Chapter 2
Privacy in Social Software

Bibi van den Berg, Stefanie Pötzsch, Ronald Leenes, Katrin Borcea-Pfitzmann, and Filipe Beato

Abstract While using social software and interacting with others on the Internet, users share a lot of information about themselves. An important issue for these users is maintaining control over their own personal data and being aware to whom which data is disclosed. In this chapter, we present specific requirements and realised solutions to these problems for two different kinds of social software: social network sites and web forums.

2.1 Scenarios and Requirements

In recent years, a new generation of the Internet has emerged, also known as ‘Web 2.0’. One often quoted definition of Web 2.0 states that this term refers to a “set of economic, social, and technological trends, that collectively form the basis of the next generation of the Internet – a more mature, distinct medium characterised by user participation, openness, and network effects” [MSF09]. Web 2.0 has four fundamental characteristics that set it apart from the first generation of the Internet (‘Web 1.0’):

- Internet users have changed from passive consumers of information (searching for information and reading materials provided by others) into active creators of content [Tap09, How08, Lea08]. In Web 2.0, users can share their knowledge and information via a wide range of channels. Blogs, YouTube movies, wikis, file-sharing and consumer reviews are examples in case.
- In Web 2.0, social interaction plays a central role. This is why Web 2.0 is also called ‘the social web’.
- In many Web 2.0 environments, sharing and creating content and knowledge is not a solitary enterprise, but quite the reverse: the production and dissemination of information and entertainment services has a highly co-operative character. Participation and co-creation are key aspects of Web 2.0.
Web 2.0 also differs from the first generation of the Internet in a technical sense: technology developers now create applications that are *embedded* into the Internet, and are accessible via any browser. Thus, the Internet has become the central platform for users to access different types of software [O'R07]. Moreover, software is offered to users as a service rather than as a product to buy separately.

Since Web 2.0 is a new phenomenon – the term was first coined in 1999 but the massive take-off of this latest generation of the Internet is only a few years old – much is still to be learned with regards to both the benefits and the risks for users, businesses and governments in this new domain. Privacy issues relating to modern technologies have been high on the agenda of both government officials around the world, researchers, and the broader public, and for good measure, since it is obvious that the emergence of Web 2.0 currently generates a wide range of new issues relating to privacy and security.

As said, the success of the social web is based on the active participation of users, and on their willingness to contribute to the creation and improvement of content on the Internet by sharing data and knowledge [SGL06, O’R07]. By using social software, a lot of personal data is disclosed either directly – think of real names and birth dates on social networking sites – or indirectly, for instance through editing specific topics in a wiki, commenting on blog entries or posting statements in a forum [GA05, EGH08]. Furthermore, personal data can be generated by establishing connections with, or disclosing information by, second parties with or without the consent of the respective person. While the possibilities of the social web may enrich people’s lives on the one hand, there are also privacy risks involved. Five central privacy issues can be distinguished with respect to information and communication technologies in general, and Web 2.0 applications in particular:

- When sharing data and knowledge in social software, users lack an overview of who has access to this information – they cannot adequately judge the size and makeup of their *audience* [PD03, Tuf08].
- Information and communication technologies enable anyone to collect, copy, link and distribute the (personal) data of others, thus allowing for the creation of extensive profiles of individual persons. Information may also easily be copied outside the original domain, thus making it even harder for users to know who has access to their information [Hou09].
- Information and communication technologies allow storage of data for a nearly indefinite time period, thus making it impossible to erase or forget this information [MS09].
- Participatory information and communication technologies such as social software enable anyone to publish another individual’s personal data, which may have serious consequences for the other’s reputation [Sol07].
- Individuals’ lack of privacy-awareness when using social software may lead to information leaks and leaving unintended and/or unobserved virtual traces.

To find out which guises privacy issues take in the new generation of the Internet, much research has been conducted with regards to privacy in social software in