Privacy: Personal Information, Threats, and Technologies
Richard Beckwith and Scott Mainwaring
People and Practices Research Lab, Intel Research
{richard.beckwith,scott.mainwaring}@intel.com

Abstract

The three primary thrusts of this paper are: first, that people do not think enough about their own privacy, in particular, they may not know enough about their privacy that they can really make informed decisions about sharing information; second, that technologies exist that can mitigate some of the problems associated with information sharing; and third, that services (in addition to technologies) might be a reasonable way to think about addressing the privacy problem.

1. Why we can’t get over it

A cheap allusion to a well-known quote – “You have zero privacy anyway. Get over it.” – is a good way to start a section on just how little people think of sharing data. This is, after all, a claim that we have allowed to slip away whatever privacy we might have had. More than that, this quote also suggests that whatever consequences might be attendant to this lack of privacy are not sufficient to demand a change. The research we’ll report suggests that people think otherwise and that technologies exist to address the issue.

On first blush, many people seem rather blasé about personal information (PI). In our research we have found that most people don’t think much about PI sharing. It just isn’t something that they usually are asked to slip away whatever privacy we might have had. More than that, this quote also suggests that whatever consequences might be attendant to this lack of privacy are not sufficient to demand a change. The research we’ll report suggests that people think otherwise and that technologies exist to address the issue.

Personal information is disclosed, recorded, verified, and often itself generated in the course of making a commercial exchange. Credit card numbers, signatures, purchase records, credit limits, and related types of PI all have important roles within this commercial ecosystem. Of course, we rarely explicitly and directly “share” this information, instead thinking in terms of “using” a credit card or “saving” a receipt. But businesses have developed whose sole purpose is to provide others with our PI so as to grease the wheels of commerce. And this leads us to underestimate its importance or the frequency with which it is accessed.

Clearly, not everyone takes this information sharing for granted. Privacy and consumer rights activists try to raise awareness, change behaviors, and mobilize political constituencies. Mass media highlight particular disasters and debates. Other “activists” in this space include a wide variety of legitimate and illegitimate stakeholders with a financial interest in encouraging consumers to divulge and share their PI, legally or illegally. Between these partisan camps are what we’ll call the civilians.

In our interviews about privacy and technology, many – perhaps most – civilians really don’t imagine that they could have much to worry about. “We don’t have anything to hide, so… I don’t feel [my privacy] is a big concern. We are not that exciting.” Or “My husband always says, ‘As long as you have nothing to hide, it doesn’t matter,’ which is true, too.” Some even seem to go so far as to say “Bring it on!”: “A lot of the e-mail I get, have at it. If you can figure out a way for them to get it … that would be fine.” What difference could it make?

If this were really what people thought, then they wouldn’t have to “get over” losing their privacy because they simply wouldn’t think that they need it. However, the fact that people will sometimes say that they have nothing to hide does not mean that they do not have privacy concerns. These people weren’t saying that they didn’t care if people had access to their PI. What they were saying was that they really don’t have a lot of juicy goings-on in their lives. In fact, when they understood what could be done with such information, people were much more likely to resist sharing data.

As noted, purchasing history, income, existing personal debt, and debt history are traded. These pieces of data may seem harmless. Personal information also includes many things: name, address, phone number, mother’s maiden name, SSN, gender, medical history, even current location. Many of these pieces of information also seem innocuous enough. Given that each piece of data seems safe, perhaps it would be fine to let others “have at it.”
The individual items may actually be safe. It is when these items are combined that a threat may emerge. Knowing, say, someone’s address, phone number, and first pet’s name would enable searching through other databases with name and phone number to find out more about a person, say, the person’s mother’s maiden name (genealogical databases are rich sources of such information). Another database might have the person’s social security number (which is sometimes used for different forms of ID, like drivers’ licenses or health insurance). A search for large banks near their employer might allow someone to guess which bank an individual uses. Given all these individual pieces of information, someone might be able to do more than simply steal an identity, they might even be able to steal whatever resources a person has in the bank.

