Privacy Policy Disclosures of Behavioural Tracking on Consumer Health Websites

Jacquelyn Burkell
Faculty of Information and Media Studies
The University of Western Ontario
London, Ontario, Canada N6A 5B7
jburkell@uwo.ca

Alexandre Fortier
Faculty of Information and Media Studies
The University of Western Ontario
London, Ontario, Canada N6A 5B7
afortie@uwo.ca

ABSTRACT
Many Internet users are seeking health information online, encountering significant privacy risks in the process. Historically, these risks are associated with personally identifiable information, but behavioural tracking presents a new and increasing threat to privacy. In this paper, we analyze the disclosure, in a set of website privacy policies, of the collection of non-personally identifiable information by consumer health information websites. The websites all engage in first and third party behavioural tracking using cookies and web beacons, and are among the sites recommended by consumer health sections of the Medical Library Association or the Canadian Health Libraries Association (see Burkell and Fortier, 2012, 2013). Our analysis reveals that while the majority of these sites disclose both first party (6/7) and third party (5/7) behavioural tracking, the language used in these disclosures is difficult to understand, tending to minimize behavioural tracking and obfuscate agency in the tracking process. These results suggest that consumer health information website privacy policies do not provide optimal disclosure of behavioural tracking practices. Library and information science professionals should work with users to ensure they are aware of the behavioural tracking practices of the websites they visit, assisting them to interpret the disclosures provided in website privacy policies.

Keywords
Privacy, consumer health information, behavioural tracking, non-personally identifiable information, privacy policies

INTRODUCTION
People have long been interested in health information, and libraries have long been involved in providing such information (Rubenstein, 2012). Increasingly, consumers are seeking health information online (Fox, 2011; Statistics Canada, 2011), and this has brought about attendant changes in the role of information professionals. One issue that is increasingly significant for information professionals is 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. Our focus in this paper is on one particular risk of increasing importance: the privacy risk associated with behavioural tracking measures used on health information websites. This paper forms part of a larger project that investigates behavioural tracking on these sites (see Burkell and Fortier, 2012, 2013). In this report, we focus on a set of health information websites, recommended by LIS professionals, that engage in third-party behavioural tracking, focusing on the degree to which behavioural tracking practices are reflected in the website privacy policies. Our question is simple: Do these website privacy policies effectively inform of behavioural tracking practices?

BACKGROUND
Information professionals have developed guidelines to help consumers identify high quality information (Fox and Rainie, 2002; Medical Library Association, n.d.). Information professionals are also aware of the privacy risks associated with the collection and use of personal information in the online context (see, e.g., American Library Association Privacy Resources for Librarians, Library Users, and Families). Historically, these privacy concerns have focused on the collection, use, and retention of the identifying information (PII) that is explicitly provided by users who are accessing online information (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-Personal Information (NPII) that cannot be associated with a specific identifiable individual,
including IP address, browser configuration information, and details of browsing behavior (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 strategies 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 this 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). These threats 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 OECD 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). One important aspect of these regulatory frameworks is the requirement for notice: users should be given notice of website information practices before any data is collected from them. This notice is typically provided in privacy policies that identify what information is collected, how it is used, and with whom it is shared.

It is important to note that regulatory frameworks did not originally contemplate the collection of NPII, and thus did not apply to the collection of this type of personal information. Regulatory bodies have noted this problem, and are beginning to respond. Thus, for example, the European Data Protection Directive 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), and under PIPEDA coverage has been extended to at least some NPII through the argument that this information can be personally identifying, and thus is subject to PIPEDA regulation (e.g., IP addresses can be linked to an individual). Moreover, although there is no 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 personally identifiable information and NPII becomes “less and less meaningful and should not, by itself, determine the protections provided for consumer data.” (p. 21-22). Thus, there seems to be general agreement that users should be informed of behavioural tracking measures active on the websites they visit, and librarians in particular should ensure that these practices are acknowledged in the privacy policies of websites they recommend. These measures are particularly important with respect to behavioural tracking, since many Internet users are unaware of the practices, and even those who are familiar with the issue may not fully understand the range and power of behavioural tracking mechanisms (McDonald & Cranor, 2010).

The agreement that users ‘should’ be informed of behavioural tracking does not, especially in the absence of enforceable regulation, necessarily imply that notice is provided. Moreover, 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.

This paper explores whether and how the collection of NPII is disclosed in the privacy policies of consumer health websites, recommended by information professionals, that engage in behavioural tracking. This forms part of a larger project that undertakes an inquiry into the behavioural tracking practices of health websites that consumers are
likely to encounter. In this analysis, we focus on two questions:

1) Are users of these websites informed of the collection of NPII, and in particular, of the specific information that is being collected?

2) Are users of these websites informed of the technologies, present on these websites, that are being used to collect this information?

