mskcc.org/cancer-care/integrative-medicine/about-herbs-botanicals-other-products

http://druginfo.nlm.nih.gov/drugportal/drugportal.jsp

http://www.needymeds.org

http://www.pdrhealth.com/home/home.aspx

http://www.rxlist.com/script/main/hp.asp

http://www.ama-assn.org/aps/amahg.htm

http://www.centerwatch.com/

http://ClinicalTrials.gov/

http://www.emedicinehealth.com/

http://ghr.nlm.nih.gov/

http://hpd.nlm.nih.gov/

http://www.quackwatch.com/

http://www.aboutkidshealth.ca/

http://www.caringforkids.cps.ca/

http://www.cancer.ca/

http://canadasafetycouncil.org/

http://www.dietitians.ca/

http://www.heartandstroke.com/

http://www.heretohelp.bc.ca/

http://sexualityandu.ca/

http://passeportsante.net/

http://www.womenshealthmatters.ca/

Appendix II: Websites Returned by Google (English)

http://allergicliving.com/
http://allergies.about.com/
http://allergyuk.org/
http://bodyandhealth.canada.com/
http://canadiansleepsociety.com/
http://chealth.canoe.ca/
http://curegoutpainnow.com/
http://diabetes.about.com/
http://diabetesshop.ca/
http://en.wikipedia.org/
http://findprivateclinics.ca/
http://forums.parentscanada.com/
http://health.nytimes.com/
http://heartdisease.about.com/
http://helpguide.org/
http://inspection.gc.ca/
http://lung.ca/
http://lupuscanada.org/
http://lupusuk.org.uk/
http://pollen.com/
http://ruk.ca/
http://skin-disorders.net/
http://sleepclinic.org/
http://sleepdisorders.about.com/
http://sleepmed.to/
http://sleepnet.com/
http://thechart.blogs.cnn.com/
http://thelupussite.com/
http://www.aftershingles.com/
http://www.agingincanada.ca/
http://www.allergyfoundation.ca/
http://www.arthritis.about.com
http://www.arthritis.ca/
http://www.bbc.co.uk/
http://www.bhf.org.uk/
http://www.canada.com/
http://www.canadiansleepsociety.com/
http://www.cbc.ca/
http://www.ctv.ca
http://www.dailyglow.com/
http://www.diabetes.com/
http://www.doctorq.ca/
http://www.everydayhealth.com/
http://www.fascrs.org/
http://www.gallbladderattack.com
http://www.gallbladderdetox.com/
http://www.gallbladdersymptomsz.com/
http://www.gout-aware.com/
http://www.gout.com/
http://www.hc-sc.gc.ca/
http://www.healthcommunities.com/
http://www.heart.org/
http://www.heartandstroke.com/
http://www.hemorrhoids.org/
http://www.hemorrhoidtreatment.info/
http://www.imhr.ca/
http://www.lupus.org/
http://www.lupusontario.org/
http://www.medicalnewstoday.com/
http://www.medicinenet.com
http://www.natcm.ca/
http://www.naturalskinrepair.com/
http://www.netdoctor.co.uk/
http://www.nhs.uk/
http://www.patient.co.uk/
http://www.pilex.com/
http://www.publichealthgreybruce.on.ca/
http://www.rheumatology.org/

http://www.rolingstone.com/
http://www.sages.org/
http://www.skincarecentre.ca/
http://www.skinsight.com/
http://www.sleepdisorders.com/
http://www.sleepdisorderssleepapnea.com/
http://www.sleepfoundation.org/
http://www.umm.edu/
http://www.washingtonpost.com/
http://www.webmd.com/
http://www.womenshealth.gov/
http://www.youtube.com/
http://yorkregionsleep.com/
http:/www.hemorrhoid.com/

Appendix III: Websites Returned by Google (French)