At this point we should say little bit about how the concerns about privacy by individuals are often conceptualized by researchers. The standard vision of privacy concerns, at least as they play out in the commercial domain, assigns people to one or another category. That is, privacy concerns are considered static. The pre- eminent categorization is the Westin Privacy Segmentation. Alan Westin – a long time privacy researcher – developed his segmentation in 1995 [1] by asking people how they strongly they agreed or disagreed with three statements:

1) Consumers have lost all control over how personal information is collected and used by companies.
2) Most businesses handle the personal information they collect about consumers in a proper and confidential way.
3) Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

Westin found that the answers produced three groupings. Privacy-oriented responses to all three placed respondents in the “Privacy Fundamentalist” segment. Privacy-oriented responses to one or two items placed respondents in the “Privacy Pragmatist” segment. No privacy-oriented answers placed respondents into the “Privacy Unconcerned” segment. While one benefit of this approach is that it is easy to administer, that is not the primary benefit. These categories have been found to categorize respondents well on a majority of the privacy-oriented questions in privacy surveys. They are useful in framing certain public policy decisions, and in tracking public opinion as it evolves over time with changes in technology, business practices, and media coverage.

But these surveys and segmentations may not be the best way to understand how people actually reason about privacy to make data sharing decisions. They assume a stability in level of concern, an insensitivity to context, and a principled rationality that often simply is not the case in real-world purchasing and data sharing decisions.

When people consider what data are available and are asked to make judgments about data access, their judgments often assume that no harm could come from sharing the data. In fact, initial responses often look like the Privacy Unconcerned. When they are presented with more information about how their data could be used, they frequently move into the Privacy Pragmatist space and sometimes close to Privacy Fundamentalism. Most significantly, their judgments are often at odds with the way things work today. This disconnection between perception and reality suggests not only that risk education is a common unmet need, but that the systems currently in place – the current reality – is broken and we might want to think about how to fix it.

Decisions are rarely founded only on perceived risks, but complex weightings of risks and benefits, not to mention emotional and habitual responses. For example, Bellotti and Sellen have found that people are more likely to accept even potentially invasive technology if they think its benefits will outweigh its potential risks [2]. Conversely, proponents of RFID technologies in retail environments have been surprised by consumer backlash, despite offering little in terms of potential benefit to balance the imagined risks of what RFID opponents call “spychips.”

People’s decision making processes are complex, as are, increasingly, the technological domains in which they must act (with correspondingly decreasing understanding). We don’t pretend to have solutions to these issues. What we would like to sketch findings from two exploratory studies looking at how privacy plays out in real-world settings, findings which support some general directions for future technologies to support decision-making in data sharing.

2. Residential care study

This study looked at privacy concerns the denizens of a study of a residential care facility [3]. This facility was instrumented with a variety of sensor technologies in order to track the activities of the various people who spent their time there. The idea was that older adults with dementia who might otherwise have to be in a locked facility or monitored closely would be able to be more free yet still safe. We spoke not only with the residents...
themselves but also the family of the residents (who make decisions for their loved ones) and staff and management at the facility (who were also tracked by the system). One might question the generalizability of the findings given the uniqueness of the setting and individuals involved but we believe that generalization is warranted because neither the staff nor the family were different in the relevant domains from the residents themselves.

In this study, when we analyzed users’ privacy assessments, we used a model borrowed from Anne Adams [4]. In particular, we focused on three aspects of personal information that Adams found determined people’s reasoning about privacy.

1) Information receiver. Who will use or have access to the data?
2) Information usage. How will the information be used, and what do I stand to gain and lose from its use?
3) Information sensitivity. How sensitive is the data?

