Collaborative Privacy - A Community-based Privacy Infrastructure

Jan Kolter, Thomas Kernchen and Günther Pernul

Abstract The landscape of the World Wide Web with all its versatile services heavily relies on the disclosure of private user information. Service providers collecting more and more of these personal user data pose a growing privacy threat for users. Addressing user concerns privacy-enhancing technologies emerged. One goal of these technologies is to enable users to improve the control over their personal data. A famous representative is the PRIME project that aims for a holistic privacy-enhancing identity management system. However, approaches like the PRIME privacy architecture require service providers to change their server infrastructure and add specific privacy-enhancing components. In the near future, service providers are not expected to alter internal processes. In this paper, we introduce a collaborative privacy community that allows the open exchange of privacy-related information. We lay out the privacy community's functions and potentials within a user-centric, provider-independent privacy architecture that will help foster the usage and acceptance of privacy-enhancing technologies.

1 Introduction

Today's rich offer of services on the World Wide Web increasingly requires the release of personal user data, which poses a growing privacy threat to Internet users. Web site providers use these personal data to create and analyze profiles or to trigger personalized advertisements. At the worst, personal information is released or sold to third parties.

Jan Kolter, Günther Pernul
Department of Information Systems, University of Regensburg, D-93040 Regensburg, Germany
e-mail: {jan.kolter, guenther.pernul}@wiwi.uni-regensburg.de

Thomas Kernchen
Steria Mummert Consulting AG, Französische Str. 48, D-10117 Berlin, Germany
e-mail: thomas.kernchen@steria-mummert.de

Motivated by users who needed technical means to protect their private data, privacy-enhancing technologies emerged [6, 14]. A frequently discussed subject in this area is anonymity on network level. On application level, privacy-enhancing technologies aim for solutions that assist users in controlling and managing the disclosure of personal data. However, most approaches rely on the compliance of service providers who are required to reveal their data handling practices truthfully.

The goal of this paper is the introduction of a collaborative privacy community that facilitates a service-provider-independent privacy management. We propose a user-centric privacy architecture and show the functions and the potentials of an inherent collaborative privacy community. Finally, we present a prototypical implementation of our solution.

The remainder of this paper is structured as follows. After describing related work in Section 2, we present an overview as well as the components of a user-centric privacy architecture in Section 3. In Section 4 we introduce the content, functions and the implementation of our collaborative privacy community. Section 5 concludes the paper with an outlook on future work.

2 Related Work

The Platform for Privacy Preferences (P3P) [7] represents an early privacy-enhancing technology system. Offering a suitable policy language that allows service providers to express machine-readable privacy policies, P3P enables a privacy agent on user-side to indicate deviations from previously-specified privacy preferences.

Weaknesses of P3P have been subject to frequent discussions in the past [10, 15]. As P3P assumes complete and truthful privacy policies, most service providers' hesitation to offer P3P privacy policies is a main reason for P3P's lagging acceptance.

Aiming to support users' ability to maintain their privacy, the European PRIME project[1] (Privacy and Identity Management for Europe) developed a privacy-enhancing identity management system, containing a privacy architecture with different design guidelines, protocols and prototypical scenarios [18].

The PRIME architecture (see Fig. 1) allows users to control the disclosure and the usage of their personal data [18, 25]. A significant element of the architecture is the PRIME Toolbox, which needs to be installed both on client-side and on user-side. The PRIME Toolbox incorporates all necessary components for privacy-enhancing identity management and enables users to manage and use different digital identities with varying personal data.

