Privacy Codes of Practice for the Social Web: The Analysis of Existing Privacy Codes and Emerging Social-Centric Privacy Risks

Girma Nigusse and Bart De Decker
Katholieke Universiteit Leuven, Department of Computer Science, Leuven, Belgium
Girma.Nigusse,Bart.DeDecker@cs.kuleuven.be

Abstract
Privacy codes of practice developed in the 1980s in response to the concerns about the introduction of computerized databases, automated personal information processing systems, and easy transmission of personal information across national and international boundaries. The current data-practice guidelines, laws, and model-codes have been modeled based on these early codes. Starting from the early 2000, the socialization of the Web introduced a new kind of private information flow model and new privacy risks. In this paper, we analyzes the core principles of existing privacy codes with respect to emerging social-centric privacy risks. Our major contribution in this work is twofold. Legally - this evaluation is useful to enhance current privacy codes for the Social Web. Technically - this evaluation is useful to design the future social-centric privacy policy languages and their data handling specifications.

Introduction
Current information privacy codes are built on the 1970s and the early 1980s Fair Information Practice codes (Rotenberg 2001) (Winn 2006). In the early 1970s, concerns regarding the rise of computerized databases and automated personal information processing systems have led to the introduction of the Code of Fair Information Practice principles (HEW 1973). At the beginning of the 1980s, the advance of automatic data processing systems made it easy to transmit personal information across national and international boundaries and so triggered a number of data privacy challenges including the unlawful storage of personal data, the storage of inaccurate personal data, and the unauthorized disclosure of such information. To meet these challenges, the Organization for Economic Co-operation and Development (OECD) codified data privacy guidelines that set benchmarks for the framing of privacy policies and legislation within the OECD member states (OECD 1980) (Rotenberg 2001). The Canadian Standard Association’s (CSA) Model Code for the Protection of Personal Information built on the OECD guidelines and it translated the concept of Fair Information Practices principles into technical codes (CSA 1996).

In the last two decades, the principles of Fair Information Practice have been substantially incorporated into a number of international, regional, and national privacy codes and laws, such as the United Nations Guidelines for the Regulation of Computerized Personal Data Files (UN 1990), the European Unions Data Protection Directive (EU 1995), and the Asia-Pacific Economic Cooperations Privacy Framework (APEC 2005).

The First multiple-site computer network, the ARPAnet, was actualized in 1969. However, until the mid-1970s, most people were unaware that computer networks existed. Using the Transmission Control Protocol (TCP), the first successful message was sent across the ARPAnet, ALOHAnet, and SATnet in 1977. The first computerized Bulletin Board system, Usenet, was opened to the public in 1979. In the late 1980s, more and more people were on-line through Bulletin Board systems. Of them, the majority (estimated 60,000 users) were still affiliated with military installations, universities, or corporations.

In 1986, the introduction of the Free-Net paved the way for public networking regardless of affiliation (Moschovitis et al. 1999). In 1990 the name World Wide Web (Web) is coined for Tim Berners-Lee’s hypertext system, and the following year he completed the original Web software (which includes URL, HTTP, and HTML) (Berners-Lee and Fischetti 1999). In the following years Gopher (1991: a simple network browser), Mosaic (1993: a Web browser, later renamed to Netscape), and Internet Explorer (1995: a Web browser) were introduced. In 1994, eCommerce services, eBanking services, Pizza Hut, and many others went online; and thus started the commercialization of the Web. Tim Berners-Lee envisioned a read/write Web. However, until the early 2000s, the Web was predominantly a read-only medium.

The socialization of the Web starts in the late 1990s. In 1997 the first social network site, the SixDegrees.com, was launched (boyd and Ellison 2008). In the early 2000, people started using the Web as a communication medium, a socialization platform, and a storage device for their diaries (Vossen and Hagemann 2007). LiveJournal was launched in 1999, Friendster in 2002, MySpace in 2003, and Facebook in 2005. The Social Web shifts the way the Internet operates into an interactive read/write medium. Currently, the size of a single day average-users’ blog posts is equal to the size of the entire Web in the mid-1995 (Golbeck 2009). At the time of this writing, Facebook alone has more than 300 million
active users. Very effectively hyped, the Social Web has changed the way in which personal information is created, disclosed, collected, shared, and stored, thus giving rise to new data privacy challenges.

