Social Networking and Canadian Privacy Law:

*Jurisdiction, Retention, and Disclosure*

Christopher Parsons
PhD Candidate, University of Victoria

Brief to Parliamentary Access to Information, Privacy and Ethics Committee
Prepared on December 23, 2012
Introduction
Since the summer of 2012, I have been a co-investigator on a project examining how social networking companies comply with aspects of Canadian privacy law. Our project examines how the expectations of social networking websites and environments, whose raison d'etre is the facilitation of the sharing of personal information about and by users, can be reconciled with prevailing understandings about “reasonable expectations of privacy” and the existing Canadian regimes that are designed to protect personal data. This research is funded through the Office of the Privacy Commissioner of Canada’s Contributions program. The use of these funds is independent of the Commissioner; as such, evidence presented to this committee reflects work that emerges from independent academic research and does not necessarily reflect the Privacy Commissioner’s own position(s).

In this submission, I highlight some of our analyses of 20 social networking sites’ privacy policies and findings about Canadians’ ability to access their own personal information that social networking sites store. These findings let us understand how the companies running these services understand their legal jurisdictional obligations and the retention of personally identifiable information. Moreover, these discoveries let us ascertain the actual access that Canadians have to profiles that they and the networking services that they associate with are developing. Together, these points reveal how social networking companies understand Canadians’ personal information, the conditions of data sharing, and the level of ease with which Canadians can access the information that they themselves contribute to these services. I conclude this submission by suggesting a few ways that could encourage these companies to more significantly comply with Canadian privacy laws.

Methodology
In our research, we examined a host of social networking services, not just the high-profile organizations like Facebook and Twitter that already receive large amounts of public and regulatory attention for their privacy practices. The choice of social networking sites was driven by the services that Canadians have adopted. Based on a survey of marketing research that evaluated the relative popularity of social networks, we examined the following: Blogger (Google); Club Penguin; Facebook; Flickr (Yahoo!); Foursquare; Google+; Instagram (immediately after acquisition by Facebook); LinkedIn; LiveJournal; MySpace; Nexopia; Ping (Apple); Plenty of Fish; Reddit; Tumblr; Wikimedia Foundation; Wordpress.com; World of Warcraft (Blizzard); YouTube (Google); and Zynga.

We focused on the companies that provide the social network (e.g. Twitter), not on the companies who provide applications to communicate with social networks (e.g. Tweetdeck, a desktop client that lets individuals post to, and read from, the Twitter social networking service). Thus, we have analyzed these networking companies’ privacy policies and tested access to the information they collect about Canadians; we have not done the same for the clients that Canadians use to access those companies’ services.

The project elements that I discuss here rely primarily on documentary analysis. Our team has analyzed the content of a sample of privacy statements and corporate data disclosure policies and performed judicial, policy, scholarly, and governmental analyses of these policies. After evaluating sample statements, we composed a matrix of the social networks
under study and the relatively common key disclosure practices that emerged from our initial evaluation. The matrix helped us develop a comparative framework to categorize and differentiate between privacy statements and disclosure agreements associated with the social networking services under study.

Our document analysis was supplemented by testing the services’ compliance with access requests against Section 4.9 Schedule 1 of Canada’s federal privacy legislation, the Personal Information Protection and Electronic Documents Act (PIPEDA). Our requests were filed where members of the research team had a pre-existing relationship with a network; as such, only the following services were tested for Section 4.9 Schedule 1 compliance: Twitter; Facebook; LiveJournal; Ping (Apple); Tumblr; Google Domains Services; Wordpress.com; Flickr; Google+/Google Services (non-domain services); and LinkedIn.

In what follows, I outline some of the key findings that we derived from the analysis of the services’ privacy policies; I then discuss the findings surrounding our access requests.

