Designing Privacy Into Online Communities

Catherine Dwyer
Pace University
cdwyer@pace.edu

Starr Roxanne Hiltz
New Jersey Institute of Technology
hiltz@njit.edu

ABSTRACT
Participation has exploded in online communities such as social networking sites, media sharing sites, and blogging sites. This widespread participation has generated substantial concern about the privacy implications of use. Although most online communities include privacy management features, it is not well understood how members of online communities take advantage of privacy management. To explore this issue, a study was conducted that compared privacy management on Facebook and MySpace, two popular social networking sites. Out of 222 subjects, 18.9% or nearly one in five had suffered a privacy incident within the past year. In a particularly striking result, less than half of those who suffered a privacy incident either reviewed or changed their privacy settings. So even though privacy incidents occur regularly, these findings indicate privacy management tools are not extensively used to protect privacy. This provides evidence of poor design of privacy management on social networking sites. To improve design, we argue that privacy must be conceptualized as a quality of an online space, rather than as a collection of access settings to be managed by individual members. This moves privacy from an individual consideration to the level of a structural component of a system. We offer an analysis of current privacy management practices and offer suggestions for improving the design of privacy management in social software.

Keywords
Online privacy management, social software, system design

INTRODUCTION
Social networking sites have grown to be the most popular type of computer-mediated communication system to support online communities. This paper examines privacy issues and behavior on the two most popular social networking sites and suggests guidelines for improved privacy management requirements.

An important first step in software design is the development of requirements. In software design, a requirement defines a promised component of a system. Requirements fall into two general categories: tasks the system will perform, and the system’s general qualities or universal properties. The first type is referred to as functional requirements, and the second as non-functional requirements.

Non-functional requirements are the emergent qualities of a system, used to judge its overall performance. This term is used in contrast to functional requirements, which define exactly what tasks a system will carry out. Examples of non-functional requirements include usability and security. A system that does not deliver its non-functional requirements is more likely to fail or not be effective (Sommerville 2001).

When designers consider privacy, do they believe it to be a functional requirement or a non-functional requirement? Is privacy an option that members of online communities can select, or is it a property of the community itself? We argue here that as we design new online communities, we must think of privacy as a non-functional system level requirement, rather than a group of access settings for each member. In other words, privacy needs to belong to an online space, not be a collection of settings attached to each individual member.

To understand this more clearly, consider how we perceive privacy levels in offline social spaces. For offline social spaces, privacy is signaled by physical characteristics: low lighting, enclosed spaces, and relative isolation from others. People who want to conduct a private conversation can recognize the privacy levels of an offline space based on physical properties. Online, privacy levels are not signaled by the inherent properties of the online social space in any clear way, except for the common assumption that nothing is private.

Following a review of the literature on privacy and social networking sites, this paper reports on data collected from an empirical study of online privacy management. The next section provides a review of privacy research in social networking sites. This is followed by a description of a study that compared privacy management on two social networking sites,
Facebook and MySpace. The next section presents an analysis of how to design privacy into online communities. This is followed by a concluding section which describes the study’s limitations and plans for future research.

PRIVACY AND SOCIAL NETWORKING SITES

Boyd and Ellison define social networking sites as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system,” (2007).

Social networking sites are rapidly evolving socio-technical systems. How does technology change the nature of social interaction? Boyd and Heer suggest that “the architectural structure of digital life alters the ways in which conversations can and do occur . . . digital communication now incorporates multiple forms of media bridging the physical and digital,” (Boyd and Heer 2006). The components of one’s profile, for example blogs and photo albums, can be thought of as elements of an ongoing conversation. Social interaction also changes because the simultaneous private and public nature of profiles is not consistent with traditional (i.e., offline) understanding of communication and self presentation.

Creating an interesting profile gives someone status within a social networking site. It enables active connections with friends. It also brings up the challenge of negotiating connections with an unknown audience, leading to uncertainty as to the real boundary between public and private. This becomes especially difficult when members try to communicate context as they present themselves. Offline, “friend” relationships have many levels. People develop relationships with work colleagues, neighbors, school friends and family. Within these relationships there are also degrees of closeness. But within social networking sites, the friend status is binary: friend or not. Since the friend status is the primary technical boundary used to control access to information, members have to consider that the information they intend to share only with close friends is likely to leak into other contexts. While on the surface members seem blasé about unknown others viewing their profile, there are two people that members greatly fear will view their profile: “boss and mother” (Boyd 2006).

