OPEN ACCESS

Privacy policies for health social networking sites

Jingquan Li

Correspondence to
Dr Jingquan Li, School of Business, The Texas A&M University-San Antonio, One University Way, San Antonio, TX 78224, USA;
jli@tamusa.tamus.edu

Received 16 November 2012
Revised 29 March 2013
Accepted 30 March 2013
Published Online First
18 April 2013

ABSTRACT

Health social networking sites (HSNS), virtual communities where users connect with each other around common problems and share relevant health data, have been increasingly adopted by medical professionals and patients. The growing use of HSNS like Sermo and PatientsLikeMe has prompted public concerns about the risks that such online data-sharing platforms pose to the privacy and security of personal health data. This paper articulates a set of privacy risks introduced by social networking in health care and presents a practical example that demonstrates how the risks might be intrinsic to some HSNS. The aim of this study is to identify and sketch the policy implications of using HSNS and how policy makers and stakeholders should elaborate upon them to protect the privacy of online health data.

INTRODUCTION

Health social networking sites (HSNS) such as Sermo (http://www.sermo.com) and PatientsLikeMe (http://www.patientslikeme.com), virtual communities where people connect with each other around common problems and share relevant health data, have been increasingly adopted by medical professionals and patients. Online physician communities provide an online platform for doctors to share clinical insights, observations, and medical knowledge. Sermo is the largest physician community in which members share ideas about clinical cases, drugs, devices, and treatment options.[1] Online patient communities emphasize direct patient support, promoting disease awareness, and positive and proactive behaviors to stay healthy while living with a disease. PatientsLikeMe is a patient community that enables users to share details about their conditions, treatments, and symptoms, and provide support for one another.[2] Both community types can greatly improve health outcomes by embracing the wisdom of crowds.

Although powerful, social networking also opens the door for inappropriate access, misuse, and disclosure of personal health data. Users are sharing incredible amounts of health data on diabetes-focused social networking sites, although few of these sites offer either scientific accuracy or data protection.[3] Openness in the virtual community, combined with the growing availability of technologies supporting the secondary use of health data,[4] can significantly increase the likelihood of misuse of health data. Considering the sensitivity of health data, people may not wish for their data to be revealed to unauthorized individuals and entities because such disclosure may negatively affect reputation,[5] relationships, job opportunities, and insurance choices.[6] In the USA, electronic health records are legally protected by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) privacy rule and the enhanced Health Information Technology for Economic and Clinical Health (HITECH) Act. However, when individuals voluntarily post the same or similar data on various websites, their privacy is not protected by the HIPAA/HITECH Act.

There is an urgent need to address these concerns as the use of HSNS to deliver healthcare becomes increasingly common. While the privacy and security issues on HSNS are well recognized by previous research,[3 6 7] the literature specifically on the behavioral and policy issues is quite limited. This paper addresses the gap by sketching the key policy implications of using HSNS.

PRIVACY RISKS

A HSNS can have a number of privacy and security issues. First, the site may maintain a vast repository of users' profiles and keep it permanently. Users are increasingly sharing their private details on such sites and, for some people, privacy takes a back seat to the hope that some exchange will help them find a better treatment, manage their condition, or improve overall health.[8] Some people may reveal their health data for the sake of the greater good. Medical professionals also post sensitive data about their patients, community, and organizations in order to share advice on clinical situations or practice management.[9 10] Some professionals are even willing to provide personal data in exchange for the utility of the services and the occasional rewards provided by the site.[11] However, once users share their health data with the site, they typically have no control over retention periods for the data or associated metadata that will be maintained in perpetuity.

Second, the content produced by users may be revealed to both intended and unintended audiences. Since anybody can register on the website, anybody can view the content on the site. For example, any person or entity may create fake accounts in order to obtain data from unsuspecting users. Another related issue is that the website may exchange data with third parties without explicit consent.[12] For example, websites like checkMD (http://www.checkMD.com) may disclose users' personal information to its business partners and other third parties.[13] The site may also allow its users to incorporate features created by third parties and let users log into third-party sites using their profile data,[14] which implies that health data available within the protected site might be leaked to the web. Consequently, health data may be exposed to various data recipients without users' knowledge.