http://afriquinfos.com/
http://agora.qc.ca/
http://arthritisbroadcastnetwork.org/
http://asthme-allergies.org/
http://asthmeallergies.com/
http://ca.loccitane.com/
http://combattrelagoutte.ca/
http://cusm.ca/
http://dejouerlesallergies.com/
http://diabete.fr/
http://drdupied.com/
http://fr.healthierchoices.ca/
http://fr.wiktionary.org/
http://grmo.ca/
http://immunize.ca/
http://infotheque.muhc.ca/
http://jointhealth.org/
http://lecoeurtelquelles.ca/
http://lesexploitsducoeur.ca/
http://lyon-sud.univ-lyon1.fr/
http://muhcpatienteducation.ca/
http://naitreetgrandir.com/
http://pilule.telequebec.tv/
http://pourquoi-docteur.nouvelobs.com/
http://qualita.ca/
http://quebec.huffingtonpost.ca/
http://rire.ctreq.qc.ca/
http://sante-az.aufeminin.com/
http://sante-guerir.notrefamille.com/
http://sante-medecine.commentcamarche.net/
http://sante.canoe.ca/
http://sante.journaldesfemmes.com/
http://sante.lefigaro.fr/
http://santecapitalenationale.gouv.qc.ca/
http://santecheznous.ca/
http://santenature.over-blog.com/
http://selection.readersdigest.ca/
http://styledevie.ca.msn.com/
http://survivornet.ca/
http://tf1.fr/
http://tvanouvelles.ca/
http://www.aaia.ca/
http://www.accu-chek.fr/
http://www.afd.asso.fr/
http://www.alfediam.org/
http://www.allodocteurs.fr/
http://www.alzheimerestrie.com/
http://www.antiphlogistine.com/
http://www.anusol.ca/
http://www.aqaa.qc.ca/
http://www.arthrite.ca/
http://www.atoute.org/
http://www.avogel.ca/
http://www.babycenter.ca/
http://www.bayerdiabetes.ca/
http://www.bd.com/
http://www.brainpop.fr/
http://www.canalvie.com/

http://www.cancer.ca/
http://www.carenity.com/
http://www.carevox.fr/
http://www.ceed-diabete.org/
http://www.cfpc.ca/
http://www.chirurgie-digestive-bizet.com/
http://www.chu-sainte-justine.org/
http://www.chumontreal.qc.ca/
http://www.cihr-irsc.gc.ca/
http://www.clarisonic.ca/
http://www.cliniquealthea.com/
http://www.cliniquedrdanielbarolet.com/
http://www.cliniquevaccinationrivesud.com/
http://www.collabopm.com/
http://www.comprendrechoisir.com/
http://www.coupdepouce.com/
http://www.cpa.ca/
http://www.crc.chus.qc.ca/
http://www.crisedegoutte.com/
http://www.crisegoutte.com/
http://www.crulrg.ulaval.ca/
http://www.cssslaval.qc.ca/
http://www.curel.ca/
http://www.dermatonet.com/
http://www.dermatoveto.com/
http://www.diabete.qc.ca/
http://www.diabetelaval.qc.ca/
http://www.diabetes.ca/
http://www.docteurclic.com/
http://www.doctissimo.fr/
http://www.dolfino.tv/
http://www.douglas.qc.ca/
http://www.douleurchronique.org/
http://www.dumaisnd.qc.ca/
http://www.e-diabete.org/
http://www.e-sante.fr/
http://www.eatrightontario.ca/
http://www.educationnutrition.org/
http://www.elle.fr/
http://www.ellequebec.com/
http://www.enfant-encyclopedie.com/
http://www.entraidediabetique.org/
http://www.entrepatients.net/
http://www.esantementale.ca/
http://www.estelledaves.com
http://www.eurekasante.fr/
http://www.extenso.org/
http://www.familiprix.com/
http://www.femmeactuelle.fr/
http://www.femmesensante.ca/
http://www.fissureanale.com/
http://www.fmcoeur.qc.ca/
http://www.fmoq.org/
http://www.futura-sciences.com/
http://www.germannewmedicine.ca/
http://www.gralon.net/
http://www.groupeproxim.ca/
http://www.gsk.ca/
http://www.hc-sc.gc.ca/
http://www.healthycanadians.gc.ca/
http://www.hemoroidetraitement.com/
http://www.herbes-medicinales.ca/
http://www.inflammgen.org/
http://www.info-sante.info/
http://www.infobebes.com/
http://www.inserm.fr/