Adams found that these three subjective aspects determined how people perceived privacy and potential violations. Obviously, the extent to which people understood these three aspects would determine the quality of their reasoning. With respect to the information receiver, most of the people we interviewed were unaware of who had access to data or the fact that the data had rarely been shared. With information usage, everyone involved—the residents, family, caregivers, and management -- was aware that the goal of data collection was to enhance the residents’ lives. However, how this goal would be reached, and how the target information would be used, was not something that people concerned themselves with. In fact, researchers were working on how to use the data most effectively and even they could not have told those involved how things would be used exactly. The point to be made, however, was that our interviewees had no idea and didn’t think they needed to be concerned. Information sensitivity is a function of what information is shared. In this case, the information includes the person’s physical location, weight, some health-related data. Such data would be generally considered quite sensitive, but in this study we found that people’s lack of understanding of the technology rendered them unable to judge. They didn’t know what was collected or how, neither did they know how the data were used. One resident summed up the general consensus when he said that the purpose of the technology was so that “someone can come and help.”

In this study, we found that when participants discussed their “analysis” of risk and benefit they did not consider much about the actual risks or benefits that the technology offered. It must be noted that the technology was rather complex and would take some background to really understand. Still, we were surprised to see that not even the staff whose movements were being tracked had come to understand what the technology did so that they could make a personal decision about compliance. The technology was treated as one large black box with a limited set of inputs and outputs and outside of the area of their concerns. When Arthur C. Clarke [5] said that “any sufficiently advanced technology is indistinguishable from magic” it is unlikely that he had reasoning about privacy in the front of his mind. However, it is easy to see that we cannot expect normal civilians to have sufficient understanding of all technology so that they can reason well about it. Much of the technology may as well involve magic.

3. Nuclear family study

For this study we interviewed people in the Portland, Oregon area. We were interested in people’s expectations of security and privacy, where they see things now, how technology affects that, what their worries are. We recruited families with children and Internet connections to the home. The interview pool was divided: 50% had children all of whom lived at home and 50% had children at least some of whom lived outside the home. Most of the interview pool had spent some time thinking about these issues but usually not very deeply. They simply did not know the information to which they should be attending. That is, they wouldn’t have known what questions to ask. Even if they did ask the questions, it’s not clear that they would have had access to the information that they would need to make decisions. Media, often the source for their information often gets the story wrong by describing technologies incorrectly or missing the mark on the real threats. (More below.)

The interview focused on 3 different areas: (1) information arriving and leaving the home (mail, magazines, email, etc); (2) school related information (for example, the dreaded “permanent record”); (3) health information (from physical medical records, to electronic records, to trending material). These semi-structured interviews began with relatively free-form discussions of information in each of the categories above. People were encouraged to think about areas where they might have privacy or security issues. This was, in terms of the psychology of memory, a recall task. This free recall was followed by a recognition task. The strategy that was used asked whether if people were provided with more information, would they see places where they had privacy concerns? For this section of the interviews,
people were presented with scenarios for information use that presented a variety of threat models.

Some of the overall trends included:

- Shifting information leads to shifting judgments. Privacy concerns are dynamic since information about privacy is often sparse or incorrect. For example, people will often say that their current location is not so sensitive. When it’s pointed out that a burglar could know how far they are from home, the concerns are typically stronger.

- People saw privacy judgments as requiring a “balance.” Most believed that the benefits of intrusions usually outweigh the negatives or downsides. Although almost everyone said you need to have a balance. One woman said, “you have to ask yourself, can you think of a way it can be misused and then design it with this in mind.”

- People were clearly reading things in the newspapers/news media that hyped issues of privacy and security and it was an issue/source of material/anxiety. All these people referred to news stories in explaining their beliefs.

- People had the same reactions about their medical records as they did with education records. ‘I own them, they’re my possession. I should control where they go”

- People included SPAM and junk mail as being an invasion of their privacy – it was something coming into their space that they didn’t control (issues with people trafficking in your identity feels like a violation).

- People were more concerned about the ways in which information was being aggregated in “virtual space” than they were about real aggregations of information.

We’ll now address a few of these trends.

