Evaluating Privacy Practices in Web 2.0 Services

Konstantina Vemou
University of the Aegean, Greece, kvemou@aegean.gr

Maria Karyda
Dept. of Information and Communication Systems Engineering, University of the Aegean, mka@aegean.gr

Follow this and additional works at: http://aisel.aisnet.org/mcis2015

Recommended Citation
http://aisel.aisnet.org/mcis2015/7

This material is brought to you by the Mediterranean Conference on Information Systems (MCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MCIS 2015 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
EVALUATING PRIVACY PRACTICES IN WEB 2.0 SERVICES

Completed Research

Vemou Konstantina. University of the Aegean, Department of Information and Communication Systems Engineering, Samos, Greece, kvemou@aegean.gr

Karyda Maria. University of the Aegean, Department of Information and Communication Systems Engineering, Samos, Greece, mka@aegean.gr

Abstract

This paper discusses the effectiveness of privacy practices and tools employed by Web 2.0 service providers to facilitate users protect their privacy and respond to public pressure. By experimenting on three recently introduced tools, which claim to offer users access and choice on the data stored about them, we analyse their privacy preserving features. Research results indicate their limited effectiveness with regard to user privacy. We discuss discrepancy between stated goals of these privacy enhancing tools and actual goals these tools accomplish.

Keywords: Privacy practices, privacy communication game, effectiveness, fair information principles.

Introduction

In the Web 2.0 era users are the centre of internet services, providing and sharing vast amounts of information. Whether publishing their status on Social Networking Services (SNS), or browsing websites, communicating or conducting commercial transactions, users entrust service providers with large amounts of personal data. This data can be used to assist users in their internet experience, by providing personalized services (Berger 2011) or optimizing the results of internet searches (Cooper 2008). However, gathered user data also expose users to a set of privacy threats, such as price discrimination (Gross and Acquisti 2005), out of context use of information (Nissenbaum 2009) and revelation to unintended audiences (Van den Berg and Leenes 2010). Users’ concerns over privacy are rising (Acquisti and Gross 2006; Karyda and Kokolakis 2008; Boyd and Hargittai 2010; Pavlou 2011); however they are continuing to share their personal data, mainly on SNS. It has also been pointed that users are not aware of several privacy invading practices, such as third-parties tracking their online activity (ENISA 2012). Lately several incidents of privacy violation have reached the spotlight, alarming online users. For instance, the introduction of the Newsfeed feature in Facebook, which delivered profile updates to other users in headline news with limited audience settings (Hoadley et al. 2010) and the Google’s circumvention of barriers in Apple’s Safari browser to track users’ online activity have reached the mass media, creating a public outcry from privacy advocates (Gorman 2012). Another example in the long range of published privacy violation incidents was the leak of Snapchat user photos (Snapchat is a mobile phone app to send self-deleting photos), via hacking in “Snapsaved.com”, a service offering use of Snapchat on a desktop computer. As reported, Snapsaved.com had failed to adequately inform users on the saving practice of their pictures (The Guardian 2014). Digital media providers are profit organizations with an interest in maintaining highly active users (Easley and Kleinberg 2010; Krasnova et al. 2009). They manage users’ personal data to achieve business goals, including enhanced products and services and more effective and efficient operation. However, with privacy advocates reproving their privacy practices after each violating incident on the
news, companies are forced to seek for solutions to appease awakened privacy concerns. Popular social media platforms have lately introduced a variety of settings, including access controls of posted information and activity logs and applied changes on their privacy policies. According to Pollach (2005) such initiatives mainly occur as a response to international law compliance needs and public pressure from privacy advocates and the media.

The effectiveness and motives of these practices have been the subject of debate. For instance, in May 2015, 80 academics sent an open letter to Google requesting more transparency in the process of citizen requests for “their right to be forgotten”, which is requests of delisting links to personal information that is inaccurate, inadequate, irrelevant, or excessive for the purposes of data processing (The Guardian 2015). Hong (2014) argued on the privacy placebos; tools or settings that allay privacy concerns, but gain limited user acceptance or offer little actual value on protecting user privacy. Bonneau and Preibusch (2010) refer to this as the “privacy communication game”, where service providers are struggling to provide privacy practices that will appease privacy concerns for privacy-aware users, without adequately addressing them, and without alarming the non-privacy-aware, thus creating a back door to continue taking advantage of user personal data.