http://www.inspection.gc.ca/
http://www.inspq.qc.ca/
http://www.jeancoutu.com/
http://www.jetaide.com/
http://www.kidney.ca/
http://www.lapresse.ca/
http://www.laroche-posay.fr/
http://www.lemonde.fr/
http://www.linternaute.com/
http://www.liver.ca/
http://www.lupus-reference.info/
http://www.lupuscanada.org/
http://www.lupusreunion.com/
http://www.madmoizelle.com/
http://www.marlene-morin.ca/
http://www.maxisciences.com/
http://www.medecine-et-sante.com/
http://www.medscape.fr/
http://www.medtronic.fr/
http://www.medtronicdiabete.ca/
http://www.merial.ca/
http://www.moietcie.ca/
http://www.momes.net/
http://www.montreal-diabetes-research-center.org/
http://www.msss.gouv.qc.ca/
http://www.msss.gouv.qc.ca/
http://www.nospetitsmangeurs.org/
http://www.notretemps.com/
http://www.novartis.ca/
http://www.oiiq.org/
http://www.oncologik.fr/
http://www.ordrepsy.qc.ca/
http://www.orpha.net/
http://www.osrmedical.com/
http://www.ottawaheart.ca/
http://www.parlonsdiabete.com/
http://www.passeportsante.net
http://www.peau.net/
http://www.phac-aspc.gc.ca/
http://www.plaisirssante.ca/
http://www.pourlascience.fr/
http://www.pq.poumon.ca/
http://www.procto.ca/
http://www.protegez-vous.ca/
http://www.psychologies.com/
http://www.rhumatismes.net/
http://www.santedesfemmes.com/
http://www.santeprivee.ca/
http://www.santevoyagehorizon.com/
http://www.saveurs-sante.com/
http://www.savoirlaitier.ca/
http://www.sdhu.com/
http://www.securite-allergie.ca/
http://www.sfdiabete.org/
http://www.shepellfgi.com/
http://www.skinceuticals.fr/
http://www.skinpatientalliance.ca/
http://www.snfge.asso.fr/
http://www.sommeil-mg.net/
http://www.sos-hemorroides.fr/
http://www.specialisteschirurgie.ca/
http://www.topsante.com/
http://www.uniprix.com/
http://www.virtuogym.com/
http://www.vivre-mieux-naturellement.com/

http://www.vulgaris-medical.com/
http://www.who.int/
http://www.zostavax.ca/
http://www1.pharmaprix.ca/
http://yoopa.ca/
https://fr.wikipedia.org/
https://sommeil.univ-lyon1.fr/

Appendix IV: Trackers

[X+1]
24/7 Media*[14]
5min Media
Accuen Media
Acuity Ads
Acuityads.com
Acxiom
Ad Decisive
ad6media
AdAction
ADAOS
Adap.TV
Adblade
adBrite
Adconion
AddThis
AdForm
AdGear
Adify
Adition
AdJug
Adknowledge
Adloox
AdMeld*
Adnologies
Adobe Adlens
Adobe Digital Marketing
Adometry
AdoTube
Adroit Digital Solutions
AdRoll
AdScale
AdTech*
Advert Stream
Advertising.com
AdXpose
Adyoulike
Aggregate Knowledge
Akamai
Alenty
AlmondNet
Amazon Associates*
AMP Platform
AOL Advertising*
AOL OBA Notice
AppNexus*
areyouahuman
AT Internet
Atedre
Atlas
atmda.com
Audience Science
Aweber
Banner Connect
Bazaarvoice
BBElements
bidswitch.net
Bizo
Blink New Media
BlueCava
BlueKai
Bluelithium
BlueStreak
Boldchat
BrandScreen
Bright Tag
BrightCove
BrightRoll
Brilig
Burst Media
Buysight
BuzzFeed
C-Col.com
CanWest Global
Casale Media
Cedexis Radar
Centro
Chango
Chartbeat
ClearSaleinG
Clearsaleing
Clickbank
Clickbooth
ClickTale
Clicky

[14] The trackers followed by an asterisk are those identified as advertisers.

ClixGalore
Collective
Commission Junction
CompeteXL
Connexity
ContextIn
ConvertMedia
Convio
Conviva
CoreAudience
CPX Interactive
Crazy Egg
Criteo
Crowd Science
Dataium
Datalogix
DataXu
delvenetworks.com
Demandware Analytics
DemDex
DG Mediamind*
Didit Maestro
Didit Maestro
Digilant
Disqus
Dotomi
Dotsub
DoubleClick Bid Manager*
DoubleClick DART
DoubleClick Floodlight
Doubleclick Spotlight
DoubleClick*
DoubleVerify
Drawbridge
Dynamic Logic
EchoSearch
Effective Measure
Ensighten
EQ Advertising
eStat
EverydayHealth
Evidon Notice
eXelate Media
Experian Marketing Services
Exponential*
eXTReMe Tracker
eyeReturn Marketing
Eyeview
Ezakus
Facebook
Facebook Beacon
Facebook Connect
Facebook Conversion
Facebook Exchange (FBX)
Facebook Social Graph
Facebook Social Plugins
Flashtalking.com
FluidSurveys
FlyerTown
ForeSee
FreeWheel
FruitFlan
GDN Notice
Gigya
Gigya Socialize
Google +1
Google Adsense*
Google AdWords*
Google AJAX Search API
Google Analytics
Google Custom Search Engine
Google JSAPI Stats Collection
Google Tag Manager
Gravity.com
gsk.com
hearst.co.uk
Hello Bar
HIRO
Histats
Hit-Parade
i-Behavior
Impact-ad
Improve Digital
Impulse
InfoLinks
Infolinks
Integral Ad Science
InviteMedia
iPerceptions
Jumptap
Kaltura
Kenshoo
Kintera
KissInsights
Komli