The rest of the paper is organized as follows. First, we describe the privacy risks in social network sites. Next, we compare the private information flow models of the Web and the Social Web. We then briefly enumerate traditional privacy codes and data-handling policy languages for social network sites. Finally, we recommend key points that must be addressed by future privacy codes and data privacy challenges.

Data Privacy Risks in Social Network Sites
The booming popularity of social network sites has created new degrees of interactivity, user control, and user involvement. It also brings a new type of social-centric data privacy problems.

Social network profiles mostly represent genuine identities (Gross, Acquisti, and Heinz 2005) (Lampe, Ellison, and Steinfield 2006) (Hoadley et al. 2009) and are regularly updated by users (Acquisti and Gross 2006) (Donath 2007). The default profile visibility setting of most social network sites are public (Bonneau and Preibusch 2009) and most users do not change the default privacy settings (Gross, Acquisti, and Heinz 2005) (Acquisti and Gross 2006) (Strater and Richter 2007) (Krishnamurthy and Wills 2008). This makes social network sites a gold mine of genuine, up-to-date, public/semi-public, and mostly comprehensive private information. Malicious individuals could mine such information to carry out endless social threats, such as identity theft, social engineering, social phishing, off-line/on-line stalking, cyber-bullying, denigration, impersonation, or sexual exploitation.

It is easy to crawl public social network profiles. For instance, in 2005 (Gross, Acquisti, and Heinz 2005), in 2007 (Mislove et al. 2007), in 2008 (Lam, Chen, and Chen 2008), and in 2009 (Lindamood et al. 2009) have successfully crawled the public profiles of several social network sites. Individual pieces of personal information represent a mere pixel of a person's life, but, when pieced together (e.g., like a social network profile), they can reveal a detailed picture of the person. In the past, compiling a person's digital dossier required the aggregation of scattered pieces of private information from different parts of the Web. Currently, such data can easily be harvested from hundreds of social network sites.

Most social network sites integrate third-party applications into their Web site. In most cases, after installation, these applications have privileged access the users profile, the user friends profiles, and the profiles of other members in the same network. Given such profile information, third-party application providers can easily build a database of users' profiles and social graphs (Felt and Evans 2008). In the Social Web, even if people do not disclose their private information, friends and others may well do so (Lam, Chen, and Chen 2008) (Naessens et al. 2009). First, the myriad of complete and up-to-date profile information in social network sites attracts various stakeholders and introduces a number of data privacy risks. Second, the emerging shift in the flow of private information poses a challenge for existing privacy codes.

Private Information Flow Model in the Web and the Social Web
We use the term “Web” to refer to the first generation Web (mid 1990s - early 2000), “Social Web” to refer to the second generation Web (early 2000 - present), and “social network site” as defined by boyd (boyd and Ellison 2008) and as an application of the Social Web.

The Social Web evolved from the Web, which was largely based on a client/server model, but the Social Web includes user as a core part of its model. In comparison to the traditional request/response based Web, the Social Web is more interactive and puts control of content into the hands of users. The Social Web patterns - such as participation-collaboration, collaborative tagging (folksonomy) and Declarative Living (Governor, Hinchcliffe, and Nickull 2009) - facilitate user interaction and information sharing.

As shown in Figure 1, until recently the model of private information flow in the Web was largely unidirectional. This model assumes that only the user can disclose her private information and a service provider may share her data with third-parties, such as payment or shipping services. Thus, the present privacy codes largely composed of service providers’ responsibilities and users’ right. The responsibilities of the services providers are specified as regards purpose specification, data quality, use limitation, and security specify, and the rights of the users are specified as regards transparency, and user participation. However, neither the responsibilities of the users nor the rights of the service providers are mentioned in the present codes.

Figure 1: The flow of private information (pi) in the Web


2boyd “defines social network 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.”
The Social Web comes with a renaissance of participation, sharing, and content re-generation. Instead of reading static Web pages, users are now actively blogging, bookmarking, tagging, sharing, or be-friending. In this new model, users can be or become the provider of capabilities, like the ability to share private information, ones own or others, which still others consume (Governor, Hinchcliffe, and Nickull 2009). In an article titled "Social Networking: Anybody is a Data Controller," Wong also argues that given the definition of the EU Data Protection Directive, the data controllers are not only the social-network site providers but also individuals who share information about friends and others (Wong 2008). The recent EU Opinion on social network sites compliance with the EU Data Protection Directive similarly notes that social network users can assume the role of a data subject or data controller (EU 2009).