A further element of the PRIME architecture is the PRIME Middleware that integrates all PRIME components and coordinates the communication between PRIME interaction parties. The PRIME console serves as a graphical interface enabling users to set privacy-related preferences that are used to negotiate data handling practices with service providers. Furthermore, an overview of already disclosed data is

[1] https://www.prime-project.eu/

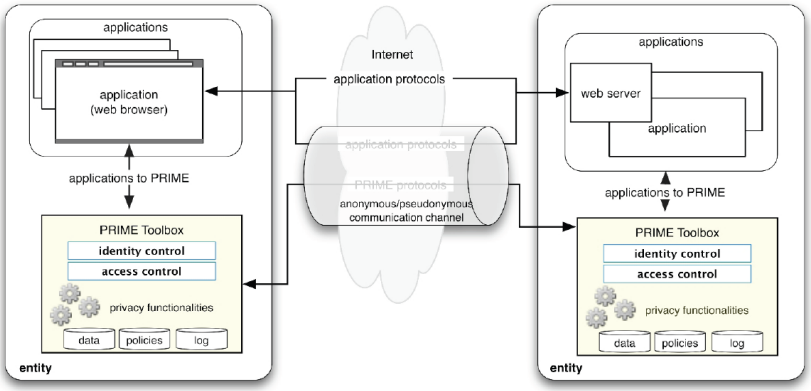

Fig. 1 High Level PRIME Architecture [18]

provided. The architecture is capable of enforcing negotiated policies, utilizing the installed PRIME components of service providers.

In order to make use of the described PRIME functionality, both users and service providers need to install the PRIME Middleware and the PRIME Toolbox. From a user perspective the attractiveness of PRIME rises, if the majority of service providers adapt their service infrastructure. Hence, the success of PRIME highly relies on the service providers' willingness to integrate the described PRIME components into their applications.

3 User-centric Privacy Architecture

In the last section we described existing privacy solutions that strongly rely on the compliance of service providers. From today's perspective it seems unlikely that service providers will fundamentally change their proven back-end services. Rising privacy threats of users will not convince service providers to adopt a comprehensive and complex privacy infrastructure. Furthermore, conflicting with their own interests, Web site providers will not contribute to the accuracy and quality of machine-readable privacy policies voluntarily.

Addressing these facts, we introduce a user-centric, provider-independent privacy architecture, employing a collaborative privacy community to share and exchange privacy-related information among Internet users. Unlike provider-dependent solutions our proposed architecture does not require service providers to set up additional components or functions. Accepting today's service landscape of the World Wide Web, we enable Internet users to control the disclosure and management of personal data.

