

COMSM2005
Law and IT Assessment

Data Privacy on Web 2.0

VERSION 1.1

Vasileios Lampos [vl7342]

Essay Title 2

“You have no privacy. Get over it.”

Scott McNealy, then CEO of Sun Microsystems, at a conference in 1999.

***Discuss** McNealys comment in relation to Web 2.0 technologies and legal protection of data privacy.*

Data Privacy on Web 2.0

Vasileios Lampos

January 13, 2008

1 Introduction

“You already have zero privacy; get over it.”¹ These were the exact words of Scott McNealy during the presentation of Sun’s Jini.² In fact, the insufficient U.S. legislation for web privacy at that time pointed to that direction. Nowadays, the amounts have changed significantly; the world wide web is an extension and in some aspects a replacement of the real world. Internet is a tool for many purposes such as entertainment, shopping, financial management, investing, and socializing. Unfortunately, this massive invasion of Internet in the modern societies exposed weaknesses in their legal systems; their legislation was not ‘ready’ for this new medium.

In this article we focus on the modern aspect of the world wide web, known as Web 2.0,³ and especially on the legal framework that secures the privacy of its users. The article is organized as follows: firstly, we define Web 2.0 and comment its current general impact; then we focus on user personal data privacy on Web 2.0 and we present possible ways of invoking his privacy; afterwards we draw our attention on web privacy legal frameworks in the United States (hereinafter U.S.) and the European Union (hereinafter EU); in the end, we refer to ways able to enhance user’s privacy.

2 Discussion

2.1 Web 2.0 definition and expressions

What is really Web 2.0? In [OR05] Web 2.0 is defined as the part of the world wide web which uses state of the art technologies in order to interact with or offer a specific service to a user. This dynamically linked environment makes massive use of databases for supporting and enhancing its applications. Consequently,

¹ John Markoff, *Growing Compatibility Issue: Computers and User Privacy*, N.Y.TIMES, Mar. 3, 1999, p.5.

² “Jini provides an environment for creating dynamically networked components”, <http://www.sun.com/software/jini/>.

³ Web 2.0 is a definition originated by Tim O’Reilly.

a large amount of private or non-private data is stored, while storage costs are decreasing. Some examples of Web 2.0 applications, which require and process user’s private or semi-private data, are provided below:

- Social networks (*e.g.* MySpace,⁴ and Facebook⁵).
- Email, instant messaging, and organizer web services (*e.g.* Google⁶ or Yahoo⁷ email, Google Talk,⁸ and iGoogle⁹).
- Blogs and bulletin boards (*e.g.* LiveJournal,¹⁰ and Blogger¹¹).
- Video, photo, bookmark, and web site rating sharing services (*e.g.* Youtube,¹² Flickr,¹³ del.icio,¹⁴ and StumbleUpon¹⁵).
- Electronic shops, and online auction-like systems (*e.g.* Amazon,¹⁶ and Ebay¹⁷).

2.2 The impact of Web 2.0 applications and services

A quick look on the current web sites traffic (by using the tool provided by Alexa.com¹⁸) indicates that the majority of users tends to browse Web 2.0 enabled sites. Yahoo.com, Google.com, and Youtube.com are ranked 1st, 2nd, and 4th respectively.¹⁹ MySpace.com counts more than 120 million registered users and is

⁴ <http://www.myspace.com/>.

⁵ <http://www.facebook.com/>.

⁶ <http://mail.google.com/>.

⁷ <http://mail.yahoo.com/>.

⁸ <http://www.google.com/talk/>.

⁹ <http://www.google.com/ig/>.

¹⁰ <http://www.livejournal.com>

¹¹ <http://www.blogger.com>

¹² <http://www.youtube.com/>.

¹³ <http://www.flickr.com/>.

¹⁴ <http://del.icio.us/>.

¹⁵ <http://www.stumbleupon.com/>.

¹⁶ <http://www.amazon.com/>.

¹⁷ <http://www.ebay.com/>.

¹⁸ Alexa.com, the Web Information Company, <http://www.alexa.com/>.