In this paper we further explore this debate, and investigate whether recently adopted privacy practices are targeted to protecting users’ privacy or if they serve different purposes, as part of a communication game, for instance. Using a set of criteria derived from the Fair Information Principles (Schwaig et al. 2006), we study the effectiveness in protecting user privacy of three tools recently introduced by popular providers: Google search history, Google takeout and Facebook download your data archive. The remaining paper is structured as follows: in chapter 2 we present the debate on privacy practices effectiveness. In chapter 3 and 4 respectively, we present the criteria as well as the analysis we performed on the above mentioned tools, followed by a discussion on their effectiveness in chapter 5 and conclusions in chapter 6.

2 Privacy Practices in Web 2.0 and their Effectiveness

Responding to negative publicity and user concerns over personal data practices, Web 2.0 service providers apply a wide range of privacy practices, including publishing privacy policy documents and introducing privacy settings. Pollach (2005) argues that the introduction of privacy practices is driven mainly by the need to comply with privacy laws and the need to gain users’ trust, by recognizing their concerns over privacy. Starting from the early 1990s when e-commerce began to flourish, privacy policies have been used as a means for companies to inform online users of their privacy practices. The quality of privacy policies in terms of guaranteed privacy levels (adhering to fair information principles) has been linked to users’ trust and engagement with the web service providers (Lauer et al. 2003; Meinert et al. 2006). However, the effectiveness of privacy policies in protecting users’ privacy has been questioned by many researches. There may be many reasons explaining why users ignore the website’s privacy policy, including its length, obscure language used (Lauer et al. 2003; Pollach 2005), or the difficulty to access it in the website (Bonneau and Preibusch 2010). Other studies found that providers tend to describe certain privacy practices in their policies, such as the collection of login information, but omit documenting other practices, such as informing users of third-parties their data will be shared with (Pollach 2005; Stuart 2007).

Several platforms employed the Platform for Privacy Preferences (P3P) (Cranor 2003), which, however, did not gain widespread acceptance and has been gradually abandoned. The idea behind P3P was that websites would express their privacy practices in a standard format and this would automatically be compared to users’ privacy preferences, to assist them in their decision to browse or not to browse the website. P3P’s fall backs include partial address of the fair information principles as it did not provide any mechanism to assure companies process personal data according to stated privacy policies (enforcement) (Jaatun et al. 2011) and the ability of companies to circumvent user’s preferences by providing erroneous or conflicting P3P compact policies (Leon et al. 2010).
Privacy seals, such as TRUSTe, BBBOnline and WebTrust, are third-party programs assuring that an organization’s handling of personal data is aligned to privacy practices declared in its privacy policy. In general they are a forcing task for organizations to provide a privacy policy, as well as to align it to fair information principles and are used to convince users that their privacy is not violated by using their online services. They have also received criticism, on the grounds that privacy seals institutions do not oblige organizations to explicitly mention all types of information collected and how they will be used. Also, as they are targeted at ensuring compliance with the company’s privacy policy, there are workarounds to using personal data, by declaring them as assets in the privacy policy (Moores and Dhillon 2003). What is more, in the event of privacy breaches, the assuring institutions can only revoke the seals and cannot help users in regaining their lost privacy (Shapiro and Baker 2001; Hui et al. 2007).

Opt-in strategies requiring that profilers obtain consent prior to collecting or sharing user data (Berger 2011) have also been debated as they are applicable to only binary policies (accept or not accept to provide data), while the privacy needs of users differ according to the purpose of personal data use (Spiekermann and Novotny 2015). It was also argued that companies can coerce users in accepting their practices by refusing to provide their services in any other case (Berger 2011; Cooper 2008). At the same time, as privacy threats on certain web 2.0 services, such as SNS, are not only based on the platform’s tracking of user behaviour but also on the difficulty to protect semi-public information published by the users themselves, opt-in or opt-out practices are only partially covering users’ privacy protection needs.

Finally, privacy settings, such as access controls, introduced in several SNS platforms, are usually judged as insufficient or unusable and were found to be neglected by users (Boyd and Hargittai 2010). This is alarming, as default privacy settings were found to be non-conforming to privacy-by-design principles (Vemou and Karyda 2014) and tend to become progressively more liberal over time (Michota and Katsikas 2014).