Boyd explores how technical features of social networking sites change the nature of social interaction in (2006) and (2007). Specifically, social networking sites enable profiles to be searched. They also allow profiles to persist, as well as be copied. By enabling search, this gives any other member the potential to find you. This means members can not accurately define the potential audience for their profile. The consequences of a persistent digital identity mean that actions taken in the heat of the moment can remain visible well after tempers have cooled down. The ability of profiles to be copied can undermine trust in someone’s identity.

In addition, friending is a different process online compared to developing relationships offline. Acknowledging an online friend takes one click. Once the friend status is established, that connection will likely remain unless there is an explosive end to the relationship. In addition, friending is another means of self expression. As more music and film celebrities develop profiles, members add them as friends as a way of specifying their tastes in music. The use of profiles for self expression was also found in a study conducted by Dwyer (2007).

The performance aspect of profiles, driven by the need for self-expression as well as social competitiveness, can encourage active public sharing of information. The more you share, the more attention you can get. However, all this shared information can be easily accessed by other parties. Liu and Maes found that it is relatively simple to harvest information from a profile and use it to generate recommendations (2005). This raises privacy concerns, because this information can be used for other purposes besides recommendations, such as marketing or profiling of individuals by law enforcement.

Privacy Implications

A number of researchers have studied the use of social networking sites with the aim of determining the scope of information that members are willing to reveal. Gross and Acquisti conducted a study at Carnegie Mellon University, extracting 4540 profiles, “virtually the entire CMU Facebook population at the time of the study,” (Gross and Acquisti 2005, p. 78). The vast majority of profiles included personal information directly related to identity. The study found that 90.8% of profiles contain an image (photograph), 87.8% of users reveal their birth date (a key piece of information for identity theft), 39.9% list a phone number (including 28.8% of profiles that contain a cell phone number), and 50.8% list their current residence.

The majority of users also disclose their dating preferences (male or female), current relationship status (single, married, or in a relationship), political views (from “very liberal” to “very conservative”), and various interests (including music, books, and movies). A large percentage of users who self-identify as “in a relationship” (62.9%) include a link to their partner’s Facebook profile.
Using this dataset of information available within the CMU Facebook network, Gross and Acquisti conducted an analysis of privacy risks associated with the use of Facebook. With the information that a large number of students provide in their Facebook profile, it is quite easy to determine the physical location of students for large portions of the day. This can be deduced from either a student’s class schedule or their dorm address. The study found 15.7% of female students and 21.2% of male students provide this information, making them vulnerable to real world stalking.

A larger proportion is vulnerable to cyber stalking through the use of AIM (AOL instant messenger). AIM allows users to add “buddies” without the knowledge or permission of the other party. Once added to a buddy list, a cyber stalker can keep track of when the other person is online. More than 77% of the profiles downloaded from CMU contain an AIM screen name.

A more subtle threat has to do with a technique known as re-identification, defined as the process of matching anonymous data with the specific individual who provided the data. For example, it can be used to collect an individual’s browsing activities for data mining purposes, or to uncover sensitive medical information (Rosenblum 2007). Overall, 45.8% of the members captured in the CMU dataset reveal their birthday, gender, and current residence, making them vulnerable to re-identification or potentially identity theft.

Another threat analysis of Facebook was conducted by Jones and Solten (2005), who collected Facebook data from four institutions. Noting that they were able to create and use accounts at four institutions to data mine tens of thousands of profiles in a week, the authors concluded that commercial data mining is not only possible, it is easy to do.

