The Growing Ambiguity of Personal Information
From Personally Identified to Personally Identifiable

If the librarian asks, “Show me some ID, please,” we assume that the basic address details usually written on an envelope will suffice. Some might ask for a driver’s licence in addition, but what about your vehicle licence or even your face? Are they personal information? The answer is not so obvious, and this is just the point: the meaning of personal information is changing.

There was a time when it was relatively clear what “personal information” was. It was your name, your street address, perhaps also an official government-issued ID, such as your Social Insurance Number. By and large, we also knew how and when others were using this information to identify us. No doubt, confusion and misidentification occurred from time to time, but we were typically identified in ways that were familiar—and transparent—to us.

To a large extent, we also expected, or trusted, organizations we knew to protect our privacy, which they did by protecting other information linked to our personal identifiers—our bank records, our census returns, our consumer credit histories, our library borrowing record, and so on. If we did not want to be contacted by someone we did not know, we could get an unlisted telephone number.

Times have changed.

Canada’s Public Safety minister recently defended attempts to update and extend the ability of law enforcement agencies to access the information that identifies us online by saying that the various ways in which we are
Is Your Licence Plate Personal Information?

In 2011, the Alberta Court of Appeal concluded that for information to be about an “identifiable individual” under the Alberta Freedom of Information and Protection of Privacy Act, the person must be identifiable; in other words, the information must have a precise connection to an individual. Such information must be about the individual—that is, directly related to the individual—and not about property that the individual might own (such as a car). Some information is not inherently personal but becomes so because it is associated indirectly with an individual through ownership. Thus, in Alberta, a driver’s licence number is personal information but a licence plate number is not, even though, in Alberta at least, a licence plate is connected to the vehicle and linked through a database to an individual.

What happens, then, when licence plate numbers are automatically photographed and identified by automatic licence plate recognition (ALPR) cameras? These devices use optical character recognition technology to automatically read licence plates and then search that information through assorted databases. The information and privacy commissioner of British Columbia released a finding on this subject in late 2012. She found that people do have privacy expectations associated with their licence plates because while the number, on its own, is non-identifying, the common linkage between the number and identifiable information means the numbers should be protected.

Authorities seem to vacillate on whether licence plate information is private or public. For example, some police organizations have taken the position that the photos taken of plates are not personal, which justifies collecting them. They argue, however, that the public should not be able to access the records that have been generated on the basis of these photos because they contain personal information. Even after the RCMP sent out a letter formally recognizing that the data collected by authorities’ scanning practices are personal information, the head of the RCMP’s scanning program and other Canadian police officers have continued to assert that because vehicle licence plates are shown in public, there is no reasonable expectation of privacy regarding that information.

identified online are no different from “phonebook data” that link a phone number to a name and a residential address. Just as the police can find out who the subscriber of a particular telephone number is, they should be able to find out who is behind the multiple identifiers that allow each of us to
This confusion obfuscates the political nature of wide-scale surveillance. When data collected are “nonpersonal,” it is relatively easy to convince the public and government oversight bodies of the appropriateness of the data collection. But when surveillance captures “personal” information, then legal protections and normative concerns arise, which could delay the deployment of such surveillance equipment. Whether a licence plate is, or is not, personal data has become an inherently political question.

which would require prior judicial authorization (a warrant) on a standard of reasonable and probable cause that a crime has been, or will be, committed.

Our subscriber information is not, however, the same as our phonebook listing:1 How we are identified online is complex and dynamic. Online communications involves many more identifiers than our name, phone number, and address. How many of us know about, let alone can decode, the following: the Internet Protocol (IP) address, the mobile identification number (MIN), the media access control (MAC) number, the Service Provider Identification Number (SPIN), the electronic serial number (ESN), the International Mobile Equipment Identity (IMEI) number, the International Mobile Subscriber Identity (EMSI) number, and the subscriber identity module (SIM)? Each of these identifiers can potentially be traced back to a unique user. So that is the first point. We are now identified in ways that are highly technical and largely mysterious. Most of us have no clue how we are identified online.