3 Theoretical Background

3.1 Research approach

This paper investigates if privacy practices adopted by Web 2.0 service providers are targeted to protecting users’ privacy or whether they pursue different goals. To gain further insight into this issue we performed an experimental analysis, using several accounts on three tools: Google Takeout, Google Search History and Facebook download your data archive.

These tools were selected based on the popularity of the platforms (Statistics by Alexa 2015) and on their functionality: they are mainly targeted at providing users with the ability to access/control information these two companies gather about them. Thus they are focused on providing users with control over their personal information, which has been a major privacy requirement by users. These tools have been recently launched and have not been evaluated in relevant research.

3.2 Conceptual framework

The evaluation of privacy enhancing tools was based on criteria (described in Table 1) derived from the set of fair information practices (FIPs), introduced during the 1990s by US Federal Trade Commission (Schwaig et al. 2006), as well as from relevant literature, extending them (Anton et al. 2002; Spiekermann and Cranor 2009).
### Fair information principles

<table>
<thead>
<tr>
<th>Notice/awareness</th>
<th>Evaluation criteria employed in this paper</th>
</tr>
</thead>
</table>
| Companies need to inform users if and which personal information is being collected and how such information will be used. | 1. Does the tool provide users with a full list of collected and shared personal information?  
2. Does the tool provide users with a detailed list of third-parties information will be shared with?  
3. In which form is users’ data shared with third parties (anonymized, aggregated etc.)?  
4. Does the tool inform users of privacy threats stemming from collected and shared personal data? |
| Choice/consent | 5. Does the tool enable users to declare preferences on how their data will be used?  
6. Does the tool enable users to declare preferences on whether this information will be shared with third parties (unless required by law)?  
7. Does the tool inform users of personal data processing for purposes they have not consented?  
8. Does the tool inform users of personal data processing by third-parties, in cases of their choice to prohibit it, despite prior data sharing? |
| Access/participation | 9. Does the tool allow modifications/corrections on collected personal data?  
10. Does the tool enable users to request complete deletion of their data?  
11. Does the tool enable users to access and correct data processed by third-parties? |
| Integrity/security | 12. Does the tool prevent other users from gaining access to information stored about a certain user? |
| Enforcement/redress: | 13. Is it visible to users (via the tool) that data processing is performed according to their declared choice?  
14. Is it visible to users (via the tool) that stored information is actually corrected or deleted following their requests? |

### Table 1. Evaluation criteria used

#### 4 Evaluating Privacy Tools that Provide Access to Stored Information

In this section we evaluate three tools that claim to provide users with access, transparency and choice on personal data, developed by the two most popular Web 2.0 service providers, Google and Facebook.

#### 4.1 Google Search History

“Google Search History” is a service provided to Google users, as part of Google’s Dashboard, presenting their search history in the Google Search engine. Users are presented with their searches performed by date and time, and the pages visited from the respective search results. Also, graphical rep-
resentations of search activity statistics are provided, such as daily or weekly activity rates, and other statistics, such as the most frequent searches of the user.

According to Google, this tool offers transparency and provides users with control over the data associated with their accounts\(^1\). This tool is also included in the privacy policy (in the section “transparency and choice”), as part of a set of privacy preserving services. Use of the tool requires logging into any Google service, to ensure security and privacy of presented information.

In general, this tool raises user awareness on collected data from search history. However, presented information is limited to users’ searches, although information on their other activities is gathered and, if combined with search history, can lead to their online profiling. What is more important is that such information gathering is not transparent to the users and does not always require their login to Google services (e.g. tracking on other webpages).

Other information, such as the location of the users or the time spent browsing result pages, could be of significance if correlated to a search; however users do not have access to this information. Also, users could form a false assumption that information on their location is not gathered and correlated to their searches, in case they have chosen not to share their location. In this case they would be presented with an empty locations history report (in the account history settings), albeit according to its privacy policy, Google maintains IP records from their logins. While this could automatically reveal users’ location (in area level) and could be correlated to their searches, the search history tool fails to provide users with access to this inferred information.

This tool does not inform users on how data on their search activity is collected or shared with third-parties, offering no traceability of such data and no choice on how it will be used. For instance, no privacy signalling technique is offered to declare accepted purpose of collected data use, and there is no opportunity to correct or delete data stored in third-party platforms. As for the search data stored by Google platform, users are given the opportunity to delete past searches from their search history or to disable collection of search history elements, but there is no transparency on whether collected data are actually deleted as result of the users’ request or they just do not appear in the history tool interface. Thus “Google Search History” tool addresses notice and access principles up to an extent, but fails to provide users with choice on how their data is handled.

