University of Regensburg
Faculty of Business, Economics and Information Systems
Department of Information Systems

User-Centric Privacy – A Usable and Provider-Independent Privacy Infrastructure

Dissertation

Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften
Eingereicht an der Wirtschaftswissenschaftlichen Fakultät der Universität Regensburg

Primary Advisor: Prof. Dr. Günther Pernul
Secondary Advisor: Prof. Dr. Alfred Kobsa

Regensburg, September 30, 2009

Author:

Jan Paul Kolter
Hemauerstr. 15
93047 Regensburg
Germany
jan.kolter@wiwi.uni-regensburg.de

Abstract

Today's rich service offer in the World Wide Web increasingly requires the disclosure of personal user data. Service providers' appetite for personal user data, however, is accompanied by growing privacy implications for Internet users. Targeting the rising privacy concerns of users, privacy-enhancing technologies (PETs) emerged. One goal of these technologies is the provision of tools that facilitate more informed decisions about personal data disclosures.

Unfortunately, available PET solutions that protect personal user data are used by only a small fraction of Internet users. A major reason for the low acceptance of PETs is their lack of usability. Furthermore, most PET approaches rely on the cooperation of service providers that do not voluntarily adopt privacy components in their service infrastructures.

Addressing the weaknesses of existing PETs, this work introduces a user-centric privacy architecture that facilitates a provider-independent exchange of privacy-related information about service providers. This capability is achieved by the privacy community, an open information source within the proposed privacy architecture. A Wikipedia-like Web front-end enables the collaborative maintenance of service provider information that includes multiple ratings, experiences and data handling practices.

In addition to the collaborative privacy community, the introduced privacy architecture contains three usable PET components on the user side that support users before, during and after the disclosure of personal data.

The Privacy Preference Generator component offers a user-friendly tool that captures the conditions under which users are willing to disclose personal data. The designed configuration wizard generates these so called privacy preferences for up to twelve Internet service types, using understandable language and supportive icons. A privacy cockpit presents a summary of the configured settings, including an intuitive evaluation.

When personal data are about to be disclosed, the Privacy Agent component assists users in making an informed choice. In particular, the developed browser extension provides selected reputational information of the visited service provider, which is queried from the privacy community. Moreover, a matching of machine-readable privacy policies with previously defined privacy preferences of the user is performed. The offered assistance is complemented by a mechanism that observes the usage of partial identities and a list of already disclosed personal data.

Finally, the Data Disclosure Log component enables users to review past personal data disclosures. We created a browser extension that automatically detects and stores personal data transfers. Based on the resulting transaction history, an additional tool displays logged personal data disclosures from different perspectives, employing intuitive visualization techniques. A flexible graph-based view offers a user-defined visualization of relations between involved entity types.

All introduced components were prototypically implemented and underwent several user tests that guaranteed usability and user acceptance of the final versions.

The elaborated solutions realize usable interfaces as well as service provider independence. Overcoming the main shortcomings of existing PET solutions, we make a significant contribution towards the broad usage and acceptance of tools that protect personal user data.

Contents

List of Figures

List of Tables

Listings

Chapter 1

Introduction

1.1 Motivation

The trend towards easily accessible online services in the World Wide Web is building momentum ever since the Internet gained popularity in the late 1990s. The Web significantly changed the way people obtain and interchange information and was responsible for a tremendous shift in companies' marketing strategies and service offerings, as the Internet enabled businesses to offer information and services of any kind online. The user, on the other hand, benefited from availability and accessibility of the growing variety of services in the World Wide Web.

The rise of omnipresent online services, however, was accompanied by privacy implications, as most service providers require the disclosure of personal user data, before a service can be executed. A study revealed that 71% of German Internet users take advantage of the opportunity to shop online [Sta08], which indicates that a massive amount of personal user information is transferred to and stored by online service providers.

In fact, the enormous amount of stored personal data frequently leads to privacy violations. Examples of data misuses are, for instance, unsolicited promotional offers of service providers, which are enabled by submitted e-mail addresses and telephone numbers. A more dangerous privacy threat represents the trade of personal user data. A recent study proved how simple personal user information can be purchased on the Internet [NDR09]. These data sets include sensitive data like account and credit card information, whose misuse has severe consequences for the user. Even if disclosed personal data are handled responsibly by service providers, the privacy threat of a data theft remains. An actual incident involves the telecommunication company T-Mobile that was victim of a theft of 17 million customer data sets [Mey08].

In the wake of the rising number of personal data misuses, users increasingly get concerned about personal data disclosures [JPJ05, Kob07b]. Addressing these privacy concerns, the European Union enacted the Directive 95/46/EC on the protection of personal data [Eur95]. In addition to data minimization, the principles of the Directive include the necessity of an explicit user consent before the collection of any personal data as well as the exclusive use for the purpose stated in a published privacy policy. A further principle of the Directive targets

users' control of and access to personal data already transferred to a service provider. Similar legislations arose on national levels and in other areas of the world.

In order to take advantage of effective privacy rights, average Internet users rely on technical means that protect personal user information and facilitate a more informative decision about personal data disclosures. Targeting these user needs, privacy-enhancing technologies (PETs) emerged, a dedicated field of research in the area of IT-Security [Bur97, GWB97]. An important early PET system aiming at aiding users in decisions regarding the disclosure of their personal data is the Platform for Privacy Preferences (P3P) [CLM+02b]. Offering an appropriate policy language, P3P gives service providers the opportunity to express their privacy policy in a machine-readable format. When the user visits a Web site, a dedicated P3P privacy agent matches the P3P privacy policy of the service provider with pre-defined disclosure rules of the user (so called privacy preferences). The matching process results in a recommended disclosure behavior, which is signaled by the privacy agent. Another interesting PET that was built within the scope of the European PRIME project[1] is Data Track. The tool provides an overview of already disclosed personal data transactions, which represents a prerequisite for exercising users' right to revoke past personal data disclosures [PFHB06].

Unfortunately, available PET tools designed for the protection of personal data are used by only a minority of Internet users, while most users disclose personal data with no technical assistance and no information that would aid their decision-making.

Significant reasons for the low acceptance of PETs are apparent usability deficits. Available solutions that capture privacy preferences are not understandable for the general user. Studies prove that users are misled by the language and ambiguous expressions in the process of obtaining privacy preferences [FHP04]. Likewise, existing privacy agents fail to offer intuitive user interfaces that show the consequences of the user's data disclosure behavior. Logging tools like Data Track do not utilize the full visualization potentials and are not capable of providing a clear overview of a large number of data disclosures.

A further reason that prevents the widespread use of PETs is their dependency on service providers. The P3P specification, for instance, requires service providers to generate, publish and maintain a machine-readable P3P privacy policy, which is provided by only a small fraction of service providers [RBDM07]. Similarly, the use of Data Track assumes the installation of the PRIME middleware on the provider side. However, the wide adoption of powerful PRIME components in proven back-end infrastructures of service providers seems considerably unlikely.

The examples of P3P and Data Track demonstrate the evident conflict of interest in the area of personal data protection. While users are seeking to protect their privacy, service providers have no incentives to provide resources that restrict the collection of valuable user data.

Addressing the above-mentioned deficits, this dissertation focuses on the exploration, development and evaluation of a usable, provider-independent privacy infrastructure. Offering a collaborative privacy community, our solution does not rely on the cooperation of service

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

providers. Local components of the introduced privacy infrastructure effectively support average users in managing personal data disclosures. In particular, the developed components provide the understandable capture of privacy preferences, a comprehensible browser-based privacy agent as well as intuitive visualizations of logged personal data disclosures.

1.2 Research Questions

Focusing on the goals defined in the previous section, this work will address the following research questions:

- **What are digital identities and how are they managed in the World Wide Web?**
  The digital identity, which includes user's personal data, represents the protection goal within the scope of this work. Before the discussion of privacy and privacy measures, we need to define and analyze the management of personal data in the World Wide Web.

- **To what extent do service providers collect personal data and what is the reaction of users?**
  After the specification of the protection goal, we need to identify service providers' personal data collection practices and conditions as well as the disclosure habits of users. The user analysis includes an outline of prevalent privacy concerns, privacy attitudes and disclosure behaviors.

- **How can existing privacy-enhancing technologies protect personal user data?**
  The summary of the status quo in the context of our defined goals is complemented by a state-of-the-art analysis of available policy languages and tools that target the protection of personal user data. The outcome needs to identify weaknesses that are responsible for the lagging acceptance of these technologies.

- **What are the usability requirements of privacy-enhancing technologies?**
  In the previous section we pointed out the importance of usability for the emergence of privacy-enhancing technologies. For this reason, requirements of understandable tools and user-friendly interfaces need to be identified. The requirement collection includes the conduction of several user studies.

- **How can privacy management be executed more provider-independently?**
  An additional goal of this work is provider-independent privacy protection. Hence, we need to propose a privacy architecture that does not rely on the active support of service providers.

- **What are the content and design requirements of a collaborative privacy community?**
  Over the course of this work we propose a collaborative privacy community that enables the exchange of privacy-relevant information about service providers. Consequently, we

need to specify the required content, the architecture and the design of such a central privacy platform.

- **How can privacy preferences of the user be captured and fine-tuned in a user-friendly way?**
  The user-friendly generation of privacy preferences represents a crucial factor for the quality and accuracy of the recommended disclosure behavior determined and signaled by privacy agents. For this reason, we need to develop user interfaces that facilitate the understandable capture of these disclosure rules.

- **How can a privacy agent in the Web browser improve the user's basis for decision-making, when personal data are about to be disclosed?**
  In order to make a more informed disclosure decision, the average Internet user requires assistance, when personal user data are requested by a service provider. Thus, we need to develop a browser-based privacy agent that collects, processes and presents privacy-related information and recommendations in a comprehensible way.

- **How can historic transactions of personal data be stored and visualized in a user-comprehensible manner?**
  An ex post manipulation or removal of disclosed personal data inevitably requires an overview of past personal data transfers. Hence, we need to develop a solution that records personal data disclosures in the Web browser. The generated transaction log needs to be visualized employing intuitive user interfaces.

1.3 Methodology

Analyzing research methodologies in information systems research, Hevner et al. identify two main paradigms [HMPR04]. Originating from natural sciences, *behavioral science* aims at the development of theories that explain and predict organizational or human behavior within the context of information systems. On the other hand, *design science* represents a problem solving paradigm and has its roots in the engineering sciences.

In addition, the authors propose and discuss a set of design science research guidelines, including the necessity of proposeful artifacts for a specified problem domain [HMPR04]. Furthermore, the needs for a verifiable solution of a known problem and a thorough evaluation are pointed out. Finally, design researchers are required to communicate their work to technical and managerial audiences.

Considering the research questions defined in the previous section, we followed the design science research approach. Building on earlier work in this field, Peffers et al. propose and demonstrate the process of design science research, which is divided into six sequential phases [PTG+06]. In the following, the characteristics of the individual phases are sketched and applied to the outline of this work:

- **Problem Identification and Motivation**
  The definition of a problem and the justification of the value of its solution mark the goals of the first phase of the design science research process. In particular, Peffers et al. suggest the conceptual atomization of the problem in order to capture its complexity. The justification of a solution's value helps researchers and audiences to accept and understand the developed contributions. The problem identification and motivation require knowledge about the state of the problem.

  The motivation of this work is first introduced in Section 1.1, which highlights the lacking user-friendliness and the dependency of tools that assist users in protecting their personal data. Chapters 2 and 3 present an overview of prevalent user behaviors and service provider practices in the context of personal data disclosures in the World Wide Web. Finally, Chapter 4 lays out an analysis of existing privacy-enhancing technologies and their shortcomings.

- **Objectives of a Solution**
  This phase specifies the objectives of a targeted solution, which are inferred by the defined problems. The specification of objectives can include the definition of requirements as well as the proposal of a new artifact that solves the problems identified in the first phase.

  Chapter 5 proposes an approach that overcomes the dependency on service providers and addresses identified usability deficits. In particular, the chapter introduces a collaborative privacy architecture along with the purposes of its four components.

- **Design and Development**
  After the introduction of the proposed concept, this phase develops the tentative design of the proposed solutions, including the detailed definition of the artifacts' functionality.

  In this work, Chapters 6, 7, 8 and 9 discuss the functionality of the privacy architecture's four proposed components. The chapters also introduce the design of suitable user interfaces that contribute to usability and user acceptance of the components.

- **Demonstration**
  This phase aims at the analysis of the artifacts' efficacy to solve the defined problems and demonstrates the developed solutions in the targeted areas.

  The privacy infrastructure introduced in this work was prototypically implemented and is applicable in real-world scenarios. Chapter 10 provides technical details of the fully operational prototype.

- **Evaluation**
  The evaluation phase measures the extent to which a solution solves the problems defined in the problem identification phase. This assessment includes the comparison of actual results of the demonstration phase with the goals and requirements defined in the objective specification phase. The outcome of the evaluation phase can trigger another cycle of the design science research process.

Within the scope of this work, our developed solutions were evaluated in two user tests. The tests revealed remaining deficits of initial approaches and prototypes and served as input for the refinement of final versions. The conditions of the conducted user tests are sketched in Chapter 5, while the test results are integrated into Chapters 6, 7, 8 and 9.

- **Communication**
  Finally, in the communication phase the proposed solutions are presented to selected audiences, highlighting their importance, novelty, design and effectiveness.

  Selected contributions of this work have been published at internationally recognized IT conferences. The summary of all prior phases is presented in this work, which complements the design science research process.

1.4 Chapter Structure

Structuring the results of the research questions and the phase outcomes of the design science research process, we divide this work into three parts. In Part I we discuss the status quo of personal data disclosures in the World Wide Web as well as existing privacy-enhancing technologies and their usability weaknesses. Representing the main contribution of this work, Part II introduces a user-centric privacy infrastructure that provides a collaborative privacy community and three usable client components. Finally, Part III lays out details about the prototypical implementation of our solutions as well as a concluding summary. Figure 1.1 depicts the chapter structure of this work.

Part I: Fundamentals

Chapter 2 defines basic terminology of personal data disclosures in the World Wide Web. In that context, linkability of personal data sets and frequently used identity management solutions are discussed. Chapter 3 shifts the focus to privacy and the handling of personal user data. We sketch the business value of personal data and lay out actual privacy regulations. We continue with a discussion of observed privacy attitudes and disclosure behaviors of Internet users. In Chapter 4 we present a detailed evaluation of privacy policy languages and available privacy tools that support users in making an informed choice about the disclosure of personal data. Over the course of the chapter, we highlight usability deficits of existing solutions.