Korrelate
Krux Digital
LeadBack
Legolas Media
Ligatus
Lijit
Lijit Networks
LinkedIn
Linksmart
LivePerson
LiveRail
LiveRamp
Lockerz Share
LongTail Video Analytics
Lotame
Lucid Media
Magnetic
MaxPoint Interactive
McAffee Secure
Medbroadcast
Media Innovation
Media Optimizer
Media6Degrees
MediaMath*
MediaMind
Mediaplex
Metrigo
Microsoft Atlas*
Mindset Media
MixPanel
MLN Advertising
Moat
Monetate
MyFonts Counter
Netmining
NetRatings
NetSeer
Neustar AdAdvisor
New Relic
news registry
NextAction
nRelate
Nugg.Ad
OMD (Omincom)
Omniture
Ooyala
Opentracker
OpenX*
Optimax Media Delivery
Optimizely
orlive.com
Outbrain
OwnerIQ
OwnerIQ
Parse.ly
Perfect Audience
Pingdom
Pinterest
Piwik Analytics
Piximedia
Platform161
PointRoll
Polldaddy
Prisma Media Digital
Proven Pixel
Public Ideas
PubMatic*
PulsePoint
Qualtrics
Quantcast*
Qubit
Questionmarket
RadiumOne
RapLeaf
Redux Media
Resonate Networks
Right Media*
Rivity
rnengage
Rocket Fuel
ROI trax
Rubicon*
Sas
saymedia
ScoreCard Research Beacon
scribblelive
Segmint
ShareThis
SimpleReach
Simpli.fi
SiteMeter
SiteScout
skimlinks*
skyword
SMART AdServer*
SocialReach
Soundcloud
SpecificClick

SpecificMEDIA
Spongecell
SpotXchange
Statcounter
SundaySky
surfing-waves.com
Surveymonkey.com
Switch Adserver
Taboola
Tacoda
Tapad
Targus Info
Tedemis
Telemetry
theGuardian
TidalTV
tinyurl
Topsy
TradeDesk
tradedoubler
Tradelab
Tremor Video
Tribal Fusion
Triggit
Trove
Truste Notice
Tube Mogul
TubeMogul
Tumblr Buttons
Turn*
Twitter Advertising
Twitter Badge
Twitter Button
Twitter.com
Tynt
Tynt Insigh
Typekit by Adobe
Typekit by Adobe
Undertone Networks
Unica
Userreport
UserVoice.com
ValueClick Mediaplex*
Veruta
Videoplaza
VigLink
Vimeo
Vindico
Visual Revenue
Visual Website Optimiser
Vizu
VoiceFive
WaterfrontMedia
WebMD
Weborama
WebTrends
WidgetBox
Wishabi
WordPress Stats
Wunderloop Connect
WysiStat
Yahoo Analytics
Yahoo! Ad Network
Yieldlab
Youtube.com
YuMe
Zanox
Zazzle
Zedo
Zenovia Digital Exchange

Appendix V: Scripts for the Dissemination Video

Speaker	English script	French script
S1 [LIVE]: VIDEO TEASER (VAL, AVNER, ANDREW) (about 17 seconds)		
Val	Because that kind of surveillance is invisible online, we kind of go to sleep.	Parce que ce genre de surveillance est invisible en ligne, c'est comme si nous étions endormis.
Avner	Unfortunately right now we don't have effective protection.	Malheureusement, nous n'avons pas en ce moment de protection efficace.
Andrew	We now have installed a major spying operation which has gone by completely without public debate.	Nous avons désormais installé une opération d'espionnage majeure sans que le moindre débat public ait eu lieu.
Text	They know where you've been, they know what you've done	On sait ce que vous regardez, on sait ce que vous aimez
	Online behavioural tracking	Le pistage comportemental en ligne
S2 V/O [ANIMATION]		
Narrator	When you look up medical information on the Internet, you probably know that some of it might be inaccurate or misleading. But there's another problem that's less obvious. It's called behavioural tracking, and its effects can be just as serious.	Vous savez peut-être que l'information relative à la santé sur internet peut être incomplète, biaisée, voire simplement erronée. Or, savez-vous qu'il faut aussi vous méfier d'un problème presque invisible dont les effets peuvent être tout aussi néfastes : le pistage comportemental.
Text	Online behavioural tracking	Le pistage comportemental en ligne
S3 V/O [ANIMATION]		
Narrator	Mechanisms for behavioural tracking include browser cookies as well as newer technologies such as flash cookies and web beacons.	Le pistage comportemental utilise l'agrégation de données non personnelles obtenues par des mécanismes tels les témoins et les pixels-espions.
Text	Behavioural tracking mechanisms	Les mécanismes de pistage comportemental
	Browser cookies	Témoins de navigateur
	Flash cookies	Témoins Flash
	Web beacons	Pixels-espions