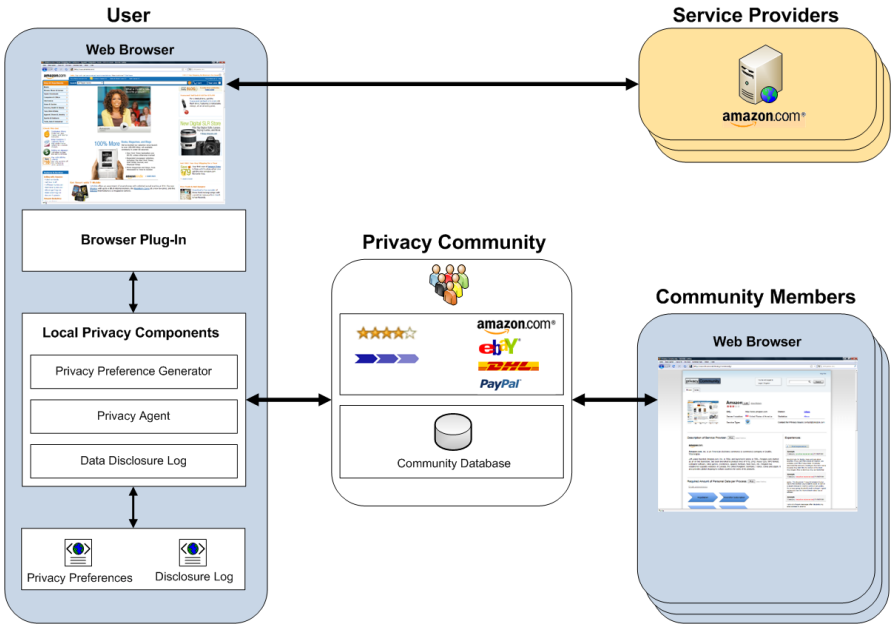

Fig. 2 Collaborative, Provider-independent Privacy Architecture

In Fig. 2 we present an overview of our privacy architecture. Seeking means to make an informed decision about the disclosure of personal data, the user is supported by a browser plug-in, which serves as the user interface. The browser plug-in displays privacy-related information and functions, which are provided by three local privacy components. The Privacy Preference Generator component assists users in controlling future information flows of personal data. The Privacy Agent component helps users check and control actual information flows. Finally, the Data Disclosure Log provides an overview of past personal information flows. All local privacy components interact with a collaborative privacy community, which provides supplemental privacy-relevant information about service providers. The community is maintained cooperatively by all participating members.

In the following we shortly discuss the main functions of each local privacy component, before the collaborative privacy community is introduced in Section 4.

3.1 Local Privacy Components

Potential information flows reflect a system's potential to disclose information [17]. From a privacy perspective, modeling users' privacy preferences, which define future disclosures of personal data, is a critical challenge.

Our user-centric privacy architecture provides a user-friendly Privacy Preference Generator component. The resulting privacy preferences reflect users' willingness to release personal data under certain circumstances and serve as basis for underlying privacy tools. APPEL [9], a privacy preference language built for P3P, provides a language to represent rule-based privacy preferences.

In our architecture, we allow users to define privacy preferences individually for different Internet service types [4], guaranteeing more realistic and practical privacy preferences.

Privacy tools that protect actual information flows help users make an informed disclosure decision, when personal data is about to be released to a service provider.

The presented privacy architecture employs a Privacy Agent component that supervises data transactions. If available, the agent reads the privacy policy of a service provider and matches it with pre-defined privacy preferences. Doing so, the agent recommends a certain behavior to the user. The P3P specification [7] provides the necessary technical means for the representation of privacy policies. The XACML standard [21] allows a more fine-grained and flexible definition of policies [2]. An example for a P3P-compliant privacy agent is the Privacy Bird [8], a browser plug-in for the Microsoft Internet Explorer.

Finally, our privacy architecture provides a tool that keeps track of all personal data transactions. Such a disclosure log allows users to manage personal data once they have been transferred to a service provider [22, 23]. A data transaction log bears the potential to present users a clear overview, which service provider stores what personal data at a certain time. This component requires both tracking functions that record and store data disclosures as well as usable interfaces that illustrate data transactions in an understandable way. Ideally, the disclosure log allows users to directly access, change or remove disclosed personal data stored by a service provider. Furthermore, a data disclosure log is capable of calculating potential linkabilities between data transactions.

4 Collaborative Privacy Community

Representing the central element of our presented privacy architecture, the collaborative privacy community facilitates the exchange of privacy-relevant information, ratings and experiences about service providers. These experiences involve, how personal data are used by certain service providers, and whether that usage is consistent with service providers' published privacy policies. These data represent a valuable, provider-independent information source for all three local privacy components, leading to a more informed disclosure behavior and enhanced privacy management of users.

The privacy community provides two access points. Internet users can browse a Wiki-like [19] Web front-end. Information about each service provider is grouped into articles, which can be viewed and edited by users. In addition, the privacy

community provides a Web service interface, allowing local privacy components on user-side to directly access necessary information.

4.1 Content and Functions

Underscoring the advantages of a provider-independent privacy infrastructure, we present the following structural and functional characteristics of our introduced privacy community.

For each service provider the community stores and offers static information, the required amount of personal data for each offered process, third parties the service provider shares personal data with, a description and evaluation of current and past privacy policies, the adherence to the published privacy policies, as well as individual experiences and ratings of community users. Additionally, the privacy community facilitates the controlled exchange of privacy preferences among connected users.