Third, the accumulated health data can be misused and/or exploited for various non-medical

To cite: Li J. *J Am Med Inform Assoc* 2013;**20**: 704–707.

uses. Some HSNS are commercial companies that have a business model based on harvesting health data for business and proprietary purposes. They may release health data to different data recipients, including doctors, pharmaceutical and medical device companies, researchers, and non-profit organizations. Aggregated health data are very valuable to commercial companies, such as drug and medical device manufacturers. Innovative data mining and health informatics technologies can link data produced from a variety of different sources to produce useful personal data aggregates or digital dossiers. Taken by themselves, certain pieces of data do not communicate much about a person, but taken together they could communicate a great deal. The digital dossiers would be immensely valuable to companies looking to market products or, in the case of insurers or employers, deny a policy or a job. The dossiers, maintained without direct government oversight, would also be an attractive target for hackers and identity thieves.

Lastly, another obvious issue is the scale of the security risk. While encrypted transmission will improve confidentiality, and authentication and access control will reduce non-authorized access, one 'hack' into the site, or one error by a site operator, or one misuse by the many other users of the site may compromise the digital profiles of numerous users.

ANALYSIS OF A PATIENT COMMUNITY

The above-mentioned risks might be intrinsic to some HSNS such as PatientsLikeMe. PatientsLikeMe is an online patient community that enables people with life-changing illnesses to share condition, treatment, and symptom information in order to monitor their health over time and learn from real-world outcomes. It was launched in 2006 when its founders recognized the need for community-based data sharing around specific diseases such as amyotrophic lateral sclerosis (ALS) and Parkinson's disease. The idea was that if users could share details about their treatments, symptoms, and conditions, better treatment plans and options could be identified by the collective wisdom on the site. As of December 2012, there were more than 175 000 registered members of PatientsLikeMe.

Like most open health communities, PatientsLikeMe has an openness policy so that users can agree to share all their health data on an ongoing basis, and users from around the world have already agreed to do so. The site collects and stores two types of data from users, shared data and restricted data. Examples of shared data may include information about biography, conditions, treatments, symptoms, outcomes, laboratory results, genetic information, survey responses, and connections to other people on the site. When a user chooses to share personally identifiable information (PII) like their name and photograph, the information will be treated as shared data. Only when a member enters personal information such as an email address and password as part of registering to use the site, is the information treated as restricted data. According to its website,[15] every piece of information users submit to the site, except for restricted data, may be shared with the community, other users, and partners. If a member chooses to designate My Profile as 'Public,' the shared data can also be viewed by non-members and linked with aggregated public reports. Restricted data are not automatically shared with, sold to, or displayed for other members or partners unless a member chooses the 'Public' privacy setting or opts in to a public registry.[15]

Since so many potential data recipients have access to the content, keeping the data confidential is a big challenge, if not impossible. As an open community, PatientsLikeMe cannot authenticate the identity of any other members with whom a member may interact in the course of using the site, or who may have access to a member's shared data.[15] Therefore, the site is especially vulnerable to identity thieves and data scrapers. This vulnerability is highlighted by a recent incident on the site. On May 7, 2010, PatientsLikeMe noticed that a new member of the site, using sophisticated software, was copying every single message off its 'Mood' discussion forum.[16] On May 20, 2010, PatientsLikeMe's president publicly disclosed the data scraping incident by a major media monitoring company Nielsen, to the site's members in a blog post.[17] This incident sparked a lively debate on the site about its data privacy practices.

Additionally, the company's business model largely relies on its ability to sell access to health data to pharmaceutical companies and others. The user-generated content helps its partners better understand the real-world medical value of therapies, drugs, and medical devices so they can improve their products and speed up the development of new solutions for patients. Health data have also been used by its own research team to conduct studies and publish reports that are accessible to the general public. Thus, users have little control over the sharing of health data within and beyond the community. The underlying assumption appears to be 'proceed at your own risk' because users consent to the site's terms of use. The problem is that users may skip over the terms of use and privacy policy when joining the community.