Privacy Policies: Legal Jurisdiction
In the course of our research, we surveyed the top social networking services (SNSes) used in Canada and analyzed them according to a range of questions relating to: the content and visibility of the policy; the procedures for the data subject (in terms of exercising privacy rights); the claims about the definition and capture of personally identifiable information (PII); the disclosure of PII to other organizations including law enforcement; commitments about security; and commitments about access and correction rights. In this section, the extent to which published privacy policies claim or reference compliance with different national and/or international legal regimes is examined. This analysis provides an understanding of where – and to what – SNS companies have explicitly agreed to be legally bound.

Many of the SNSes that we examined claimed compliance with at least some national or international privacy laws or regimes. Notably, however, Flickr, Instagram, Meetup, Noxopia, Reddit, Wikimedia Foundation, and Wordpress all fail to mention compliance with any specific national or international regime. This said, it should be noted that when we contacted Instagram, they did open a dialogue about complying with Canadian privacy laws, though without ever complying in practice.

Despite the Office of the Privacy Commissioner of Canada’s high-profile engagements with foreign social networks such as Facebook, only Club Penguin in our sample group specifically states its compliance with Canadian privacy law. Club Penguin is a Canadian company that was acquired by Disney. Most other social networks, including Blizzard, Facebook, Google, LinkedIn, LiveJournal, MySpace, Twitter, and Zynga, emphasize that they comply with some American statute, such as the Child Online Privacy Protection Act (COPPA). As a result of their (stated) compliance with COPPA, these companies avoid knowingly collecting personal information from children under the age of 13, though this does not mean that the companies avoid collecting information about children under this age: parents, teachers, and others who interact with young children and youths can and do publish information about these children. The mechanisms that these networks use to avoid collecting PII from children under 13 is often quite crude, amounting
to preventing account creation if a person selects an age of under 13 years when registering for an account. Consequently, a great deal of PII about these children can be, and is, collected on these sites by those. Further, PII about children is collected when children are knowledgeable enough to select the age of 13 or older category when signing up for a Facebook or Twitter account.

Other companies, including Google, Facebook, LinkedIn, LiveJournal, MySpace, Apple’s ‘Ping’, Twitter, Blizzard,ix and Zynga, assert their compliance with U.S.-E.U. Safe Harbour, and some also note compliance with the U.S.-Swiss Safe Harbour Framework. Significantly, between the time we surveyed these companies’ privacy policies and when I prepared this paper, MySpace modified their commitment to U.S.-E.U. Safe Harbour. Specifically, their policy now reads:

When a Member who is located in the European Union chooses to post Profile Information that will be publicly disclosed, that Member is responsible for ensuring that such information conforms to all local data protection laws. MySpace is not responsible under the EU local data protection laws for Member-posted information.

The conditions that provoked this change remain unknown, though they occurred as Europe debates their so-called “Right to be forgotten” principle, which social networking companies have widely come out against. Foursquare has not adopted Safe Harbour principles and explicitly informs its international visitors that “federal and state governments, courts, or law enforcement or regulatory agencies may be able to obtain disclosure of your information through laws applicable in the United States. Your use of this site or the Service or your submission of any Personal Information to us will constitute your consent to the transfer of your Personal Information outside of your home country, including the United States, which may provide for different data protection rules than in your country.”xii This effectively positions American laws as the preeminent laws that these networks agree to abide by.

When individuals do have a complaint concerning how one of these services is collecting, retaining, or processing personal data, the companies will often try to restrict where these complaints can be heard. Quite often, privacy policies or terms of service will state the jurisdictions and courts in which all legal proceedings must be conducted. Save for Yahoo!,xiii Nexopia,xiv and Plenty of Fish (a Canadian dating social network),xv which recognize Canadian courts, all claims must go through either American federal court or the state courts of California or New York. Only Zynga, a social gaming company, explicitly recognizes European jurisdictions, stating that non-US citizens would “agree to submit to the personal jurisdiction of the courts in Luxembourg.”xvi