¹⁹ Jan. 5, 2008, *Id.*

ranked as the 6th 'most busy' web site,²⁰ whereas Facebook.com counts more than 59 million active users as well as an average of 250,000 new registrations per day since January 2007,²¹ and is ranked 7th on Alexa.²²

The two 'parties' affected by the statistics presented above are obvious; on the one side there are simple web users and on the other internet business. Both parties try to get a benefit from the usage of the web; a helpful or essential service and economical profit respectively. Users are delighted to use effective web services of any type and not pay for them. On the other hand companies want to attract traffic in their web sites, which will be later transformed into income. Problems may occur when a 'hidden' third party tries to take advantage or use (illegally) the information shared between the initial two parties.²³

There is a continuous and growing relationship between a successful web site and the high amount of its users. As the web site grows and offers a bigger amount of services, opportunities or products, even more users will tend to use it. That was the 'story' of many successful web sites, such as Google.com, Youtube.com, Amazon.com and Ebay.com. In a web-based social network, users would appreciate easy ways of establishing relationships; this will increase the quality of service to them as well as the commercial value of the web site [AG06].

2.3 Users and privacy

According to [Hsu06] even experienced users who express concerns about their privacy, tend to 'forget' them in practice; the social contexts (*i.e.* web site categories) determine their final behavior. Furthermore, users are more willing to provide personal private information to third parties which are considered to be trustful (*i.e.* popular web sites) rather than to third parties that are unknown to the public majority [EB03]; this is logical, but the point is that users do provide their private data eventually.

Social networks, such as Facebook.com, usually contain the whole package of information, which is able to expose a user. Of course, social networks are not the only source for retrieving a person's private data;

²⁰ *Id.*

²¹ Facebook, Official Statistics, <http://www.facebook.com/press/info.php?statistics/>.

²² see *supra* note 19.

²³ initial two parties were users and the company with whom they share their information.

other Web 2.0 sites could be used either for this purpose or for 'filling the gaps' in information taken from another web source. In any case, social networks pose the greatest danger in destroying user's privacy because they gather bigger amounts of information and as a result we focus on them.

It is common sense that a person's privacy is destroyed, when this person can be uniquely identified from the online information he has provided. In that case the information disclosure has exceeded the fingerprinting threshold, and thus a fingerprinting threat has emerged [Con06]. A hidden danger for users that tend to browse and actively use more than one web sites (*e.g.* a movie rating web or a forum and a social networking web site), revealing sparse information about the same subject to each one of them, is re-identification; the correlation of user's preferences may destroy his desired privacy [FCS+06].

A research on Facebook.com social network's accounts indicated that 88% of its users list real full names on their profiles, 90.8% contain an image, 87.8% reveal their birthdate, 50.8% reveal their current residence, and 39.9% submit their phone number [GA05]. In addition 53% of the users list their political views, 59% their sexual orientation, and 28% even the name of their relationship partner [AG06]. A more recent research points to approximately the same statistic results [DHP07].

Gross and Acquisti in [GA05] describe how someone could use this information in order to make estimations about the social security number (hereinafter SSN) of person (in the U.S.). Knowing the SSN of a person could possibly lead to an identity theft. Another danger that emerges is recovering banking private data, such as customer numbers.²⁴

Many problems arise, for the reason that persons forget reality rules, when they use their personal computers. Users think that computers form a virtual environment which is unavailable offline; as a result they behave differently comparing to their real life behavior. People forget that social networks cannot be taken away from reality. On the contrary, companies admit that apart from interviews, they are searching social networks' search engines for the persons that form a job application. Extreme photos or even videos in

²⁴ *e.g.* National Westminster Bank Plc (NatWest, UK, <http://www.natwest.com/>) uses the birthdate of a customer plus a four digit number to compose an internet user id number. User accounts that are not protected with a strong password are in danger.

which a user performs something on the edge (*e.g.* is drunk or on drugs) or unethical comments and quotes in his profile will be a strong recommendation for them not to accept someone’s job application. Even the feeling of being exposed is a threat for the users; no one likes embarrassing moments in his life to be publicly available, but in a social network that cannot be always controlled [Ros07].

Another type of attack that uses the structure of social networks is described in [JJJM07]. It is a phishing²⁵ attack which first extracts users’ private data from a social network and then uses them to deceive the users. The attack was able to retrieve private data (*e.g.* password, and account information) from the users and its success percentage was more than 70%.

2.4 Privacy in the U.S. - The legal framework

United States legislation for Internet data privacy is not complete; special protections for sensitive personal data, such as health information, ethnicity, political affiliation, sexual orientation, religious beliefs, social security numbers, and sensitive financial information, do not exist [BEP04]. As a result, web privacy is not well defined and courts have to draw analogies to doctrines unrelated with the web in order to take the appropriate decisions. Such analogies are of high importance as the ensuing results may be completely different [Hod06].

A milestone U.S. Supreme Court decision was taken for *Katz v. United States* case.²⁶ The Court declared that “the Fourth Amendment²⁷ protects people, not places”,²⁸ indicating that the law actually protected values that a home represents, such as security and individual autonomy, rather than crossing inside someone’s property boundary [Wei07].

