Privacy Advisors for Personal Information Management

Rhonda Chaytor
Dept. of Computer Science
Memorial University
Newfoundland, Canada
rchaytor@cs.mun.ca

Edward Brown
Dept. of Computer Science
Memorial University
Newfoundland, Canada
brown@cs.mun.ca

Todd Wareham
Dept. of Computer Science
Memorial University
Newfoundland, Canada
harold@cs.mun.ca

ABSTRACT

A Personal Health Record is a beneficial tool, especially for managing personal health information. As the value of personal health information differs from patient to patient, patient confidentiality cannot be guaranteed in such a tool. This paper demonstrates how patients could unintentionally violate the privacy of others, just by disclosing their own personal information. Legal, ethical, and social aspects of the disclosure of personal health information is discussed and a new patient privacy tool proposed.

Keywords

Personal Information Management, Personal Health Record, Privacy

1. INTRODUCTION

Personal Health Information Management (PHIM) [4] has many challenges. For example, a patient may be too embarrassed to share details when speaking face-to-face with a physician. Moreover, a patient is typically only allowed ten minutes to visit a physician and may feel pressed to describe all symptoms and experiences. Another difficult task for patients is keeping track of who has their information; secondary physicians, specialists, and insurance providers almost certainly hear on a regular basis patient responses like "*But I just explained all of that to my doctor yesterday, do I really need to go through that again?*" See Pratt [4] for other PHIM challenges and tools that aim to overcome them.

2. PERSONAL HEALTH RECORD

One PHIM tool is a web-based Personal Health Record (PHR) [4] that includes a subset of patient data maintained by physicians and allows patients to view and maintain their own health information (e.g., medical histories, past surgeries, allergies, and medications). A PHR could alleviate patient embarrassment by allowing the patient to include all necessary, and perhaps critical, information without having to converse with anyone. Also, a PHR allows patients to modify and access information on demand so that physician visits run efficiently. Furthermore, a PHR helps patients collect, organize, and share information across distributed providers, permitting critical patient information to be available at every point of care. Some other services that may be offered to patients through a PHR include [6, 7]:

- Prescription renewal
- Appointment requests
- Access to disease information
- Access to practice information
- Self-referral requests
- Ability to store consultation information

As illustrated in [4], besides collection and maintenance, there is also need for sharing personal health information; however, when personal information is shared, there is an associated risk [2]. Even though a PHR may very well be a beneficial PHIM tool, there is a tradeoff between sharing information and guaranteeing patient confidentiality.

3. PATIENT PRIVACY

One of the most important promises a physician makes is that of confidentiality. Thus, whether on paper or online, *most* patients will expect that their health information will remain confidential. There are, however, reasons and opportunities for a patient to share their personal health information and unlike bank account and credit card information, some patients may not mind releasing *some* or *all* of their health information. Disclosure may occur if a patient is asked survey questions on the telephone by a telemarketer, or if an insurance company offers a patient a discount upon release of their PHR, or if a journalist outright offers a patient money for the information, etc. Due to the variance in how patients value their personal health information, patient confidentiality cannot be guaranteed in a PHR.

3.1 Risk of Re-identification

Figure 1(a) is a simplified example of a set of patient records for a particular region. It is desirable to make a public release of these records for medical research purposes, like the release in Figure 1(b). The original dataset cannot be