Part II: Proposed Privacy Management Infrastructure

Focusing on provider independence and usability, Chapter 5 introduces the characteristics and the components of our proposed privacy architecture. The chapter also outlines the conditions of user studies we conducted over the course of the component development. Chapter 6 defines the purpose, the composition and the design of a collaborative privacy community, which maintains information about data handling practices of service providers, serving as open

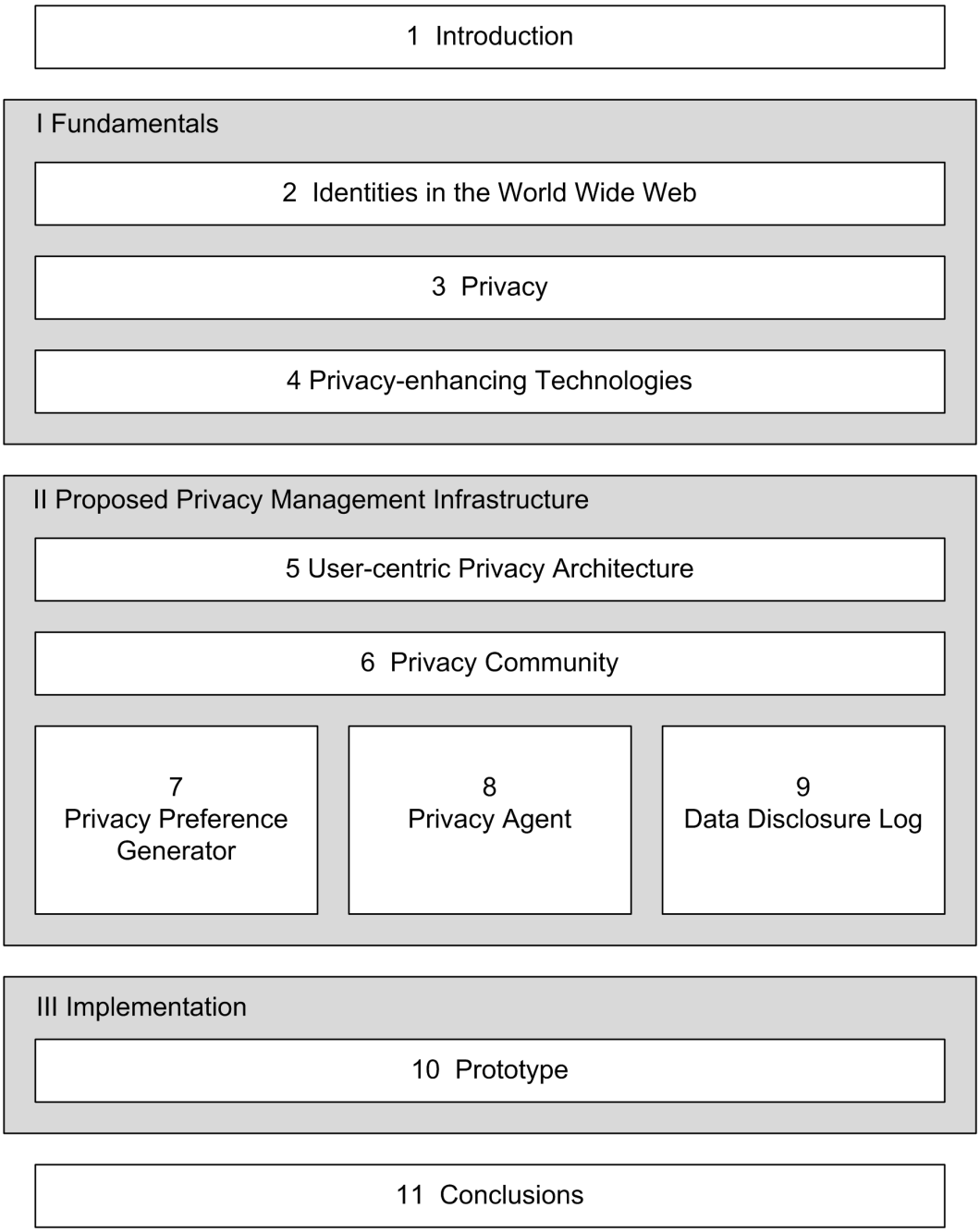


Figure 1.1: Chapter Structure

information source for the remaining local privacy components. In Chapter 7 we present a novel configuration tool that captures the preferred circumstances of personal data disclosures in a user-understandable way. Chapter 8 lays out the design of a browser-based privacy agent that offers privacy-related information about the visited service provider, when personal data are about to be disclosed. Finally, Chapter 9 discusses a logging component that keeps track of past personal data disclosures. The chapter presents a browser extension that records personal data transfers and introduces clear and intuitive visualizations of these logged transactions.

Part III: Implementation

Demonstrating the developed solutions of this work, Chapter 10 provides prototypical details of the implemented privacy infrastructure. Finally, Chapter 11 concludes this work by summarizing the results of the research questions defined in Section 1.2.

Part I

Fundamentals

Chapter 2

Identities in the World Wide Web

The disclosure of personal user data is a basic requirement for the use of many services in the World Wide Web. The transmitted personal data make up the user's digital identity and represent the user on the provider side. This chapter starts with basic definitions of the terms identity and partial identity and continues with a discussion about the potential linkability of partial identities. Finally, the chapter presents relevant identity management solutions that support users in controlling the disclosure of partial identities.

2.1 Identity

In general, user attributes that are electronically collected and stored are referred to as digital identity [CKK05].

The term digital identity is first mentioned in the context of data surveillance [Cla93] and defined as a set of user attributes bound to a distinct user in a certain domain. From the moment of creation the digital identity is used to identify the user in that domain [JZS07].

A broader definition of the term identity – without the specification "digital" – is suggested by Hansen and Meissner who define an identity as all available properties and characteristics of the user [HM07]. According to this definition, every user possesses a single identity. The digital identity represents a subset of the user's identity and includes attributes that can be captured technically.

In contrast, Pfitzmann and Hansen define an identity as all combinations of attributes that identify a user [PH08]. This implies that multiple identities can be generated and used based on the user's complete identity.

2.2 Partial Identity

A user generally does not transfer his/her complete digital identity when interacting with a service provider in the World Wide Web. Instead, only certain attributes of the identity are

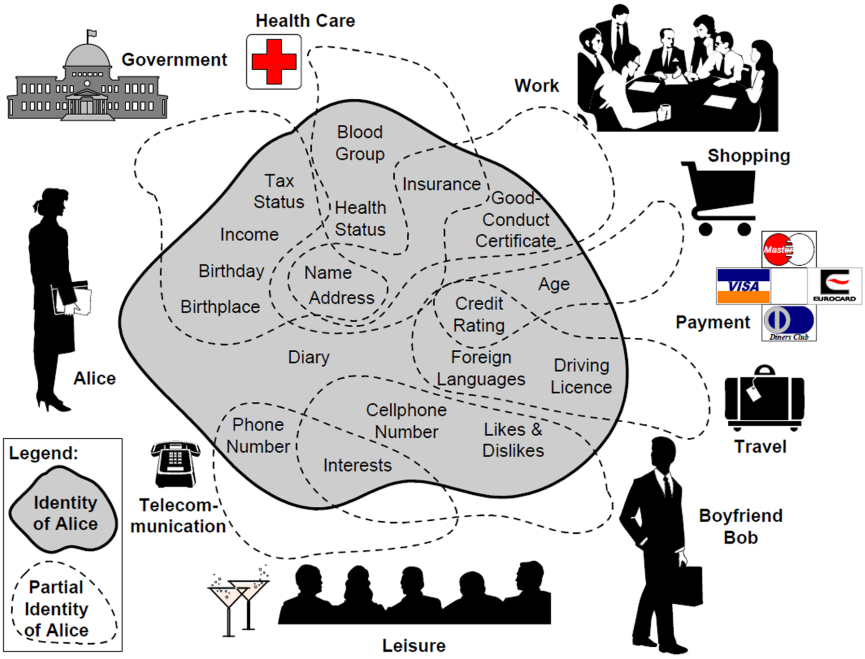


Figure 2.1: Use of Partial Identities [CKK05]

used for a transaction. This subset is generally referred to as partial identity [HM07, PH08] and represents a person in a certain context.

Partial identities are generated depending on the situation, the communication partner and the contextual requirements [CKK05] and, hence, include varying sets of attributes. This action corresponds to the natural behavior of humans in everyday life where identity management is carried out intuitively [CSS+05]. An example of such an intuitive release of personal information is the release of a credit card for a specific purchase. As the user interacts with numerous service providers, the use of multiple partial identities is a logical consequence (see Figure 2.1).

Unlike identities, a partial identity does not necessarily identify a user [CK01] and moves between anonymity and identifiability, depending on the amount, type and truth of the attribute values. If a partial identity solely consists of attributes that do not sufficiently identify a user, the user is regarded as anonymous within the set of all remaining users [PH08].

When services in the World Wide Web are used, total anonymity is rarely possible [CKK05]. Its corresponding lack of relation to the user prevents any attributability and, hence, the fulfillment of most services. Due to this fact, a partial identity is usually bound to a pseudonym, which serves as identifier for the service provider [BPFP05]. In a given domain a pseudonym represents a unique identifier that differs from the user's real names [PH08]. A pseudonym allows the attributability of user actions without identifying the user. When a pseudonym for a user is chosen, the pseudonym should not reveal any information about the user, including the linkability of the pseudonym to the user and the linkability of different pseudonyms [BPFP05].

2.3 Linkability

The linkability of used partial identities represents a serious privacy threat in the World Wide Web, as it allows cooperating service providers to combine different sets of disclosed attributes and to generate a larger partial identity of the user.

Pfitzmann et al. provide a general definition of unlinkability, specifying that two objects are unlinkable, if the attacker cannot determine whether or not a relation between both objects exists [PH08]. According to this definition two unrelated objects are also regarded as linkable, if the attacker can rule out a relation between them.

A definition of Hansen and Meissner is more application-oriented and focuses on linkable data, which exist, if different attributes bear identical values and if this is visible to humans or computers [HM07].

Both definitions show that it is not trivial for the user to recognize the linkability of used partial identities. A main reason for this difficulty is the perspective of the attacker. Without the consideration of the attacker's knowledge, linkability can only be estimated [SK03].

The protection against linkability threats requires a measurement of the possibility that disclosed personal data can be linked by third parties. In the following, we present existing approaches for the formalization and quantification of unlinkability.

Based on methods for the measurement of anonymity Steinbrecher and Köpsell specify a general information-theoretic model for the representation and quantification of unlinkability [SK03]. The model is based on a set of elements in a system that are defined by an equivalence relation. Elements share transitive relations to each other, which allows for the generation of equivalence classes. Elements are related, if they belong to the same equivalence class.

Furthermore, the authors define a degree of entropy that quantifies the unlinkability and expresses the likelihood that an attacker determines the relation of two objects. According to the model $P(a \sim b)$ represents the probability of linkability between the elements a and b. The degree of unlinkability is calculated by this probability and the converse probability. The formula results in a degree of unlinkability $d(a, b)$ ranging from 0 to 1:

$$d(a, b) = -P(a \sim b) * log_2 P(a \sim b) - P(a \nsim b) * log_2 P(a \nsim b)$$

For the calculation of the degree of unlinkability only the statistical distribution of elements within equivalence classes is used. The authors do not offer guidance, how an attacker determines those probabilities.

Building on the previously described model, Clauß defines a framework for the measurement of unlinkability in an identity management system [Cla06]. Unlike the model of Steinbrecher and Köpsell, this framework focuses on the attribute values of partial identities.

In particular, the model includes users, digital identities, messages and attributes. It assumes a finite amount of users who exchange their digital identities in messages. Moreover, possible values of a finite amount of attributes are defined in advance. An attacker intercepts messages

of interacting users and processes the contained attributes. As the model defines that attributes in a given message belong to the same user, messages can be linked by searching for identical attribute values.

The framework of Clauß also considers the statistical distribution of sent messages for the inference of the total distribution. For each intercepted message the appearance probability of all possible digital identities is calculated. Furthermore, a set of suspects is calculated, which contains all digital identities that match the attributes of the intercepted message. The set of suspects in combination with the appearance probabilities allows for an accurate calculation of the probability that a message belongs to a certain digital identity. Out of these values the linkability probability of two messages as well as a degree of entropy is calculated.

Malin presents an alternative for the above-described information-theoretic models. His model of k-unlinkability is based on the existing approach of k-anonymity and requires sensitive data to be linkable with not less than k identities [Mal08]. Applied to the World Wide Web, each service provider possesses a so called Trial Matrix. The rows of the Trial Matrix represent users, while the columns store transferred attribute values. A link occurs, if two rows from two associated Trial Matrices bear identical values. This, however, does not necessarily mean that both sets originate from the same user, as the selected rows may have additional links. The model requires that each row holds at least k links to associated Trial Matrices. In this context, k can be regarded as degree of unlinkability, as it reflects the challenge to find true links.

Unfortunately, the k-unlinkability model assumes associated Trial Matrices to contain the same set of subjects, meaning that the model requires all service providers to have identical users. This scenario, however, seems considerably unrealistic in the context of the World Wide Web.

2.4 Identity Management

Identity management is not clearly defined in literature and refers to systems with different characteristics. A frequently cited terminology of the FIDIS project[1] divides identity management systems into three groups [BMH05]. The first group defines account and access administration of computers and network services within organizations. These systems utilize directory services to store personal data for extended use. Solutions of the second group log user interactions and generate user profiles for further internal use. While the first two groups are primarily applied in large organizations, the third group of identity management systems focuses on user-controlled context-dependent role and pseudonym management, which corresponds to our examined context.

As indicated in the previous sections, users generally disclose large numbers of partial identities in the World Wide Web. Moreover, these multiple identities require the secure application of credentials, which becomes a growing challenge for most users. Offering mechanisms for the control of authentication credentials and personal attributes, identity management systems support users in the application of digital identities and their included partial identities [KP01].

[1] http://www.fidis.net/

Early identity management solutions in this area, so called identity silos, facilitate the simultaneous login for multiple service providers using the same identifier [JZS07]. This practice is also referred to as Single Sign-On. A famous representative of an identity silo is Microsoft Passport, which was later renamed to Windows Live ID. After a successful login on the central Passport identity provider, an issued token permits access to all participating service providers. Unfortunately, centralistic identity management systems like Windows Live ID provide little transparency with regard to the selection of attribute sets that are disclosed to service providers. Furthermore, Single Sign-On is limited to participating service providers. As all login actions are forwarded to a central identity provider, identity silos are frequently presumed to monitor users [PBP06].

Federated identity management systems like the Liberty Alliance[2] constitute a ”Circle of Trust” including service providers and identity providers that agreed to trust each other. Within the ”Circle of Trust” users can choose to federate identity information and share personal attributes with additional parties. However, federated identity management solutions offer little remedy for the identified deficiencies of identity silos. Specifically, these approaches are still restricted to a limited selection of identity providers and service providers [EMS07]. Moreover, users have little options to influence disclosed attribute sets [PBP06].

Addressing the deficits of identity silos and federated identity management systems, the focus shifted to user-centric identity management. In systems of this type, the control over identities and attributes originates from the user and not from service providers [RR06, EHG07]. More specifically, Kobielus postulates that in user-centric identity management systems users exclusively generate identities and users explicitly confirm all interactions with service providers [Kob07a].

Providing criteria for the evaluation of user-centricity, Cameron defined the widely recognized seven ”Laws of Identity” [Cam05]. These seven laws are collected from numerous discussions and observations and represent requirements to a user-centric identity management system. In particular, Cameron’s laws of identity require the explicit user consent for the disclosure of identifying user information. In order to mitigate the magnitude of a security breach, the amount of released personal information should be minimized. Moreover, Cameron proposes to make users aware of the interacting party. Finally, the use of multiple identifiers in favor of a unique global identifier is recommended.

Two modern identity management systems that are regarded as user-centric are OpenID[3] and Microsoft Cardspace[4].