²⁵ Phishing is a malicious attack able to deceive the user. It usually contains a redirection script that transfers the user to another web site (which most of times is identical to a popular web site). Then the user is prompted for his personal information; all the information that will be typed by the user is now available to the ‘attackers’.

²⁶ 389 U.S. 347 (1967).

²⁷ “The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.”, United States Constitution, Amendment IV.

²⁸ 389 U.S. 351.

The Supreme Court of United States in *Smith v. Maryland* case,²⁹ based on its opinion about a number of cases³⁰ including *United States v. Miller*,³¹ decided that “[a] person has no legitimate expectation of privacy in information he voluntarily turns over to third parties.”³² Third parties may use such information for legitimate business purposes; users have explicitly agreed to allow operators to use their personal data for marketing purposes. An exception arises when a person did not intend to disclose information to a specific recipient [Hod06]. The Court of Appeals for the Armed Forces in *United States v. Maxwell*³³ case stated that “[m]essages sent to the public at large in the “chat room” or e-mail that is “forwarded” from correspondent to correspondent lose any semblance of privacy”³⁴, indicating that one cannot expect privacy when his actions are not private anymore.

When an object is in “plain view”, the Fourth Amendment does not protect it anymore; there is no requirement for a warrant.³⁵ By applying this doctrine to a user’s public³⁶ profile which is part of a web 2.0 application, we may assume that being in “plain view” the protection of privacy is lost. On the other hand, a user should expect privacy when he uses a “limited”³⁷ profile; retrieving the user’s private data violates the Fourth Amendment unless the society is not prepared to recognize this privacy as “reasonable”.³⁸ However, if a user consents in searching his data, then again his expectation of privacy is destroyed, *i.e.* if a user of a social network accepts as a friend a person that turns out to be a police officer (or a person cooperating with a police officer) [Hod06].

In 2001, U.S. enacted the Patriot Act in order to strengthen their defences against terrorism.³⁹ The Patriot Act permitted the collection of records that included any tangible thing, meaning that investiga-

²⁹ 442 U.S. 735 (1979); question whether the use of a pen register constitutes a ‘search’ within the meaning of Amendment IV, U.S. Constitution.

³⁰ 442 U.S. 744, *Couch v. United States*, 409 U.S. 335-336; *United States v. White*, 401 U.S. 752; *Hoffa v. United States*, 385 U.S. 293, 302; *Lopez v. United States*, 373 U.S. 427.

³¹ *United States v. Miller*, 425 U.S. 442-444.

³² 442, U.S. 743-744.

³³ 45 M.J. 406 (C.A.A.F. 1996).

³⁴ *Id.* at IA.

³⁵ *Katz v. United States*, 389 U.S. at 361.

³⁶ *public* has the meaning of available to everyone.

³⁷ *limited* has the meaning of available to a certain group of people (known as “friends”) which the user has already defined.

³⁸ *Katz v. United States*, 389 U.S. 347, 361.

³⁹ Uniting and strengthening America by providing appropriate tools required to intercept and obstruct terrorism (U.S.A. Patriot Act) Act of 2001.

tors were allowed to examine the contents of electronic communications directly (*e.g.* interaction with a web site of the government) or indirectly (*e.g.* request a web site’s transactional logs) [JBM03]. In addition, Section 215⁴⁰ of the Patriot Act assures that every activity under this act will remain secret.⁴¹

2.5 Privacy in the EU - The legal framework

In contrast to U.S. legislation, the right to privacy is better defined in the European Union [BEP04]. EU web privacy related legislation consists of the following Directives: information Directive (95/46/EC),⁴² Directive on privacy and electronic communications (2002/58/EC),⁴³ and Directive on the retention of data (2006/24/EC) which mainly applies the previous EU Directives to private information necessary for communication services.⁴⁴

The information Directive in Article 2(a) defines ‘personal data’ as “any information relating to an identified or identifiable natural person” which in turn is defined as a person that may be identified by “a reference to an identification number” or “to his physical, physiological, mental, economic, cultural or social identity”; as a result it protects user’s sensitive information. Article 8 of the same Directive mentions the specific categories of user’s sensitive data⁴⁵ as well as the special circumstances in which their processing is allowed. Such circumstances include when the user “has given his explicit consent” or when “processing

⁴⁰ Section 215 of the Patriot Act amends Foreign Intelligence Surveillance Act of 1978 (FISA) by applying changes to the Sections 501 - 503 of FISA, Patriot Act 2001, Section 215.