The model of private information flow in the Social Web is multidirectional. In this new model (see Figure 2), a user can share her private information directly with the social network provider, with her friends, or with third-party applications or service providers. In fact, not only a user can share her private information with others, but also her friends or other members could tag, share, or disclose her private information to other members or third-parties. Third-party applications often invisibly share a users’ profile, a user’s friends’ profile, and a user’s social graph information with third-party application providers (Felt and Evans 2008) (Hull, Lipford, and Latulipe 2009). Some features of social network sites, such as the Facebook’s News Feed and MiniFeed, could boost information sharing and make users’ private information more visible than seemed reasonable (boyd 2008a) (Hoadley et al. 2009).

Dissecting Privacy Codes of Practice for the Social Web

Here, the core principles of widely-accepted privacy codes are explained briefly and are analyzed with respect to emerging social-centric privacy risks. The key issues that must be considered by future privacy codes and data handling policy languages are outlined.

I. Transparency

The code of transparency requires service providers to be open about their policies related to the management of personal information (OECD 1980) (EU 1995) (CSA 1996). To comply with this code, Web sites openly publish their privacy policies. Such policies often states, among other things, the identity of the service provider, the purpose of the data collection, the data retention period, the list of third-party recipients, and the possibility of user participation (Nigusse, Decker, and Naessens 2009).

In the Web, private information flow was unidirectional and service providers are required to be transparent. Openness was a responsibility of only the service providers. In the Social Web, the flow of private information is multidirectional so not only service providers but also users and all other stakeholders have to be transparent with respect to each other.

Befriending. Users’ profile is the backbone of every interaction in the Social Web. First, a typical befriending interaction requires an exchange of profile information between two users. Hence, it requires the openness of both users. However, before befriending, users should agree on the purpose of each others data-collection policies, data-retention policies, data-sharing policies, accountability, and the ability to access, rectify, erase, or block their own private information from each others profile. Second, after befriending, such interaction often include a myriad of both user’s connection (social graph) and both user’s friends’ profile information. Third, both user’s friends can access the friends of their friend profile information, i.e., friends as a third-party recipients of the user’s other friends profile information. At the time of this writing, an average Facebook user has 130 friends.\(^3\)

Third Parties. In the Web, third-party service providers access user’s profile information mainly through a service provider. In the Social Web, third-party service providers have various forms and mostly interact directly with users. In the traditional context, third-party service providers assist Social Web providers by hosting services at co-located facilities and may process payments. In the Social Web context, third-party application providers directly interact with users through applications and also facilitate interactions among users. Hence, they can directly access users’ and the user’s friends’ profile information and social graphs. A growing number of musicians, politicians, for-profit and non-profit organizations, and others incorporate social network sites into their public relations program (Waters et al. 2009). Most of these third-party service providers collect a large

amount of user’s profile information\(^4\) and often act on behalf of a company or use the social network sites mainly to advance their commercial, political, or charitable goals. Hence, they are data controllers (EU 2009) and they should be open about their data practices. Moreover, they should negotiate their privacy policies directly not only with service providers but also with users.

**Policy Authoring.** Preference setting tools in social network sites allow users to specify their privacy preferences. For example, using these tools users can set their profile visibility preference. The present social network sites, however, do not provide policy authoring tools to users. As noted above, users can assume the role of data controller, and, like service providers, they should be able to specify their data handling policies. In other words, they should be transparent about their data handling practices. For instance, before two users befriend, they should be able to match their data handling policies. Hence, service providers should provide users a usable policy authoring tool.

**II. Consent**

Except where inappropriate, the code of consent requires service providers to request the consent of the data subject prior to any personal information collection and use (OECD 1980) (EU 1995) (CSA 1996). Moreover, users’ secondary consent is required to further manipulate already collected private information under previously unstated conditions.