So, in aggregate, what do these findings say? They suggest that some large social networking companies are reluctant to adopt or implement European and Canadian data protection laws. Such reluctance may be based on economic reasons, such as avoiding hiring counsel in various nations; for linguistic reasons, such as wanting to defend themselves only in a language that founders understand and are fluent in; or for other business reasons. More specifically, under this final category, large social networking companies may worry that complying with data protection and privacy laws in the EU and Canada could hinder or forbid practices that the companies currently employ to benefit
commercially from collecting, processing, and retaining individuals’ personally identifiable information. Having spoken to matters of jurisdiction surrounding the networks, I now turn to address the length of time that data is stored by some of these networks.

Privacy Policies: Data Retention
A simple examination of how social networking companies state they retain data is revealing. Google recognizes that, after deleting account information, they may not immediately delete data and that they may not remove data from their backup systems. Such claims are worrying given the long-term retention problems surrounding Street View data insofar as actual retention periods remain ambiguous. While Facebook states that it typically takes a month to delete data — with some information remaining in backup logs up to 90 days — the company’s success in actually deleting data, such as photos uploaded to the site, has long been questionable. Companies such as Yahoo! and Foursquare offer commitments similar to Facebook’s. Foursquare also notes that, even after subscribers delete information, “copies of that information may remain viewable elsewhere, to the extent it has been shared with others, distributed pursuant to privacy settings, or copied or stored by other users.” Tumblr parallels this statement, informing subscribers that even when deleting their accounts’ content, public activity, such as posts that were ‘liked’ or shared, will remain stored on servers and accessible to the public.

For other services, the ‘deletion’ of subscriber data may largely amount to hiding the information from public viewers. LiveJournal, for example, recognizes that, while individuals can delete their accounts and accompanying information, data may take an unspecified amount of time to delete, and the company may choose to retain the information to the extent necessary to protect the company’s legal interests, comply with court orders, et cetera. The use of ‘et cetera’ leaves open the full range of possible motivations to retain data in contravention of a subscriber’s request. With Meetup, the company reserves the right to retain information that the user requests be removed if retention is needed to resolve disputes, troubleshoot problems, or enforce the terms of service. Regardless, the company promises “your information is never completely removed from our databases due to technical and legal constraints (for example, we will not remove your information from our backup stores).” Nexopia offers similar decrees concerning the removal of personal information as Meetup, insofar as Nexopia states that individuals ought not expect that their personal information will be completely removed from the company’s systems following a deletion request.

Given that many of these services function as platforms that allow external developers to capture, process, and retain users’ generated data, the potential exists for data that is ‘deleted’ on the platform (e.g. Facebook, Twitter, LinkedIn, Foursquare) to be retained indefinitely by third-party developers, leaving no way for the platform to enforce a users’ deletion request on the third party. Companies such as Club Penguin, Yahoo!, Google, and Apple reserve the right to share collected or contributed information within and across their corporate organizations, and most social networks include provisos that they ‘may’ (read: will and do) share information with analytics companies and associated advertisers. Significantly, when we examined the social networking services using Ghostery, a tool that identifies web trackers, we found that all services with the exception of Facebook and Google used third-party analytics and/or advertising services. Facebook and Google, of
course, use their own backend analytics and advertising systems and do not need to rely on third parties for such services.

What can be made from this information? Quite simply, these companies rarely offer reliable ways to delete information after it is added to their respective social networking services. As a result, individuals who find their information on these networks – either because they have put it there themselves or because a third-party has uploaded it – have limited ability to remove the information. So, while many of the companies who run these services have developed sophisticated systems to mine data for advertising, anti-copyright infringement, and harm prevention purposes, they have yet to develop more than rudimentary tools to let individuals confidently remove data from the corporate servers and systems.

Key Access Discoveries
As a further test of jurisdiction, we asked various SNSes to provide comprehensive records of the information they held on our research team members. Of all the companies who were contacted, only Facebook, Twitter, Google, Instagram, LinkedIn, or Tumblr responded in any way. These companies chose to provide incomplete information, refused to provide any information, or provided only basic data that was generated by the subscriber.