PRIVACY POLICIES

While HSNS like PatientsLikeMe have laudable aims, including improving health outcomes and advancing medical research, the inherent openness of social networking and self-motivated data sharing, combined with extremely valuable and sensitive health data, can make users vulnerable to myriad privacy violations. The stakes increase with the amount of health data disclosed, the number of data recipients, and the increasing use and disclosure of health data for non-medical purposes. Although users may have very different viewpoints about privacy, preserving privacy can be extremely helpful for all users, and especially those with chronic diseases and those with stigmatized illnesses. A recent survey shows that 58% of social network users restrict access to their profiles and 67% of female profile owners restrict access to friends only compared with 48% of male profile owners.[18] So far discussions about how best to protect the privacy of health data have been narrowly focused on users' consent[19] or on privacy settings.[20 21] Yet, empirical and theoretical research suggest that users often lack enough information to make privacy-sensitive decisions and, even with sufficient information, are likely to trade-off long-term privacy for short-term benefits.[11] Users' online practices are also constrained by their degree of digital literacy and by the technical design of the website, which may impede easy management of settings and consent regarding the use and disclosure of personal data.[22] Hence, it is important to ensure a comprehensive privacy framework for the social network environment.

If such a framework is to be developed, several policy implications need better evaluation. The first involves the interdependencies between data sharing and risks. The amount of the data shared by users is positively correlated with their experiences of risks. The more data users share, the more risks they encounter, and the more policy attempts to limit the risks, the more it also limits the utility of the services. The only tried and true solution to social network privacy issues is either to limit the data shared or to protect the data shared. To mitigate the risks, users must share the minimum amount of personal data to accomplish the intended purpose. For instance, users

may not provide PII such as their real name and national identification number. Yet, users often share many more personal details on HSNS than they would otherwise because complete information is pivotal for effective health care. Many users wrongly assume that the existence of a 'privacy policy' means that their health data will be protected, even when the policy is an openness policy.[19] Even with appropriate protections built into HSNS, some have not customized their settings for optimal protections.[23]

A second related concern is how to get users engaged in protecting their own privacy. Although information sharing is inextricably linked to improving health care, it is important to ensure that personal data are not inadvertently shared with an unintended audience. A particular user's data are either visible to the public, or, if the user is aware of privacy issues and able to use the settings of the respective website, to a somewhat selected group of other members. Some sites have multiple levels of privacy. It is critical to understand the privacy settings available within each of these sites and to apply the maximum level of privacy available. However, settings may change without prior notification, be difficult to fully implement, and ultimately will not change the content other members can access.[20] Users should also closely monitor how data flows on the website because the context surrounding health data or the technology may be dynamic. Since users are not necessarily aware of self-protection, privacy awareness and education is an important element of the framework. Users may not wish for their personal data to be revealed to a possible unintended audience, which can include marketers, employers, insurers, and others, but they may not have the knowledge and technical skills to protect their privacy. Even in the absence of regulation, the site has an ethical duty to minimize risks to users whose data it gathers . To prevent improper access and use of the site by an unintended audience, it makes good business sense for the HSNS provider to work toward a 'privacy by education' principle for cultivating privacy-literate users. The site should be encouraged to inform users of the dangers of inadvertently disclosing PII online. It could also provide a user-friendly way for users to protect privacy. For instance, the privacy policy could state how the individual can request removal of publicly displayed PII. Furthermore, the site could notify individuals of any material changes to its PII collection, use, or disclosure practices before making the change in the privacy policy.