OpenID offers users to select one out of numerous available identity providers [RR06]. When visiting an OpenID-enabled service provider, the user is directed to his/her identity provider to login. After a successful authentication, the user is redirected to the service provider that grants access to the service offer. Apart from the freedom to choose a preferred identity provider,

[2] http://www.projectliberty.org/
[3] http://openid.net/
[4] http://www.microsoft.com/windows/products/winfamily/cardspace/default.mspx

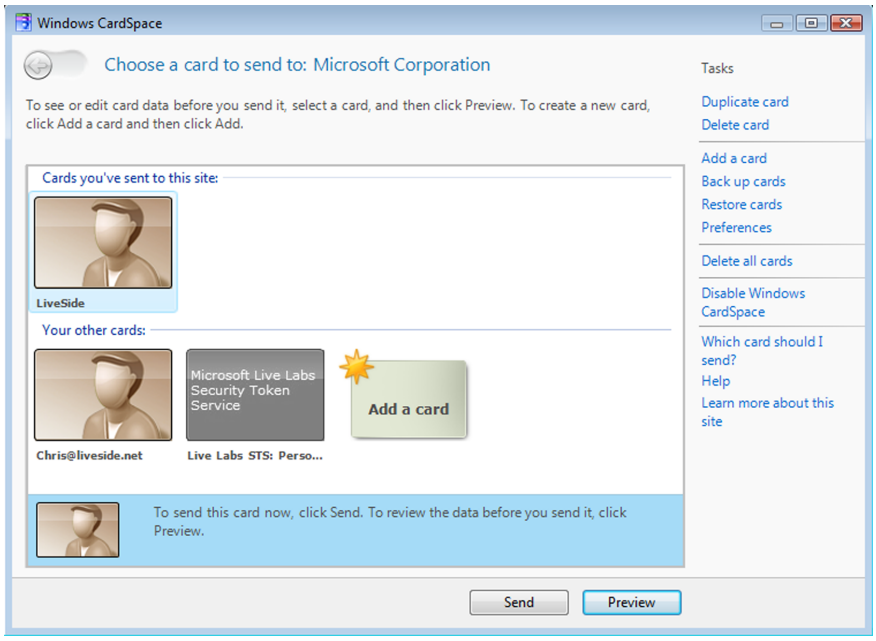


Figure 2.2: Microsoft Cardspace Interface

OpenID does not require users to disclose a context-wide identifier to service providers, which lowers the risk of linkable partial identities (see Section 2.3). In addition, OpenID identity providers facilitate the optional upload of personal attributes that can be forwarded to selected service providers.

Microsoft Cardspace implements a user-friendly identity selector for the platform-independent Microsoft Identity Metasystem [CJ06]. With the introduction of Windows Vista, Cardspace is an integral part of the Microsoft operating system. The client tool depicts generated partial identities of the user as cards in a clearly arranged user interface (see Figure 2.2, LiveSide.net[5]). Containing a set of personal user attributes, each card is either self-generated or directs to an identity provider that stores the respective attributes. If a service provider supports the Cardspace protocol, the identity selector on the client side automatically opens and asks users to select the adequate card to be disclosed. Consequently, the attribute values of the selected cards are transferred. Similar to OpenID, a context-wide unique identifier is not required. However, Cardspace depends on a locally installed client that maintains the individual card data of the user.

The discussed user-centric identity management systems highlight the advanced capabilities of modern identity management systems. While OpenID and Microsoft Cardspace provide comfortable mechanisms to manage and control the disclosure of partial identities, these systems do not support users in the decision about the disclosure of personal data and, hence, do not help protect the privacy of users. In the following chapters we shift the focus to privacy

[5] http://www.liveside.net/main/archive/2007/07/02/cardspace-infocard-and-live-id.aspx

and describe technical solutions that are specifically designed for the protection of personal user attributes.

2.5 Summary

This chapter defines the terms identities, partial identities and their relation to personal user data. We discuss the use of multiple partial identities in the World Wide Web and describe the purpose of pseudonyms. We continue with potential threats that originate from linkable partial identities and present approaches for the measurement of unlinkability. We concentrate on information-theoretic models as well as the k-unlinkability model. After a categorization of identity management systems, we characterize the development from identity silos and federated systems to user-centric identity management solutions like OpenID and Microsoft Cardspace.

Chapter 3

Privacy

In order to specify the scope of privacy within this work, this chapter starts with an analysis of privacy definitions. After summarizing the characteristics of personal user data, we weigh the business value of personal user information. We continue with an overview of effective privacy laws and regulations. Finally, we discuss user attitudes and user behaviors in the context of personal data disclosures.

3.1 Definition

One of the first definitions of privacy originates from the year 1890 and was published in the ”Harvard Law Review”. In the article ”The Right to Privacy” the authors Warren and Brandeis define privacy as

> ”the right to be let alone” [WB90].

This simple definition reflects the recognition of privacy at the end of the 19th century, when print media and new technological innovations like photography were responsible for the growing invasion of private and domestic areas. The authors express that the circulation of personal information is perceived as privacy violation and suggest prioritizing humans’ basic rights over the interest of the society to spread information.

Almost eighty years later in 1967, at the beginning of the information age, Westin proposed a novel definition of privacy. In his article of the book ”Privacy and Freedom”, Westin coined the well-known definition that privacy represents

> ”the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others” [Wes67].

A different view on privacy is provided by Goldberg et al. in the context of their work on privacy-enhancing technologies. This definition focuses on the user’s defensive abilities and specifies privacy as

> ”the ability of the individual to protect information about himself” [GWB97].

The cited definitions demonstrate the differing interpretations and recognitions of the term privacy. Considering the research questions we phrased in Chapter 1.2, the scope of this work is best summarized by the privacy perspective of Westin, which addresses the self-determined management of personal data.

In addition to the discussed views on privacy, several contributions have been published that define privacy in a variety of ways. For a comprehensive overview the interested reader is referred to [You78, Bur82, IP99, Sol08].

3.2 Personal Data

As personal user data represent the fundamental elements of our focused privacy perspective, we provide a brief overview of this sensitive information. According to the EU Directive 95/46/EC personal data are defined as

> ”any information relating to an identified or identifiable natural person (data subject); an identifiable person is one who can be identified, directly or indirectly, [...]” [Eur95]”

Analyzing this definition, we infer that personal data describe any characteristics attributable to a natural person.

The EU Directive provides six attribute groups suitable to describe a person, which include physical, physiological, mental, economic, cultural and social data [Eur95]. A broader scope of potential descriptive attributes is provided by user models. User Modeling targets the collection and structured storage of user data, which results in the accurate representation of a person [KKP01].

In addition to the characteristics of descriptive data, identifiable user data allow for the direct identification of data subjects [May06]. According to the Independent Center for Privacy Protection, Schleswig-Holstein, combinations of the following attributes have the function to grant uniqueness to a person [Ind03]: gender, first name, surname, date-of-birth, birthplace, number of birth certificate, identity of the parents, nationality, place of residence and profession. As the value of these identifying attributes has been proven in jurisdiction, many of these attributes are listed on national identification cards and passports.

The extent to which specific attributes are capable of uniquely identifying a person cannot generally be determined. For example, in a small group of persons the value of the attribute surname may be sufficient to uniquely identify each person. If the population of a whole city is considered, additional attributes are required (see also Chapter 2.3). In the context of k-anonymity, Sweeney proves that combinations of few demographic characteristics facilitate the unique identification in large populations [Swe02].

3.3 Business Value of Personal Data

In today's information age, service providers are able to easily capture, process and forward personal user data [HAF05], which leads to an increased collection of this information.

In many cases, a company's business model relies on the collection of personal user data, which are, for instance, used to tailor personalized advertisements [CS07]. Consequently, the amount of personal user data represents a significant asset for these service providers.

In the following we show the business potential of personal data for service providers.

3.3.1 Marketing

E-mail marketing represents a frequent form of direct marketing and plays an important role in the online marketing mix of companies [Lam06]. The sending of e-mails only requires valid e-mail addresses. Additional user attributes like address and personal interests allow for an individualized customer contact. Low costs and flexibility are regarded as main advantages of this type of marketing [Lam06].

Telephone marketing consists of all active and passive marketing measures that utilize the telephone [BK02]. Unfortunately, this type of direct marketing bears a considerably negative reputation. For this reason, Binder-Kissel recommends sales agents to listen 45% of the time of a telephone call [BK02]. In addition to a telephone number, effective telephone marketing relies on a full name, which is essential for the sales agent to relate a potential customer.

Personal user data also create additional options for a company's price policy. In particular, detailed information about customers allows for price discrimination. This strategy sets an individual price of a product or service for each customer [Odl03]. Inputs of these pricing methods are the individual price elasticity of customers (first degree price discrimination) or the sold quantity (second degree price discrimination) [CSW97]. Third degree price discrimination prices products or services depending on the segment or location of a customer. The required market segmentation is enabled by personal customer data.

3.3.2 Personalization

In addition to conventional marketing techniques, stored personal user data are increasingly used for personalization [Tre07], which is defined as any modification process of systems that considers information about the user [Sea03].

In general, personalization systems are divided into the following categories [Kob04]:

- **Adaptable Systems**
  Adaptable systems enable users to specify interests and preferences, which are considered in future user interactions of the system. The user manually adapts the system to his/her needs.

- **Adaptive Systems**
  These personalization systems execute fully automated adaptation processes. Behavioral user data and explicit user input serve as basis for adaptive systems.

Both types of personalization systems rely on the collection of user data, in order to determine specific user preferences and to execute adaptations to the system. For the collection of personal data user model acquisition methods are applied [KKP01].

According to Searby [Sea03] motivations for the application of personalization systems include targeted advertising, which generates increased revenue in the consumer market. Furthermore, personalization allows for the application of advanced customer relationship management that facilitates a customer-centric marketing approach. In that context, personalization has been identified as an essential instrument for customer loyalty [KKP01].

3.4 Privacy Laws

Privacy laws and regulations generally apply when identifying personal data are processed [WK08]. In the following we outline the background of existing privacy legislations.

Within the last thirty years, more than forty countries enacted national privacy laws, which define different technical and organizational requirements for the storage and processing of personal data in information systems [WK08].

A pioneering incubator for most effective privacy laws is the ”Guidelines on the Protection of Privacy and Transborder Flows of Personal Data” published by the Organisation for Economic Co-operation and Development (OECD) in 1980 [Org80]. The guidelines consist of non-binding recommendations for companies within the OECD member countries and aim at the international harmonization of privacy regulations, the facilitation of a free information exchange and the avoidance of unjustifiable trade barriers. We aggregate the OECD guidelines into the following principles:

- Notification & Consent
  In order to obtain data by lawful and fair means, users should be notified or explicitly accept the collection of personal data.

- Purpose
  Personal user data should be relevant for the purpose they are collected for. Additionally, the intended purpose should be specified and communicated to the user not later than at the time of data collection. In the consequence, personal data should exclusively be used for that specified purpose.

- Third Parties
  Likewise, personal data should not be forwarded to third parties without the prior knowledge and agreement of the user.

- Security
  Collected personal user data should be protected by the data collector from any potential misuse.

- Openness & Access
  Data collectors should follow a policy of openness with regard to privacy practices and privacy policies. Users should have the option to access stored personal data and to request the removal of personal data under certain circumstances.

- Accountability
  Finally, data collectors should be held accountable for the fulfillment of the stated principles.

In 1995 the European Union (EU) enacted the "Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data" [Eur95]. Unlike the OECD guidelines, the Directive is binding to all EU member states. Specifically, the legislation defines minimum standards that need to be implemented by national laws.

Most of the specified rules are based on the OECD guidelines. In addition to the OECD principles, the EU Directive regulates data traffic to non-EU countries. The data collector is only permitted to transfer personal user data to these countries, if privacy laws are effective that are in accordance with the EU Directive.

While the Directive dictates a national adoption in all member states, a review of the Commission of the European Communities in 2003 did not result in a positive evaluation [Com03]. The report states that privacy legislations in member states still show significant differences, which prevents multi-national organizations to develop European-wide data handling strategies.

The negative attestation triggered the foundation of a Work Programme for the better implementation of the Data Protection Directive. In 2007 a report of the Work Programme concluded that compliant privacy regulations are implemented in all member states [Com07].

In Germany the Federal Constitutional Court regards privacy as basic right, allowing persons to decide for themselves who to entrust personal information.

The Bundesdatenschutzgesetz (BDSG) defines general privacy regulations for public authorities as well as for private and corporate areas [BDS06]. The paramount purpose of the BDSG is to protect the individual from any processing of personal information that violates personal rights. The BDSG defines that the collection and processing of personal data is generally forbidden, unless a clear legal basis exists or the respective person gave explicit permission for that action. While the legislation covers personal information of a natural person, it is also applicable to anonymized and pseudonymized information.

The BDSG offers broad regulations for the protection of privacy, which may however be overridden by more specific regulations. The Telemediengesetz is an example for such a

sector-specific legislation, which regulates the processing of personal customer data collected by online service providers [TMG08].

In the United States privacy regulations are exclusively defined by sector-specific laws. Examples for these types of privacy legislations are the ”Health Insurance Portability and Accountability Act” [HIP96], which was enacted in 1996 for privacy in medical settings, or the ”Children’s Online Privacy Protection Act” [COP98], which came into effect in 1998 for the protection of children under the age of 13.

As the European Union does not regard U.S. privacy regulations as equivalent to the EU Directive 95/46/EC, the United States and the European Commission formulated the Safe Harbor Principles [U.S00], which resemble the rules of the EU Directive. If these principles are followed by U.S. companies, these ”Safe Harbors” qualify for the data import from EU countries.

The described privacy regulations only represent a summary of the worldwide privacy landscape. We pointed out that primarily the EU Directive 95/46/EC generated regulations that – given the proper implementation in all member countries – guarantee acceptable European-wide privacy protection. Unfortunately, outside Europe no binding super-national privacy laws exist. This situation implicates that in the World Wide Web users need to know the legal domicile of a service provider, in order to assess the legal privacy protection.

3.5 User Attitude & User Behavior

While privacy laws can only define general conditions for collectors of personal user data, the privacy attitude and the disclosure behavior of users represent elemental factors for effective privacy protection. We highlight surveyed privacy concerns of users and present studies that measure privacy attitudes over time. After acknowledging the fact that the actual behavior of users differs from their expressed privacy preferences, we identify criteria that influence the disclosure of personal data.

3.5.1 Privacy Concerns

In recent years numerous surveys have been conducted on privacy concerns. Within a summary of thirty surveys Teltzrow and Kobsa reveal that all studies unanimously underscore users’ concerns about the privacy implications of personal data disclosures and the monitoring of Internet activities [TK04].

In particular, the summary of surveys indicates that 70% to 95% of users are concerned about privacy and security of their personal information. Likewise, 82% to 95% refused to disclose personal information to a Web site at one time or another. Moreover, 27% of users expressed not to submit personal information to Web sites at all. The summary also points out that 6% to 40% send false or fictitious information to Web sites, when asked to register. Finally, 89% to 90% are concerned that businesses share personal user data for purposes that are different to

the originally expressed purposes. In addition to personal data misuses, users are concerned about various kinds of monitoring activities of Web site providers.

3.5.2 Privacy Attitude

Since the late 1970s Alan Westin conducted numerous privacy surveys that aim at an analysis of privacy attitudes of users [KC05]. In most of Westin's surveys a privacy awareness index is calculated that serves as indicator for trends in privacy recognition.

Participating test candidates represent randomly selected U.S. residents that are interviewed via telephone. In order to classify their privacy attitude, a set of statements about the business use of personal information is presented. Test respondents have the option to either strongly agree, somewhat agree, somewhat disagree or strongly disagree with a statement [Tay03].

After the first survey of this kind in 1991 [HW91], the wording of the statements has been adapted to actual trends and privacy threats. Since 1995 the following statements are read to test candidates [KC05]:

1. Consumers have lost all control over how personal information is collected and used by companies.
2. Most businesses handle the personal information they collect about consumers in a proper and confidential way.
3. Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

Responses to these statements are utilized to group test persons into one of the following privacy concern categories [Tay03]:

- **Privacy Fundamentalist**
  "[Privacy Fundamentalists] feel very strongly about privacy matters. They tend to feel that they have lost a lot of their privacy and are strongly resistant to any further erosion of it."
- **Privacy Pragmatists**
  "[Privacy Pragmatists] have strong feelings about privacy and are very concerned to protect themselves from the abuse or misuse of their personal information by companies or government agencies."
- **Privacy Unconcerned**
  "[Privacy Unconcerned] have no real concerns about privacy and [...] have far less anxiety about how other people and organizations are using information about them."