While Facebook has “self-download” feature to let users to access their own information, the feature is largely the result of pressure from a public advocacy campaign titled “Europe v Facebook.” This initiative meant to improve Facebook’s “transparency” with their users, as well as to enhance “control” over users’ personal information on Facebook’s platform, particularly those found outside of US legal jurisdiction. Europe v Facebook provided the groundwork for the Irish Data Protection Commissioner’s report into Facebook’s data collection and retention principles, leading to a major change in Facebook’s “Data Use Policy,” including the initiation of the self-download feature.

Though the self-download feature does let subscribers access some of the data that Facebook collects, uses, and processes, much of the data - particularly metadata – is withheld from users. Network analysis tests conducted by Privacy International reveal a more comprehensive listing of data collected by Facebook on users than Facebook discloses using the self-download tool. A comprehensive listing of data collected on users of Facebook’s services that is excluded from the self-download feature includes: user logs; IP address information including ISP; content posted on other user’s pages; meta-data associated with videos; information logs on user “likes”; browser information; information specific to user interaction with advertisements; information gathered through “conversation tracking”; information that “indicates a relationship” with other users; information about pictures that users used to be tagged in, but have since been “un-tagged” from; “Tracking information” that Facebook gathers from user interaction with other websites; search history compilations through Facebook’s “search” function; information on newsfeed settings; information on “click-flows” and user visits to individual pages of the platform; information on use of personal data in the “friend finder” function; disclosure on the uses of user data in “matching” processes associated with ad targeting or facial recognition; information on the use of pictures for Facebook’s new “face recognition” tool, or any other biometrical data that may be used to identify users; data Facebook collects on
users (e.g. phone numbers) when other people in the user’s network ‘synchronize’ a device (e.g. iPhone) with Facebook; information gathered on users’ relationships to other users (friends, brother, etc.); and information on “invitations” to groups, events, or pages that users have sent to friends in their network.xxvii

Twitter, similarly, makes some information available but withholds a considerable amount of metadata. Twitter’s disclosure of users’ PII relies heavily on identity authentication. Subscribers to the service first request a full copy of their information. They are subsequently asked to open a ticket with Twitter, and after opening the request ticket, they are asked to send the following: a statement authorizing the disclosure of the specific information being requested; a statement containing the ticket number; a document with the subscriber’s Twitter ID; the email address that Twitter has on file as linked to the account; and a scanned copy of government-issued photo identification. After providing this information, Twitter provides a downloadable copy of the user’s information. All information contains hashes to ensure that the data provided corresponds with data actually stored in Twitter’s database. At issue, however, is that not all metadata is provided to the end-user. The following five lines show the information provided to a subscriber about a single tweet:

```
user_id: 14087212
created_at: Thu Mar 06 06:03:10 +0000 2008
createdVia: web
status_id: 767404918
text: Let's learn about Twitter, eh?
```

Compare these five lines with the listing of all the fields and metadata that are actually associated with a tweet circa 2010 – 59 or 60 lines of information: tweet’s unique ID; text of a tweet; tweet’s creation date; ID of a tweet that is being replied to; screen name and ID of who is being replied to; whether the tweet has been favorited; whether the tweet has been truncated to 140 characters; the author’s user ID; the author’s user name; the author’s screen name; the author’s biography; the author’s URL; the author’s location; rendering information of the tweet; the account’s creation date; whether the account has contributions enabled; number of tweets the user has favorited; number of users the author is following; the user’s time zone and time offset; number of tweets the user has; the user’s selected language; whether the user’s account is set to protected status or not; number of users following the author’s account; whether the user has geolocational tagging enabled; place IDs; the user’s contribution ID, if he has one; URL to fetch a detailed polygon for the place location; printable names of the place; the place associated with the tweet; type of place (e.g. neighborhood or city); country the place is in; bounding CSS for the place; and the application that sent the tweet.xxviii In light of Twitter’s reluctance to provide full metadata information, one Canadian citizen has filed a formal complaint to the Office of the Privacy Commissioner of Canada; the case remains unresolved.xxix