The second point is that using the Internet is not like using a telephone. It is not just a communications medium but the basic platform through which many of us engage in essential professional, personal, and political tasks: booking hotels and flights, social networking with friends and colleagues, shopping for books and music, organizing our lives through calendars, and conducting research. This information can be far more revealing about our lives than what we may say during telephone conversations. How we are identified through digital networks, therefore, provides important insights into who we are, what we do, whom we do it with, and when and where we do it.

Thus, the scrutiny of identifiers by organizations can reveal enormous amounts about our daily lives. If you want to test who might have access to your browsing habits, install a free download program like Collusion or Ghostery. Within seconds of browsing, you will see a list of ad networks or Web analysis and reporting tools that are tracking and sharing information about your online activities. Browse around further, and the list multiplies and spreads like a spider web. In the online world, we have become “identifiable” even if we are not “identified.”

**Are IP and MAC Addresses Personal Information?**

Every device connected to the public Internet is assigned a unique number known as an Internet Protocol (IP) address that allows applications to send information—like browsing results and email—to the correct recipient. IP
addresses consist of four groups of numbers separated by periods. Since these numbers are usually assigned to Internet service providers within region-based blocks, an IP address can often be used to identify the user’s general location. But the issue gets complicated because some IP addresses are dynamic, changing frequently.

The privacy commissioner of Canada has said that an IP address is personal information:

An Internet Protocol (IP) address can be considered personal information if it can be associated with an identifiable individual. For example, in one complaint finding, we determined that some of the IP addresses that an Internet service provider (ISP) was collecting were personal information because the ISP had the ability to link the IP addresses to its customers through their subscriber IDs.²

In spite of such decisions, there is a significant and long-running battle over whether the IP address is, or is not, personal information for the purposes of privacy law. The answer to this question is crucial for determining whether the average Internet user has any personal privacy rights over his or her searches, browsing habits, blog posts, or social networking activities. Google’s official position is that an IP address is not personal information because it identifies a machine and not a person.³ Many users may share one computer with a single IP address—members of the same family, for instance, or employees within a business, or students who share a library computer terminal. An Internet service provider will be able to associate the IP address with a home or business account but not (at least not ordinarily) to any particular person using a device linked to the Internet.

The mobility of our devices means that we are continually connecting to the Internet at coffee shops, airports, and other public places through a number of IP addresses. The privacy concerns are amplified with the growing use of media access control (MAC) addresses. MAC addresses are numbers that uniquely identify mobile devices—like cellphones, iPods, laptops, or tablets—on a network.

Just because devices and addresses are not stable does not mean that the addressing protocols are not personal information. If I change my home phone number every week, is it any less personal data? Is there really no threat to privacy because specific search queries supposedly cannot be narrowed down to a single individual? Knowing what a small group is seeking
online can allow a third party to associate that behaviour with each individual member of that group, spreading the privacy risk and potential harm.

Although a MAC address or an IP address is rarely going to be directly related to one identifiable individual, it is how the MAC address or IP address is combined with other information (or could reasonably be combined with other information) about tastes, behaviours, and interests that has privacy advocates concerned. If you knew and combined enough online and offline information, you might have enough data to make a highly probable (sometimes almost perfect) guess about who was doing what, when, and where.

Identification and Re-identification

A related point is that individuals can be positively identified even when none of their personally identified information, like their name or address, is available. This is accomplished simply by combining other basic and non-identifiable information about them. A recent study of a random sample of people living in Montréal shows that almost 98 percent could be positively re-identified by name if one knew three variables: date of birth, gender, and postal code. The researchers point out that these findings have especially troubling implications for health research, because people are demonstrably more comfortable about sharing their health data if there is a low risk of re-identification.

Re-identification science works to identify unique individuals despite efforts that have been made to strip obvious identifiers from existing data sets (called “de-identification”). The sophistication with which such re-identification science is pursued in some quarters has led some researchers to conclude that the goal of de-identification can give a false sense that anonymity has been achieved. Common anonymization practices no longer protect privacy. Re-identification science disrupts basic assumptions about what is, and is not, personal data and has forced regulators and analysts to rethink essential principles about information privacy. Personal identification is not a binary choice between data being either identifiable or not identifiable. Rather, the process of identification resides on a complicated and dynamic continuum and depends on what other information may later be combined with that already collected. Risks to individuals do not disappear when personal identifiers are removed. And this is not just a scientific
Is Your Facial Image Personal Information?