Narrator	Browser cookies are small text files that are stored on your computer when you visit a website. They contain information such as your IP address, the pages you visit, the time of your visit, and your actions on the site.	Les témoins, aussi appelés « cookies », sont de petits fichiers sauvegardés sur votre ordinateur lorsque vous visitez un site web. Ils contiennent des renseignements tels que votre adresse IP, les pages web que vous visitez, le temps de votre visite et vos actions sur le site web.
	There are two types of cookies: First Party Cookies and Third Party Cookies.	Les témoins peuvent provenir du site web que vous visitez ou d'une tierce partie.
Text	Browser cookies	Témoins de navigateur
	IP address	Adresse IP
	Pages visited	Pages visitées
	Time spent	Temps passé
	Mouse clicks	Actions de l'utilisateur
	First-party cookies	Témoins provenant du site web visité
	Third-party cookies	Témoins provenant de tierces parties
Narrator	First Party Cookies are written and read by a website each time you visit. The information stored is typically used to personalize your experience on the site: for example, to show you new information related to a previous search. A history of your visits to the website can be assembled by using the information stored on these types of cookies. First Party Cookies written by one site cannot be accessed by other websites and do not generally present a privacy threat.	À chacune de vos visites sur un site web, des témoins provenant du site web lui-même peuvent enregistrer de l'information sur vos actions et cette information est normalement utilisée pour personnaliser votre expérience. Le site web peut se servir, par exemple, d'une recherche antérieure pour vous montrer des produits susceptibles de vous intéresser. Les témoins provenant directement du site web que vous visitez ne sont pas accessibles à d'autres sites et ne présentent d'ordinaire pas de menace à la protection de votre vie privée.
Text	First-Party Cookies (set directly by the website)	Témoins provenant du site web visité
	"Remember this" (Or use "Save this")	Sauvegarder ceci
	"You liked this"	Vous avez aimé ceci
	"Try this"	Essayez cela
Narrator	Sometimes, however, there are advertisers present on a site. They can write what are called Third-Party cookies. These same advertisers will be present on many other	Il peut arriver, par contre, qu'un site web contienne des témoins provenant de tierces parties, tel un annonceur publicitaire. Le même annonceur est

	sites that you visit and they have access to information about all of these visits through cookies that they write.	souvent présent sur plusieurs sites et possède donc un accès privilégié à vos comportements sur le web grâce à l'agrégation de données colligées d'un site web à l'autre.
	Third-Party cookies present a bigger privacy problem because as you surf the web, advertisers build a detailed profile on what you like, what you do, even what worries you by linking together the information from the many cookies that they write to your computer. Even though you aren't directly identified by this, a lot of information is being saved.	Les témoins provenant de tierces parties présentent une menace plus importante pour la protection de votre vie privée, car les profils qu'ils peuvent permettre d'assembler sont très précis.
Text	Sale	Solde
	Third-Party Cookies (set by ads on the website)	Témoins provenant de tierces parties
Narrator	Advertisers use more than just cookies such as newer technologies such as Flash and web beacons.	De nouvelles technologies, tels les témoins Flash et les pixels-espions sont aussi utilisés.
	Flash cookies are similar to HTTP cookies in many ways, but they are managed by Adobe Flash Player and they can 'respawn' or rewrite deleted browser cookies, creating 'zombie cookies' that can't easily be (deleted).	Les témoins Flash sont semblables aux témoins traditionnels, à la différence qu'ils sont gérés par le Flash Player d'Adobe. Les témoins Flash ont également la capacité de ressusciter les témoins que vous auriez effacés.
	Web Beacons are invisibly embedded in many web pages and emails. And they can also gather information from regular cookies at the same time.	Les pixels-espions sont des pixels invisibles insérés dans le code d'une page web. Ils sont capables de collecter le même type d'information que les témoins et sont également capables d'interagir avec eux.
Text	Flash cookies	Témoins Flash
	Web beacons	Pixels-espions
	Sale	Solde
	1 pixel x 1 pixel	1 pixel x 1 pixel
	IP address	Adresse IP
	Times & length of visit	Moment et durée de la visite
	User interaction	Actions de l'utilisateur
S4 V/O [ANIMATION]		