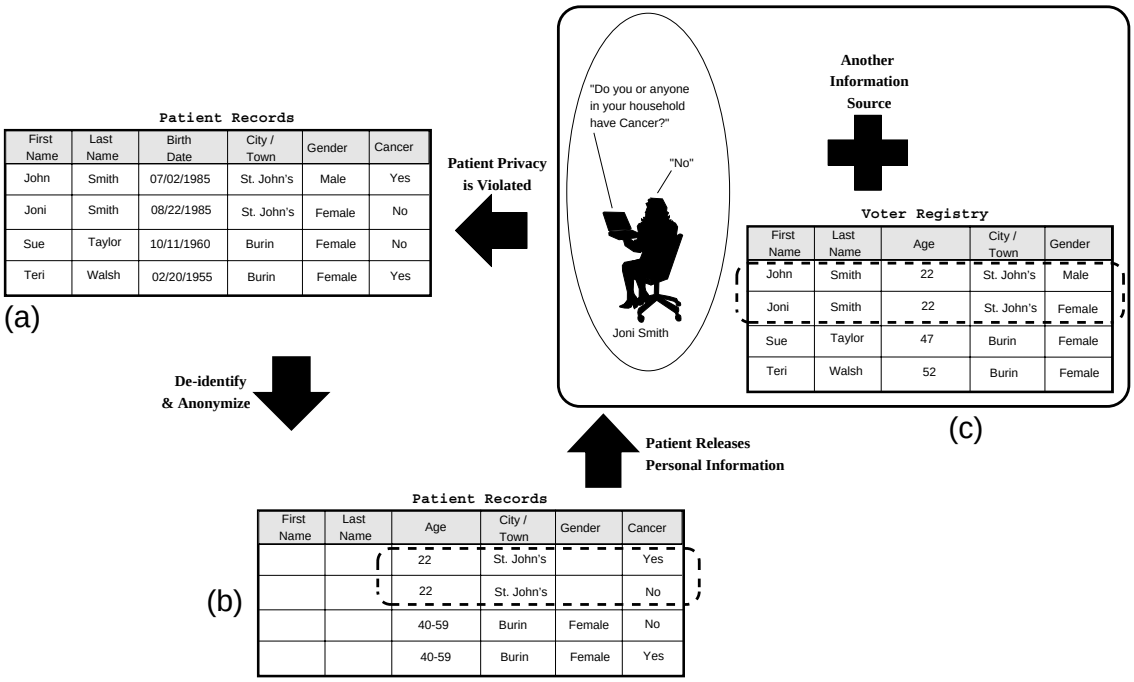

Figure 1: A privacy violation.

publicly released because when data values are unique to a particular person, it is possible to *re-identify* that person, even if the dataset is de-identified (i.e., usual identifying attributes like name and social security number have been removed from the dataset). Re-identification is accomplished by linking records containing unique data values to public databases (e.g., insurance policies, voter registries, and telemarketing profiles). Sweeney's [5] research, for example, shows the ease at which a medical record can be linked to an individual with little more then a zip code, date of birth and gender; she was able to pinpoint the governor of Massachusetts' medical record (six people had his birth date, only three of them were male, and he was the only person in his 5-digit ZIP code).

3.2 Anonymization as a Solution

Making individuals anonymous can prohibit this sort of record-linkage. One way of achieving anonymity is ensuring that groups of people look exactly the same by *suppressing* or *generalizing* the distinctive personal identifiers. For example, the birth date *July 2nd, 1985* can be generalized to age *22*, or can be further generalized to age *20-39*, or may be withheld entirely. In the literature, this approach is said to satisfy *k-anonymity* [5]. An anonymous set of patient records is illustrated in Figure 1(b). Notice that anonymization ensures that no information is distinctive to one person; in Figure 1(b) it is impossible to tell which 22 year old from St. John's has Cancer.

3.3 Anonymization is Not Adequate

Unfortunately, anonymization breaks down in a PHR. As soon as one patient decides to disclose personal health information, enough information can be obtained to re-identify other individuals. Figure 1(c) demonstrates how John Smith's medical condition is revealed as a consequence of Joni Smith's disclosure of personal information. Joni Smith did not see the harm in disclosing the fact that she did not have Cancer. On the surface, revealing one's personal information seems harmless, but there is a serious and real risk. Potentially malicious[1] information seekers may influence patients to reveal information, which may in turn, unintentionally violate the privacy of other unsuspecting patients. These unsuspecting patients may be family members (trait-based linkage), members of the community (geographic-based linkage), coworkers (specifics-based linkage), physicians or HMO's (inclusion-based linkage), etc.

3.4 Legalities

In short, there is no rule or regulation that specifically prohibits the release of information about one individual because it may be used to infer personal information about another individual.

[1]Information seekers need not be malicious; consider the recent article [1], which raises social, ethical, and legal issues associated with law-enforcement agencies searching for criminals by using the DNA collected from their relatives for other purposes.