4.1.1 Static Information about Service Providers

When accessing an unknown Web site without privacy-enhancing technologies, users generally have the option to trust a service provider's privacy statement at face value or to find information about the service provider's reputation. A survey [12] shows that many users do not look up reputational information, but rather judge service providers' trustworthiness by estimating the Web site's "Look and Feel", considering questionable factors.

As collecting information about a service provider is time-consuming, this behavior of especially inexperienced users is understandable. Addressing this fact, the privacy community gives users an overview of information about service providers, such as the server location, the service type and a short description of the service offer. That information is utilized by the local Privacy Agent component and displayed to the user on demand, enabling users to easily retrieve necessary data to judge the trustworthiness of service providers. The provider's service type enables the Privacy Agent to more accurately match privacy preferences of the user with a Web site's privacy policy. The local Data Disclosure Log component benefits from information, how to access and revoke personal information that have already been transferred to a service provider.

In particular, static information about a service provider in our privacy community include:

- the service provider's URL
- the location of the server
- the offered service type(s)
- information how to change/revoke already transferred personal data

- contact information
- a short textual description of the service provider
- overall privacy rating

A URL is required to exactly identify each service provider. The server location clarifies jurisdictional matters, as different privacy laws apply in different countries. The offered service type(s) allow the application of more fine-tuned privacy preferences. As each service provider's service type (e.g. "Web Mail" or "Online Shopping") is accessible, privacy preferences can individually be defined and applied for each service type. Helping users exercise their rights to access and control already transferred data [11], the community provides information (e.g. a link or an e-mail address) how to change or remove these disclosed information. Exact contact information facilitates prosecution, if personal data are misused, or if users want to enforce their rights to revoke their personal data. Furthermore, a short description specifies the main characteristics of a service provider. Finally, an aggregated overall privacy rating shows a quick estimate of user ratings, which are explained below.

4.1.2 Required Amount of Personal Data

Our proposed privacy community enables users to know in advance, what personal information is needed to use a certain service in the World Wide Web.

Users generally understand the necessity to disclose, for example, their name, address and payment information for a product order at an online shop. If the service provider asks for additional information, such as the marital status, the date-of-birth or the annual salary, users tend to abort the process, if they feel uncomfortable releasing this excessive data. An online survey we conducted with 350 persons revealed that 77% of all test persons cancel registration and buying processes, if too many personal data are requested. Unfortunately, with today's technical means users are unable to determine at most Web sites, what personal information is necessary to use a specific service. To find out, users have to start the process of filling Web forms. In many cases the most privacy-sensitive information is requested on the last form page. If the user decides not to proceed, the frustrated user wasted valuable time and disclosed the already transferred information with no use.

The privacy community spares users from this negative experience and offers the amount of necessary data in advance. For each process a service provider offers the community stores all required personal data. In this context, a process refers to each separate action the service provider offers, such as "Buy" or "Subscribe to Newsletter". In addition, the community stores, when a process relies on the completion of a different process. The process "Buy" could, for instance, require the completion of the process "Registration". An automatic evaluation based on the service type assists users in evaluating the required amount of personal data a service provider requests.

As the amount of collected personal data represents a fundamental element of privacy policies, the local Privacy Agent can retrieve this information from the com-

munity and match it with individual privacy preferences, if no machine-readable privacy policy is available from the service provider.

4.1.3 Third Party Releases

The decision to disclose personal information to a service provider not only relies on the amount of data, but also on the service provider's data handling practices. Here, the release of user data to third parties is a considerably privacy-sensitive factor.

For each service provider the privacy community stores third parties the service provider shares personal data with. These parties could be affiliated companies or corporate networks. This information can be displayed to the user by the local Privacy Agent on demand. Again, information about third party releases can be utilized to replace a machine-readable privacy policy of the service provider.

4.1.4 Collecting and Explaining Privacy Policies

For many users the service provider's textual privacy policy is the only available information about data handling practices. Studies show, however, that privacy policies are not understandable to and are read by only a small fraction of Internet users [16, 24]. A privacy community facilitates experienced users to write an understandable description of privacy policies. As privacy experts comprehend all aspects of a policy, they have the ability to paraphrase important elements of the privacy policy in a form that - compared to automatic privacy policy summaries [3, 8] - is easy to understand.

