How can technologists relate to privacy?

Privacy Tradeoffs in the Information Age
Swinburne University of Technology
29 April 2013

Stephen Wilson
Lockstep Group

The technology-privacy gap

I'm a technologist who stumbled into privacy.

Some 12 years ago I was doing a big security review at a utility company. Part of their policy & procedures suite was a privacy statement posted on the company's website. I was asked to check it out.

The privacy statements said things like 'We the company collect the following information about you [the customer] ... If you ever want a copy of the information we have about you, please call the Privacy Officer ...'.

I had a hunch this was problematic, so I took it to the chief IT architect. He said he had never seen the statement before! So that was the first problem. And he confirmed my suspicion there was no way they could readily furnish complete customer details, for their CRM databases were all over the place.

How could such a document, totally at odds with reality, come to be written? Clearly there was a lot going on in the world of privacy that we technologists needed to know about.

So with an inquiring mind, I took it upon myself to read the Privacy Act. And I was amazed by what I found. In fact I wrote an academic paper in 2003 about the ramifications for IT of Australia's 10 National Privacy Principles, and that kicked off my privacy career.

Ever since I've found, time and time again, a shortfall in the understanding that "technologists" as a class have regarding data privacy. The gap between technology and the law is perpetuated to some extent by the popular impression that the law has not kept up with the march of technology. As a technologist, I am more optimistic. I actually find that principles-based data privacy law anticipates almost all of the current controversies in cyberspace (though not quite all).

LOCKSTEP

Technology & Privacy mindsets

Privacy laws are written with some naivety about the mechanics of computers, especially in respect of how information flows and accumulates in modern information systems. Many crucial parts of the law seem to assume that records are still paper-based and can be annotated in the margins.

On the other hand, engineers find the law to be impenetrable.

And when IT professionals hear the well-meaning slogan "Privacy Is Not A Technology Issue" they tend to think 'thank god – that's one thing I don't need to worry about'.

Of course privacy is a technology issue, insofar as there are endless ways that privacy impacts tech, and vice versa. But the two disciplines are on different wavelengths.

Personal Information

Information or an opinion, whether true or not, about an individual whose identity is apparent, or can reasonably be ascertained, from the information or opinion

Privacy Act 1988

Collection

An organisation must not collect PI unless the information is necessary for one or more of its functions or activities

National Privacy Principle 1

Some legal technicalities continue to catch technologists off guard.

People tend to think intuitively that Personal Information is the stuff of forms and questionnaires and call centres. But like many intuitions, our personal feelings about privacy can be misleading. Which is a bit ironic because engineers like definitions, they like precision, and so they should have a look at the technical definitions in the Privacy Act.

Technologists can be surprised that the definition of Personal Information covers a great deal more than data provided directly.

For instance, if metadata or event logs in an IT system are personally identifiable, then they constitute Personal Information regardless of whether they are untouched by human hands.

Our privacy legislation is technology neutral with regards the manner of collection. Indeed, the term "collection" is not specially defined in Australian privacy law.

So if Personal Information has wound up in an information system, it doesn't matter whether it was gathered directly from the subject, or imported, or found in the public domain, or generated almost from scratch by some algorithm … it has been *collected* and as such is covered by the Privacy Act.

So now let's look at some famous recent technology missteps …

Lessons from Google's StreetView Wi-Fi collection

Google StreetView cars collect the locations of Wi-Fi hubs for the geolocation database; we can assume that the location data and hub names are not personal.

It was found in 2010 that the StreetView software was also inadvertently collecting unencrypted Wi-Fi network traffic, some of which contained Personal Information. Australian, Dutch and other Privacy Commissioners found Google was in breach of privacy laws, for collecting Personal Information without consent and without a reasonable business need.