⁴¹ “No person shall disclose to any other person (other than those persons necessary to produce the tangible things under this section) that the Federal Bureau of Investigation has sought or obtained tangible things under this section”, Section 215, (new) Section 501 of FISA, (d), Patriot Act 2001.

⁴² Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

⁴³ Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector.

⁴⁴ Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC.

⁴⁵ “Member States shall prohibit the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life.”, 95/46/EC, Article 8.

is necessary to protect the vital interests of the data subject”.⁴⁶

As far as ‘cookies’ are concerned,⁴⁷ the Directive on privacy and electronic communications declares that where “are intended for a legitimate purpose, their use should be allowed on condition that users are provided with clear and precise information” about their purpose.⁴⁸ The same EU Directive specifies how user’s private data should be handled for marketing purposes. In this case storing and using them “may only be allowed if the subscriber has agreed to this” and these data “should also be erased or made anonymous after the provision of the service”.⁴⁹

2.6 Binding legislations - The Safe Harbor Principles

The worldwide spread of e-commerce urged for a unified legislation framework which could make it operate ‘smoothly’ through different countries and regions. This requirement led to the establishment of the ‘Safe Harbor’ arrangement. Safe Harbor principles intended to fill up ‘conveniently’⁵⁰ the inadequacies that U.S. data protection and privacy legislation had comparing to EU information Directive (95/46/EC) [LQ02].

Amongst others, Safe Harbor includes the following Principles:⁵¹

- *Notice.* Users must be informed about the purposes for which their information is collected, and the third parties to which their information will be disclosed to.
- *Choice.* Users must be provided with the opportunity to choose whether their personal (opt-out)⁵² or their sensitive (opt-in)⁵³ information is to be disclosed to a third party or to be used for purposes non compatible with the original one.

⁴⁶ ‘data subject’ is the term used to refer to a person - user.

⁴⁷ “HTTP cookies are used by Web servers to differentiate users and to maintain data related to the user during navigation, possibly across multiple visits.”, definition from Wikipedia.org, http://en.wikipedia.org/wiki/HTTP_cookie/.

⁴⁸ 2002/58/EC, Whereas: (25).

⁴⁹ 2002/58/EC, Whereas: (20).

⁵⁰ meaning that no additional legislation was required.

⁵¹ Safe Harbor Principles, http://www.export.gov/safeharbor/SH_Privacy.asp.

⁵² user must have the opportunity to be removed from a service or to deselect an option.

⁵³ user must give his explicit consent for the use of his personal information.

- *Onward Transfer.* Organizations, which follow the pre-mentioned Principles, are allowed to disclose information to third parties if they can ensure the same level of privacy.⁵⁴
- *Security.* Organizations must take reasonable precautions to ensure the protection of users private information.

2.7 Enhancing user’s privacy

The use of privacy policies on the web sites, as we have previously mentioned, is generally enforced by the legislation. However, a small percentage of the users is able to understand them comprehensively [JP04]. Another medium used to indicate that a web site is safe for users is a private seal. Privacy seals are provided by third parties such as BBBOnline⁵⁵ and TRUSTe⁵⁶ [BRDM07].

According to [Wri07], the European Commission tries to enforce the use of Privacy Enhancing Technologies (PETs, May 2, 2007). These are frameworks able to protect or prevent unauthorized access to private data. Several PETs examples could be automatic anonymisation of data, encryption systems that secure the transmission of information over the Internet, and software able to control every operation of cookies installed on a system (referred as “cookie cutters”). A more advanced example is the implementation of Platforms for Privacy Preferences (P3P). P3P project enables web sites to state their privacy policies in a standard format; user agents, holding user’s private data preferences, interpret these policies and perform an automated decision masking when appropriate. Thus, users can adjust their personalized privacy policies once for every web site they visit [CLM+02]. Unfortunately, P3P technology is not enforced through legislation, and as a result only 13.59% of the most famous web sites, and 10.46% of Google search engine results are P3P-enabled [ECC06].

3 Conclusions

It will always be difficult or impossible to have a legislation that covers every new technology aspect. Technology is always open to the fresh and unpredictable and as a result a legal framework for the unpredictable

⁵⁴ which means that third parties must be subject to an equivalent legislation.

⁵⁵ <http://www.bbbonline.org/privacy/>.

⁵⁶ <http://www.truste.org/>.

cannot always exist. Even when the missing legislation seems to be ‘obvious’, it takes a long time until it is composed and enacted.