Furthermore, as privacy policies change over time, the community keeps a history of privacy policies, containing both textual policies as well as machine-readable P3P policies, if available. This enables users to compare ex post, what privacy policy has been valid, when personal data have been disclosed to a service provider.

The privacy community also allows users to rate each stored privacy policy. A calculated privacy rank [1] supports inexperienced users to recognize and compare data handling practices of service providers.

4.1.5 Adherence to Privacy Policies

As a privacy-friendly privacy policy is no guarantee that a service provider will follow this expressed policy, our community enables users to rate the policy adherence of service providers. Based on their individual experiences users can evaluate, whether a service provider processes personal data as stated in a privacy policy. For example, if not expressed in the privacy policy, a personalized e-mail offering a product would justify a negative policy adherence rating of this service provider. Displayed by the local Privacy Agent, this information considerably influences users' decision to disclose personal data.

4.1.6 Sharing Privacy Preferences with Connected Users

In Section 3.1 we pointed out the purpose and usage of individual privacy preferences. The Privacy Preference Generator component allows the definition of these disclosure rules, which are in turn used by the Privacy Agent component to calculate disclosure recommendations. The quality of these recommendations strongly relies on the accuracy of the defined privacy preferences. Even though the Privacy Preference Generator component should alleviate this challenge by offering a usable and understandable user interface, building accurate privacy preferences is a critical task. This especially applies to inexperienced users, as they are not familiar with service providers' data handling practices and the privacy-related language used.

For these users the privacy community offers means to adopt privacy preferences from experienced users. Offering a social networking component [5], the privacy community allows users to upload and share privacy preferences with connected users. Privacy preferences of a trusted privacy expert represent valuable input for the local Privacy Agent of inexperienced users, resulting in improved disclosure recommendations.

4.2 User Management

The privacy community manages three different user roles. The basic user role is assigned to every user and allows the access of all information about service providers. Furthermore, it permits editing articles collaboratively. Basic users are able to create articles of new service providers. In order to prevent vandalism, the privacy community provides backup and archive functionality. An overview of all existing articles is available.

If users want to connect to other members of the privacy community, a simple registration is necessary. Registration only requires a username and a password. The community does not request any additional personal information. Registered users have the opportunity to upload their privacy preferences. Offering a social networking component, it is possible to look up and connect to friends who can share privacy preferences. We point out that this social networking component does not have the purpose of maintaining social contacts but only to exchange privacy experiences and privacy preferences. Users can self-assess their level of knowledge and experience in the area of privacy, helping inexperienced users to estimate the quality of their opinions and preferences.

Finally, users holding the administrator role define vocabularies of service providers' offered processes as well as personal data types. If necessary, administrators are able to block users/members.

4.3 Prototype

We implemented a prototype of our proposed privacy community. The Web front-end is available following this link[2].

4.3.1 Architecture

Figure 3 depicts the privacy community's architecture. As both the community's Web front-end and the local privacy components on user-side, access the community, we employ a service-oriented architecture (SOA) [20]. A SOA loosely couples client applications from the back-end and provides a high degree of interoperability. This enables a variety of clients to access the community database. Web services that provide a machine-readable WSDL definition encapsulate all information pieces of the community. Furthermore, the interaction via SOAP messages guarantees a consistent data exchange format.

For the Web front-end we utilize an Ajax-based [13] Web architecture, allowing asynchronous, interactive communication between the Web front-end and the community server. The client-side Ajax engine transforms JavaScript requests into SOAP requests, which are sent to the community server. The Web service server receives and processes requests querying the community database, before requested data are sent back to the client via SOAP. On client-side the Ajax engine transforms the SOAP messages to a user-friendly GUI, employing html and css. The local privacy components of our presented architecture - the Privacy Preference Generator (PPG), the Privacy Agent (PA) and the Data Disclosure Log (DDL) - directly access the Web service server via SOAP messages.