Narrator	In some cases behavioural tracking can be beneficial. Websites can use tracking mechanisms to personalize your experience, delivering the information you want the way you want it. Tracking also allows advertisers to provide ads of interest.	Il arrive que le pistage comportemental ait des effets positifs. Les sites web utilisent ces mécanismes pour personnaliser votre expérience. Ils permettent également aux annonceurs de vous montrer des publicités qui sont près de vos intérêts.
Text	First-Party Cookies (set directly by the website)	Témoins provenant du site web visité
	"Remember this" (Or use "Save this")	Sauvegarder ceci
	"You liked this"	Vous avez aimé ceci
	"Try this"	Essayez cela
	These might be of interest to you…	Vous pourriez être intéressé par ceci
S5 [LIVE]: STREETER CLIPS		
Streeter 1	The store, knowing what I want and being able to recommend and personalize it for me... I would see it more as a benefit than a detriment.	Qu'un magasin connaisse ce que je veux et soit capable de me faire des recommandations personnalisées… je vois cela plus comme un avantage que comme un inconvénient.
Streeter 2	Sometimes they can help you directly search when shopping...	Parfois, on vous aide à trouver ce que vous voulez.
Streeter 3	I just do it for the sake of quicker browsing...	Je le fais parce que c'est simplement plus efficace.
Streeter 4	I like the benefit of having things recommended to you, saving you time.	J'aime profiter des suggestions qu'on me fait, cela fait gagner du temps.
	If they're utilized well, then there's a lot of benefit, right?	Si elles sont bien utilisées, il y a plusieurs avantages, non?
	'Cause from my perspective, I'm going to see ads anyways, I might as well see ads that are appealing to me.	Je vais voir des publicités de toute façon, aussi bien voir des publicités susceptibles de m'intéresser.
S6 V/O [ANIMATION]		
Narrator	But what's gathered through tracking could be used to discriminate against people. Credit card companies, for example, could deny credit to individuals who have searched for credit counselling; insurers could raise rates for people who have searched for diabetes treatment. These are only a couple of examples that show how behavioural tracking could put	Il arrive par contre que l'information collectée serve à discriminer un utilisateur. Une compagnie de crédit, par exemple, pourrait traiter différemment quelqu'un ayant cherché de l'information sur le recouvrement de dettes. Un assureur pourrait offrir un prix différent à quelqu'un ayant cherché de l'information sur telle ou telle maladie. Ces deux

	you at a disadvantage: making services or products more expensive, or even resulting in denial of service.	exemples ne servent qu'à illustrer à quel point le pistage comportemental pourrait jouer contre vous sans que vous le sachiez.
Text	Online profile	Profil de l'utilisateur
	Credit denied	Demande de crédit refusée
	How much?	Combien?
	An additional 10% for your condition	Un supplément de 10 % en raison de votre maladie
	Behavioural tracking	Pistage comportemental
S7-1 [LIVE]: VAL (24 SECS)		
Val	This type of information really does constrain us because it's taken out of context.	Ce genre d'information nous contraint vraiment, car elle est prise hors contexte.
	When I'm talking to my friends online, I'm not talking to my mother. When I'm talking to my mother, I'm not talking to my employer.	Quand je m'adresse à des amis en ligne, je ne m'adresse pas à ma mère. Quand je m'adresse à ma mère, je ne m'adresse pas à mon employeur.
	And my behaviour online depends on who I'm interacting with. Once I lose the ability to keep all of those various lines between my roles in place, it becomes really uncomfortable for all of us.	Et mes comportements en ligne vont dépendre de la personne à qui je m'adresse. Si nous perdons la capacité de garder ces séparations en place, cela devient très inconfortable pour nous tous.
Text	Valerie Steeves	Valerie Steeves
	University of Ottawa	Université d'Ottawa
	Department of Criminology	Département de criminologie
S7-2 [LIVE]: JACQUIE		
Jacquie	The kind of targeting that we're talking about, the shaping of the world around you on the basis of information that you've essentially leaked in your practices online, has implications for the choices that you can make, it has implications for the options that are presented to you.	Le genre de ciblage dont nous parlons — la définition de notre environnement en fonction des informations que nous avons disséminé avec nos comportements en ligne — a des conséquences sur les choix que nous faisons. Il y a des conséquences sur les options qui nous sont présentées.
Text	Jacquelyn Burkell	Jacquelyn Burkell
	University of Western Ontario	University of Western Ontario
	Faculty of Information & Media Studies	Faculty of Information and Media Studies