3.4.1 American Legislation

The development of de-identification legal requirements in the U.S. is most extensive under the *Health Insurance Portability and Accountability Act (HIPAA)*, which governs the electronic transmission of health information. Revisions to the accompanying *Code of Federal Rules*, generally known as the *HIPAA Privacy Rule*, came in to effect in 2003, including a de-identification rule. Datasets that are adequately de-identified can, in effect, be shared with researchers for secondary purposes. Acceptable de-identification takes two alternative forms:

1. Removal of a prescribed set of 18 attributes (the so-called *safe harbor* method)
2. Certification from a qualified statistical expert that the risk of identifying an individual is *very small* under the circumstances.

Since the safe harbor method (1) does not anticipate an attacker using information from one record to help re-identify another record, this is implicitly not a concern under the rule. The risk analysis under rule (2) would also typically use the safe harbor method as its standard for risk. In other words, consideration of the risk of inferences that can be drawn about a second person may not be necessary under either alternative. Although the rule also says that actual knowledge of re-identification risks precludes the release of information (which could conceivably include the risk of one person's information helping identify other records), since no analysis of vulnerabilities in the datasets is required, such actual knowledge seems improbable.

3.4.2 Canadian Legislation

In Canada, no specific legal rules for de-identification currently exist. A mix of federal (*Personal Information Protection and Electronic Documents Act (PIPEDA)*, which has been in place since 2000) and provincial statutes (e.g., Ontario's *Personal Health Information Protection Act*, 2004) govern privacy in the public, private, and health arenas, which are fairly consistent in defining personal information to refer to an identifiable individual. This is commonly interpreted to mean that if personal identifiers are removed from the data, the information is no longer within the scope of the pertinent legislation. Although obtaining the information and manipulating it might be legally problematic, current legislation does not seem to impose anything more circumspect than removing attributes that might indicate the identity of the individual prior to its release.

3.5 Ethics

The lack of legal rules does not mean regulators and courts will not eventually recognize that releasing one patient's de-identified information can impact another's patient's privacy. Under appropriate circumstances, a civil claim or complaint would force an interpretation of general principles of negligence and privacy rights in this context. In the meantime, it is largely an open question to what extent an information provider should consider such implications. On the one hand, there is a limit to an information holder can be expected to do to predict or prevent unlawful activities of third parties, and the prevention of all possible inferences from released data is not feasible. On the other hand, the consequences of intrusion into personal information is so significant for the individuals involved, that there seems to be an ethical imperative (even if the legal obligation is unclear) to take serious precautions beyond simple elimination of obvious identifiers.

4. A NEW PIM PRIVACY TOOL

Karat [3] says that with new tools, privacy policies and legislation will improve. Karat's PIM tool for privacy enforcement, SPARCLE, is a step in the right direction. It allows individuals to be informed participants when dealing with organizations. What is still missing is the ability to let individuals know how their actions may unintentionally infringe upon the rights of others. In addition to privacy enforcement, what is needed is a new tool called a **PIM Privacy Advisor (PIM-PA)**.

Individual privacy has been discussed and researched (e.g., in [2, 3, 4, 5]), but to date, no research addresses the privacy of others. Figure 2 shows how a PIM-PA could be consulted prior to release of personal information in order to prevent the violation of one's own, and other patients' privacy. It is envisioned that a PIM-PA will solve a diverse set of computationally interesting problems that build upon current k-anonymity-based problems; however, the focus has shifted from protecting the personal information of all patients in a dataset to assessing the risk associated with an individual patient's disclosure of personal information. All problem definitions stem from the following basic problem:

- *Given an individual patient from a set of patient records, is there is a set of attribute values that the patient may release to other (potentially malicious) parties, without disclosing the identity of anyone in the population?*

There are at least two natural variations of this problem. First, instead of trying to protect the entire population, a patient may only care about protecting the privacy of a particular group of patients, which may consist of the individual patient, the patient's family or neighbors, *etc.* (**group-membership**). Second, a patient may not consider all attributes to be equally private and may prefer full or partial disclosure over no disclosure at all (**attribute-disclosure-partition**). For example, consider the following scenarios:

- A patient receives a call from a telemarketer who is requesting selected information regarding the patient's health. Obviously, the telemarketer already has the patient's phone number and possibly the patient's name and address, since that information is publicly available in phone books. Now what additional information may the patient reveal to the telemarketer without disclosing his or her own identity?
- A patient requests a life insurance quote and is asked about his medical history. His wife's medical condition has caused him problems with some insurance companies in the past, so what information can the patient

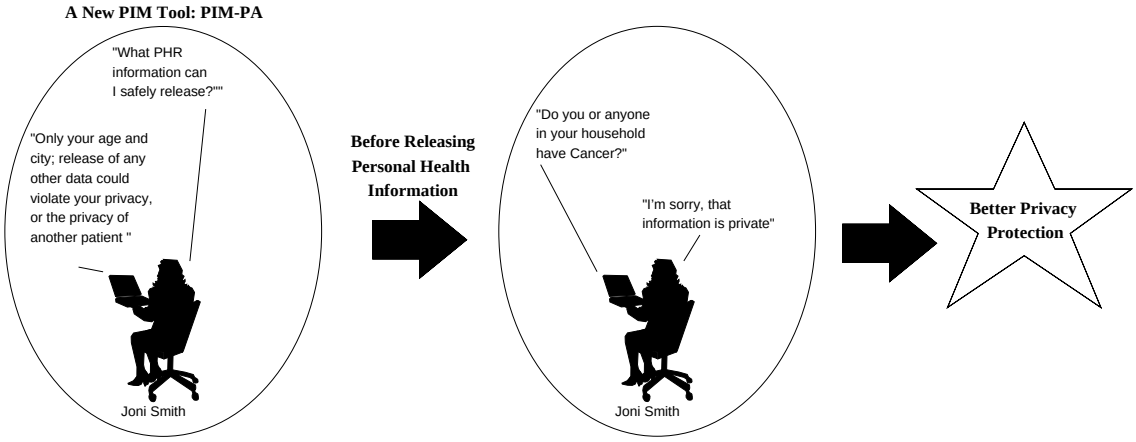

Figure 2: PIM-PA facilitating in the prevention of privacy violations.

release without disclosing the identity of his wife; i.e., a patient who has the same address attribute?

- A patient lives in a public figure's neighborhood. She has been offered money by a journalist in exchange for information in her PHR and is not overly concerned about releasing her own information. How much information can the patient reveal to the journalist without disclosing the identity of her neighbors?

5. CONCLUSIONS

A PIM-PA could council patients on *all* the risks of disclosing their personal information and is intended to coexist with the PHR system. Providing PHR users with a tool like a PIM-PA would free them from having to worry about legal, ethical, and social repercussions of sharing their personal information, giving them more time to concentrate on managing their health care. To be an useful tool, it must be efficient, user-friendly, and customizable. Moreover, given that legal issues surrounding privacy protection have not yet been interpreted, a PIM-PA will have to be extendable and easily modified.

6. REFERENCES

[1] F. R. Bieber, C. H. Brenner, and D. Lazer. Finding criminals through DNA of their relatives. *Science*, 312(5778):1315–1316, June 2 2006.

[2] T. Erikson. From PIM to GIM: personal information management in group contexts. *Communications of the ACM*, 49(1):74–75, January 2006.

[3] C. Karat, C. Brodie, and J. Karat. Usable privacy and security for personal information management. *Communications of the ACM*, 49(1):56–57, January 2006.

[4] W. Pratt, K. Unruh, A. Civan, and M. Skeels. Personal health management. *Communications of the ACM*, 49(1):51–55, January 2006.

[5] L. Sweeney. Achieving k-Anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):571–588, 2002.

[6] J. S. Wald, B. Middleton, A. Bloom, M. Gleason, E. Nelson, Q. Li, M. Epstein, L. Volk, and D. W. Bates. A patient-controlled journal for an electronic medical record: Issues and challenges. In *MEDINFO 2004 Conference Proceedings*, pages 1166–1170. IMIA, 2004.

[7] M. Wang, C. Lau, F. A. Matsen, III, and Y. Kim. Personal health information management system and its application in referral management. *IEEE Transactions on Information Technology in Biomedicine*, 8(3):287–297, September 2004.