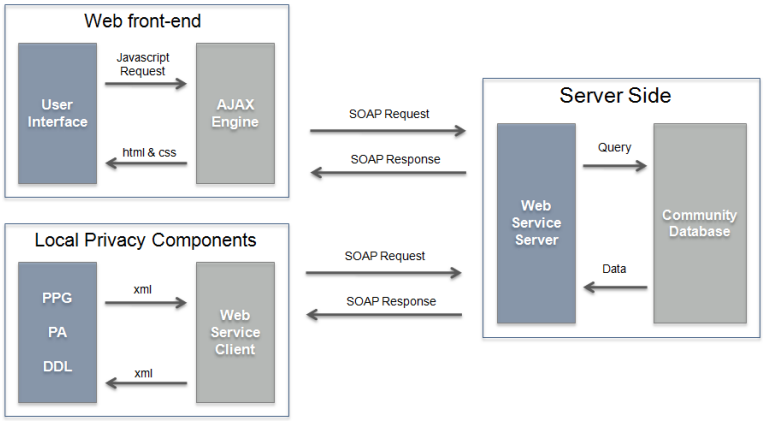

Fig. 3 Architecture of the Privacy Community

[2] http://www-ifs.uni-regensburg.de/Privacy/Community

4.3.2 Implementation

For the Web front end we utilize the JavaScript framework Yahoo! UI Library[3] (YUI), which offers the necessary drag & drop and autocomplete functions as well as overlays and browser history handling.

The back-end employs NuSOAP[4], a PHP-based Web service server that provides the required functionality for our proposed solution. The Web service interface definitions can be accessed following this link[5].

For the sake of brevity the interested reader is referred to the hyperlink above for a detailed review of the front-end design.

5 Conclusions

In this paper we present the concept and design of a collaborative privacy community. Marking a central element of our underlying user-centric privacy architecture, the privacy community allows a provider-independent exchange of privacy-relevant information and ratings about service providers. Moreover, our solution enables users to know in advance, what personal data is required for a specific service. Benefitting from the knowledge of experienced users, the privacy community facilitates a more informed decision about the disclosure and management of personal data. Provider independence as well as the collaborative character will contribute to a broader acceptance of privacy-enhancing technologies.

Future work will involve user tests as well as the integration of local privacy components.

References

1. R. Agrawal, W. I. Grosky, and F. Fotouhi. Ranking Privacy Policy. In *Proceedings of the 23rd International Conference on Data Engineering Workshops (ICDE 2007)*, pages 192–197. IEEE Computer Society, 2007.
2. A. Anderson. The Relationship Between XACML and P3P Privacy Policies, November 2004. http://research.sun.com/projects/xacml/XACML_P3P_Relationship.html.
3. F. Arshad. Privacy Fox - A JavaScript-based P3P Agent for Mozilla Firefox. Technical report, 2004.
4. M. Bergmann. PRIME Internal Privacy Preference Survey About Privacy Concerns and Condidtions. Technische Universität Dresden, Technische Berichte, TUD-FI07-04-Mai-2005, May 2005.
5. D. M. Boyd and N. B. Ellison. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

[3] http://developer.yahoo.com/yui/

[4] http://sourceforge.net/projects/nusoap/

[5] http://www-ifs.uni-regensburg.de/Privacy/Community/server_side/soap_server.php