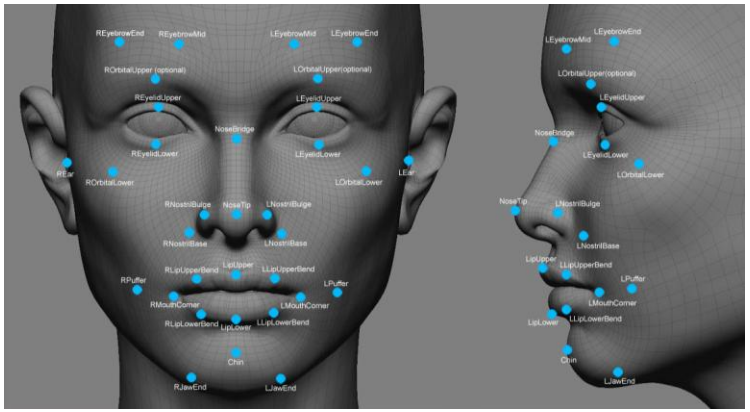
Many technologists argued with me at the time that Wi-Fi data in the "public domain" is categorically not "private", and that Google was within its rights to do whatever it liked with it. But the argument fails to grasp the technicality that our privacy laws basically do not distinguish public from "private". **In fact the words "public" and "private" are not operable in the Privacy Act** (indeed there is a view that the Privacy Act is really a data protection law).

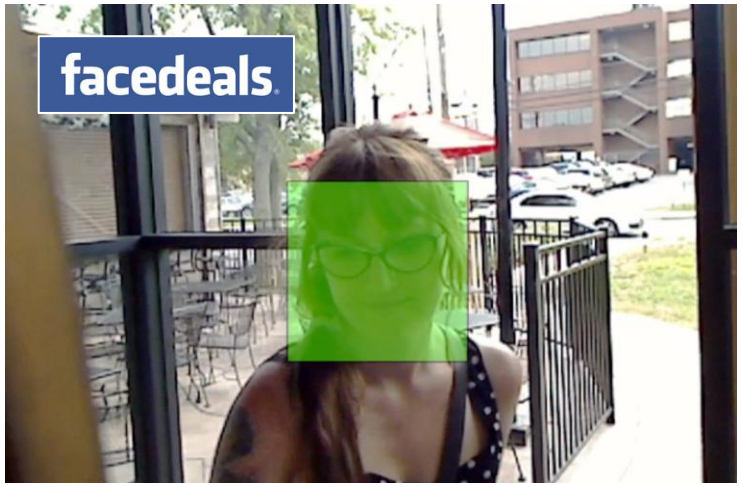
Lesson: it doesn't much matter if data is sourced from the public domain: you are still subject to Collection and Use Limitation principles.

And it is essential to distinguish between casual notions of "private" vs "public" and the technical definition of "Personal Information".

LOCKSTEP

Lessons from Facebook's facial recognition





Facebook's photo tagging creates biometric templates used to subsequently generate tag suggestions. Before displaying suggestions, Facebook's facial recognition algorithms run in the background over all photo albums. When they make a putative match and record a deduced name against a hitherto anonymous piece of image data, they have *collected* Personal Information.

European privacy regulators in 2011 found biometric data collection without consent to be a serious breach, and forced Facebook to shut down facial recognition and tag suggestions in the EU.

Lesson: it doesn't much matter if you generate Personal Information using sophisticated algorithms: as opposed to collecting it directly, you are still subject to Collection and Use Limitation principles.

And this matters a great deal as we start to see commercial applications of face recognition. The start up "Facedeals " for instance created a lot of noise when a beta was launched in mid 2012. The idea is to match (with consent) images captured on CCTV against templates taken from Facebook. It is not clear yet how Facebook will benefit from third parties using the templates, nor if users have clearly consented to the commercial exploitation of metadata that originated from simple photo tagging.

LOCKSTEP

Lessons from Target's Big Data pregnancy detector

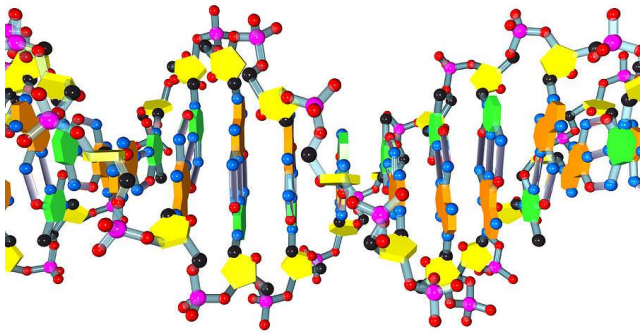
Target in the US was found to be experimenting with statistical methods for identifying that a regular customer is likely to be pregnant, by looking for trends in her buying habits.