Throughout this article, we discussed about the privacy issues that the use of Web 2.0 applications brings to surface. Do we have privacy after all? We think that we should have a respected amount of privacy. A fair conclusion is that U.S. legislation does not protect users as much as EU legislation does, especially after the enactment of the Patriot Act (2001), which allows U.S. government agencies to retrieve information from every web company, as long as they have a significant reason to do so [JBM03].

However, the privacy level cannot be assured only by the applied legislation. Users should first know how to secure their privacy, while they browse and use web sites. As a result, the knowledge of how to ‘behave’ online, should become a part of people’s education; users should be raised to be more sceptical, learn to read an agreement before they accept it, know the categories of information whose online publication may have bad consequences.

A recommendation would be the development of a more flexible and robust legal framework which would be able to instantly adapt itself to new technology expressions. P3P is an example; not a complete one though. Semantics should be used to describe the operations of everything online and on the other side, legal systems should have a semantic representation. A complete application of semantics could be a primary solution not only to the adaptation of legislation to new technologies, but also for deciding if web applications are compatible with the law.

Adapting McNealy’s quote to modern reality, would result to something like “You have some privacy. Try to secure it.”.

References

- [AG06] A. Acquisti and R. Gross. Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. *Privacy Enhancing Technologies*, pages 36–58, 2006.
- [BEP04] D.L. Baumer, J.B. Earp, and JC Poindexter. Internet privacy law: a comparison between the United States and the European

- Union. *Computers & Security*, 23(5):400–412, 2004.
- [BRDM07] P. Beatty, I. Reay, S. Dick, and J. Miller. P3P Adoption on E-Commerce Web sites: A Survey and Analysis. *IEEE Internet Computing*, 11(2):65–71, 2007.
- [CLM⁺02] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. *W3C Recommendation*, 16, 2002.
- [Con06] G. Conti. Googling considered harmful. *Proceedings of the 2006 workshop on New security paradigms*, pages 67–76, 2006.
- [DHP07] C. Dwyer, S.R. Hiltz, and K. Passerini. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. *Proceedings of AM-CIS 2007*, 2007.
- [EB03] J.B. Earp and D. Baumer. Innovative web use to learn about consumer behavior and online privacy. *Communications of the ACM*, 46(4):81–83, 2003.
- [ECC06] S. Egelman, L.F. Cranor, and A. Chowdhury. An analysis of P3P-enabled web sites among top-20 search results. *Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet*, pages 197–207, 2006.
- [FCS⁺06] D. Frankowski, D. Cosley, S. Sen, L. Terveen, and J. Riedl. You Are What You Say: Privacy Risks of Public Mentions. *29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 565–572, 2006.
- [GA05] R. Gross and A. Acquisti. Information revelation and privacy in online social networks. *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, 2005.
- [Hod06] M.J. Hodge. The Fourth Amendment and privacy issues on the "new" Internet: Facebook.com and MySpace.com. *Southern Illinois University Law Journal*, 31:95–122, 2006.
- [Hsu06] C.J. Hsu. Privacy concerns, privacy practices and web site categories: Toward a situational paradigm. *Online Information Review*, 30(5):569–586, 2006.
- [JBM03] P.T. Jaeger, J.C. Bertot, and C.R. McClure. The impact of the USA Patriot Act on collection and analysis of personal information under the Foreign Intelligence Surveillance Act. *Government Information Quarterly*, 20(3):295–314, 2003.
- [JJJM07] T.N. Jagatic, N.A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, 2007.
- [JP04] C. Jensen and C. Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. *Proceedings of the 2004 conference on Human factors in computing systems*, pages 471–478, 2004.
- [LQ02] W.J. Long and M.P. Quek. Personal data privacy protection in an age of globalization: the US-EU safe harbor compromise. *Journal of European Public Policy*, 9(3):325–344, 2002.
- [OR05] T. O'Reilly. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, 2005.
- [Ros07] D. Rosenblum. What Anyone Can Know: The Privacy Risks of Social Networking Sites. *IEEE Security and Privacy*, 5(3):40–49, 2007.
- [Wei07] D.J. Weitzner. From Home to Home Page: New Challenges to Basic Notions of Privacy and Property. *IEEE Internet Computing*, 11(2):90–93, 2007.
- [Wri07] T. Wright. Promoting Data Protection by Privacy Enhancing Technologies. *CTRL-OXFORD-*, 13(5):147, 2007.

4 Appendix

4.1 Word count

- Words used in the main text of the article after the removal of headings and footnotes/citation/references pointers: 2690 *approx.*
- Words used in the footnotes that contribute to the main article: 374 *approx.*
- **Words to be counted: 3064 *approx.*** < 3300 words.