METHODOLOGY
The seven websites selected for this study are among the 88 consumer health information websites recommended by library associations including the Canadian Health Libraries Association (2010), and the Medical Library Association (2010; for additional details, see Burkell and Fortier, 2012, 2013). These sites represent a purposive sample of the recommended sites: third party trackers were observed on all sites, and the set included government and commercial sources, and 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).

Each of the selected sites was verified by the researchers to be engaged in NPII collection using first party cookies, third party cookies, and web beacons (see Burkell and Fortier, 2013 for details of the methods that were used to identify NPII collection). It is important to note that the reported cookies and web beacons were encountered in the course of regular browsing on the publicly available parts of each website: we did not register for any services, provide information to access restricted or ‘member only’ parts of the website, follow any links to outside sites, or click through on any advertisements.

The privacy policies of the selected sites 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’) that downplay frequency;
- conditional verbs (e.g., ‘may’) that introduce uncertainty;
- nominalizations (e.g., ‘the collection of data’) and the passive voice (e.g., ‘data are collected’) that obscure agency.

RESULTS
NPII Collection Practices
All of the web sites selected for this study use first party cookies, third-party cookies, and web beacons or bugs (see Table 1).

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

Table 1. Behavioural Tracking Mechanisms on the Selected Websites

First party cookies are set and read by the site itself, and are used to track web site visits, usually in order to personalize the online experience. These cookies are of two types: session cookies and persistent cookies. Session cookies last for the length of the session only, while persistent cookies remain active for some pre-determined length of time after the website visit. First party cookies support tracking of the user as they move through the site and potentially across multiple visits (for persistent cookies), but first party
Third party cookies are set and read by an outside party (typically an advertiser such as DoubleClick or a web analytics company such as Google Analytics). Not surprisingly, third parties typically set persistent cookies, since these support integration of user data across multiple visits and across multiple websites and thus allow for more detailed user profiling. We did, however, also observe a small number of session cookies (valid for the session only, used for contextual advertising) set by third parties. Third party cookies typically track the following information:

- the date and time of website visit;
- IP address;
- the type of browser;
- the web page address;
- advertisement clickthrough (or not).

Web beacons appearing on web pages (also called ‘web bugs’, or ‘invisible gifs’) send information to an outside party (typically an advertiser or web analytics company). The information transmitted by web beacons can include the identity of the page where the beacon was located, and this can be augmented by additional information such as whether an ad has been viewed, IP address, and the time of access. This information can be aggregated (through matching on IP address) with information collected by other web beacons used by the same organization, thereby allowing tracking of user behaviour across sessions and across sites.

Each of the selected websites sets first-party cookies, both sessional and persistent. 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 advertising agencies using both cookies and web beacons to monitor user behaviour (see Burkell and Fortier (2013) for additional details on tracking behaviour).

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 behaviour tracking in their privacy policy. The remainder of this analysis addresses the six policies that included some discussion of privacy issues.

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' (emphasis added) '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 use 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.

**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) emphasizes the benefits that users experience as a result of the use of cookies. MedicineNet.com is the only website that conditionalizes 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 this 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.**

... 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.

CONCLUSION/DISCUSSION
In this paper, we analyze the disclosure, in a set 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, and are among the sites recommended by CAPHIS and/or MLA (see Burkell and Fortier, 2012, 2013). Our focus in this paper, however, is not a detailed discussion of NPII collection practices, but rather 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), and newer privacy regulations (e.g., the 2009 EU directive 2009/136/EC). It is relevant to ask, therefore, whether privacy policies effectively disclose behavioural tracking practices.

The majority of the privacy policies we analysed (6 of 7) include at least some disclosure of first party behavioural tracking, and only slightly fewer (5 of 7) acknowledge behavioural tracking on their websites by third parties. The effectiveness of this disclosure, however, is limited by the use of complex language, and passive and sometimes conditional grammatical constructions (see Pollach, 2005). The majority of the policies also provide some information about behavioural tracking mechanisms, including the fact that users can opt-out of cookies. 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.

Behavioural tracking is here to stay, and as time passes and technology changes, it becomes ever more sophisticated. With respect to these practices and the privacy risks they present, library and information professionals face a variety of challenges. First, we must ourselves 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, we must monitor the behaviour tracking practices of the websites we recommend to consumers, identifying those websites that use behavioural tracking mechanisms and therefore present to users the privacy risks associated with profiling. This research examines the efficacy of website privacy policies as a means of understanding behavioural tracking practices, revealing yet another challenge for library and information science professionals: while we must read these policies carefully for indication of behavioural tracking, we cannot assume that these policies effective disclose the tracking practices of the websites to which they apply. Ultimately, the Internet is a place where the user must protect themselves, and understanding the practices of disclosure in website privacy policies is one way to achieve that protection.

ACKNOWLEDGEMENTS
This project was funded by the Office of the Privacy Commissioner of Canada for funding through the Contributions Program. The authors would like to thank Ruby Lavallee for her assistance with this research.

REFERENCES