Privacy risks within social networking sites were discussed in an article published in IEEE Security and Privacy (Rosenblum 2007). For students who spend four years using social networking sites to document wild outings, boorish behavior, and insensitive or racist remarks, the idea of a prospective employer viewing one’s profile is quite frightening. As described in the studies by Gross and Acquisti and Jones and Solten, access to social networking sites is quite porous, and it is naïve to assume that a profile public to some people is private to everyone else. Companies now use search engines and also access social networking sites to conduct background checks on prospective employees. This is especially valuable because federal fair hiring practices have restrictions on what questions can be asked in a job interview. As one hiring officer explained, “You really do get a lot of information you can’t ask for in the job interview, but you go on the Web and it is all there,” ibid, p. 46. As companies seek to choose employees based as much on their values as their raw abilities, reviewing these sites is a way to determine if there is “something about their lifestyles that we might find questionable, or that we might find would go against the core values of our corporation,” ibid.

The implications of Internet information playing a role in the hiring process has been incorporated into a Harvard Business Review case study entitled “We Googled You,” (Coutu, Joerres et al. 2007). This case study presents the hypothetical case of a recent Chinese American college graduate, Mimi Brewster, applying for a job at a luxury goods provider who hopes to grow its market in China. While Mimi speaks Chinese fluently, a search of the Internet finds that she played an active role in protests of China’s treatment of a dissident journalist. This case illustrates the dilemma faced by companies that use the Internet to investigate candidates. The company in question here faces one risk by passing on an otherwise excellent candidate, and a different risk by hiring someone who has publicly protested the actions of the Chinese government.

**COMPARISON OF PRIVACY MANAGEMENT ON FACEBOOK AND MYSPACE**

In response to the growing participation in social networking sites despite heightened privacy risks, a study was designed to gain more insight into how individual members control their online privacy. A survey was designed to capture data on privacy management within two sites, Facebook and MySpace. A detailed report on an earlier version of this study can be found in (Dwyer, Hiltz et al. 2007). Two versions of the survey were created, with the same questions re-worded to refer to either Facebook or MySpace. The survey was administered using the online survey tool Zoomerang (www.zoomerang.com), and consisted of 61 questions. Subjects were offered five dollars in compensation for completing the survey. The target population for this study was members of the New Jersey Institute of Technology (NJIT) community who participate in Facebook or MySpace. This includes students, faculty, and alumni. Using the tool Friend Blaster Pro (www.friendblasterpro.com), the MySpace site was searched for NJIT subjects. This was accomplished by first looking for members who mention NJIT in their profile. Additional searches were made for groups on MySpace with an NJIT affiliation. The results from these searches were combined, and duplicates were dropped. This resulted in a target population of 986 members on MySpace with an NJIT affiliation identified through these search mechanisms.

At least one attempt was made to contact each identified MySpace subject, although 22 subjects blocked messages from unknown MySpace members. Every subject that could be contacted was sent at least one invitation to participate, and 927 out of the 964 received two invitations. The survey was fully completed by 115 MySpace subjects, resulting in a response rate of 11.9%.
The target Facebook population for this study was members of the NJIT Facebook network. As of September 27, 2007, the number of members in the NJIT Facebook network was 6,467. From this network, potential subjects were identified with a random search tool provided by Facebook. Using this process, about 900 profiles were collected. Once duplicates were eliminated, this resulted in a sample frame of 778 potential subjects. The survey was completed by 107 Facebook subjects, for a response rate of 14.2%.

**Results**

A total of 222 subjects completed a valid response to the survey, 107 completing the Facebook survey and 115 completing the MySpace survey. There are significantly more male (160) than female (62) subjects for this study, which is consistent with the target population of NJIT students, staff and alumni. There is no significant difference in the gender distribution between the Facebook subjects (30 female and 77 male) and the MySpace subjects (32 female and 82 male).

The ages of the subjects ranged from 18 to 69. The mean value is 24.53, the median is 23, and the mode is 22. There is a small but significant difference in age between the Facebook and MySpace subjects. The Facebook age mean is 23.07, and the MySpace age mean is 25.89 ($t$-value = -2.911, sig. = .002, df = 220).

The largest group of subjects was comprised of undergraduates, with 58 from Facebook and 55 from MySpace. This is followed by graduate students (28 Facebook and 17 MySpace), faculty or staff (2 Facebook and 3 MySpace), and 57 indicating they are currently not a student (18 Facebook and 39 MySpace). There was a higher number of non-students that are part of the MySpace population (34%) versus the Facebook population (17%). The differences in school status between the two populations is significant (Pearson Chi-Square =19.92, df = 10, sig. = 0.029).