As Ayan strolls down Montréal’s Sherbrooke Street, her friends and acquaintances recognize her face, and if they know her reasonably well, they can associate her face with her name. This situation has led some to assert that she, like anyone else in a public place, has no reasonable expectation of privacy. As soon as Ayan chooses to be out “in public,” she immediately surrenders her privacy rights.

But things are not so simple. If the clothing store that Ayan enters captures her image on its surveillance cameras, does that mean that she has surrendered any rights to control how that image might be used, or who might access it? Our laws say no. Even if her image cannot be immediately associated with Ayan herself as “an identifiable individual,” organizations must, by law, protect those data and only use them for legitimate purposes. The fact that Ayan, and most of the people she knows, use social media further complicates this situation. Social media sites use facial-imaging software that allows her and her friends to tag images with identifiers and share them widely.

Such software uses algorithms to measure facial features such as the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. Those data can then be used to search for other images with matching values on other databases. “Tagging” images on social networks such as Facebook is controversial and has inspired significant protests from privacy advocates. In late 2012, Facebook promised to forego its facial-tagging program in response to protests from regulators and advocates, but the practice continues.

The privacy implications of being able to associate Ayan’s face to her name are huge. Aside from the uses that may be made by law enforcement, these technologies have also been described as a “stalker’s dream.” Yes, Ayan’s face, and your face, are personal information. And yes, we should have rights over how that information is captured and shared.
Internet usage, or by installing spyware directly on individual computers. Third-party advertisers do not need to know your real life “identity” so long as you can be identified by a technically specified address and thus targeted with personalized ads.

Processing personally related information online is therefore fundamental to the business models through which “Big Data” companies actually make money. Advertising is the lifeblood of the Internet economy. To the extent that companies can discover more detailed and extensive information
about personal preferences and behaviours, they will make more money. To some extent, privacy laws constrain that ability. Rules about notification, informed consent, access, correction of personal data, and so on are not just an important limit on the ability of an organization to monitor consumers; they also have profound economic consequences. So, too, does the definition of personal information and the argument over what is, and is not, personal information. If information is “personal,” organizations are constrained, but if it is not, they are unregulated.

User-Generated Content

Another source of confusion around traditional understandings of personal information relates to social networking. Traditionally, we conceived privacy concerns as stemming from personal information about individuals being collected and processed by organizations. Big organizations primarily control personal data, which they analyze using the latest technologies in order to make decisions about individuals in their capacities as consumers, clients, students, employees, and so on.

In the world of social networking, however, the individual generates most of those data. User-generated content (UGC), also known as consumer-generated media (CGM), refers to any material created and uploaded to the Internet by users themselves, whether that is a comment left about a book on Amazon.com, or a video uploaded to YouTube, or a profile on Facebook. UGC has been around in one form or another since the earliest days of the Internet. But in the past few years, thanks to the growing availability of high-speed access and search technology, it has become one of the fastest-growing forms of content and has revolutionized how users interact with each other and how advertisers reach those individuals.

If we produce user-generated content, does that personal information belong to us or to the companies whose platforms host it? Do these organizations have a responsibility to apply all the privacy principles to the data we provide? Our regulators tend to say yes, insisting that social-networking services are data controllers, whatever the source of the personal data processed.6

Companies tend to see things differently, which is apparent from the definitions of “personal information” contained in their official privacy policies, as documented by a recent study of the most popular twenty-four social-networking sites used in Canada.7 Predictably, conceptions of which
characteristics accurately define personally identifiable information vary across these sites. Here are some examples:

- **Google** (for Blogger and Google+): Information that the user provides to Google which personally identifies that person, such as your name, email address, or billing information, or other data that can be reasonably linked to such information by Google.
- **Facebook**: Name, profile pictures and cover photos, network, gender, username, and user ID. Facebook may collect IP address, GPS location, Internet service provider, location, type of browser, or the pages you visit.
- **Flickr**: Name, gender, birthdate, postal code, and email address. Flickr collects information about users’ transactions with Yahoo and with their business partners, including information about users’ use of financial products and services that they offer.
- **Instagram**: The amount and type of information that Instagram gathers depends on the nature of the interaction.
- **Plenty of Fish** (a Canadian dating site): Contact information, personal preferences (e.g., language preferences), marketing information (e.g., photographs), other information provided in your personal profile (e.g., interests, marital status, height, weight, occupation).
- **Zynga**: Name, profile picture or its URL, user ID number, your friends’ user ID numbers and other public data, login email, physical location and that of access devices, gender, birthday.