In Australia, the privacy implications would be amplified by the fact that tagging someone in a database as pregnant [even if that prediction is wrong!] relates to health and therefore represents a collection of *Sensitive Information*. Our Privacy Act *requires express consent in advance* of collecting Sensitive Information. Stores may therefore need to disclose to their customers up front that Big Data processes can produce health information by mining their buying habits. Express consent to this may be required.

Note a present problem in Australia for grocery stores that sell medicinals online. St Johns Wort for example seems innocuous but it indicates that a customer has (or believes they have) depression. IT security managers and database admins might not have thought about the implications of logging mental health information in regular web servers and databases.

LOCKSTEP

“DNA Hacking”



In February 2013 research was published where a subset of anonymous donors to a DNA research program were identified by cross-matching genes to data in public genealogy databases.

All of a sudden, the ethics of re-identifying genetic material has become a hot topic. A lot of attention is focusing on the nature of the informed consent; different initiatives (like the Personal Genome Project and 1,000 Genomes) give different levels of comfort about the possibility of re-identification. Absolute anonymity is typically disclaimed but re-identification is usually said to be difficult.

But a nice legal problem is that regardless of the consent given by a Subject (1st party) to a researcher (2nd party), when a 3rd party takes anonymous data and re-identifies it without consent, they have collected Personal Information, as per the principles discussed above.

Lockstep believes, following the European facial recognition precedent, that re-identification of DNA without consent is likely to be ruled problematic (if not unlawful) in some jurisdictions, and is therefore unethical in all jurisdictions.

We aint seen nothin yet

Consider augmented reality glasses. They transmit extraordinary amounts of raw data to the cloud for processing where object recognition and content addressable image banks will allow system operators to figure out what you like doing, without you ever needing to "Like" anything!

So the deep privacy issue with Google Glass is not sexting or peeping toms. It's the fantastic amount of intimate Personal Information they will be able to extract from our everyday data stream.

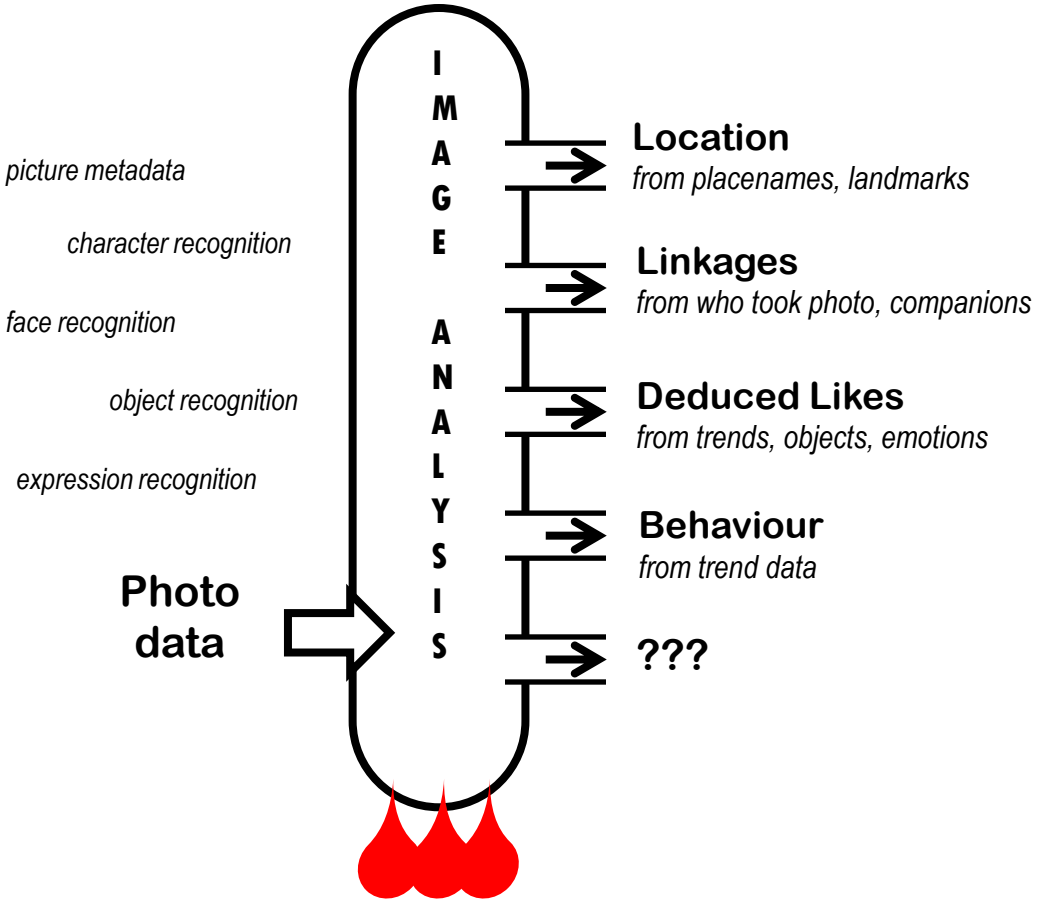
And here is the big problem in Big Data …