The subjects of this study report quite active use of social networking sites. 45 Facebook (42.1%) and 39 MySpace subjects (33.9%) use the site every day or several times a day, and another 47 Facebook (43.9%) and 47 MySpace subjects (40.9%) use the site at least one a week. Only about 18% (42 out of 222) use the site infrequently (once in a while). There is no significant difference in the frequency of use when comparing Facebook subjects to MySpace subjects.

About 45% of subjects (48 from Facebook and 52 from MySpace) report use of social networking sites about as frequently as a year ago. 24% (19 from Facebook and 33 from MySpace) report they are using these sites less frequently than a year ago, and 29% report they are using the sites more frequently (38 from Facebook and 27 from MySpace).

<table>
<thead>
<tr>
<th>Over the past year did you experience any incidents that led you to be concerned about privacy when using [name of social networking site]?</th>
<th>Facebook</th>
<th>MySpace</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Yes</td>
<td>16</td>
<td>15.0%</td>
<td>26</td>
</tr>
<tr>
<td>No</td>
<td>91</td>
<td>85.0%</td>
<td>89</td>
</tr>
<tr>
<td>Total</td>
<td>107</td>
<td></td>
<td>115</td>
</tr>
</tbody>
</table>

(only those who answered "Yes" were asked the next two questions).

<table>
<thead>
<tr>
<th>Did you review your privacy settings after this incident?</th>
<th>Facebook</th>
<th>MySpace</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>8</td>
<td>50.0%</td>
<td>12</td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>50.0%</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td></td>
<td>26</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did you make any adjustments or changes to your privacy settings after this incident?</th>
<th>Facebook</th>
<th>MySpace</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>8</td>
<td>50.0%</td>
<td>11</td>
</tr>
<tr>
<td>No</td>
<td>8</td>
<td>50.0%</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td></td>
<td>26</td>
</tr>
</tbody>
</table>

Table 1 Summary of Prior Issues With Privacy
Report of Privacy Incidents

An important indication of the value of privacy management tools is whether they are used in response to a privacy incident. To test this, subjects were asked whether they had a personal experience with respect to privacy problems on these sites. The results of those questions are summarized in Table 1. In the first question, the text [name of social networking site] was replaced by either Facebook or MySpace, depending on the version of the survey.

The subjects were asked the following question: “Over the past year did you experience any incidents that led you to be concerned about privacy when using [name of social networking site]?” A total of 16 Facebook subjects (15.0%) and 26 MySpace subjects (22.6%) answered yes to this question. Combining both totals, this shows that nearly one in five (42 out of 222) subjects reported a problem with privacy over the past year. Although more MySpace subjects reported incidents, the difference is not significant.

Those 42 subjects reporting a privacy incident were asked two follow up questions. The first question was: “Did you review your privacy settings after this incident?” Eight Facebook subjects said Yes, and eight said No. A total of 12 MySpace subjects said Yes and 14 said No, for a total of 20 reporting Yes (47.6%) and 22 reporting No (52.4%).

The second question was: “Did you make any adjustments or changes to your privacy settings after this incident?” Once again eight Facebook subjects said Yes, and eight said No. There was a wider gap in MySpace, with 11 reporting Yes and 16 reporting No. Adding these together, 19 out of 42 (45.2%) reported making adjustments in their privacy settings. Note that despite Facebook’s heavy emphasis on privacy management (Facebook 2008), only half of Facebook subjects reported accessing those tools in response to a privacy event.