### 4.2 Google Takeout

“Google Takeout” is a tool provided to Google users, in order to download data they have imported or created in several Google websites/services. According to Google, it provides users access to their information, as stored by the company and can also serve for backup of user data or for importing it to other services\(^2\). It can also serve as an awareness raising tool, as users’ access to their personal data is of critical importance for privacy protection. Other platforms, such as Facebook, offer similar services as privacy preserving practices.

Users need to be logged in their accounts, in order to access “Google Takeout” from the account settings and when they try to access/download the archive they are prompted to re-enter their account credentials, although they may follow a link sent to their email account.

“Google Takeout” offers users the opportunity to download a copy of their imported data from several Google services such as Gmail, Contacts, Google+ and Google Latitude. In the created archives, users are presented with the information they have inserted or directly created as a result of using these services. For instance, users are presented with a history of YouTube videos they have watched. However, provided information does not include other information the company has stored on users, such as

---


2. [https://support.google.com/accounts/answer/3024190?hl=en](https://support.google.com/accounts/answer/3024190?hl=en)
log data. An exception to this is the location history of the users, which presents them with a history of the places they have been (in coordinates), along with inferred activities, e.g. being in a vehicle or walking. Even in this case, other inferred and stored information, such as identification of work or home places, is not provided.

Despite offering a presentation of users’ stored data, “Google Takeout” fails to communicate which entities have access to this data, other than the Google Platform and the users themselves. Users are not presented with information derived from their data, as a result of processing and combining different services use data, while relation of user data to privacy threats, such as price discrimination or publishing to unwanted audiences is omitted.

“Google Takeout” does not provide users with an interface to report their preferences on how their data is used. Also, although “Google Takeout” serves as single point users can access stored information from different Google services, it is ineffective in directly requesting correction or deletion of some data. Instead, users need to access the different services websites to modify or delete them. Concluding, “Google Takeout” enhances users’ awareness.

4.3 Facebook download your data archive

“Facebook download your data archive” is a tool provided by Facebook for users to get a copy of their data as stored by the platform. As stated in Facebook’s privacy policy, this is a feature to enable users’ access to stored information about them. The downloaded archive contains a list of user created and posted data, such as profile information, photos and friends list, but also other categories of data stored without users’ notice, such as log data and advertisement activity (e.g. which advertisements the user has clicked).

The tool makes users aware that data not directly created by them are also stored by the company. For instance, the archive of user data contains friend requests, either generated by the users themselves or by other users, regardless of acceptance. Also login data (device and IP addresses) and information on advertisements users have clicked are presented. In the same way, users are presented with some basic information on data inferred by their activity and posts, such as advertisement categories they are targeted with. However, there is no reference to other types of inferred data from the users’ activity, which audiences got access to users’ activity (e.g. wall posts), which data were used in advertisements displayed to the users’ network and which were shared to third-parties. Furthermore, the network created by friends of friends is not graphically or in any other way presented, although it is stored and can be exploited by the company.

Although the feature is mentioned in the part of privacy policy concerning users’ access and correction of data stored about them, the only automated option to correction/deletion from the data archive is to delete user account. Even in this case, as stated, some parts of information, such as logins will be maintained by the platform, for at least a year. In case of content deletion from the account, e.g. removal of a photograph or video, this will be removed from the downloadable archive, but users have no direct means to confirm that the item itself, as well as metadata gathered by the platform are deleted and also have no means to enforce such deletion on third-parties that have accessed it. “Facebook download your data archive” covers a range of notice requirements, by providing users with secondary information stored about them, but fails to provide effective choice and access mechanisms.

3 https://www.facebook.com/privacy/explanation
5 Discussion

The tools analysed in this study aim to provide users with access, transparency and choice with regard to their personal information gathered by Web 2.0 providers. Research results, based on Fair Information Principles showed that these tools succeed in informing users of stored data about them, which is a step towards notice and raising awareness. However, we found that not all information stored about users is presented, and users are not presented with the outcome of processing such data. For instance, in Google’s tools, log data were omitted, although the knowledge of IP address of users’ connection enables the company to presume their location. Also, inferred information on the users’ networks, such as proximity of friends based on the frequency of their communications or the combination of their locations was not presented in Facebook’s tool. Information on third-parties that gained access to users data were not presented via the studied tools, and there was no report on how stored or published information could lead to privacy threats for the users.