Principles-based data protection laws have proven to be relevant and powerful in the cases of Google's StreetView Wi-Fi collection and Facebook's facial recognition. And these regulations seem to apply to data mining for health information, and to DNA re-identification. Conventional privacy management involves telling individuals *What* information you collect about them, *Why* you need it, *When* you collect it, and *How*.

But with Big Data, even if a company wanted to be completely transparent, it may not know what Personal Information lies waiting to be mined and discovered in the raw data, nor when exactly this might be done.

LOCKSTEP

More than data mining



An underlying theme in Big Data business models is data mining, or perhaps more accurately, data *refining*.

An increasing array of data processing techniques are applied to vast stores of raw information (like image data in the example) to extract metadata and increasingly valuable knowledge.

There is nothing intrinsically wrong with a business model that extracts value from raw information, even if it converts anonymous data into Personal Information.

But the privacy promise enshrined in OECD data protection laws – to be open with individuals about what you know about them and why – can be hard to honour when we are still discovering new ways to refine raw data into actionable insights.

The trade in Personally Identifiable Information

There is a bargain at the heart of most social media companies today, in which Personal Information is traded for a rich array of free services.

The bargain is opaque; the "infomopolies" are coy about the value they attach to the Personal Information of their members. A bit like the early settlers of Manhattan Island, the Infomopolies pay a pittance for the riches we give them.

If OSNs were more open about their business models, it seems likely that most of members would still be happy with the bargain. After all, Google, Facebook, Twitter et al have become indispensable for many of us. They do deliver fantastic value.

But the Personal Information trade needs to be transparent.

If the trade is fair and open, and if secondary uses of PII are restrained, then the *privacy* of end users can be maintained.. That is, we can and should be able to trade PII without losing privacy.

LOCKSTEP

"Big Privacy"

- Exercise constraint
- Meta transparency
- Fair value for personal data
- Engage customers in the deal
- Dynamic consent models

So I think technologists have to rise to the challenge of Big Data and understand "Big Privacy". Remember privacy is essentially about restraint. If a business knows me, then privacy means they are restrained in how they use that knowledge.
We're at the very start of Big Data. Who knows what lies ahead. "Meta transparency" means not only being open about what Personal Information is collected and why, but also being open about the business model and the emerging tools.
Most savvy digital citizens appreciate there is no such thing as a free lunch; they already know at some level that "free" digital services are paid for by trading Personal Information. Some individuals manage their own privacy in an ad hoc way by deliberately obfuscating or manipulating the details they divulge. Ultimately consumers and businesses alike will do better by engaging in a real deal that sets out how PI is truly valued and leveraged.
One of the most important areas for law and policy to catch up with technology seems to be in consent. As businesses discover new ways to refine raw data to generate value, individuals need to be offered better visibility of what's going on, and new ways to opt out and opt back in again depending on how they gauge the returns on offer.

We need to get more sophisticated in the way we deal with privacy. We should not sugar coat privacy. We often hear that "Privacy Is Good For Business" but that's too trite. Privacy as an objective sits in a complex multidimensional requirements space. We should recognise frankly that privacy is in fact at odds with many other requirements, like security, performance, cost, and new business opportunities. Engineers need to treat privacy as another source of requirements, work out where tensions are, and resolve the conflicts. That's what engineers do.