These definitions have implications for privacy. For instance, Nexopia, advertised on its site as “Canada’s largest social networking site for youth,” advises users that “to help members find and communicate with each other, you may submit and post additional profile data, including but not limited to the following: weight, height, sexuality (i.e., sexual orientation), dating and living situation, and information regarding your interests.” To be sure, this information is not mandatory for using Nexopia, yet all information provided in one’s profile is not identified as “collected personal information” and may thus be shared accordingly.
Ping, Apple's social networking site (SNS) for music, ostensibly provides a category of protected personally identifiable information to its users but limits this category to contact and payment information. The category does not include information gathered about a user's family and friends: when a Ping user shares his or her favourite music with others, “Apple may collect the information you provide about those people such as name, mailing address, email address, and phone number.” Put simply, Apple collects the personally identifiable information of third parties, and, because Apple's privacy policy does not apply to these third parties, Apple does not consider this information to be personally identifiable.

And then there is the question of metadata—the data about the data, typically including identifiers such as users' IP addresses, their operating systems, and any information gained from cookies: information that can subsequently be used not only to identify individuals and their personal browsing habits but also to track their physical location. Of the twenty-four SNSs surveyed in this research, not one identified any element of metadata as personally identifiable information, nor did any of them give users any expectation of privacy regarding their metadata. Unsurprisingly, the motivation for this treatment of metadata is overwhelmingly couched in the language of the SNS's efforts to improve the user experience. IP addresses or cookie information are necessary, it is reasoned, to combine services, to prevent problems, to keep products safe, and, generally, to tailor one's use for a more “personalized” approach. The broader privacy implications are rarely addressed.

Many social networks (indeed, many websites) also permit access through pseudonyms that conceal a user's identity but allow them to be recognized on a return visit. These are sometimes referred to as unique "handles" and are designed to be deliberately opaque—but clearly linkable to a particular individual. People rely on this form of identification in multiple scenarios and contexts on the Internet because pseudonyms often encourage more candour and openness. However, people also tend to choose the same pseudonyms for different sites, making it easy for them to be re-identified.

Since online companies make money with these data, should we not have some rights over their use? But how, then, would one exercise those rights if a condition of using a service is to authenticate one's identity? There is circularity here: one has to reveal one's real identity to exercise rights over personal data that were originally shrouded.
So What Does the Law in Canada Say About Personally Identifiable Information?

Over the past thirty years, the federal and provincial governments across Canada have gradually passed privacy legislation. Initially, most of these laws regulated the public sector; only later were they extended to private corporations. With few exceptions, most organizations in Canada, both public and private, are expected to follow a set of common information privacy principles. Not surprisingly, however, legal definitions of what constitutes personal information are not uniform.

Most laws tend to use the word “identifiable information.” Thus, the federal law governing the private sector (PIPEDA) states, “Personal information means information about an identifiable individual.” This is very flexible, but it can also be quite circular.

Other laws define specific types of personal data exactly and include long lists of categories of data to which the legislation applies. Here, for instance, is the list in the Freedom of Information and Protection of Privacy Act in Ontario:

(a) information relating to the race, national or ethnic origin, colour, religion, age, sex, sexual orientation or marital or family status of the individual;
(b) information relating to the education or the medical, psychiatric, psychological, criminal or employment history of the individual or information relating to financial transactions in which the individual has been involved;
(c) any identifying number, symbol or other particular assigned to the individual;
(d) the address, telephone number, fingerprints or blood type of the individual;
(e) the personal opinions or views of the individual except where they relate to another individual;
(f) correspondence sent to an institution by the individual that is implicitly or explicitly of a private or confidential nature, and replies to that correspondence that would reveal the contents of the original correspondence;
(g) the views or opinions of another individual about the individual; and
(h) the individual’s name where it appears with other personal information relating to the individual or where the disclosure of the name would reveal other personal information about the individual.11