4. Concerns are dynamic

As alluded to above, as interviewees were presented with more relevant data, their reasoning about privacy shifted. This is not uncommon in the literature. Cranor, Reagle and Ackerman [6] reported that in their study, the level of concern reported by respondents did not predict their “reactions to scenarios involving online data collection”. In a study where subjects had to make decisions in context (that is, they were subject to the variables of real life) Consolvo, Smith, Matthews, LaMarca, Tabert, and Powledge [7] reported that the Westin segmentation “was not a good predictor of how [subjects] would respond to
requests for location from their social relations.” In these studies, scenarios and real-life show that people are not tied to their general answers about concerns. People are not static.

Health and wellness data will serve to present the overall responses. For example, people were told that some services might collect health information in the home. The benefits of these systems, they were told, would be that they would allow better home-based rehabilitation or even allow the elderly to be more independent. This very nearly defines the kind of situation where a balance between benefit and threat would be pointed out. Then people were asked what issues might arise if other people could get at the information. When people are told that medical information can be used for making decisions about whether an insurance company would handle their account, they often say that this seemed fair. When they are told that some people will refuse to have tests done (AIDS tests, for example) because they don’t want insurance companies (or others) to know about a preexisting condition, their ideas may change. When they are told that some genetic markers could be found for diseases like Alzheimer’s in newborn infants and might affect the insurability of a newborn, this acceptance of risk was usually rethought.

A child’s health record is especially sensitive to many people. Even in areas where people feel one way about their own data, their views are typically very different when they need to make decisions about the sharing of their children’s data. Even in areas where they thought a child’s data should be shared, when negative consequences were pointed out, their decisions were far less certain. For example, many people thought they would want classroom teachers to know about health issues of students in the classroom. However, when potential bias in treatment (sometimes known as the “Pygmalion in the Classroom” effect) is pointed out and the deleterious effects of this bias noted, the privacy concerns were much stronger.

5. Media (mis-)defines concerns

Popular media are the source of much of our respondents’ information about privacy and technology. Media are good at popularizing but often wrong at technology. Most of the people interviewed were aware that sharing personal information (either intentionally or by mistake, such as putting sensitive materials in a recycling bin) opens the door to identity theft. When asked about this, they were quick to point out that they learned about it through the media. Media-driven awareness is one likely culprit for the increased levels of privacy concerns among
the general population. To the extent that the media are responsible for broader concern with protecting some personal information, as evidenced by the now widespread ownership of shredders, they may have done a valuable service. However, the media are not always so well-informed.

While we are not here taking a position on RFID and privacy, we will note that many media outlets have misreported on the threats of particular RFID implementations. A recent plan to use RFID by a school in California was scrapped. The plan would have had students carry an ID that would register when they passed through a doorway in the school. This would automate attendance. There was community outrage that about “tracking” students. Of course, this system would only work when a particular badge passed through a portal (a doorway) with a reader. Students wouldn’t be otherwise “tracked”. But some sources have popularized the idea that RFIDs may be read from a distance (even from satellites). Would the community have been as concerned if the technology were better understood? We don’t know.

The New York Time’s (03/02/05) discussion of a young socialite’s “cell phone hacking” made it seem as if a radio (Bluetooth or otherwise) on her phone led to a social catastrophe for her and her many friends. This coverage in other sources was occasionally combined with directions suggesting that people with Bluetooth telephones should turn off the radio when they aren’t using it. This suggestion is a fine one that will protect people from threats like “Bluejacking” however the socialite’s phone was not hacked. The internet-based server that stored the information was. There is nothing she could have done – short of not storing the information – that could have saved her (or her friends) this embarrassment.

ChoicePoint, a company that sells personal information, was recently in the news. Some news reports said that their internet site had been “hacked”. What these reports seemed to miss was that this was not a case of a person breaking into a website (as with the socialite) but rather with a company being convinced to sell personal information to the wrong kind of person. This was “social engineering”.

On March 9, 2005, USA Today’s E-Legal column discussed privacy violations and MP3 players and was another example of scaremongering that got the technology wrong. The presumed threat was information leaking from a player due to “wireless capability”. Most players don’t even have this capability and those that do present little to no risk.