The subjects provided the following detail about these incidents:

- “Constant spam from adult content type accounts has basically driven me away from MySpace and more towards the use of Facebook.”
- “I found out my Facebook page was on a suspected sex offender's computer in another state last year (along with many other girls) which made me more cautious of what personal content I put on Facebook. Also, I don't think (this may be ignorance) you can remove your email from your profile page.”
- “My 17 year old cousin's MySpace account was hacked into and was filled with pornographic images, comments, and videos.”
- “My profile was hacked by the 'Create Your Own South Park Character' website. While using this site, I gave up my MySpace password to 'automatically post to profile.' Once the South Park site had my password, it was used it to leave many, many advertisements for their site in all my friend's comments. I changed my password & the hacking stopped. Lesson learned.”
- “Not really except a person whom I was stalked by before tried to contact me through Facebook again which was scary but that is about it.”
- “Not with Facebook, but someone I helped at work tracked me down on MySpace once. That was pretty creepy.”
- “Random people that I don't know would message me asking me very personal questions that I did not feel comfortable answering.”
- “Yeah the girl who committed suicide because of online bullying. And I knew someone who met up with people she met online, it made me very uncomfortable.”
- “Yes, but I've only had one person contact me that I did not know. And rumors just started this month about a guy that apparently hacks into your profile through friend's pages. I find it highly unlikely, and it's nice that after this long on Facebook, it's stayed a secure place. It's so refreshing after MySpace has just turned into a complete spamfest.”

These subjects were very willing to share their thoughts about their use of these sites, as well as their concerns and fears. They are very enthusiastic users, but do recognize there are risks to this use. Their comments indicate they are not apathetic about privacy, a common excuse for low levels of use of privacy management tools. For both sites, these results expose weaknesses in the effectiveness of existing privacy management tools.
HOW TO BUILD PRIVACY INTO ONLINE COMMUNITIES

If a site’s members suffer a privacy incident, then you would expect them to apply privacy management tools to strengthen their privacy. The weak results found in this study for subjects who suffered a privacy incident suggest that the design of online privacy management needs to be improved.

As described above, non-functional requirements refer to fundamental qualities or properties of a system. In order to build privacy into online communities, we argue that designers should think of privacy as a non-functional system level requirement. This moves privacy from an individual consideration to the level of a structural component of a system. In other words, privacy needs to become a quality of an online space. The conclusion to be drawn from this study is that privacy cannot be designed as a group of settings individually adjusted, as if privacy were the same as a preference for a certain font or text color.

How does approaching privacy management as a non-functional requirement change the design of privacy management? How can privacy be designed as a quality of an online space that can be easily recognized? Methods must be developed that signal both the individual’s privacy status, as well as the social norms with respect to privacy of other groups and members of the site. Three suggestions follow on how to accomplish this.

1) Evaluate the Privacy Level of Each Component

Usability and security are examples of non-functional requirements. Just as each component of a system can be evaluated as to its usability and security, so should each component be evaluated as to its privacy. Specifically, the privacy level of each component needs to be measured and signaled through the interface to members. Social networking sites have many components in addition to individual profiles. But what are the privacy levels of these other components? For example, Facebook has pages for groups as well as consumer products, bands, and politicians. Participation in groups and pages for bands is public to everyone on Facebook. However, there is no clear indication in the appearance of a group page that signals the public nature of participation.

An illustration of this problem is the case of Caroline Giuliani, whose father Rudy was a candidate in the 2008 Republican presidential primaries. Like most other college freshmen, she had a Facebook profile, and joined the network of Harvard University where she is a student. She also joined a Facebook group in support of Barack Obama, a Democratic candidate for president. Her support of a Democratic candidate while her father was a Republican candidate became the talk of blogs and the online media, especially since Caroline was estranged from her father and had refused to appear publicly with him at campaign events. Within 24 hours of the revelation, Caroline had dropped her membership in Facebook (Caldwell 2007). In this case, although Caroline’s profile was private, her membership in a Facebook group was public to any other Facebook member. What is particularly confusing here is that the same piece of information – her membership in the Barak Obama group – had both public (within the group) and private (within her profile) levels of visibility. This needs to be clearly signaled in the site’s design.

2) Provide Privacy Feedback

Before heading out to work in the morning, most people get feedback on their appearance by looking in a mirror. But you can not hold up a mirror to your profile on a social networking site to see what others can see. What is really needed in order to provide feedback is a privacy WYSIWYG (“what you see is what you get”). Providing an option that shows what is visible to friends versus strangers can reinforce better privacy habits. Without a mirror, you can not tell for sure if you have something in your teeth. Without a privacy WYSIWYG, can you be sure your boss won’t see that silly picture you just posted?