In particular, a test candidate is classified as privacy fundamentalist, if he/she agrees with the first statement and disagrees with statement two and three [KC05]. Conversely, if a test person

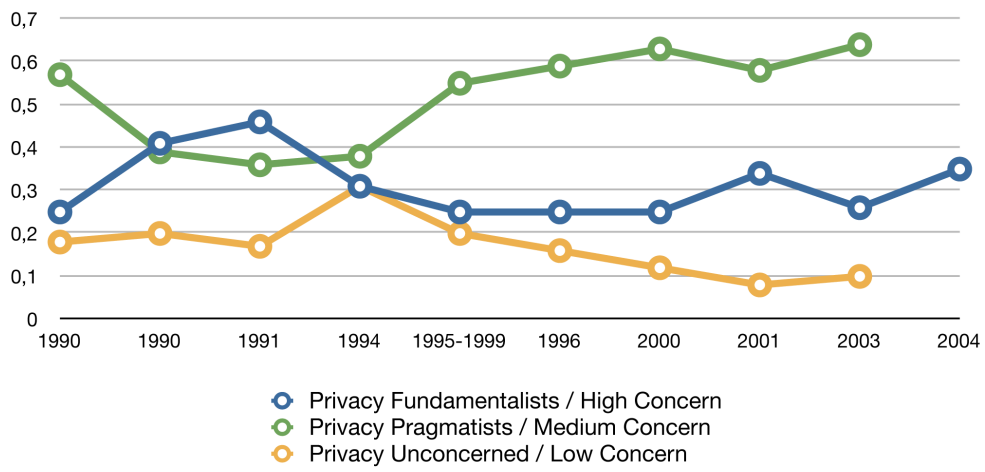


Figure 3.1: Development of Privacy Attitudes

disagrees with statement one and agrees with the following statements, he/she is regarded as privacy unconcerned. Test respondents with all remaining combinations of answers are classified as privacy pragmatists. We point out that for the evaluation no differentiation is made between the answers ”strongly agree” and ”somewhat agree” as well as ”strongly disagree” and ”somewhat disagree”, respectively.

Kumaraguru and Cranor lay out a summary of all Westin surveys conducted between 1990 and 2004 [KC05]. Their work allows for an analysis of privacy attitudes over time, which is illustrated in Figure 3.1. The distribution of privacy categories indicates a clear majority of privacy pragmatists. Over time a decreasing proportion of unconcerned users is observed, while the share of privacy pragmatists is rising constantly. The deviations before 1995 are explained by the differing statements used for the surveys. The outlying value of privacy fundamentalists in the year 2001 is regarded as consequence of the 9/11 attacks [HI02].

In an effort to refine the privacy attitude classification of Westin, Jensen et al. suggest the following set of revised statements that are built to detect distinctive properties of the Westin privacy concern categories [JPJ05]:

1. I am concerned about online identity theft.
2. I am concerned about my privacy online.
3. I am concerned about my privacy in everyday life.
4. I am likely to read the privacy policy of an ecommerce site before buying anything.
5. Privacy policies accurately reflect what companies do.

In addition to the four answers offered by Westin, test respondents have the option to express a neutral opinion to a statement. A test candidate is classified as privacy fundamentalist, if at least four privacy-oriented answers and no negative answer are given. If no statement is answered

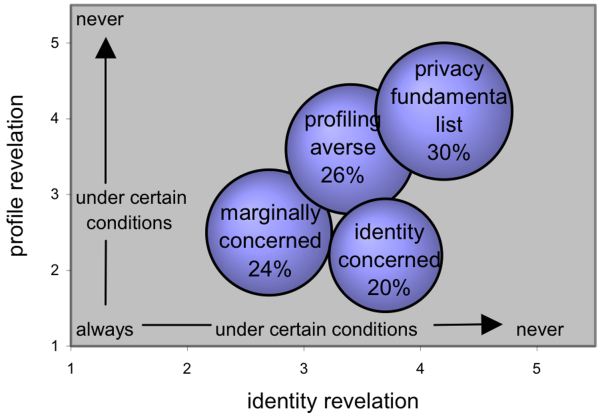


Figure 3.2: Four Clusters of Privacy Concerns [SGB01]

privacy-oriented and if at most one neutral answer is given, a test person is categorized as privacy unconcerned. The remaining combinations of answers identify privacy pragmatists. A survey conducted in 2005 resulted in shares of 34% privacy fundamentalists, 43% privacy pragmatists and 23% privacy unconcerned. Even though different statements have been used, the distribution of user groups approximately resembles the prior discussed results of the Westin surveys.

A survey of Ackerman et al. analyzes the privacy behavior of 381 online users [ACR99]. The study focuses on privacy concerns in specific scenarios, accounting for the situational factor of privacy. Similar to Westin, the classification of users results in three clusters. 17% of all test persons are categorized as privacy fundamentalists, while 56% build the pragmatic majority. Finally, 27% are classified as marginally concerned about privacy.

Spiekermann et al. examine the behavior of 206 test customers [SGB01] and identify four user clusters (see Figure 3.2). In addition to privacy fundamentalists (30%) and marginally concerned users (24%), the remaining pragmatic users are clustered into two additional groups, which differ within the scope of privacy concerns. Identity concerned users (20%) worry about the potential revelation of identifying data like name, address, or e-mail. Profile averse users (26%) are concerned about the potential user profile generation and the entailed inference of interests, hobbies and other personal information.

3.5.3 Deviations from Expressed Preferences

Contrary to prevalent privacy concerns proven by numerous user studies (see summary in Section 3.5.1), Ackerman et al. identify mismatches within the test results of a conducted survey (see Section 3.5.2), which hint at discrepancies between self-reported preferences of users and their actual behavior [ACR99].

Based on Ackerman's findings Spiekermann et al. compare self-reported privacy preferences to observed behaviors in a fictitious Web shop [SGB01]. The authors prove that users are

willing to disclose an increasing amount of personal data, even if stated differently. The most inconsistent behavior is found in the cluster of privacy-aware test candidates.

This observed distortion is supported by Behrens who found that 20% of all persons who stated they have placed orders at online shops expressed they would never disclose personal data like name and address on the Internet [Beh01].

In his work about privacy-enhanced Web personalization, Kobsa concludes that due to these apparent discrepancies Web developers should not be discouraged about the increasing privacy concerns revealed by consumer surveys, as in daily practice users are preoccupied with other factors that override their privacy concerns [Kob07c].

Instead of acting according to their privacy concerns, users seem to perform a cost-benefit trade-off, weighing potential privacy threats against a short-term use [Tay03, AG05]. Such a benefit can, for instance, be a personalized Web offer, a financial reward or a social benefit [Kob07c].

3.5.4 Factors Influencing the Disclosure of Personal Data

As the Internet is regarded as a place of incomplete and inconsistent information, users tend to apply heuristics and subjective criteria in decision making processes [BGS05], including decisions about the disclosure of personal data.

Cranor et al. [CRA99] conducted the first comprehensive user study that analyzed the disclosure behavior of users. The study's participants were statistically selected and represented a demographic average of the population. Specifically, the following main factors have been identified as relevant for the disclosure of personal data:

- Sensitivity of Personal Data
- Decision Factors for the Disclosure of Information
- Identifiability
- Tracking

The study confirms that every person has an individual *sensitivity* when it comes to the disclosure of a certain personal data type (see also [CS05]). This sensitivity, however, is hard to measure. Cranor et al. detect this sensitivity by asking users about their level of comfort, assuming that users' comfortableness decreases, if sensitive data are disclosed. The outcome allows an approximate calculation of each data type's sensitivity.

As illustrated in Figure 3.3, personal data like interests and preferences tend to be disclosed with little concern. Personal information that facilitates service providers to directly contact users is disclosed more reluctantly. The authors suspect that the differing level of potential annoyance of these data types impacts the disclosure behavior. For this reason, many users are willing to disclose their e-mail address, while most candidates do not feel comfortable

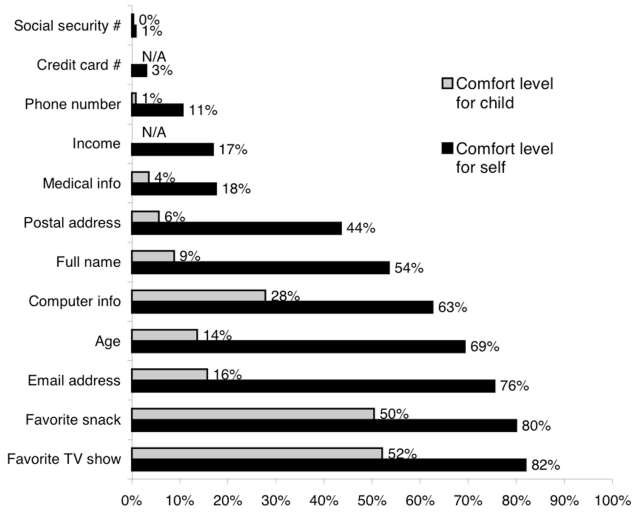


Figure 3.3: Level of Comfort for the Disclosure of Individual Data Types [CRA99]

providing a telephone number. Personal data like the social security number, a credit card number and the personal income show the lowest level of comfort.

Evaluating privacy policies of Web sites, Cranor et al. identify *decision factors* that impact a decision about the disclosure of personal data. The results are charted in Figure 3.4. Specifically, the authors identify data sharing with third parties, the identifiable use, the purpose of information collection and the kind of information as most important factors for the disclosure of personal data. Factors like the presence of a privacy seal and the presence of a data retention policy play a secondary role, which is explained by the lacking technical understanding of most users and their mistrust in service providers' ability to erase all user data. In addition, *identifiability* influences the disclosure behavior of users. As expected, users who are not identified – or believe that they are not identified – tend to disclose more personal information. *Tracking* refers to the observation of user's browsing behavior by one or more service providers, which can be achieved by the use of a personal identifier in cookies. Even though most users are able to associate tracking with cookies, they lack the technical knowledge to understand the involved processes [JPJ05]. Cranor et al. showed that users accept tracking for a specific use or value. They do not ban personalized advertisements in general, but do not accept that practice across multiple sites.

In a similar user experiment Jensen et al. offer test persons variants of Web shops that differ in privacy-relevant factors like contact data, effective privacy policies and the presence and use of graphical icons [JPJ05]. The findings of the survey show that unconcerned users value the existence of privacy-friendly policies. For pragmatists the authors derive a complex model that includes several factors. Privacy seals and the specific content of privacy policies are identified as most significant criteria. For fundamentalists no statistical dependency on any factor can be determined.

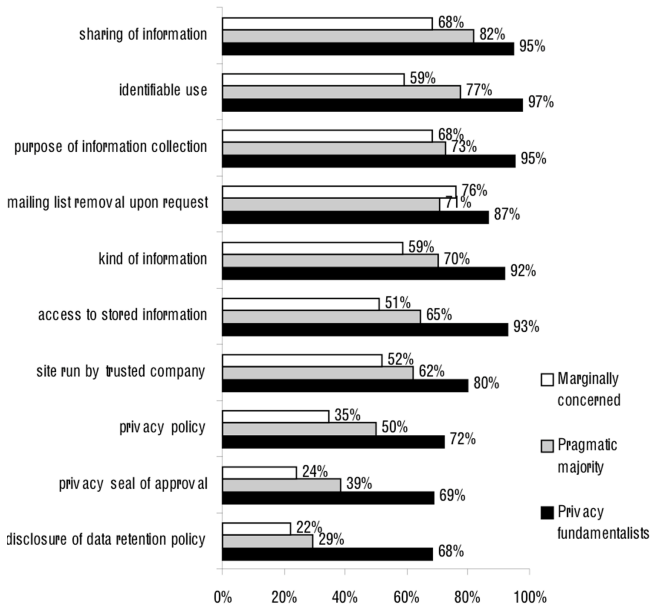


Figure 3.4: Decision Factors for the Disclosure of Personal Data [CRA99]

3.6 Summary

In this chapter we present differing recognitions and interpretations of privacy and focus on the self-determined management of personal data. After summarizing the main characteristics of personal data, we discuss the business value of collected personal customer information for marketing and personalization activities.

Furthermore, we lay out the development of supra-national and national privacy laws and regulations in Europe and the United States. While the European Union enacted the binding Directive 95/46/EC, which implements the OECD guidelines, the United States defined sector-specific privacy laws. In general, privacy regulations set basic rules for data collectors, but do not guarantee effective privacy protection.

We continue with an analysis of surveyed privacy concerns and privacy attitudes. Presenting the classification of Westin, we highlight the increasing majority of privacy pragmatists, while the fraction of privacy unconcerned is fading. Furthermore, we point out mismatches of expressed user preferences and observed behaviors, before we outline factors that impact the disclosure of personal data.

Chapter 4

Privacy-enhancing Technologies

In this chapter we focus on technologies and tools that assist users in protecting their privacy. As machine-readable privacy policies play an essential role in most approaches, we start with an analysis of suitable policy languages in the context of privacy. In the remainder of the chapter, we present an overview of existing privacy-enhancing technologies, focusing on solutions that support users in making an informed decision about personal data disclosures. Finally, we evaluate usability efforts made in this field of research.

4.1 Privacy Policy and Privacy Preference Languages

In general, a privacy policy of a service provider summarizes the usage of collected personal user data [KD08]. Most Web sites in the World Wide Web describe these data handling practices in a textual form, which, however, provides little clarity and understandability to users [Pol07]. Targeting this shortcoming, many privacy-enhancing technologies (PETs) aim at a clear presentation and intuitive evaluation of privacy policies, which requires a machine-readable representation of that information.

Based on a pre-defined categorization scheme this section focuses on suitable languages and frameworks for the machine-readable definition of privacy policies. This analysis includes the examination of corresponding privacy preference languages.

4.1.1 Categorization Scheme

The above-cited definition of privacy policies does not allow a clear identification of application scenarios. Seeking a precise categorization scheme for policy languages, we present a model that defines potential types of privacy polices. We consider a scenario where a data subject (user) discloses personal data to a data consumer (service provider) in the World Wide Web. The data consumer may forward these personal user data to a third party. Figure 4.1 illustrates the participating entities along with possible types of privacy policies, which are introduced in the following:

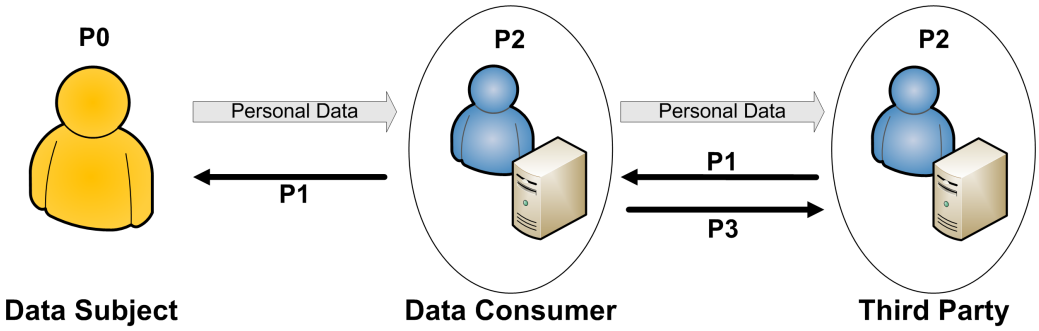


Figure 4.1: Policy Types in the Context of Personal Data Disclosures (adopted [MCMW06])

- **P0: Privacy Preferences of the Data Subject**
  This policy type allows data subjects to specify the conditions under which they are willing to disclose certain personal data. Since they are machine-readable, PETs can thereupon match them with published privacy assurances of the data consumer (see P1) and signal a resulting recommendation.