6. H. Burkert. Privacy-enhancing Technologies: Typology, Critique, Vision. In P. Agre and M. Rotenberg, editors, *Technology and Privacy: The New Landscape*, pages 125–142. MIT Press, Cambridge, MA, USA, 1997.
7. L. Cranor, B. Dobbs, S. Egelman, G. Hogben, J. Humphrey, M. Langheinrich, M. Marchiori, M. Presler-Marshall, J. Reagle, M. Schunter, D. Stampley, and R. Wenning. The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. *W3C Working Group Note*, November 2006.
8. L. Cranor, P. Guduru, and M. Arjula. User Interfaces for Privacy Agents. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(2):135–178, 2006.
9. L. Cranor, M. Langheinrich, and M. Marchiori. A P3P Preference Exchange Language 1.0 (APPEL 1.0). *W3C Working Draft*, April 2002.
10. Electronic Privacy Information Center. Pretty Poor Privacy: An Assessment of P3P and Internet Privacy. Technical report, 2000.
11. European Parliament. EU-Directive 95/46/EC. Official Journal of the European Communities No L 281 31, October 1995.
12. B. J. Fogg, J. Marshall, O. Laraki, A. Osipovich, C. Varma, N. Fang, J. Paul, A. Rangnekar, J. Shon, P. Swani, and M. Treinen. What Makes Web Sites Credible?: A Report on a Large Quantitative Study. In *CHI '01: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 61–68, New York, NY, USA, 2001. ACM.
13. J. J. Garrett. Ajax: A New Approach to Web Applications. http://www.adaptivepath.com/ideas/essays/archives/000385.php, February 2005.
14. I. Goldberg, D. Wagner, and E. Brewer. Privacy-enhancing Technologies for the Internet. In *Proceedings of the 42nd IEEE Spring COMPCON*, San Jose, CA, USA, February 1997. IEEE Computer Society Press.
15. G. Hogben, T. Jackson, and M. Wilikens. A Fully Compliant Research Implementation of the P3P Standard for Privacy Protection: Experiences and Recommendations. In *ESORICS '02: Proceedings of the 7th European Symposium on Research in Computer Security*, pages 104–125, London, UK, 2002. Springer-Verlag.
16. C. Jensen, C. Potts, and C. Jensen. Privacy Practices of Internet Users: Self-reports versus Observed Behavior. *International Journal of Human-Computer Studies*, 63(1-2):203–227, 2005.
17. S. Lederer, I. Hong, K. Dey, and A. Landay. Personal Privacy through Understanding and Action: Five Pitfalls for Designers. *Personal and Ubiquitous Computing*, 8(6):440–454, 2004.
18. R. Leenes, J. Schallaböck, and M. Hansen. Privacy and Identity Management for Europe, PRIME white paper, version 3. https://www.prime-project.eu/prime_products/whitepaper/PRIME-Whitepaper-V3.pdf, May 2008.
19. B. Leuf and W. Cunningham. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley Longman, Amsterdam, 2001.
20. C. M. MacKenzie, K. Laskey, F. McCabe, P. F. Brown, and R. Metz. Reference Model for Service Oriented Architecture 1.0. *OASIS Standard*, October 2006.
21. T. Moses. eXtensible Access Control Markup Language (XACML) Version 2.0. *OASIS Standard*, February 2005.
22. J. Pettersson, S. Fischer-Hübner, and M. Bergmann. Outlining Data Track: Privacy-friendly Data Maintenance for End-users. In *Proceedings of the 15th International Conference on Informations Systems Development (ISD 2006)*. Springer Scientific Publishers, 2006.
23. J. Pettersson, S. Fischer-Hübner, M. Casassa Mont, and S. Pearson. How Ordinary Internet Users Can Have a Chance to Influence Privacy Policies. In *Proceedings of the 4th Nordic conference on Human-computer interaction (NordiCHI '06)*, pages 473–476, New York, NY, USA, 2006. ACM Press.
24. I. Pollach. What's Wrong With Online Privacy Policies? *Commun. ACM*, 50(9):103–108, 2007.
25. D. Sommer, M. Casassa Mont, and S. Pearson. PRIME Architecture version 3, Deliverable D14.2.d. https://www.prime-project.eu/prime_products/reports/arch/pub_del_D14.2.d_ec_WP14.2_v3_Final.pdf, July 2008.