Other Canadian laws include subtly different categories of sensitive and nonsensitive forms of information. But such lists can never be exhaustive, and the definition of what is, and is not, sensitive is invariably subjective and inherently related to the context. For instance, having our names and addresses in the phonebook might be in our interests, but that same information on a blacklist, a no-fly list, or a file of bad credit risks would be incredibly sensitive. In other words, the same information in different contexts and used for different purposes can affect the risk to privacy dramatically.

Many other laws, like the privacy law in Ontario, specify that the information has to be “recorded.” But what does that mean? Can one have rights over one’s personal data even if they are not recorded? The law covering the private sector in Québec is a bit different: personal information is “any information which relates to a natural person and allows that person to be identified.”12

Other laws include lists of information to which the legislation does not apply: basic business contact information, for example, or, more controversially, “work product” information produced by individuals in the course of their employment, business, or profession. Controversially, this exemption has been extended to include medical prescriptions written by Canadian doctors. The work-product exemption also tends to exclude the data submitted about a business on consumer reporting websites like www.travelocity.com or www.yelp.com. It would be totally unreasonable to ask a business to consent before a consumer posted a critical review of his experience at a hotel or restaurant. But then what about evaluations of teachers or professors on www.ratemyprofessor.ca? Is this the personal information of the professor or of the student, or both?

The Canadian privacy commissioner often struggles with whether personal information, as defined in the federal laws governing the public and private sectors (PIPEDA and the Privacy Act, respectively), is being processed, and thus whether its legal provisions apply. In many cases, the question of whether privacy is at risk often rests on tricky questions of probability. Our commissioners and courts struggle with an evolving legal framework, which always seems to be one or two steps behind the technology.
Conclusion

The contentious and confusing definition of personal information exposes a basic problem with trying to use privacy laws to address the entire range of social problems captured by the word *surveillance: surveillance can occur even when personal information is not collected*. The examples above demonstrate that the information available about us online cannot be split into two neat categories, some of it personal and some of it nonpersonal. Rather, the risks to privacy tend to depend on what organizations assume about us when they collect information about us and on how likely it is that they will be able to use our information to identify us individually. Analysis of the risks may just as likely be based on subjective judgments about organizational motivations. And just because an organization can identify an individual does not mean that it will do so.

This trend also confronts us with a larger question about how to understand this looming social problem in political terms. Privacy analysis and privacy law tend to begin and end with the existence of personally identified or identifiable information. If no claim can be made about the actual or potential linkage between a surveillance practice and a specific individual, then the privacy regime cannot help.

One major contribution of surveillance scholarship is the insistence that power relations are present between the watcher and watched even when personal information is *not* captured. Video surveillance cameras do not need to be working or monitored to change behaviour: the prospect or potential for surveillance is often enough. Individuals might not be monitored at any one time, but they would be well advised to behave as if they were. Similar dilemmas plague the capture of information by ubiquitous computing devices, remote sensors, drones, or radio frequency identification (RFID) tags, which allow data to be transferred wirelessly using electromagnetic fields and are used by many industries to track the physical location of products. And on the Internet, your browsing behaviour might not be monitored, but many of us now know enough about the potential for surveillance to be careful and to take protective steps, or perhaps not to browse on certain topics.

Surveillance technologies structure power relations and imbalances between individuals and between individuals and organizations, whether personal data are captured or not. If no personal data are collected, it is difficult to contend that a “privacy problem” per se exists. Yet power is and can be exercised without any personally related data being captured, anonymized or
The growing ambiguity and complexity of these questions brings into focus the range of surveillance problems that lie outside the very broad realm of personal privacy protection.  

Notes


9 See Apple Privacy Policy, last updated 1 August 2013, www.apple.com/privacy/.

10 Canada, Personal Information Protection and Electronic Documents Act (PIPEDA), 2000, s. 2(1).

11 Canada, Freedom of Information and Protection of Privacy Act, R.S.O. 1990, c. F.31, s. 2 (1).