- **P1: Privacy Assurance of the Data Consumer**
  This classic form of a privacy policy is used to summarize privacy practices of the service provider in natural or machine-readable language [LLCL98]. Such a privacy assurance is usually published on the service provider's Web site. In the context of privacy and access control policies, Anderson names the consuming identity, the involved data, the action and the purpose/context as mandatory parts of such a policy [And04]. Anderson also mentions conditions and obligations, which apply to parts of policy types P2 and P3.

- **P2: Internal Use and Release of the Data Consumer**
  Policies of this kind allow data consumers to *internally* monitor the prior defined and published privacy assurances (see P1).

- **P3: Processing Rules for the Re-use of Data**
  If the data consumer forwards personal data to third parties, this policy type regulates that personal data are processed in line with originally published purposes and contexts.

Further aggregating the defined policy types, Kumaraguru et al. suggest the classification into languages that communicate the privacy policy of a company and languages that allow the specification of user preferences [KCLC07]:

- **Privacy Policy Languages** Privacy policy languages cover policy types P1, P2 and P3. External privacy policy languages, like P1 and P3, define policies that are published or exchanged between actors, while internal languages, like P2, describe internal rules of a data-consuming actor [WK08].

- **Privacy Preference Languages** Privacy preference languages correspond to policy type

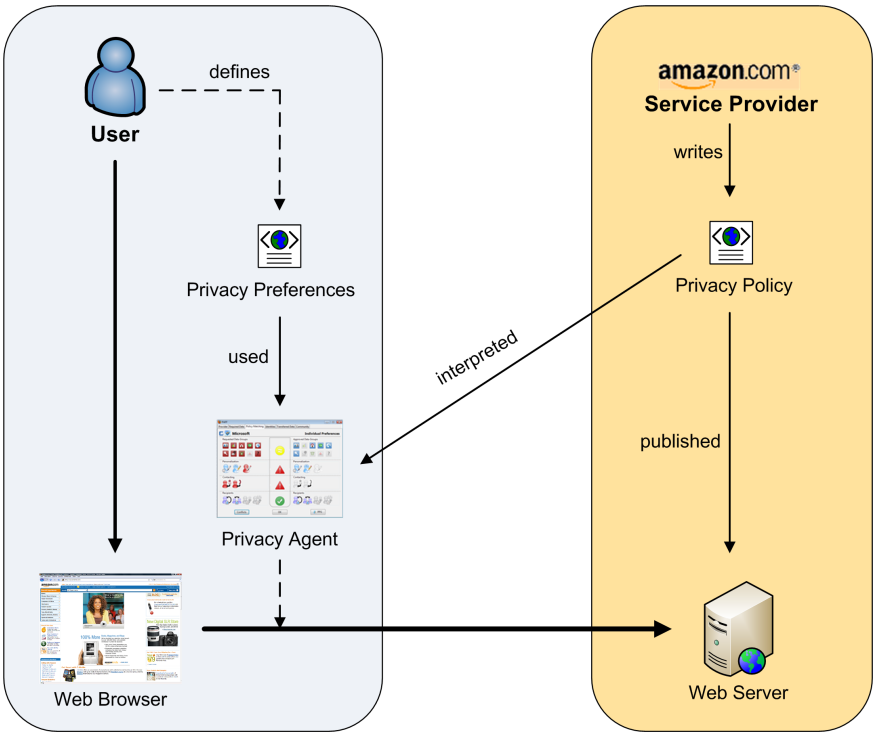


Figure 4.2: Schematic Overview of P3P

P0 and facilitate the machine-readable representation of privacy preferences of a data subject.

In the following sections we evaluate the capabilities of existing privacy policy and privacy preference languages.

4.1.2 Privacy Policy Languages

In this section we analyze policy languages that are suitable for the representation of external and internal privacy policies. Privacy policy languages should provide a high degree of functionality, in order to cover all defined policy types. Moreover, we weigh the expressiveness of a language, which guarantees the definition of all mandatory parts of a privacy policy (see Section 4.1.1, [And04]). An additional criterion is the provision of policy enforcement on the provider side.

We start with the well-known P3P specification and continue with the XACML and EPAL frameworks.

P3P – Platform for Privacy Preferences

The Platform for Privacy Preferences (P3P) was recommended by the W3C in 2002 and provides a vocabulary that defines the syntax and semantics of P3P privacy policies [CLM+02b]. The current version 1.1 was published as W3C Working Group Note in 2006 [CDE+06].

P3P allows service providers to express their textual privacy policies in a standardized, machine-readable format [RC99]. The resulting P3P policy is published at a known subdirectory on the service provider's Web server. When the user visits a Web site, the presence of a P3P policy allows a dedicated privacy agent tool on the user side to retrieve the P3P policy and match it with pre-defined privacy preferences of the user. Those user preferences are defined using the preference language APPEL (see below). Figure 4.2 depicts a schematic overview of the P3P policy matching process. Unlike textual privacy policies, P3P facilitates the automatic analysis and clear presentation of a service provider's privacy policy.

Before we focus on the elements of the P3P vocabulary, we list the following exemplary P3P policy:

```
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1" xml:lang="de">
 <POLICY name="example_policy" discuri="/privacy.html">
  <ENTITY>
   <DATA-GROUP>
    <DATA ref="#business.name">Name of Servie Provider</DATA>
    <DATA ref="#business.contact?info.online.email">mail@provider.org</DATA>
   </DATA-GROUP>
  </ENTITY>
  <ACCESS><none/></ACCESS>
  <STATEMENT>
   <DATA-GROUP>
    <DATA ref="#dynamic.clickstream"/>
    <DATA ref="#dynamic.http"/>
   </DATA-GROUP>
   <PURPOSE>
    <EXTENSION optional="yes">
     <PPURPOSE xmlns="http://www.w3.org/2006/01/P3Pv11">
      <communicate/>
     </PPURPOSE>
    </EXTENSION>
    <current/>
    <admin/>
    <tailoring/>
   </PURPOSE>
   <RECIPIENT><ours/></RECIPIENT>
   <RETENTION><indefinitely/></RETENTION>
  </STATEMENT>
 </POLICY>
</POLICIES>
```

Listing 4.1: P3P Policy Example

As the P3P specification enables service providers to assign an individual P3P policy to each segment of a Web site (e.g. a specific subpage or Web form), the element `POLICY` marks a distinct P3P policy, which is identified by the attribute `name` [CDE+06]. An additional attribute `discuri` links to the respective textual privacy policy. All policies of a P3P policy file are contained in the element `POLICIES`.

The element `ENTITY` consists of subelements that specify a mandatory name of the Web site's host. Additional elements allow the specification of contact information such as an e-mail

address, address information or a telephone number. The ACCESS element indicates the ability of a user to access disclosed personal information. In our example policy, this ex post access of disclosed personal data is not offered. An optional DISPUTES element allows the definition of dispute-resolution procedures. This tag could, for instance, name a service representative who is responsible for the handling of privacy disputes.

The STATEMENT element provides information about data handling practices of the service provider and contains the elements DATA-GROUP, PURPOSE, RECIPIENT and RETENTION.

Also used for the description of service provider information within the ENTITY element, DATA-GROUP specifies personal data types collected by the service provider. For the representation of data types the P3P standard defines a basic data type schema, which can be extended by the service provider. Furthermore, P3P provides 17 data type categories (like, for instance, financial or health). These aggregations can be used instead of fine-grained data types. The example policy indicates the collection of clickstream and http information.

After the specification of collected personal data, the PURPOSE element names purposes for the request of personal data. In addition to administration and development purposes, five elements focus on personalization, defined with regard to the persistence of the personalization activities and the identifiability of the used personal data. Two additional elements are devoted to contacting purposes and differ from the defined communication channels. Finally, the current purpose indicates that the requested data types are used for the fulfillment of the offered service. With the introduction of version 1.1 of the P3P standard an optional extension allows a more precise specification of the current purpose. For this reason, 23 primary purpose (PPURPOSE) elements are defined (for example sales or education). The presented example policy uses the requested personal data for a communication service. In addition, personal data are used for administration purposes (admin) and personalization activities of the current session (tailoring).

The RECIPIENT element lists groups of third parties the service provider shares personal user data with. The six provided elements include exclusively internal usage, distribution to partners that apply the same data handling practices and distribution to providers with unknown data handling practices. In the above example policy, the service provider states to use the requested data exclusively internally.

The P3P standard enables service providers to refine PURPOSE and RECIPIENT elements with an optional required attribute, which indicates offered opt-in or opt-out options of the user.

Finally, the RETENTION element reveals the time the service provider plans to store collected personal data. In the example policy, the provider signals an indefinite storage of the requested user data.

Currently, P3P along with APPEL represents the only comprehensive framework for the definition and matching of machine-readable privacy policies. However, while P3P exists for several years, most service providers refused to add P3P policies to their Web sites [ECC06].

Studies of Egelmann et al. analyzed the top results of frequently used search queries. The studies reveal that the vast majority of Web sites does not offer a P3P policy. Furthermore, most of the identified P3P policies showed syntactical policy errors. A survey of Reay et al. compared the adoption of P3P between 2003 and 2005 [RBDM07]. Similar to the studies of Egelmann et al., the survey found that the adoption of P3P is stagnant and that frequent errors occur in P3P policies. The comparative survey also unveiled the low maintenance of published P3P policies. Due to the apparent adoption rate and quality of P3P policies, most Internet users are not familiar with the term P3P [JPJ05].

Main reasons for the lagging acceptance of P3P are frequently discussed weaknesses of the P3P specification.

Focusing on functionality, the P3P standard is designed to communicate privacy practices to the user [CLM+02b]. Applied to Figure 4.1 this definition corresponds to policy type P1. P3P does not provide means for the definition of the remaining policy types P2 and P3.

In terms of expressiveness, the syntactically correct combination of P3P elements allows the definition of more than 36.000 P3P policies [CGA06]. Unfortunately, the standard does not provide understandable and semantically clear explanations of the element values [CGA06, Coy99], which limits the accurate translation of textual privacy policies [ECC06].

In general, it seems unlikely that the offered vocabulary is capable of exactly defining any privacy practices of service providers. For example, a service provider's specific purpose for a data collection may not be representable by a combination of only twelve purpose elements. In this case, the P3P standard suggests to choose the vocabulary terms that most closely match the privacy practices and, in addition, to provide a human-readable explanation [CLM+02b]. Such a remedy, however, seems barely acceptable in the sensitive topic of privacy. In this context, Stufflebeam et al. conclude that P3P neither has the scope nor the breadth to define complete company-wide privacy policies [SAHJ04].

Even though not claimed by the standard, the non-existent enforcement of privacy policies represents a further weakness of P3P. While a published P3P policy asserts privacy practices of the service provider, users are not able to control the usage of their personal data [SSA06]. This lacking enforcement requires users to trust the accuracy and truth of the service provider's privacy statements.

Moreover, P3P offers users to either accept or reject a published privacy policy [WK07]. The standard does not offer the possibility to negotiate privacy policies with service providers.

XACML – eXtensible Access Control Markup Language

The most recent version 2.0 of the Extensible Access Control Markup Language (XACML) was ratified by the OASIS in 2005 and standardizes the management of security policies, access requests and authorization decisions [Mos05a].

A main characteristic of XACML is the policy enforcement model, which is defined by IETF and ISO [YPG00, WSS+01, ITU95]. The components of the model are depicted in Figure 4.3.

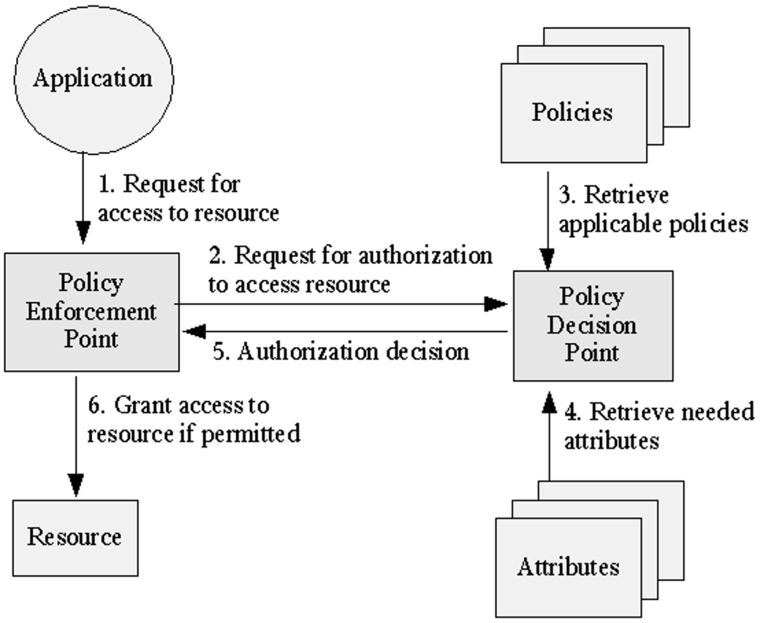


Figure 4.3: Components of the XACML Policy Enforcement Schema [And06]

According to the model access to a protected resource is rerouted to a Policy Enforcement Point (PEP) that formalizes the data request (1). Thereupon, the access request is diverted to a Policy Decision Point (PDP) (2). The PDP collects applicable policies for the access request (3). In addition, attributes relevant for the access request are retrieved (4). After an evaluation of the policies in combination with the received attributes, an authorization decision is calculated and returned to the PEP (5). According to the XACML standard this decision results in the values *Permit*, *Deny* or *Not applicable*. The latter value indicates that no suitable policy is available for the access request or that the available attributes are not sufficient to evaluate the policy. Based on the authorization decision the PEP grants or rejects access to the protected resource (6).

XACML policies, requests and decisions are expressed using a specific vocabulary whose elements are referred to as attributes. Unlike P3P, attributes are built domain-specifically; the syntax – including a unique identifier and a data type – and the semantics are defined by the involved applications. For an authorization decision the PDP requires the domain-specific attribute syntax, while it is not required to know the attribute semantics.

An XACML access request contains mandatory attributes about the accessing subject, the protected resource and the requested action [Mos05a]. Optional attributes provide additional environmental information. In the context of a privacy-related data request, the specification of a purpose is a mandatory part of the access request. For this reason, the XACML 2.0 extension ”Privacy Policy Profile” provides two additional attributes that name a purpose the data resource was collected for, and a second element that indicates the purpose of the request [Mos05b].

For the expression of policies the XACML standard provides attribute data types and matching functions. A policy set element contains all policies of an XACML policy file. Each policy

includes one or more rules that define access restrictions. The evaluation sequence of policies and rules are controlled by a set of policy- and rule-combining algorithms, which are expressed as attributes of the policy and rule tags. Listing 4.2 shows an exemplary XACML policy. Consider that the rules of the first policy evaluates to *Permit* and *Deny*. The rule-combining algorithm of policy 1 (`permit-overrides`) defines that any rule that results in a *Permit* decision overrides all remaining rules. Consequently, policy 1 evaluates to *Permit*. Policy 2 results in a *Deny* decision, as the first applicable rule evaluates to *Deny*. As the results of policy 1 and policy 2 contradict, the policy-combining algorithm of the policy set (`deny-overrides`) leads to a final *Deny* decision.

```
<PolicySet PolicyCombiningAlgId="urn:oasis:names:tc:xacml:1.0:deny-overrides">
  <Policy RuleCombiningAlgId="urn:oasis:names:tc:xacml:1.0:rule-combining-
          algorithm:permit-overrides">
    <!-- Policy 1 -->
    <Rule>
      <!-- Permit -->
    </Rule>
    <Rule>
      <!-- Deny -->
    </Rule>
  </Policy>

  <Policy RuleCombiningAlgId="urn:oasis:names:tc:xacml:1.0:rule-combining-
          algorithm:first-applicable">
    <!-- Policy 2 -->
    <Rule>
      <!-- Deny -->
    </Rule>
    <Rule>
      <!-- Permit -->
    </Rule>
  </Policy>
</PolicySet>
```

Listing 4.2: XACML Policy Example

Rules represent the most elemental parts of a XACML policy. Each rule contains conditions that specify functions to be applied for an attribute matching. The following example shows a rule that compares two attribute values of the access request. The rule evaluates to *Permit*, if the purpose the data resource is requested for equals the original purpose it was collected for.

```
<RuleEffect="Permit">
  <ConditionFunctionId="function:regexp-string-match">
    <ResourceAttributeDesignatorAttributeId="resource:purpose">
    <ActionAttributeDesignatorAttributeId="action:purpose">
  </Condition>
</Rule>
```

Listing 4.3: XACML Privacy Rule [Mos05b]

Even though XACML is primarily designed as an access control solution, the standard can be extended to cater the needs of privacy-related scenarios.