Like Facebook and Twitter, Google offers a download service through their ‘Data Liberation Front.’ After requesting data using an automated form through their tool - which requires users to request discrete data from major Google services, instead of automating a full download of all information attached to a Google account – the data is made available in a cacophony of different formats, depending on the data type. Contact information is
formatted to be incorporated into contact book programs, discussions on Google+ and ‘+1s’ are provided as strict HTML, and the profile page is in the JSON format. While this does provide a better machine-readable formatting of data than some other services, it still lacks comprehensive metadata information: IP address information is missing, location (where appropriate) is missing, and so forth.

In aggregate, metadata itself constitutes content. It can provide geolocational information, information about social networks and broader communications patterns that are not evidenced in a single statement, tweet, or Facebook message. It can reveal the activity of a user on any specific social network and times of activity, as well as relative affluence based on devices used to communicate with the social network, technical sophistication based on client software that is used, and it can be used in conjunction with other users’ metadata for commercial data mining purposes. Consequently, given that metadata often constitutes personal information, these companies have all failed to fully account for the personally associated data generated by the users.

LinkedIn, Instagram, and Tumblr each responded when we requested access to our personal information. Unfortunately, data was not ultimately provided. Both LinkedIn and Instagram engaged us in discussion - LinkedIn opened a ticket, and Instagram negotiated to provide information - but neither ever actually provided us with the personal information that their networks had collected, used, or processed about us. Tumblr’s legal staff stated that the company “will not be providing the information you requested. Tumblr is a U.S.-based company with its headquarters in New York. It does not have a corporate presence in Canada and, therefore, it does not fall under the jurisdiction of PIPEDA or Canada's Office of the Privacy Commissioner.” In a subsequent follow-up, after we had further explained the company’s obligations under PIPEDA, the company reiterated: “We appreciate your interest in engaging in a legal discussion about the scope and reach of PIPEDA, but our prior correspondence stands.” The stated requirement to work through New York courts is interesting, given that Tumblr’s privacy policy recognizes only the California Civil Code (S. 1798.83-1798.84) and acknowledges that California residents are entitled to ask for information about the categories of subscriber data the company is sharing with affiliates and third-parties.

In addition to these difficulties accessing their personal data, subscribers to these services may encounter challenges when alerting a social networking company to their concerns about how the company is retaining, processing, or disclosing their personal information. Of our sample, only three companies - Plenty of Fish, Reddit, and World of Warcraft - published their privacy officers’ contact information. Most other companies had somewhat ambiguous contact forms or address information. Few companies had clear complaints or resolution processes. This said, two services, LiveJournal and MySpace, recognize the uniqueness of EU subscribers, with the former providing an EU mailing address for complaints and the latter encouraging Europeans to submit questions using the company’s online form or by mail. Tumblr also stands out, insofar as the published mailing address is exclusively for California residents. Only Instagram lacked a complaints mechanism entirely. However, subsequent research revealed that its staff was willing to discuss, if not act on, personal information related concerns. Instagram’s processes for dealing with these kinds of requests may change over time, given their recent acquisition by Facebook.
So, what can we say about subscribers accessing the personal information that these services retain? To begin, it can be incredibly challenging to access one’s own personal data. Save for the limited disclosures of information provided by Facebook, Twitter, and Google, fully accessing our information would require a formal complaint to the Office of the Privacy Commissioner of Canada. Having to contact a government ombudsperson to extract personal information seems like an overly onerous requirement. Moreover, even when data was provided it was limited and, arguably, not comprehensive in that the metadata associated with social networking communications was not provided. Furthermore, companies such as Tumblr explicitly flout their dissent with non-American law. Finally, even trying to complain about the services - or contacting a privacy officer to learn about how personal information is captured and can be downloaded by the subscriber - is challenging given the relative lack of effective complaints mechanisms. Such high levels of friction in accessing one’s personal information speak poorly of these companies’ practices, given that the companies themselves are ostensibly designed to promote (relatively) frictionless sharing of personal information. It seems as though when subscribers want to know all the personal information that exists on social media company servers, the company makes it difficult, sometimes to the point of denying access, for subscribers to access that data. The situation is so fraught that it appears citizens can only learn what comprehensive information the companies have been collecting if they involve a national ombudsperson, a task that few citizens have an appetite for.