Are these mis-reports a problem? Only if you think you can learn about the threats inherent in technology from the media and thereby make informed decisions.

6. Who owns my data?

Once we understand its value and potential abuses, personal information is something that we are loath to lose the ability to control. Because of its value, and especially because of the risks, we need to control when and with whom our personal information (PI) is shared.

Interestingly, Americans do think that they own their personal information and, in some ways perhaps, they do. U.S. law does offer consumers some rights over these data. According to the Digital Millennium Copyright Act, one of the few reasons individuals are allowed to circumvent media DRM (Digital Rights Management) mechanisms is to protect against unauthorized collection of personal information. HIPPA legislation made law the patient’s belief that medical records belong to the patient – sort of. HIPPA allows patients to get copies of health records (for a fee) and a patient has the right to ask for restrictions on disclosures (although providers are not required to comply). If the provider agrees, they are legally bound to comply. Patients can also request that records be amended (although providers can deny this request under some circumstances). Health care providers can only disclose data if it is to be used in connection with healthcare operations (treatment, payment, and evaluations). Certain legal requirements for disclosure also exist (suspected abuse, national security, and to individuals with a custodial relationship).

Researchers have investigated ways in which individuals can own and control their own data [8]. These researchers point out that information about a typical person in the US might be found in a thousand different websites and it is not the responsibility of the database owner to ensure that data are kept up to date. By moving the data into a single source under the person’s control, people could overcome more easily issues with keeping data up to date (consistency). They could also ensure that they are involved in making decision about sharing data (privacy). Not all data would be written to by the individual (e.g., social security number) and this among other data perhaps couldn’t be changed by the individual (e.g., birth date). Some data (e.g., commentary about one’s medical records) perhaps couldn’t even be read by
the individual. Nevertheless, the individual could control it.

There are issues with how the data are used once released (see comments below on using DRM for consumers) and how to control the number of times use might be transferred, if at all. Nevertheless, these are solvable problems and the framework should be considered.

One point needs to be made about the value of personal data, especially because who “owns” this value is a potential issue. The ChoicePoint story so recently in the paper, informs us that for about 100 dollars a query, ChoicePoint lets customers query its database about our PI and even provides those data to customers. DoubleClick built a business out of letting its customers know about the various websites a specific browser has been to. If a history file is worth money, it’s no wonder that spyware has proliferated. The violation people feel is that “their data” are being sold and they aren’t making the decision or making a profit. Technologies could address both these issues.

7. Adequacy of existing technologies

Notice and consent have been prominent in many applications where potentially sensitive information will be shared. But by and large, this is an astoundingly inadequate means of protecting privacy. While interacting online, persons are often asked to set up preferences to be used in future sharing (cf. the client side of P3P). The fact that concerns shift with both context and information means that we really cannot establish beforehand exactly what we are willing to share. Together with the dynamic nature of our willingness to disclose, the facts a) people often are unaware that data are being shared at a given time, and b) the likelihood of forgetting that assent has been given conspire to suggest that our recorded preference would not be our current preference.

In many “traditional” cases, such as with HIPPA-compliant medical information sharing, a person is asked to consent to sharing a particular set of data with a third party. There are additional problems here because methods of fusing data from various sources often will allow information recipients to reconstruct PI that the person thought was obscured when they granted the access.

Finally, ubiquitous and embedded systems present the most difficulty for the standard regime. The ideal of ubiquitous systems, to disappear into the fabric of one’s life, means that forgetting about their presence is a greater risk than for the previously mentioned technologies. Furthermore, as many applications of ubiquitous systems rely on personalization and location-based technologies, the potential for disclosing PI – including one’s location – is quite high.

8. Privacy protection technologies

An array of technologies exist that do, or could, mitigate some of the risks associated with sharing PI.