According to Madsen et al. XACML is solely suitable for the definition of policy types P2 and P3 [MCMW06]. As the syntax and the semantics of the standard are extensible, the application for a definition of a policy type P1 is possible.

Berthold considers P3P as a query language, as a P3P policy essentially queries attributes from the user [Ber04]. The author shows that the addition of the privacy-related purpose attributes to XACML facilitates the definition of queries similar to those in a P3P policy. While XACML does not yet provide the necessary vocabulary to match the expressiveness of P3P, the definition of additional privacy-related attributes could simply be integrated into the ”Privacy Policy Profile”. After such an extension XACML would emerge to a complete policy framework that covers policy types P1, P2 and P3.

Designed as access control solution, the XACML standard is fully capable of enforcing policies (see Figure 4.3).

EPAL – Enterprise Privacy Authorization Language

In 2002 IBM Research published a model for the formalization of privacy policies in enterprises [KS02]. The model identifies six necessary elements of a privacy policy. Based on this work, Ashley et al. introduce the ”Platform for Enterprise Privacy Practices” framework (E-P3P), which allows enterprises to automatically enforce privacy practices [AHKS02]. Finally, E-P3P served as foundation for the development of the ”Enterprise Privacy Authorization Language” (EPAL) as part of IBM’s Enterprise Privacy Management solution [AHK+03].

EPAL offers a formal language that – like XACML – resembles an access control solution [SAHJ04]. The framework is built on the same policy enforcement model applied in XACML (see Figure 4.3). Likewise, EPAL also provides a domain-specific definition of attributes (see above).

According to Bhattacharya and Gupta, EPAL represents a list of rules of differing priority [BG05]. Each EPAL rule lists an action a certain user group is allowed to perform on certain data [AHK+03]. Rules also define conditions that consider external context information. Moreover, obligations describe limitations of the performed action that need to be executed by the EPAL environment. Such an obligation could, for instance, involve the deletion of the requested data after a certain period of time.

In the following we show a simplified example of an EPAL rule. The example corresponds to a privacy policy that states to collect contact information for research purposes and to comply with strict legal restrictions. An EPAL policy used for the internal enforcement of such a statement is listed in the following:

```
1 <rule>
2   <ruling>ALLOW</ruling>
3   <userCategory>organization name</userCategory>
4   <action>DISCLOSE</action>
5   <dataCategory>ContactInformation </dataCategory>
6   <purpose>ResearchPurpose </purpose>
```

```
7   <condition>Disclosure subject to strict legal restrictions</condition>
8   <obligation/>
9 </rule>
```

Listing 4.4: Example of an EPAL Rule [Mos05b]

As the vocabulary used in an EPAL policy is specified in a separate file, EPAL allows the definition of slim rules. This design enables the definition of a vocabulary, which is valid for a whole corporate network or even an industry sector [AHK+03].

If more than one rule is applicable for a data request, EPAL evaluates the first applicable rule of a policy; the following rules are ignored. Compared to XACML, EPAL does not provide rule-combining algorithms.

As EPAL was built by IBM to facilitate the internal enforcement of privacy statements, the framework exclusively covers policy type P2. However, efforts to translate EPAL rules into P1 policies have been pursued [SVHW03]. In addition, the presence of a publishable vocabulary definition potentially facilitates the transfer of EPAL rules to third parties, which corresponds to a policy type P3 scenario.

With regard to expressiveness, EPAL offers all necessary policy elements to define internal privacy policies. As the specification was designed for the internal use in enterprises, it does not offer an element for the definition of disclosures to third parties. Furthermore, the lack of rule-combining algorithms complicates the definition of a foreseeable evaluation behavior.

Like XACML, EPAL is designed as an access control solution and, hence, enables the automatic enforcement of policies.

4.1.3 Privacy Preference Languages

After an analysis of policy languages that focus on data consumers and third parties, this section discusses languages for the expression of the data subject's privacy preferences. These preferences allow a dedicated user tool to automatically calculate a disclosure recommendation based on a service provider's machine-readable privacy policy.

Like for privacy policy languages, expressiveness represents a crucial evaluation criterion. In this context, a high degree of expressiveness facilitates the generation of rules that accurately reflect a data subject's data disclosure behavior. A suitable privacy preference language should also be semantically consistent, meaning that preferences that differ syntactically but match semantically should not result in ambiguous policy matchings. Finally, we consider the syntactic and semantic compatibility to privacy policy languages, which is a prerequisite for a successful policy matching.

In addition to the P3P preference language APPEL, this section analyzes XPref and presents an overview of the semantic preference language Rei.

APPEL – A P3P Preference Exchange Language

The P3P standard (see Section 4.1.2) proposes the idea of an automatic matching of privacy policies with user's individual privacy preferences [CLM+02b]. For this purpose, along with the P3P standard "A P3P Privacy Preference Exchange Language" (APPEL) was introduced by the W3C [CLM02a]. APPEL facilitates the rule-based definition of privacy preferences and complements the concepts of the P3P standard.

In particular, APPEL rules contain conditions that define acceptable or unacceptable content of a P3P policy. An applicable APPEL rule triggers a defined action, which is specified by the `behavior` attribute of the rule element. APPEL defines three activity values for `behavior`. The `request` activity indicates that the user accepts the privacy policy of the service provider, while `block` signals the opposite. The `limited` activity specifies that the user partially accepts the processed privacy policy. A supplemental `prompt` attribute implies that the user wishes an explicit notification about the application of the respective rule. A corresponding agent could, for instance, implement such a notification with a pop-up window.

Demonstrating the structure and the composition of APPEL, the following listing presents a simple APPEL rule set that includes two rules:

```
1  <appel:RULESET xmlns:appel="http://www.w3.org/2002/04/APPELv1"
2                 xmlns:p3p="http://www.w3.org/2000/12/P3Pv1">
3
4    <appel:RULE behavior="block">
5      <p3p:POLICY>
6        <p3p:STATEMENT>
7          <p3p:DATA-GROUP>
8            <p3p:DATA>
9              <p3p:CATEGORIES appel:connective="or">
10               <p3p:physical/>
11               <p3p:demographic/>
12             </p3p:CATEGORIES>
13           </p3p:DATA>
14         </p3p:DATA-GROUP>
15         <p3p:RECIPIENT appel:connective="or">
16           <p3p:same/>
17           <p3p:other-recipient/>
18           <p3p:public/>
19           <p3p:delivery/>
20           <p3p:unrelated/>
21         </p3p:RECIPIENT>
22       </p3p:STATEMENT>
23     </p3p:POLICY>
24   </appel:RULE>
25
26   <appel:RULE behavior="limited" prompt="yes"
27               promptmsg="Suspicious Policy. Do you want to continue?">
28     <appel:OTHERWISE/>
29   </appel:RULE
30
31 <appel:RULESET/>
```

Listing 4.5: Simplified APPEL Rule Set [CLM02a]

The first rule (line 4) defines a block rule that specifies elements of a P3P privacy policy, which are not acceptable for the user. The example shows that this rule includes the appearance of specific data categories and recipients. APPEL provides a set of `connective` attributes for the definition of logical connections of child nodes. The APPEL connective `or` in the `CATEGORIES` element, for instance, specifies the occurrence of either the physical or the demographic data category.

The second rule (line 26) employs the `OTHERWISE` element to capture those P3P policies no applicable rule is defined for. The `limited` attribute shows that the user partially accepts those privacy policies. Additional attributes signal the necessity of a prompt message along with the content of that message.

The sequence of the rule evaluation is defined by the order of the rules within the rule set. Similar to EPAL, no additional rules are considered after an applicable rule has fired.

In summary, APPEL was tailored as a slim and understandable privacy preference language. Designed as part of the P3P framework, it is semantically compatible with P3P privacy policies.

Unfortunately, the language's expressiveness is significantly limited by the provided connectives [AKSX03]. If connectives are defined within an element, all child nodes are affected by that single connective; the combination of different connectives in a set of child elements is not possible and requires the definition of multiple rules.

Additionally, the matching semantics represent a further deficiency of APPEL and prevent a widespread use in practical scenarios. In particular, the language does not provide elements that allow the prioritization of rules. Instead, rules are evaluated, as they appear in the rule set. The first applicable rule is fired, while the remaining rules are ignored. This evaluation leads to an unexpected and erratic behavior in certain cases.

The identified deficiencies prove an insufficient semantic consistency, especially when rules that express what is acceptable are evaluated. The proper evaluation of APPEL rules can only be guaranteed, if unacceptable parts of a privacy policy (block rules) are defined [AKSX03]. This practice, however, complicates the generation of rule sets.

XPref

Seeking remedy for the identified shortcomings of APPEL, Agrawal et al. propose the preference language XPref [AKSX03], which utilizes the XML query language XPath [BBC+07]. XPath defines the standardized addressing of parts of an XML document and seems an obvious choice for the evaluation of XML-based P3P policies. Compatible to P3P, XPref offers an interesting alternative for APPEL.

XPref reuses several elements of APPEL, such as `RULESET`, `RULE` and `behavior`. Representing the core of XPref, the rule conditions are not expressed with rule subelements, but with XPath expressions, contained by the attribute `condition`. Listing 4.6 shows an exemplary XPref rule:

```
<RULESET>
  <RULE behavior="block"
     condition ="/POLICY/STATEMENT/PURPOSE/*
         [(name(.) = "contact" or
           name(.) = "telemarketing")]" />
</RULESET>
```

Listing 4.6: Exemplary XPref Rule [AKSX03]

The `condition` attribute starts with a path that addresses all child nodes of a P3P policy statement's `PURPOSE` subnode. Subsequently, these child nodes are matched with the purposes `contact` and `telemarketing`. The condition of the XPref rule is satisfied, if at least one of these purposes matches. Additional XPath operators facilitate the evaluation of multiple statements of a P3P policy.

Unlike APPEL, the characteristics of XPref enable users to define both acceptable and unacceptable conditions [KCLC07], which contributes to the expressiveness of the language. Unfortunately, targeting P3P policies and based on APPEL, XPref inherits several weaknesses of these specifications, such as the limited vocabulary for the definition of preferences [KDJ+05].

While Agrawal et al. present a detailed overview of APPEL's weaknesses [AKSX03], they do not prove the semantic consistency of XPref rule sets [BJW08]. As XPref follows the same syntax orientation as APPEL, the same inconsistencies are likely to occur during the evaluation of rules [LYA03].

Rei

Addressing the limited expressiveness and semantic inconsistency of syntax-oriented languages, Kagal et al. introduce the flexible policy language Rei [KFJ03], which is built on deontic concepts [MW93].

Specifically, Kolari et al. discuss the application of Rei in the context of user preferences [KDJ+05]. A domain-specific ontology provides relevant classes and properties for the definition of privacy preferences. Many elements of the ontology correspond to elements of APPEL. Additional preconditions allow the filtering of rules, before conditions are evaluated. Like described in prior preference languages, conditions allow the reference to elements of a P3P policy. Moreover, statements of additional ontologies can be used, which facilitate the expression of fine-tuned privacy practices. A variety of connectors further increases the expressiveness of Rei.

Additionally, the language allows the definition of rule priorities, which assist users in determining the evaluation behavior. Rei also provides elements for the specification of obligations that are bound to granted actions.

Compared to syntactically oriented languages, the ontology-based concept of Rei offers a maximum variety of privacy preferences. Even though not covered by the authors, the semantic orientation of Rei contributes to the consistency of the language.

4.1.4 Summary

After the presentation of a privacy policy categorization scheme, this section analyzes and evaluates capabilities of privacy policy languages (policy types P1, P2 and P3) and privacy preference languages (policy type P0).

The P3P specification provides a complete framework for the definition of privacy policies. However, frequent criticism of the standard includes its limited expressiveness, imprecise semantics and lacking negotiation and enforcement capabilities. XACML, on the other hand, offers an access control solution that has the potential to serve as powerful privacy policy framework, featuring a policy enforcement model and an extensible vocabulary. To this day, the XACML privacy profile, however, does not provide all necessary elements required for the definition of privacy policies. Offering similar capabilities as XACML, the privacy policy language EPAL is primarily designed for the internal enforcement of privacy policies in enterprises. Further efforts towards the transformation of EPAL policies to external privacy policies would allow a more widespread application.

The privacy preference language APPEL facilitates the simple definition of rules that list acceptable and unacceptable content of P3P privacy policies. Major shortcomings of APPEL are its evaluation semantics and – like the underlying P3P specification – its limited expressiveness. Weaknesses of APPEL are targeted by XPref that utilizes XPath expressions for a more accurate definition of rule conditions. In an effort to overcome the deficiencies of syntax-oriented preference languages, Rei proposes a semantic approach for the specification of privacy preferences. While the integration of ontologies maximizes the potential expressiveness of the language, semantic technologies – including the generation of ontologies – are far from reaching the necessary level of maturity.

4.2 Existing Privacy Tools

Motivated by users who required technical means to protect their privacy, so called privacy-enhancing technologies (PETs) emerged [Bur97, GWB97], which aim for the limitation or prevention of personal data misuse [PK99]. PETs make personal data transactions more transparent and facilitate users to exercise their right of informational self-determination.

4.2.1 Categorization

As laid out in Chapter 3.1, differing recognitions and interpretations of privacy exist. Likewise, versatile PETs have been developed in recent years that focus on different aspects of privacy. The three cited privacy definitions of Chapter 3.1 allows for the classification of PETs into the following categories:

- **Category 1**
  "the ability of the individual to protect information about himself" [GWB97]

In the context of PETs, we interpret this definition as the protection of the user's identity and his/her physical data. From that perspective this definition applies to more technical, IT security-related PET approaches like cryptographical encryption and anonymization. The goal of tools belonging to this category is the modification of the user's identity in such a way as to guarantee the exclusive access of personal data by authorized recipients. For an overview of technical implementations in this area the interested reader is referred to [Fed05].

- **Category 2**
  "the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others" [Wes67]
  The definition of Alan Westin highlights the user's ability to decide about the amount, the recipients and the conditions of personal data disclosures. This interpretation is frequently referred to as "informed choice", which accurately summarizes the definition of Westin.
  The purpose of tools considering this privacy aspect is the presentation of all available information required to support the user in a disclosure decision.

- **Category 3**
  "the right to be let alone" [WB90]
  Originating from the year 1890, this definition at first glance looks too simple to be regarded in modern times. However, considering the aggressive marketing strategies of many companies – including unsolicited mail and telephone offers – the cited definition has not lost anything of its actuality.
  PET tools belonging to category 3 offer means that protect the user from these unwanted contacts.

Analogous to Chapter 3.1, the goals of this work target PETs that enable users to make an informed choice about the disclosure of personal data. For this reason, we focus on solutions that offer privacy protection as defined in category 2. In the remainder of this work, when mentioning PETs, we refer to solutions of this category.