Providing users with choice over their personal information was also among the stated goals of these tools; however our analysis revealed that the tools did not offer any method for users to present their preferences on how their data should be treated. For instance, users may wish their interests or hobbies information to be used for improved search results or for friend recommendations, and not for marketing purposes, but there would be no choice to express it via the analysed tools.

Moreover, users were not presented with the choice to request correction for data stored of presumed about them. For instance, while Facebook’s archive informed users of the advertisement categories they were enlisted, based on their activity, they were not provided with the ability to opt-out of a certain category of advertisements. In some cases users were provided with the capability to delete some parts of stored information (e.g. from Google history), but were not in the position to control whether this information or related information (e.g. from cookies) were actually deleted from the company. Apart from this, in some cases, such as in Facebook archive, users were only offered with the choice of complete deletion of the account.

Typically, platforms provide several privacy enhancing tools to address complementary privacy requirements of users. Based on our analysis, even if more than one tool were used, certain aspects of user privacy would still be unaddressed. For instance, the application of Google search history along with Google Takeout would not cover requirements related to choice of data use purpose. In the same context, Google’s Dashboard provides no options to declare preferences on how their data can be used. Finally, a combination of the above tools could not provide a more detailed list of third-parties gaining access to user data, thus notice/awareness principle would still be partially addressed.

Another issue derived from our analysis is the presentation of such tools to users. In Facebook, the tool to download user archive was placed at the end of the account settings list, in smaller font than the other settings. It seems that it was intended to separate this tool from other settings, alas highlight its existence. This could be viewed as a usability issue, posing questions on the platform’s motives of introducing the tool. Also, our experience with these tools shows that usability in terms of presentation of data could be significantly improved. An example of such improvement is to present formed networks of friends or the content of advertisements users have clicked. Also, Google Takeout used open archive templates (JSON etc.) to help users import their data to other platforms, but it would be preferable to also provide a way of accessing stored data in webpages, for users with less technical skills.

In conclusion, our analysis revealed a discrepancy between stated privacy goals and achieved goals of the mentioned tools. They were found to cover only partial requirements on notice and awareness, while superficially treating other privacy preserving principles, such as access/participation. The last was mainly stated as their goal. Also, the analysed tools did not address choice/consent and enforcement principles. Usability preservation led to questions on actual reasons the tools were introduced for, taking into account that the majority of users are not IT specialists and non usable or difficult-to-locate tools could lead to their avoidance. In fact, companies do not seem to be paying much
attention on usability, but this could be part of their strategy playing the “privacy communication game”.

6 Conclusions and Further Research

This paper explores a “privacy communication game”: companies adopting privacy practices or introducing tools to help users in their efforts to protect their privacy and appease public outrages, while in fact their effectiveness or usability is limited. By evaluating three recently introduced tools that promise users access and choice on data stored about them, and analysing their features based on the Fair Information Principles, we reveal their limited effectiveness with regard to privacy protection, even in cases of their combined use. Our analysis shows that there may be a discrepancy between the stated goal, which is to protect users and offer them rights on their data, and actual goals of the companies, such as appeasing public pressure and marketing privacy to temper users’ privacy concerns.

Web 2.0 service providers employ a business model that partly depends on the exploitation of user data to derive its profits. Privacy practices and tools, while presented as solutions to users’ privacy concerns, are largely used as a reactive “damage control” instrument, after privacy breaches or invading practices reach the spotlight and receive criticism. We also found the usability of the privacy tools we studied rather low, which can be explained, as they are meant to offer some control and appease concerns of privacy-aware users but in such a way that non-privacy-aware users do not understand privacy threats they are exposed to.

As a result, available privacy tools are in some cases extending the privacy illusion users are under, despite promising control over their personal data. In an attempt to raise users’ awareness and assist them, a set of criteria to evaluate if tools meet their privacy requirements, regardless of what is stated or communicated by the platforms should be offered. Also, as awareness of PETs existence is found to be of importance for users to engage with them, and as it is a common practice for offered tools to be “hidden” in account settings, actions to raise awareness of the tools existence are also necessary.

References