- Encryption technologies can protect data from unauthorized access. However, they can be notoriously difficult to use directly (see [9]).
- Anonymization or pseudonymization technologies can automatically blur personal information so that, although authenticated, the information is “just personal enough” for current needs; for example, just enough to verify that a channel is secure [10].
- Interface technologies can call attention to privacy threats and remind people how they have set their privacy parameters. For example, web browsers may be augmented to display usually hidden information about cookies and P3P settings (see [11,12]). [It should be pointed out that the earlier critique of the client side of P3P does not impact the server side which in every case will be necessary.]
- Digital Rights Management (DRM) technologies can allow conditional access to encrypted information, tracking and allowing usage on a per-user and per-device basis [13].

In particular, DRM software, which has been developed in support of content owners’ rights, could be repurposed to protect those whose PI would be consumed (Consumer Rights Management). CRM could be enhanced to enforce time limits on access and use of data, and possibly enabling micro-payments to be made back to the individual for each use of her or her PI. Alerting individuals that a request for information has been made and allowing real time assent would be a further enhancement.

9. Privacy protection services

Services could be developed to address many of the issues we have discussed here. Indeed, challenges presented in trying to scale in a completely automated manner the mechanisms above may in fact require an “organization in the loop” to be practically and financially
viable. For example, allowing individuals to assign privacy advocates who would intercept and respond to requests would be of interest to many. Delegated advocates would be virtually required for some individuals, e.g., children. Indeed, the opacity inherent in advanced technologies suggests that most people might be better served by an advocate than they would by trying to understand the implications of their data sharing. A trusting relationship would be of utmost importance.

An advocate could serve a large number of customers by, for example, having a credit card service. Just as with other card companies, data on purchases could be stored but the difference would be that if someone were to want to buy some data, they would have to go through a brokerage system provided by the advocate.

A personal information brokerage could be developed to help consumers not only to hold onto their PI but also to extract value from that PI when it is in their interest to do so. Customer loyalty cards, which give customers discounts or rebates on purchases, are an example of where a business finds value in PI and is willing to provide something to the consumer for sharing. Another example is credit card companies. Many people get frequent offers for new credit cards. What should be apparent is that the companies offering these cards usually have paid for information about the recipient of these offers. In fact, a social engineering hack of ChoicePoint (someone pretending to be a company needing this kind of information) is what led to the recent troubles at ChoicePoint. It is also reasonable to consider how much data are worth in different scenarios. If a credit card company is told that a consumer is willing to look at their offer, or if a boat manufacturer knows that someone is interested in hearing their pitch, the information about that person is worth much more than if they simply match some abstract profile. Adding interest adds value. These customers are more than simply demographically qualified. They are qualified by desire.

We present these ideas for potential services as fodder for debate and discussion, and don’t want to imply that it would be easy or straightforward for such services to be designed, deployed, and to be economically sustained (without, for example, slipping into an exploitative relationship with the individuals for whom they are intended to be advocating). As is the case for many privacy technologies, the lack of civilian understanding often translates into a lack of demand. Nevertheless, we feel there is a possibility that services of this sort could bootstrap themselves by both educating consumers about and reimbursing consumers for the value and use of their PI, thereby both creating demand and serving it. They could thus begin to open up the black box.

9. Summary

In this paper we addressed three primary issues. First (and foremost) we attempted to demonstrate that the complexity of advanced technologies and the variety of concerns in a modern life make it difficult for individuals to make good privacy decisions. We described two studies laying out some of the concerns and focused our discussion on four areas, the dynamic nature of privacy preferences, media misinformation, data ownership, and the complexity of technology. We also briefly reviewed some technologies that could be used to mitigate some of these issues and also discussed the potential of a personal information brokerage and a privacy preserving affinity card.

Before closing we must point out that the two legs we have discussed (1) understanding the human side and (2) a collection of advanced technologies and services is not enough. To truly protect our personal information we need at least these but also require that the legal system develop in such a way that our personal information is protected.

11. Acknowledgements

We’d like to thank our research participants; Scott Lederer, who was involved in the study at the residential care facility; and Genevieve Bell, who (frighteningly) may remember giving comments on this long ago.

12. References