Similar to the terminology used by Lederer et al. [LHDL04], we further divide PETs into groups that consider temporal activities of users. The basis of this refined categorization is the use of PETs before, during or after the disclosure of personal data.

- **Potential Information Flows**
  The first group of PETs assists users in protecting potential personal information flows [LHDL04]. These information flows contain all privacy-related activities a user executes before the transfer of personal data, such as the definition of disclosure rules.

- **Actual Information Flows**
  The second group focuses on actual personal information flows. These PETs cover solutions that support and protect users, when personal data are about to be disclosed. A

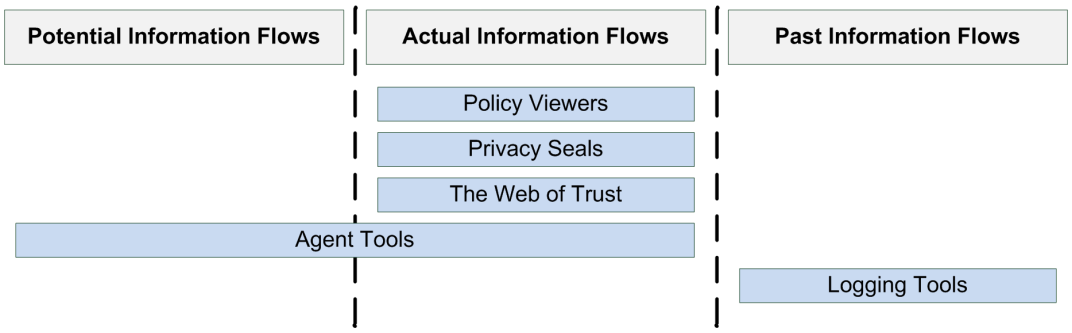


Figure 4.4: Categorized PET Solutions

typical PET tool of that group is a privacy agent that offers relevant information to the user and recommends a specific behavior.

- **Past Information Flows**
  Finally, we define a third group of PETs that enables users to manage past personal information flows. This group includes approaches that concentrate on the presentation and ex post modification of already disclosed personal data.

The defined information flows contribute to an approximate placement of PET solutions and serve as basis for a comprehensive PET tool set introduced in the second part of this work.

4.2.2 Available Solutions

This section evaluates available PET tools and technologies that offer users an informed choice about the disclosure of personal data. Figure 4.4 divides the discussed tools into the refined categories of the previous section. Policy viewers, privacy seals and the Web of Trust exclusively assist users in protecting actual information flows. Agent tools proactively capture disclosure rules, which are matched with privacy policies of visited service providers. Hence, these tools contribute to the protection of both potential and actual information flows. Finally, logging tools allow for the management of past information flows.

4.2.2.1 Policy Viewers

As mentioned earlier in this chapter, service providers publish privacy policies that communicate planned data handling practices with regard to collected personal user data. In addition to the requested data types, the privacy policy lists intended purposes and third parties personal data may be forwarded to. As a consequence, the published privacy policy of a service provider represents an essential information source for the user's decision-making process.

In Section 4.1.2 we introduced the P3P specification that offers service providers to express a machine-readable version of a textual privacy policy. This digital representation allows so

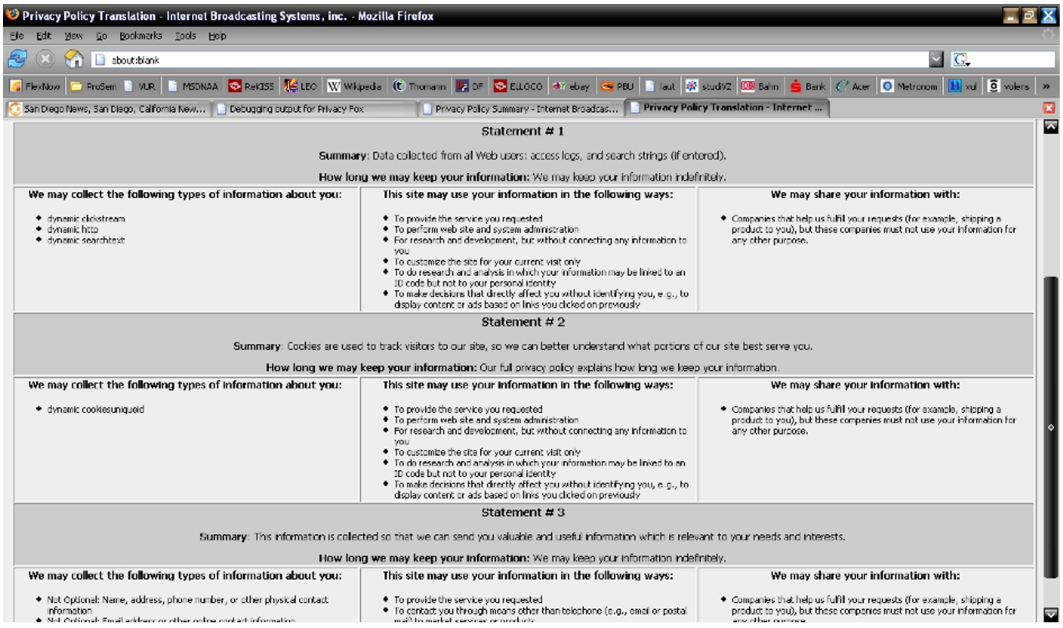


Figure 4.5: Policy Translation of the Privacy Fox

called policy viewers to read and display the content of P3P policies in a way that – compared to textual privacy policies – is more understandable to users.

The P3P Policy Viewer[1] provides a simple solution of such a tool. The project's Web site allows users to enter the URL of any P3P policy. Subsequently, the P3P Policy Finder reads and lists the content of the specified policy in a structured and human-readable way.

Unfortunately, the approach requires the user to manually check the existence of a P3P policy file. Furthermore, it seems unlikely that average users on a regular basis will copy the policy's exact physical location and visit the project's Web site. Hence, this solution will not contribute to the frequent use of PETs.

A policy viewer that provides a more advanced functionality and usability is offered by the Privacy Fox, an extension for the Mozilla Firefox browser [Ars04]. Like the P3P Policy Viewer, Privacy Fox aims at a user-friendly representation of P3P policies, including detailed but clear information about a service provider's privacy practices. In particular, Privacy Fox translates a P3P policy into natural language that is understandable for the average Internet user. Such a translation is generated by transforming P3P elements into standardized text blocks.

Integrated into the Web browser, the extension offers a "Translated P3P Policy" menu item that opens browser tabs with the generated policy translation and a tabular policy summary. Figure 4.5 presents a typical policy translation that separates the content of each detected policy statement of the processed privacy policy. Three categories divide the displayed information into requested personal data, the intended purpose and third party recipients. If no P3P policy is detected at the known locations of a Web site, the user is notified.

[1] http://www.webentrust.com/p3ptools/viewer/

Figure 4.6: TRUSTe Privacy Seal

Compared to the P3P Policy Viewer, the Privacy Fox provides a convenient solution for the graphical presentation of P3P policies. However, the user interface of the tool does not offer the required design to attract average Internet users. Furthermore, the textual translations do not fully explain the consequences to inexperienced users. Nevertheless, the Privacy Fox displays a privacy policy summary that potentially enables users to make a more informed disclosure decision.

4.2.2.2 Privacy Seals

Showing privacy policies in a more understandable way, policy viewers represent a first step towards more transparent privacy policies. While the discussed solutions accurately display elements of the processed privacy policy, they do not offer any evaluation or recommendation. Instead, inexperienced users need to find out for themselves, if the presented privacy practices are acceptable.

Addressing this fact, certain non-profit organizations and private companies inspect privacy practices of service providers. If the data handling practices meet the criteria of these third parties, a privacy seal is issued, which is placed on the service provider's Web site. If the user finds a privacy seal on a Web site and if the user is familiar with the criteria of the issuing institution, such a privacy seal signals a basic level of privacy protection to the user. In theory, this procedure gives inexperienced users a reliable orientation about applied privacy practices.

One of the frequently found privacy seals is TRUSTe[2], a non-profit privacy program that allows Internet users to gain trust in service providers [Ben99]. The issuance of a TRUSTe seal (see Figure 4.6) is bound to certain privacy practices of the service provider.

In particular, these mandatory practices include the publication of a complete privacy policy, the presence of opt-out mechanisms for third party disclosures, the implementation of procedures that protect personal data from unauthorized internal access as well as the user access to already disclosed information. TRUSTe continuously performs checks that control the persistent validity of the certified criteria.

While privacy seals were introduced to certify a minimum level of trust to the user, the sole presence of a privacy seal should not induce users to fully trust a Web site. The user needs to know the specific criteria of the certifying third party in order to weigh the significance of a privacy seal. In that context, Kobsa argues that several privacy seals in the United States are only

[2] http://www.truste.org/

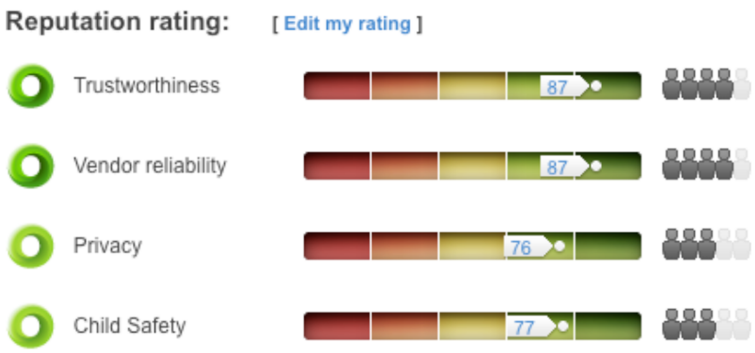


Figure 4.7: Available Web of Trust Reputation Ratings

bound to a service provider's affirmation to adhere to its published privacy policy [Kob07b]. The privacy practices themselves are not criteria of the privacy seal issuance. Likewise, Kolari et al. mention that the issuance of a privacy seal is highly coupled to the mere presence of a P3P privacy policy [KDK+04]. Hence, the existence of a privacy seal could mislead users and pretend a non-existent level of privacy.

Even though privacy seals perform a first evaluation of service providers' privacy practices, the above-cited criticisms prove that these certifications do not guarantee reliable privacy protection.

4.2.2.3 The Web of Trust

The concept of the prior discussed privacy seals was built on the assumption that users trust the judgment of a single third party. In contrast, the Firefox extension Web of Trust[3] builds on the evaluation transfer within user networks and enables users to exchange reputational information about Web sites.

The browser extension displays four rated categories of the visited Web site (see Figure 4.7). The category Trustworthiness reflects potential fraudulent activities. Vendor Reliability, on the other hand, focuses on general service experiences with a Web site and resembles ratings of known reviewing systems on eCommerce platforms. According to the Web of Trust manual, the Privacy rating answers the question whether a user can entrust a service provider with a personal e-mail address, while Child Safety signals inappropriate content.

The extension allows users to submit an own rating to the network by scrolling the horizontal slider on the respective bar.

The analysis of the tool reveals that the four evaluated categories are broadly defined, which complicates the differentiation for users. Furthermore, the Web of Trust does not seem to focus on privacy, as only one rating covers privacy practices. Unfortunately, this privacy rating does not measure certain privacy aspects, but captures a general, considerably subjective user impression.

[3] http://www.mywot.com/

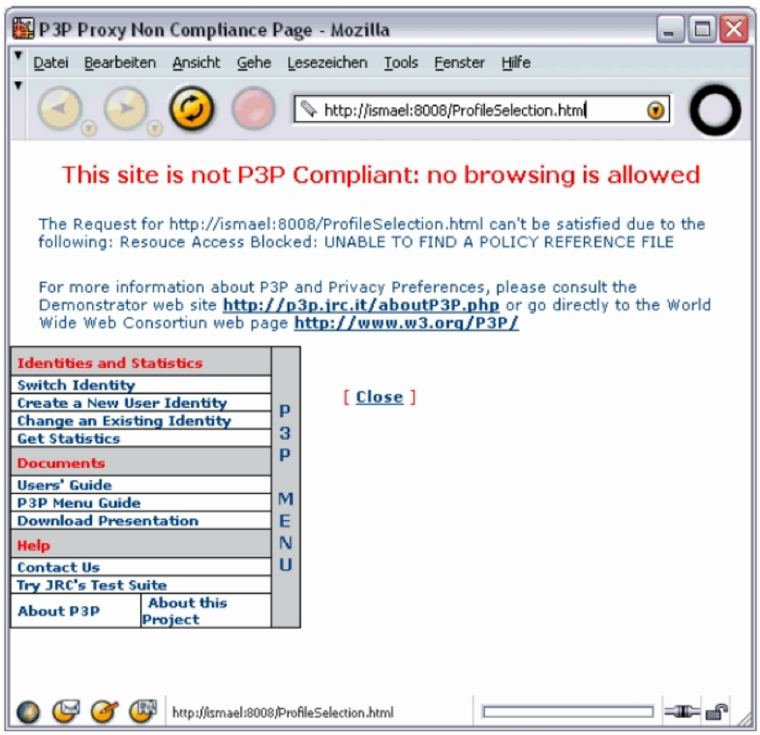


Figure 4.8: JRC Personal Proxy – User Notification

While the actual use of the Web of Trust ratings for a disclosure decision is hard to measure, the tool demonstrates the capabilities of exchanged reputational information. The second part of this work utilizes this powerful potential for the introduction of a collaborative privacy infrastructure.

4.2.2.4 Agent Tools

Privacy agent tools perform an automatic privacy policy evaluation, by dynamically calculating a recommended disclosure behavior based on individual user preferences.

An early approach of a privacy agent represents the W3C Privacy Evaluator, which sketches the design of a specific user agent [Nel98]. According to the W3C Note, this class of agents informs users about detected privacy policies and warns about non-existent or insufficient policies.

The corresponding prototypical implementation, the Privacy Jigsaw Proxy Server (PJPS), supervises all HTTP transactions and interrupts the interaction, whenever a P3P privacy policy does not provide essential information, such as intended purposes or recipients. Moreover, the tool scans HTTP form elements and warns the user, when a service provider collects identifiable information.

A more advanced solution of a privacy agent is provided by the JRC Personal Proxy[4], which is

[4] https://www.datenschutzzentrum.de/selbstdatenschutz/p3p/p3p_jrcp.htm

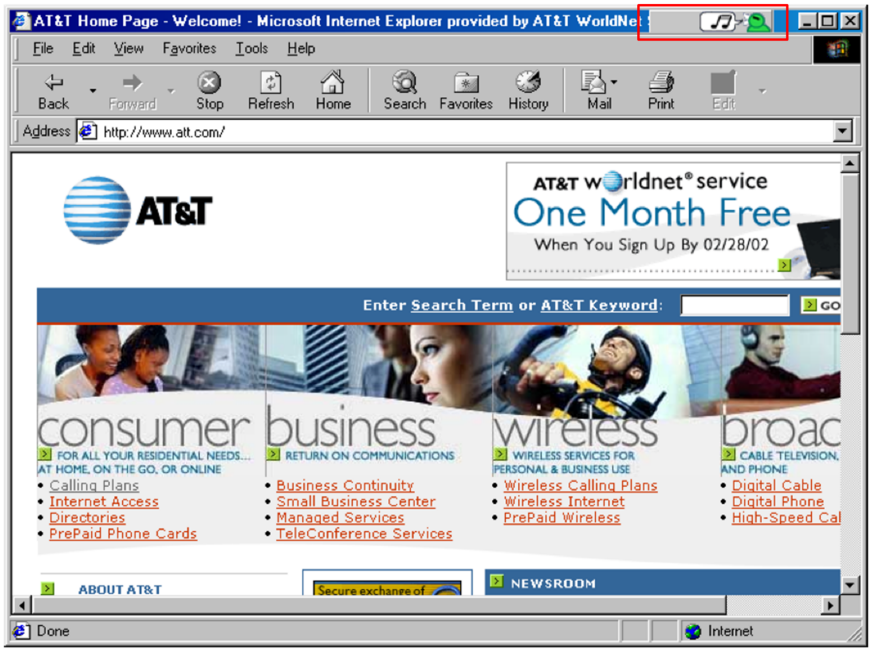


Figure 4.9: Privacy Bird – Embedded into the Microsoft Internet Explorer

regarded as the first solution that fully implements the P3P specification. Hence, the tool was frequently utilized to assess the capabilities of the P3P standard [HJW02].