**Conclusion**

This submission has explicated factors that social networking companies explicitly recognize as legally shaping their services’ privacy and data retention aspects. Moreover, when we tested data disclosure compliance in relation to data collected about Canadians, we found the companies lacking at best, and entirely negligent at worst. In aggregate, this submission reveals where these companies have been inattentive to Canada and its privacy laws.

Jurisdictionally, few companies recognize the need to comply specifically with Canadian law, a deficit that may contribute to their poor behaviour. Moreover, with regard to data deletion, section 4.5.3 of PIPEDA states “personal information that is no longer required to fulfill the identified purposes should be destroyed, erased, or made anonymous. Organizations shall develop guidelines and implement procedures to govern the destruction of personal information.” Given the vague deletion commitments that most of these companies offer, the data retention and deletion policies they have developed lack clear governance towards data removal. Clear guidance is needed here; companies must know what a good model of data deletion looks like, and what would comply with Canadian law. In terms of Canadian privacy law, when companies did provide us with our requested personal data, it was not comprehensive; metadata must be recognized as constituting personally identifiable information, or Canadians will forever be in the dark about the full range of data that companies are collecting and how it might be being used.

This committee would, ideally, consider ways of strengthening the ‘bite’ of Canadian privacy law, in order to get foreign companies to consider our laws when developing and deploying their services. Such ‘bite’ need not slow innovation so much as encourage rapid development that accords with Canadian privacy law. Administrative fines might be
appropriate to levy against companies found to willfully violate our privacy laws, or these companies might be required to have a clear statement of data retention and deletion processes that offers a defined way of removing data from social networking services. Metadata collected by these services might be recognized as constituting personally identifiable information and, as such, lend weight to efforts by Canadians to access all of the information these services collect about Canadian users. By increasing the nation’s relative stature in the eyes of SNS companies, a more ‘privacy-friendly’ set of service options may emerge - ones that reflect how these systems ought to operate - and once again reveal Canada’s ability to influence the development of popular and highly-used tools in positive ways that affect not just Canadians, but the entire global base of these services’ users.

CITATIONS

\[i\] The author would like to thank Joyce Parsons for her editorial assistance with this submission.
\[xi\] “Blizzard Entertainment® Online Privacy Policy.”
\[xii\] https://foursquare.com/legal/privacy


[http://www.nexopia.com/privacy](http://www.nexopia.com/privacy)

[http://europe-v-facebook.org/FAQ_ENG.pdf](http://europe-v-facebook.org/FAQ_ENG.pdf)

‘Facebook told to stop indefinitely holding users’ advertising data’, Charles Arthur, The Guardian, December 21, 2011,  


To see this information in visual format, see: Christopher Parsons. (2010). “Twitter, Mobile Browsers, and Metadata Privacy,” *Technology, Thoughts, and Trinkets* (blog). Published April 22, 2010. Last accessed November 13, 2012. Available at:  

Based on personal correspondence between Christopher Parsons and the complainant.

To date, no full traffic analysis of Google metadata has been performed. Doing so would require performing a man-in-the-middle attack to decrypt data transmitted between Google and a client computer and is presently outside the scope of our research project.

Michael Sussmann, Personal e-mail with Christopher Parsons.
