Abstract: This project examines behavioural tracking on consumer health websites recommended by LIS practitioners and returned by Google searches for the most commonly searched conditions. Analysis shows that over three quarters of the websites in these groups employ tracking technologies, potentially aggregating information across websites and allowing the assembly of detailed user profiles.

Résumé : Ce projet examine le suivi comportemental par rapport aux sites Web sur la santé des consommateurs recommandés par les professionnels de la BSI et retournés lors d’une recherche sur Google, pour les problèmes de santé les plus couramment recherchés. L’analyse montre que plus de la moitié des sites .com dans ces groupes utilisent des techniques de suivi pour recueillir de l’information à partir des sites Web et ainsi assembler un profil détaillé de l’utilisateur.

Introduction
Consumers who access health information online are advised by LIS professionals to evaluate online health information with respect to factors such as intended audience, sponsorship, and information accuracy, currency, and completeness. They are also advised to consider the privacy practices of the sites that they visit, and pay attention to whether the site includes advertisements (Medical Library Association. A user's guide to finding and evaluating health information on the web). Typically, guidelines suggest that consumers ensure that they understand website practices with respect to the collection and disclosure of personally identifying information, while concerns about advertising are usually tied to the balance, coverage, and objectivity of the information delivered on the site.

Recently, however, privacy advocates including the Privacy Commissioner of Canada have focused on a new concern: that of behavioural tracking (Office of the Privacy Commissioner of Canada, 2011). Behavioural tracking techniques allow website providers and advertisers to develop sophisticated and detailed user profiles without the collection of personally identifying information. First party behavioural tracking techniques are used by the websites themselves to track the behaviour of website users, keeping a record of actions such as the pages they visit and the time they spend on each. These records, if stored as cookies on the users’ computer, can persist across website visits. In some cases third parties, such as companies that serve advertisements on multiple sites, set tracking cookies and collect behavioural data. The privacy risks associated with third party behavioural tracking are more significant, because these data can be aggregated across the multiple websites on which the third party is active (DoubleClick, for example, serves ads on many different sites, and can aggregate the behavioural tracking data they collect from all of these sites to create a more detailed user profile).

Many uses of behavioural tracking data are relatively innocuous, including user analytics to improve website functionality and personalization of website experience. Behavioural tracking data can also be used to serve targeted ads to users, and third-party companies can target ads to the same user on multiple sites. The detailed profile that results from
behavioural tracking is also of potential interest to employers, insurers, and providers of financial services — in fact, to anyone who would benefit from segmenting Internet users according to their online behaviour.

In the realm of health information, the potential applications are legion, and many raise significant privacy issues. Online health information searches can reveal health care concerns and interests, potentially before these are even discussed with a health care provider. While in some cases the potential consequences are positive (e.g., website personalization, or targeted advertisements for treatments of interest to the user), this information could also be used to segment consumers for price discrimination or even denial of service (e.g., insurance applications; Center for Digital Democracy et al., 2009).

The current research project examines online consumer health information sources recommended by LIS practitioners and returned by Google searches to identify the prevalence of behavioural tracking on these sites. At this point, our analysis does not address the uses of this information, which are often difficult to determine. Instead, our goal here is to provide a descriptive overview of the behavioural tracking encountered by consumers who access health information online.

**Methods**

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); 2) the consolidated set of websites returned on the first page of Google searches of the ten most commonly searched conditions, as identified by the Pew Research Center’s Internet & American Life Project (Fox 2011). 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. We used the application *Ghostery* (www.ghostery.com) to identify the ad networks, behavioural data providers, web publishers, and other companies that use invisible technologies (web beacons, web bugs, etc.) to track user behaviour on the site. We visited each site, noting the tracking companies that were identified by *Ghostery* on the site. Privacy policies, where available, were retrieved and examined for disclosure of these tracking technologies.

**Results**

The recommended lists from CAPHIS and CHIPIG yielded a total of 85 distinct recommended sites for health information. The first pages of the Google searches for the ten most commonly searched conditions yielded a total of 81 websites that did not appear on the recommended lists: one of these links provided by Google was no longer active, leaving 80 sites in the ‘Google only’ set. We examined each of these sites to identify whether they collect behavioural tracking data.

<table>
<thead>
<tr>
<th>Cookie/Beacon</th>
<th>Number in Recommended Sites (% of sites)</th>
<th>Number in ‘Google Only’ sites (% of sites)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Third Party Advertising</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DoubleClick</td>
<td>8 (10%)</td>
<td>19 (23%)</td>
</tr>
<tr>
<td>Google Adsense</td>
<td>6 (7%)</td>
<td>17 (21%)</td>
</tr>
<tr>
<td><strong>Analytics</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The presence of behavioural tracking was identified in large the majority of websites. Among the 84 recommended websites, 71 (85%) were associated with at least one tracking company. Among the 80 active ‘Google only’ sites, 72 (90%) were associated with at least one tracking company. Table 1 shows the number and percent of websites that showed evidence of tracking by the most commonly identified companies, separated into three broad categories: third party advertisers, website analytics companies, and social media connectors. Google Analytics, a service that is free to the website and uses first party cookies for tracking behaviour, was by far the most commonly identified tracking company/service, present on over half of recommended and ‘Google only’ sites. Recommended sites were no less likely to use behavioural tracking than were ‘Google only’ sites ($\chi^2 (1)=0.35$, n.s.).

Sites that use tracking technologies typically provided some notice of this fact in their privacy policies. Among the 71 recommended sites that use some form of tracking, 58 (82%) had at least an oblique reference to tracking in their privacy policy, while 6 (8%) did not acknowledge tracking in their privacy policy, and a further 7 (10%) did not have a privacy policy. Among the ‘Google only’ sites, 55 (76%) have some acknowledgement of tracking in their privacy policy, 6 (8%) have a policy that does not mention tracking, and 11 (15%) do not have a privacy. There were no significant differences in the notification practices of recommended sites and ‘Google only’ sites ($\chi^2 (2)=0.96$, n.s.).

Among the tracking companies, third-party advertisers present the most significant privacy risks, since they can and do aggregate behavioural tracking data across multiple websites. We examined each of the sites, therefore, to identify whether they were associated with any third party advertising (most commonly DoubleClick and Google Adsense, although other companies such as Casale Media were also included in this group). Half of the ‘Google only’ sites that included trackers had a third party presence (37 sites, or 51%), and almost one-quarter (17, or 24%) of the recommended sites that tracked had a third party presence on their site. Recommended sites were significantly less likely to include tracking by third party advertisers ($\chi^2 (1)=10.32$, p<.01).

**Conclusions**

The results of this research suggest that sites that deliver health information are likely to use cookies or web beacons that allow the tracking of user behaviour, potentially across multiple sessions and across multiple sites. Furthermore, this practice is not always disclosed to consumers in the website privacy policies. The sites recommended by CAPHIS and CHIPIG are not significantly less likely than those returned only by a Google search to use behavioural tracking measures. Given that behavioural tracking represents a significant and growing privacy concern with the potential to expose website...
users to negative consequences associated with user profiling, LIS professionals should work to ensure that their clients have minimal exposure to this risk by vetting recommended sites to ensure they do not use third party tracking cookies, and by educating consumers regarding the risks of these cookies.

References