S8 V/O [ANIMATION]		
Text	And there's not much you can do about it under current laws.	Et les lois actuelles ne nous protègent guère.
S9 [LIVE]: AVNER (22 SECS)		
Avner	Unfortunately right now we don't have effective protection. That's the sad state of where we are with all of our regulators, and we have a system of privacy commissioners.	Malheureusement, en ce moment, nous n'avons pas de protection efficace. C'est un triste état où nous sommes avec nos législateurs, même si nous avons un système de commissariats à la protection de la vie privée.
	That's not an effective way for somebody to get action in their favour if they're facing a problem. If they were, let's say, denied coverage by their insurance company, the privacy commissioner is not the answer to their woes.	Ce n'est pas un moyen efficace pour quelqu'un lorsqu'il est confronté à un problème. Si quelqu'un, par exemple, se voyait refuser d'être assuré, le commissariat à la protection de la vie privée n'y pourrait rien.
Text	Avner Levin	Avner Levin
	Ryerson University	Ryerson University
	Ted Rogers School of Management	Ted Rogers School of Management
	Director of the Privacy Institute	Directeur du Privacy Institute
S10 V/O [ANIMATION]		
Narrator	So what does it take for people to realize there need to be limits? The revelation of the US National Security Agency's wholesale monitoring of internet traffic served as a wake-up call to some.	Que devons-nous attendre pour réaliser que certaines limites ne doivent pas être franchies? Les révélations d'Edward Snowden sur les activités de la National Security Agency ont réveillé certains d'entre nous.
Text	NSA Revelations	Révélation sur les activités de la National Security Agency
	Online surveillance	Surveillance en ligne
S11 [LIVE]: VAL (22 SECS)		
Val	People might be uncomfortable if they think, oh, someone might be watching me, but because that kind of surveillance is invisible online, we kind of go to sleep.	Les gens peuvent être inconfortables à l'idée qu'on les surveille, mais parce que ce genre de surveillance est invisible en ligne, c'est comme si nous étions endormis.
	We kind of ignore it, because it isn't in our face. Typically what mobilizes public debate around these issues are these	C'est comme si nous l'ignorions, parce qu'elle n'est pas flagrante. Typiquement, ce qui mobilise le débat public, ce sont les

	incidents, or episodes, like the NSA revelations, where people kind of go 'Hey, wait a minute, that's not alright.	événements comme les révélations sur les activités de la National Security Agency. Les gens se réveillent alors et se rendent compte que ce n'est pas normal.
S12 [LIVE]: ANDREW (40 SECS)		
Andrew	Great expansion of the state surveillance, often in collaboration or collusion with the large corporate internet providers is a challenge to our democratic system of government.	Une grande expansion de la surveillance, souvent en collaboration ou en collusion avec les grands joueurs internet, est un défi pour nos systèmes démocratiques.
	We now have installed a major spying operation which has gone by completely without public debate.	Nous avons désormais installé une opération d'espionnage majeure sans que le moindre débat public ait eu lieu.
	If you think that's ok, then I'm afraid that you have problems with democracy, because democracy would insist that it's the individual rights that take paramount, and that the state needs to be accountable to the citizens, not the other way around.	Si vous croyez que c'est correct, j'ai peur que vous ayez un problème avec la démocratie, car ce sont les droits individuels qui priment en démocratie. C'est l'État qui doit rendre des comptes aux citoyens et non le contraire.
Text	Andrew Clement	Andrew Clement
	University of Toronto	University of Toronto
	Faculty of Information	Faculty of Information
	Identity Privacy and Security Institute	Identity Privacy and Security Institute
S13 [LIVE]: JACQUIE		
Jacquie	When you're health information online, you can be seeking information about really sensitive issues. And often it's not even about you - you can be seeking information about somebody else.	Quand nous cherchons de l'information relative à la santé en ligne, nous pouvons chercher de l'information hautement privée. Et, souvent, ce n'est même pas à propos de nous.
	But the profile that you're developing says an awful lot potentially about who you are, about what you're concerned about, about what you're doing, about what your health is like.	Mais le profil que nous développons ce faisant peut en dire long sur nous, sur ce qui nous concerne, sur ce que nous faisons, sur notre état de santé.
	And that information can be used for the kind of social sorting that many people have talked about that has financial and professional and other kinds of implications.	Et cette information peut être utilisée pour stratifier les gens, une pratique qui peut avoir des implications financières, professionnelles ou autre.
S14 V/O [ANIMATION]		