The third issue is how to build privacy and security into the platform while still tapping the value of user-generated content. Although some HSNS are designed in part to provide personalized health feedback to users, their business model largely depends on sharing the content with commercial entities for research and other purposes. The provider may not willingly offer too much privacy because this makes it harder for users to put their disease experiences in context and impedes the attempt to fulfill business and proprietary objectives. However, without effective privacy and security controls, the platform can be a tempting target for malicious individuals and entities. To overcome the problem, privacy and security have to become properties of the architectural components of the 'privacy by design' system.[24] This principle means that the site must implement reasonable controls and processes to protect PII from unauthorized access, use, disclosure, or distribution by default, without requiring a user's action. Data anonymization techniques enable health data to be used for a wide range of purposes while minimizing the risks to individual privacy. The data exposed could be de-identified by either the safe harbor method, which requires the removal of enough PII, or the statistical method, requiring a qualified statistician to attest that the data raise very low risk of re-identification.[25] Furthermore, it could build and maintain adequate data security. Embedding privacy and security into the site's design allows for improving privacy settings for authorized users, preventing unauthorized access, use, and disclosure of PII, and providing transparency about uses of health data.

The fourth issue is how to hold individuals and entities accountable for non-medical uses of health data. Widespread use of health data for business and proprietary purposes heightens the urgency to engage the public in a policy dialog about data privacy. In the event that individuals and entities violate users' privacy, new legislation, if not the HIPAA/HITECH Act, is needed to protect the privacy of online health data. The legislation should mandate that the provider must give individuals options to control how their health data are used for non-medical purposes. The legislation should further prohibit inappropriate commercial uses resulting, for example, in discrimination, even with express consent. The provider should be encouraged to enforce adequate data de-identification mechanisms against risks such as the inappropriate use and exploitation of data sharing. The legislation should also enact prohibitions for the unauthorized re-identification of anonymized data.[25] Legal and financial remedies must exist to address any privacy violations or security breaches. Legislators and stakeholders could continually press the provider to adopt a privacy by regulation principle for building a privacy-sensitive site.

These recommendations (table 1) should inform deliberations about the privacy of online health data including HSNS and other e-health technologies. Some of the principles have been partly suggested for other online settings, including electronic medical records,[25] personal health records,[19 24 26] general social networks,[7] weblogs,[8] etc. Because potential privacy holes exist not only in the platform but also in users' behavior and our legal system, 'privacy by education', 'privacy by design', and 'privacy by regulation' are three inter-related principles that complement each other for privacy protection. In policy discussions we often focus on one of the three key principles and forget about the others. All three must be balanced to ensure a private and secure social network environment.

CONCLUSION

The growing use of HSNS presents significant risks for individual privacy. Users themselves play a critical role in helping to safeguard their own data. However, users often are unaware of

Table 1 Recommendations for protecting the privacy of health data

Privacy policy	Recommendations
Privacy awareness	Sharing the minimum amount of person-specific data to accomplish the intended purpose. When in doubt, err on the side of providing less data
Privacy by education	Privacy-awareness education; user-friendly way of setting privacy; use and protection of personally identifiable information (PII) policy; advance notice of any material changes to the privacy policy
Privacy by design	Building data protection and privacy by design into the platform; sharing anonymized data within and beyond the community
Privacy by regulation	Ensuring consent to non-medical uses before users' data are used; banning unauthorized re-identification of anonymized data; prohibiting inappropriate uses of health data