The importance of providing privacy feedback is reinforced by remarks from one of our Facebook subjects. She reported little or no privacy concern because “if I don’t accept someone as a friend they can’t look at my profile so it doesn’t bother me.” However this subject had it all wrong, because her profile is quite visible to members of the Facebook NJIT network (i.e., the authors), and the information available includes her full birthday, her instant messenger screen name, as well as access to her “wall,” i.e., the messages posted from Facebook members.

Besides knowing your current privacy status, it is also important to be able to test the effect of changes to privacy settings. If a member makes changes to their privacy settings, those changes should be signaled in the appearance of their profile. Currently, when members adjust their privacy settings, they receive no feedback as to the consequences of their actions. They may think they have improved their privacy, but they have no way to know for sure. As one subject explained, “The privacy settings are way too complicated. Sometimes I am unsure if I have achieved what I wanted after doing some changes.”
3) Publish Privacy Norms

Social norms do have the ability to influence behavior, but in order to do so they must be visible. While all sorts of popularity metrics are available on social networking sites, such as the size of a group or the number of wall postings someone has, no metrics are available that capture norms with respect to privacy settings. Unless norms are calculated and clearly displayed, they will not have any influence on others’ behavior.

An example of a privacy norm could be the percentage of members who display their email address to friends as well as strangers. In addition, privacy norms can also be determined for different populations, such as the entire site, as well as different sub-groups. An especially valuable set of metrics would be those calculated for a member’s set of online friends. Because this list likely contains close friends and/or like minded individuals, it has more ability to provide a strong normative effect. By comparing the range of settings for site as a whole, versus a group of friends, each member should be able to determine where their settings lie with respect to community norms.

Research does show that social software influences the structure of social spaces and the development of norms of use (Humphreys 2007). Just as offline social attitudes with respect to privacy have evolved over time, so must online conceptions of privacy evolve. This will require a more concerted effort to build privacy into the structure of social software, so that privacy is as apparent for online social spaces as the privacy levels of offline social spaces. This will require the conceptualization of privacy as a non-functional requirement that is pervasive within each component of social software.

CONCLUSION

This paper presents data from a study showing a weak connection between members’ experience with privacy incidents and the use of privacy management tools within social networking sites. The implications of this result are that the realm of privacy design be shifted to the social space itself, and away from the individual. This is more in line with how people negotiate privacy in offline social spaces. As we move about offline, we look for cues as to the privacy level of the spaces we enter. Therefore, we need to build cues into our online spaces, in order to define and reinforce new norms and standards of online privacy.

Limitations to this study are related to the administration of the study. Subjects for this study were members of the NJIT community that participate on Facebook and MySpace. The NJIT community is not representative of the populations of either site. In addition, due to technical constraints the subjects were recruited through one way in Facebook (through email) versus MySpace (by using the site to contact subjects). This difference could have introduced an unknown bias in the type of subjects who responded.

It is also important to keep in mind that the nature of these sites is in constant flux. New features are rolled out on a frequent basis. Privacy policies can change. Therefore these results only report on a snapshot in time. Extending any findings to the future, or to other sites, must be done with caution.

The results of this study show there is much to be learned with respect to how individuals manage their online privacy. An additional qualitative study is being planned to uncover greater detail as to the strategies members employ to protect their privacy on these sites. In addition, more work will be carried out to further develop and test the privacy design concepts introduced in this paper.

It seems the use of technology becomes more social every day. This is despite a palpable disrespect for privacy and lack of basic civility. These anti-social forces must be addressed by designers of socio-technical systems, or their use will eventually wither. The long term success of social software will depend on the ability of designers to build agile, reliable, and of particular importance, easily understandable and modifiable privacy protection mechanisms.

ACKNOWLEDGMENTS

The authors wish to thank the members of the NJIT community who enthusiastically shared their experiences with social networking sites. We also wish to acknowledge the important contributions of Marshall Scott Poole, George Widmeyer, Katia Passerini, and Naomi Rotter who provided invaluable assistance in the design and implementation of this research.

REFERENCES