Narrator	Advertisers are always coming up with new ways to collect data about you and if you want to protect your privacy, you need to keep up to date on mechanisms and be aware of how to manage them. What can we do to limit what advertisers know about us?	Vos comportements sur internet sont une mine d'or pour les annonceurs et ceux-ci ne renonceront pas à l'exploiter. Si vous désirez protéger votre vie privée, vous devez vous tenir au courant des différents mécanismes de pistage comportemental et de la manière avec laquelle vous pouvez les gérer.
Text	Sale	Solde
tS15-1 [LIVE]: STREETER 5 (YOUNG MAN) (22 SECS)		
Streeter 5	There are various tools and plug-ins on your browser that you can use to wipe away your digital trail, to mask your online usage, so that various companies don't know all these things about you.	Il y a plusieurs outils et modules d'extension qu'on peut ajouter à nos navigateurs qui peuvent effacer notre empreinte en ligne de manière à ce qu'on ne puisse pas en savoir autant sur nous.
	There are often negative consequences – you might not get cool services, but that is a decision an individual should be able to make.	Il y a souvent des conséquences négatives, comme la perte de fonctionnalité. Mais il s'agit d'une décision qu'un individu devrait pouvoir prendre.
S15-2 [LIVE]: JACQUIE		
Jacquie	It's a matter of being aware, of thinking about this as a possibility, of knowing the world in which you're operating and taking the steps you need to take in order to minimize your risk.	Il s'agit d'être au courant, de penser à cette possibilité, de connaître le monde dans lequel on opère et de prendre les mesures nécessaires pour minimiser les risques.
S16 V/O [ANIMATION]		
Narrator	Here's a few things that you can do to protect yourself from behavioural tracking – always remembering there's a trade-off between convenience and security.	Voici quelques gestes que vous pouvez faire pour vous protéger contre les conséquences du pistage comportemental. Prenez note, par contre, qu'il y a un équilibre à atteindre entre la facilité d'utilisation d'un site web et la protection de votre vie privée.
Text	Convenience	Facilité d'utilisation
	Security	Protection de la vie privée
Narrator	READ: Protect yourself by knowing what might be collected about you and how it might be used by reading privacy policies and terms of service. You should know, however, that	Sachez quelles informations peuvent être collectées et comment elles peuvent être utilisées en lisant les politiques de protection de la vie privée des sites web que vous visitez.

	disclosure of behavioural tracking is not required by regulation, so if there is no disclosure you can't be absolutely sure that there is no tracking.	Prenez note, par contre, que la réglementation actuelle n'oblige pas les sites web à divulguer le pistage comportemental.
Text	READ: Privacy policies & Terms of Use	Lisez les politiques de confidentialité
	Disclosure not required	Divulgation non obligatoire
	Privacy policy	Politique de confidentialité
	"Nothing here about tracking"	Rien ici sur le pistage
Narrator	DELETE/SET: Periodically delete cookies in all the browsers that you use OR set your browsers to refuse them.	Effacez régulièrement les témoins dans chacun de vos navigateurs ou réglez vos préférences pour refuser les témoins ou les témoins provenant de tierces parties.
Text	DELETE: cookies periodically in all of your browsers.	Effacez régulièrement les témoins dans chacun de vos navigateurs
	SET: browsers to refuse cookies	Réglez vos préférences pour refuser les témoins ou les témoins provenant de tierces parties
Narrator	BLOCK: Install applications such as Ghostery that identify web beacons and allows you to block them.	Installez une application comme Ghostery qui identifie les pixels-espions et vous permet de les bloquer.
Text	BLOCK: Install software that will block web beacons.	Installez une application qui vous permet de bloquer les pixels-espions
Narrator	MANAGE: flash cookies by changing Adobe Flash Player settings.	Gérez les témoins Flash en apprenant à ajuster les paramètres du Flash Player d'Adobe.
Text	MANAGE: flash cookies by changing Adobe Flash Player settings.	Gérez les témoins Flash en apprenant à ajuster les paramètres du Flash Player d'Adobe
Narrator	There are many resources online that will help you keep on top of behavioural tracking.	Plusieurs ressources sont disponibles sur internet pour vous aider à gérer les mécanismes de pistage comportemental.
	Most of all, being aware that these mechanisms exist, understanding what information is collected and how it might be used is the first step to protecting your privacy online.	Être au courant de ces mécanismes, comprendre quelle information est collectée et comment elle peut être utilisée est, en somme, la première étape pour protéger votre vie privée sur internet.