the risks and do not have the skills and ability to protect their privacy. The provider is reluctant to offer protections because they may reduce the benefits of open communication and data sharing. But even if privacy mechanisms were built into the platform and even if users were aware and competent in optimizing their privacy settings, users would still be exposed to potential privacy violations by the provider and its partners. Addressing these pressing challenges ultimately requires a policy framework for the access, use, and disclosure of health data for non-medical purposes. This study identified the policy implications of social networking that should inform efforts to protect the privacy of health data. 'Privacy by education,' 'privacy by design,' and 'privacy by regulation' are three key principles that lay out the groundwork for future research. Policy makers and stakeholders should elaborate upon the principles through discussions that will produce, over time, user awareness and understanding, appropriate public policies, and supportive technologies that adequately protect the privacy of online community inhabitants.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Glenn B. A tangled web. *Med Econ* 2009;86:30–6.
2 Wicks P, Massagli M, Frost J, *et al*. Sharing health data for better outcomes on PatientsLikeMe. *J Med Internet Res* 2010;12:e19.
3 Weitzman ER, Cole E, Kaci L, *et al*. Social but safe? Quality and safety of diabetes-related online social networks. *J Am Med Inform Assoc* 2010;18:292–7.
4 Safran C, Bloomrosen M, Hammond W, *et al*. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. *J Am Med Inform Assoc* 2007;14:1–9.
5 Solove DJ. The end of privacy? *Sci Am* 2008;299:100.
6 var der Veldan M, El Emam K. "Not all my friends need to know": a qualitative study of teenage patients, privacy, and social media. *J Am Med Inform Assoc* 2013;20:16–24.
7 Adams SA. Blog-based applications and health information: two case studies that illustrate important questions for Consumer Health Informatics (CHI) research. *Int J Med Inform* 2010;79:e89–96.
8 Li J. Improving chronic diseases self-management through social networks. *Popul Health Manag* 2013 (in press).
9 Moubarak G, Guiot A, Benhamou Y, *et al*. Facebook activity of residents and fellows and its impact on the doctor-patient relationship. *J Med Ethics* 2011;37:101–4.
10 Thompson LA, Black E, Duff WP, *et al*. Protecting health information on social networking sites: Ethical and legal considerations. *J Med Internet Res* 2011;13:e8. http://www.jmir.org/2011/1/e8/ (accessed 25 Mar 2013).
11 Acquisti A, Grossklags J. Privacy and rationality in individual decision making. *IEEE Secur Privacy* 2005;3:26–33.
12 Williams J. Social networking applications in health care: threats to the privacy and security of health information. *Conf Proc SEHC* 2010;2:39–49.
13 Terillion Privacy Policy. http://www.terillion.com/privacy-policy/ (accessed 25 Mar 2013).
14 Mooradian N. The importance of privacy revisited. *Ethics Inf Technol* 2009;11:163–74.
15 PatientsLikeMe privacy policy. 2012. http://www.patientslikeme.com/about/privacy (accessed 25 Mar 2013).
16 Angina J, Stecklow S. 'Scrapers' dig deep for data on Web. *Wall Street J* Oct 11, 2010. http://online.wsj.com/article/SB10001424052748703358504575544381288117888.html (accessed 25 Mar 2013).
17 Heywood B. Transparency, openness and privacy. 2010. http://blog.patientslikeme.com/2010/05/20/bentransparencymessage/ (accessed 25 Mar 2013).
18 Pew Internet & American Life Project. Privacy management on social media sites. 2012. http://pewinternet.org/~/media//Files/Questionnaire/2012/Privacy_management_on_social_media_survey_questions_excerpt.pdf (accessed 25 Mar 2013).
19 McGraw D, Dempsey J, Harris L, *et al*. Privacy as an enabler, not an impediment: building trust into health information exchange. *Health Aff* 2009;28:416–27.
20 Landman MP, Shelton J, Kauffmann RM, *et al*. Guidelines for maintaining a professional compass in the era of social networking. *J Surg Educ* 2010;67:381–6.
21 Dowdell EB, Burgess AW, Flores JR. Online social networking patterns among adolescents, young adults, and sexual offenders. *Am J Nurs* 2011;111:28–36.
22 Livingstone S, Brake DR. On the rapid rise of social networking sites: new findings and policy implications. *Children Soc* 2010;24:75–83.
23 MacDonald J, Sohn S, Ellis P. Privacy, professionalism and Facebook: a dilemma for young doctors. *Med Educ* 2010;44:805–13.
24 Williams JB, Weber-Jahnke JH. Social networks for health care: addressing regulatory gaps with privacy-by-design. *Conf Proc PST* 2010:134–43.
25 McGraw D. Building public trust in uses of Health Insurance Portability and Accountability act de-identified data. *J Am Med Inform Assoc* 2013;20:29–34.
26 Collins SA, Vawdrey DK, Kukafka R, *et al*. Policies for patient access to clinical data via PHRs: current state and recommendations. *J Am Med Inform Assoc* 2011;18 (Suppl 1):I2–7.