**Third Parties.** Social network sites often request users’ consent during registration. Profile information collection by third-party applications in most social network sites, however, is largely invisible and these applications do not provide descriptive software license agreement, terms of use, or privacy policies (Hull, Lipford, and Latulipe 2009). Nor do they request users’ consent. In fact, in addition to the social network sites, third-party service providers, like third-party application providers, friends, interest groups, or fan page administrators, should also request their users’, friends’, or fans’ consent before data collection.

**Features.** Currently a growing number of social software features boost profile-information sharing, linking, and aggregation. Very often, they make users’ private information more visible than seemed reasonable (boyd 2008a) (Hoadley et al. 2009) without users’ explicit consent. Prior to applying new features, that could boost profile information sharing, linking or aggregation than seemed reasonable, service providers and third-party application providers should ask users’ secondary consent.

**Spillovers.** Existing privacy codes assume the data subject as the only revealer of her own private information. However, in the Social Web peoples’ private information is often disclosed by friends and other individuals. This phenomenon has been called in the literature "disclosure by a friend to the public", "involuntary information leakage", "unauthorized disclosure" and as "spillover" (Lam, Chen, and Chen 2008) (Naessens et al. 2009) (Grimmelmann 2009) (Levin and Abril 2009). Since the present pri-

\(^4\)In Facebook alone more than 10 million users join fan-pages each day. http://www.facebook.com/press/info.php?statistics

**II. User Participation**

The code of user participation gave users the right to access their own private information, challenge its correctness and to rectify, amend, erase, or block it where appropriate (HEW 1973) (OECD 1980) (EU 1995) (CSA 1996).

In comparison, user participation in the Web had been very limited. In the Web, service providers solely maintain users’ private information and provided very limited options for users to participate. After data collection users had no intuitive means to access their own private information. In the Social Web, however, users can use intuitive profile-editing tools to access and also rectify, erase, amend, or block their own private information.

**V. Data Quality**

The data-quality code requires service providers to take reasonable steps to ensure that the personal information collected from users is accurate, complete, and up-to-date (HEW 1973) (OECD 1980) (EU 1995) (CSA 1996).

Social network site users update their profile information regularly (Donath 2007). Previous research has shown that 50.4% of Facebook’s users update their profile at least once a month (Acquisti and Gross 2006). Moreover, social network profiles largely represent genuine identities (Gross, Acquisti, and Heinz 2005) (Lampe, Ellison, and Steinfield 2006) (Hoadley et al. 2009). In addition, it is possible to clone fake characters (Fakesters) and to impersonate or fraudulently represent oneself (boyd 2003). Fakesters are “Collectors” (also called “gateway friends” or “social network hubs”) with many friends who are not connected with each other (boyd 2006). Random photos from the Web and data that mimics real profiles has been used by the Pretendster.com to automate Fakesters creation in Friendster (boyd 2008b). In comparison to the Web, the Social Web consists of accurate, complete, and up-to-date personal information.

**Denigration.** Anyone can set up profile pretending to be someone in order to humiliate or damage other peoples’ reputation. Denigration affects peoples’ self-presentation and also degrades the accuracy of private information in social network sites. Hence, service providers should take reasonable steps to minimize denigration.

**VI. Security**

The security code requires service provider to take appropriate technical and organizational measures to avoid unauthorized access, use, alteration, erasure, or disclosure of private information collected from users (HEW 1973) (OECD 1980) (EU 1995) (CSA 1996).

**Visibility.** In the Web, confidentiality has been used to make sure that private information is accessed only by authorized entities, such as authorized employees of a service provider or a third-party recipient.
In the Social Web, however, in addition to service providers and third-parties, profile information can also be accessed by, or shared with, friends, friends of a friend, people in the same network or even by unintended audiences. In most social network sites the default users’ profile visibility setting is public (Bonneau and Preibusch 2009) and most users do not change default settings (Acquisti and Gross 2006) (Strater and Richter 2007) (Krisnamurthy and Wills 2008). As a result, large scale and automated public profile crawling (accessing) is not difficult (Gross, Acquisti, and Heinz 2005) (Mislove et al. 2007) (Lam, Chen, and Chen 2008) (Lindamood et al. 2009). In the Social Web, service providers should provide users usable visibility management tools and also set fair default profile visibility options, rather than applying strict confidentiality measures.

Complete Account Deletion. In the Social Web, complete account deletion is problematic. Even if users terminate their account, secondary information they have posted in friends and others account will remain online. Moreover, a research by (Bonneau and Preibusch 2009) shows that only 3 out of 29 surveyed popular social network sites mentioned explicit data retention limit in their privacy policies. In the Social Web, not only service providers and third-parties, but also friends, fan-pages, third-party application providers and others could also retain users’ private information. Hence, they should also need to specify explicit data retention limits. Using the agreed data-retention limit, service providers should take a reasonable step to enforce complete account deletion.

VII. Enforcement and Accountability

An effective execution of the above five privacy codes requires legal and technical enforcement schemes. In addition, service providers should be accountable for complying with the promises they make (OECD 1980) (EU 1995) (CSA 1996). Without enforcement and accountability measures, these codes are difficult to realize.

Discussions and Recommendations

The present privacy codes were formed based on the 1970s and the early 1980s Fair Information Practice codes. Thus, they are inadequate as regards the social-centric privacy risks. For instance, the Code of Fair Information Practice principles predates the first message exchanged across the Internet’s first three distinct networks and the OECD guidelines predates the Web (Moschovitis et al. 1999).

Table 1 summarizes how the shift in the private information flow model also creates a shift of data handling responsibilities in the Social Web. In the Social Web, users (user in the Social Web context includes the profiles of ordinary users, fan-pages, for-profit/non-profit organizations, politicians, and the like) should be responsible to communicate their data handling policies. Users should enforce their policies and should be accountable for their data handling practices. Users should request the consent of the data subject prior to sharing other users’ private information or before using collected information for another purpose. Users should maintain the quality of the private information they had collected from other users and also let other users to access, rectify, amend, erase, or block their own private information. Users should also take appropriate measures to safeguard other users’ private information. In Table 1, service providers, third-party service, and third-party application providers are generally considered as service providers.

Table 1: The responsibilities (Resp.) and rights of users (U) and service providers (SP) in the Web and the Social Web.

<table>
<thead>
<tr>
<th>Codes</th>
<th>Web</th>
<th>Social Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Right</td>
<td>Resp.</td>
<td>Right</td>
</tr>
<tr>
<td>Transparency</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Consent</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Participation</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data quality</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Security</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Enforcement</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

In the Social Web, privacy protection must consider the emerging social-centric privacy risks and the changes in the model of the private information flow. We strongly recommend that future privacy codes to consider user-to-user transparency, user-to-user consent, and user-to-user participation. They should impose user-side responsibilities with respect to data quality, security, policy enforcement, and accountability. They should also exclusively address emerging issues such as spillovers, denigration, visibility, default privacy settings, and complete account termination.

In parallel, we recommend that future data handling policy languages should provide usable policy authoring, consent negotiation, and user participation tools to users. In addition, these languages should also provide usable and onsite data quality, security, enforcement, and accountability management protocols.

Conclusion and Future Work

This paper provide a foundation for further discussion on the enhancement of existing privacy codes for the Social Web. We showed that the present privacy codes are inadequate to address emerging social-centric privacy risks. As a result, we outlined key issues to enhance the present privacy codes and the future Social Web data handling policy languages. The technical aspects of policy authoring, consent negotiation, and user participation tools and data quality, security, enforcement, and accountability management protocols will be investigated by our future research.

Acknowledgement

This research is partially funded by the Interuniversity Attraction Poles Programme Belgian State, Belgian Science Policy and the Research Fund K.U.Leuven and the IWT-SBO project (ADAPID) Advanced Applications for Electronic Identity Cards in Flanders.
References


boyd, D. 2006. Friends, friendsters, and myspace top 8: Writing community into being on social network sites. First Monday.


boyd, D. 2008b. None of this is Real, volume Structures of Participation in Digital Culture. New York: Social Science Research Council., chapter 8.

CSA. 1996. Canadian standards association’s model code for the protection of personal information.


EU. 1995. Directive 95/46/ec on the protection of individuals with regard to the processing of personal data and on the free movement of such data.


UN. 1990. Un general assembly, guidelines for the regulation of computerized personal data files.