Once the application is installed, the user is asked to choose from five APPEL rule sets with privacy levels ranging from very low to very high. Alternatively, the user can opt to import an externally generated APPEL rule set. The tool itself does not capture fine-grained privacy preferences.

The core module of the application is devoted to the matching of APPEL privacy preferences and P3P privacy policies. Additionally, a central component implements the proxy component, which actively filters content and notifies users about the results of the matching process.

After the start of the proxy, the user is asked to enter a URL. Subsequently, the tool searches for a P3P policy on the chosen Web site. If the policy matches the preference rule set of the user, the page is loaded and forwarded to the Web browser. If no P3P policy is detected or if the policy does not comply with the user preferences, the Web page is blocked and the user is notified (see Figure 4.8).

The JRC Personal Proxy represents a considerably radical solution of a privacy agent. If no P3P policy is provided, the user is required to deactivate the application in order to visit the entered Web site. As most service providers do not offer P3P policies, the chosen proxy approach does not seem suitable for the continuous use.

Targeting a practicable P3P agent and focusing on usability aspects, Lorrie Faith Cranor in cooperation with AT&T developed the Privacy Bird, which is implemented as browser plug-in for the Microsoft Internet Explorer 5.01 [CGA06].

Figure 4.10: Privacy Bird – Calculated States [CGA06]

According to the authors, the design of the Privacy Bird benefited from experiences made with four preceding prototypes. In 1997 an early W3C prototype for the first time demonstrated the concept of P3P and initially unveiled the challenges of building a suitable user interface. In the consequence the proxy-based Privacy Minder as well as the P3P Browser Helper Object followed. Out of these efforts, the AT&T Usability Testing Prototype was developed, which is regarded as direct predecessor of the Privacy Bird. Finally, the initial version of the Privacy Bird was introduced in 2002.

Following the concept of a P3P privacy agent, the Privacy Bird captures privacy preferences that serve as foundation for subsequent P3P policy evaluations [CGA06]. For the definition of privacy preferences the application offers pre-defined configuration options in favor of fine-tuned preferences. This design decision accounts for the complexity of APPEL rules, which stands in contradiction to the inexperience of average Internet users.

In the context of the introduction of a user-friendly tool that captures disclosure rules, Chapter 7.3 discusses and evaluates the configuration of privacy preferences in the Privacy Bird.

Figure 4.9 depicts the embedded Privacy Bird in the Microsoft Internet Explorer. The application scans visited Web pages for available P3P policies, which are matched with the pre-defined privacy preferences. A bird icon in the header of the browser reflects the results of the matching process. The designed icons utilize the metaphor of traffic lights (see Figure 4.10). If a detected P3P policy meets the requirements defined in the individual user preferences, the bird icon is colored green. A red exclamation point next to the green bird indicates that the visited Web site contains embedded content that does not provide an adequate privacy policy. If the Privacy Bird does not find a P3P policy on a visited Web site, the bird icon in the browser header turns yellow. A red bird icon signals a detected conflict, implying that the found P3P policy contains privacy practices that are specified as unacceptable in the privacy preferences. If the application is deactivated, the bird icon is colored gray. All calculated recommendations are supported by a twitter sound.

In addition to the policy matching, the Privacy Bird – similar to the presented policy viewers – offers a human-readable overview of all statements found in a detected P3P policy. The summary of the applied privacy practices is complemented by reasons for an adequate or inadequate privacy policy. Figure 4.11 shows a policy summary for a Web site that does not match the privacy preferences of the user.

Unlike the JRC Personal Proxy, the Privacy Bird does not interrupt the user's surfing behavior. Instead, the application provides decent but clear recommendations. A user-friendly interface

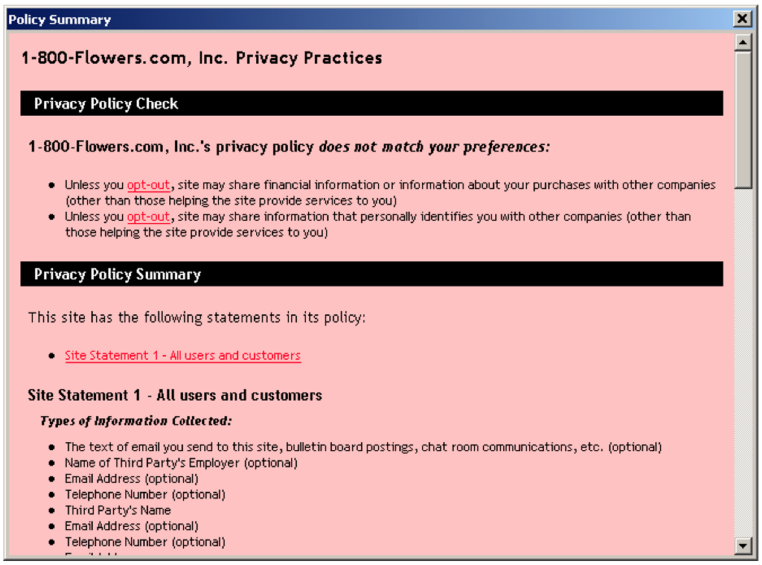


Figure 4.11: Privacy Bird – Policy Summary [CGA06]

addresses the needs of inexperienced Internet users. In Chapter 8.2 we present a usability evaluation of the Privacy Bird, which serves as foundation for the design of our proposed privacy agent tool.

4.2.2.5 Logging Tools

While the presented tools concentrate on potential and actual data disclosures, logging tools aim for an ex post evaluation of already transferred personal data.

A prototypical application of this kind was developed within the PRIMA project[5] of the Technical University of Darmstadt. Specifically, the PRIMA Datamanager offers an HTTP proxy that records personal data disclosures and lists an overview of conditions and details of past data transactions [BSTW03]. Unfortunately, the application does not provide suitable user interfaces that take advantage of the logged information.

The European project PRIME[6] (Privacy and Identity Management for Europe) implements the concept of a powerful and user-friendly logging tool.

The PRIME project introduces a privacy-enhancing identity management system that contains a privacy architecture with several design guidelines, protocols and prototypical scenarios [LSH08]. A central element of the architecture marks the PRIME Middleware, which integrates all PRIME components and coordinates the interaction between users and service providers. This complex middleware forms the basis for advanced privacy mechanisms and enables users to accurately control the disclosure and the usage of personal data.

[5] http://www.ito.tu-darmstadt.de/projects/prima/datamanager/index_en_html
[6] https://www.prime-project.eu/

Among other capabilities, PRIME facilitates users to record and access past personal data disclosures. This functionality is provided by the PRIME Data Track [PFHB06].

Pettersson et al. list a variety of Data Track functions that target the protection of personal user data. Representing a significant part of the application, the logging function records various details of a data disclosure. In addition to disclosed data types and the disclosure date, Data Track logs the purpose and the conditions under which personal data are disclosed to service providers. The logging function benefits from Data Track's integration into a comprehensive privacy infrastructure that features the user-transparent negotiation of privacy policies and their enforcement on the provider side [HFHPB07].

In addition to the mere recording functionality, Data Track enables users to actively manage past personal data disclosures. If requested by the user, Data Track presents a clear overview of past data transactions. Moreover, Data Track offers an ex post modification and removal of these personal data disclosures [PFHB06].

Chapter 9.2.1 presents an evaluation of the Data Track user interface (see Figure 9.1) within the scope of our proposed visualization tool.

PRIME's Data Track provides a considerably powerful client tool that facilitates the logging and the management of past data disclosures. However, Data Track is built on the PRIME architecture that needs to be installed both on the client side and the provider side.

4.2.3 Summary

Considering differing recognitions and interpretations of privacy, this section starts with a categorization of PETs. Relating to the goals of this work, we continue with a presentation of available PETs that assist users in protecting their personal data.

We start with an analysis of policy viewers that present machine-readable privacy policies in a clearly arranged form. Especially Privacy Fox offers a suitable solution that contributes to the understandability of privacy policies. The main weakness of policy viewers, however, is the lacking evaluation mechanisms inexperienced users rely on.

A privacy seal, on the other hand, signals the positive inspection of a service provider's privacy practices. This certification is issued by third parties who check privacy practices based on specific criteria. If the user is aware of the evaluation process, privacy seals provide a quick orientation to the user, when an unknown Web site is visited. However, issuance methods of privacy seals are subject to wide criticism.

In contrast, the Web of Trust utilizes a user network to collect service provider ratings. Unfortunately, the ratings do not focus on privacy practices and are not accurately defined. Compared to privacy seals, the experiences of numerous participants provide a more reliable direction to users.

Offering advanced evaluation capabilities, privacy agents capture individual privacy preferences of the user and perform a matching with machine-readable privacy policies found on visited Web sites. Signaling the matching result as a bird icon in the browser, the Privacy Bird offers the

most user-friendly interface of all available privacy agents. Like policy viewers, the discussed privacy agents depend on the presence of P3P policies, which are maintained by only a small fraction of service providers.

Finally, we analyzed existing logging tools that allow the management of past personal data disclosures. In this group, Data Track offers a clear visualization and the retroactive modification of already disclosed personal data. Data Track, however, requires the installation of the complex PRIME middleware.

4.3 Usability in Privacy-enhancing Technologies

The previous sections demonstrate that most PET solutions bear the potential to significantly improve the privacy of users. However, even though the concepts of the presented approaches have been developed several years ago, PETs are barely used in the World Wide Web.

Apart from the lagging willingness of service providers to adopt privacy mechanisms, a major reason for the resistance of average Internet users to accept PETs is the inferior role of privacy [WT99]. The primary user goal represents, for instance, the purchasing process at an eCommerce Web site or the opening of a Web mail account, while the protection of privacy is regarded as secondary goal. Whenever tools obstruct users in achieving their primary goals, users avoid such technologies [COB06]. Hence, usability in PETs plays a decisive role for the acceptance of these solutions. Unfortunately, most PET implementations fail to meet these requirements of the user interface.

ISO 9241-11:1998[7] provides a general definition of usability. According to the standard usability refers to the

> ”Extend to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”.

In the following, we briefly summarize research contributions in the area of security and privacy.

Titling ”Why Johnny Can’t Encrypt”, a relatively early work of Whitten and Tygar focuses on usability in the context of a data encryption application and represents a frequently cited example of usability research in the security field [WT99]. The authors emphasize that compared to conventional applications different criteria exist for usability in security software, hinting at the distinctive characteristics of security tools. Specifically, in order to prevent dangerous errors, security-related applications should offer understandable feedback. However, due to the complexity of security configurations, the design of clear feedback is considerably challenging. The authors suggest that the use of intuitive interfaces, visual metaphors and consistency contributes to understandability and helps reduce the perceived complexity of

[7] Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability (ISO 9241-11:1998)

security mechanisms. As operating errors lead to severe consequences, security applications should capture irreversible actions.

Within the scope of TOR anonymization networks, Clark et al. recommend a set of usability guidelines [CvOA07]. According to the authors, users should always be aware of the steps required to complete a core task. In addition to the necessity of a clear interface, the authors stress the importance of understandable terminology.

Focusing on privacy-related tools, Patrick and Kenny introduce the Privacy Interface Analysis [PK03], which defines Human Computer Interaction (HCI) requirements and solutions derived from the EU Directive 95/46/EC. For instance, the authors transform the privacy principle Transparency into an HCI requirement stating that ”users must be aware of the transparency options, and feel empowered to comprehend and control how their Personally Identifiable Information is handled”. As solution for this exemplary privacy principle, the authors recommend the explanation of transparency information as well as the provision of tutorials.

In contrast to numerous PET implementations, usability research of PETs is limited to a small number of contributions. Addressing this unbalance, the remainder of this work concentrates on intuitive and understandable PET solutions that assist users before, during, and after personal data disclosures. Integrated into a collaborative privacy infrastructure, the proposed solutions are introduced in the following part.

Part II

Proposed Privacy Management Infrastructure

Chapter 5

User-centric Privacy Architecture

In the first part of this work we outlined fundamentals of personal data disclosures in the World Wide Web along with the involved privacy implications. Moreover, we evaluated existing privacy-enhancing technologies and stressed the lack of usability in today's solutions.

Considering these findings this work proposes a user-centric privacy architecture that features a collaborative privacy community and three usable privacy management components on user-side. The presence of a privacy community creates an open, provider-independent information source for the accessing local components. This part starts with the introduction of our proposed architecture, followed by a discussion and presentation of suitable solutions for each architectural component.

In particular, this chapter presents a conceptual overview of our modeled privacy architecture. We introduce the role and the main purpose of each component within this architecture. As the design of all presented solutions highlights usability, we sketch the conditions of the user studies we conducted over the course of the component development.

5.1 Overview

Most of today's privacy-enhancing technologies that claim a comprehensive protection of user data significantly rely on the support of service providers (see Chapter 4.2). Such an involvement could, for instance, require the provision of a machine-readable privacy policy or the integration of external privacy components on the provider side.

This dependency, however, is responsible for the low practical applicability of many promising solutions. Even though threats of personal data misuse are growing, the example of the P3P specification shows that service providers do not contribute to the widespread availability of accurate P3P policies voluntarily. Likewise, from today's perspective it seems unlikely that service providers will fundamentally change their internal back-end infrastructures, as required for realizing the ideas of the PRIME project.

In addition, we identified usability deficits as a major factor for the lagging acceptance of

privacy-enhancing technologies (see Chapter 4.3). While many solutions offer advanced functionality for the protection of personal data, they are not designed for the majority of inexperienced Internet users and lack understandable user interfaces and explanations.

Acknowledging the conflicting interests of service providers as well as the need for usable tools, we introduce a user-centric, service provider-independent privacy architecture [KKP09], which is depicted in Figure 5.1. A collaborative privacy community facilitates Internet users to share privacy-related information about service providers. Three privacy components on the user side offer user-friendly tools that assist users in controlling potential, actual and past information flows (see Chapter 4.2.1), utilizing service provider information of the privacy community.

Unlike provider-dependent privacy technologies our proposed privacy architecture does not require the direct support of service providers. We rather accept today's service landscape of the World Wide Web and offer users a more applicable privacy infrastructure.

5.2 Components

In the following, we briefly specify the main purpose and functions of each architectural component. In the remainder of Part II the privacy community and the three local privacy tools are discussed in detail.

Privacy Community

The privacy community marks the central element of our proposed privacy architecture and enables users to collaboratively exchange privacy-relevant information, ratings and experiences about service providers. This information includes, for instance, the required amount of personal data for the fulfillment of a service and third parties the provider shares personal user data with. A collaborative privacy community represents a valuable data source for all three local privacy components and facilitates provider-independent privacy protection. For the collaborative editing of collected data, our solution provides a Web front-end that maintains privacy-related information about each service provider.

Chapter 6 introduces the content and functions as well as the design of our solution of a collaborative privacy community.

Privacy Preference Generator

Enabling users to control potential personal data flows, the proposed privacy architecture provides a Privacy Preference Generator component that captures individual privacy preferences. Privacy preferences define individual conditions of personal data disclosures. As the Privacy Agent component matches user's privacy preferences with a service provider's privacy policy, the resulting recommendation highly depends on the accuracy of individual preferences.

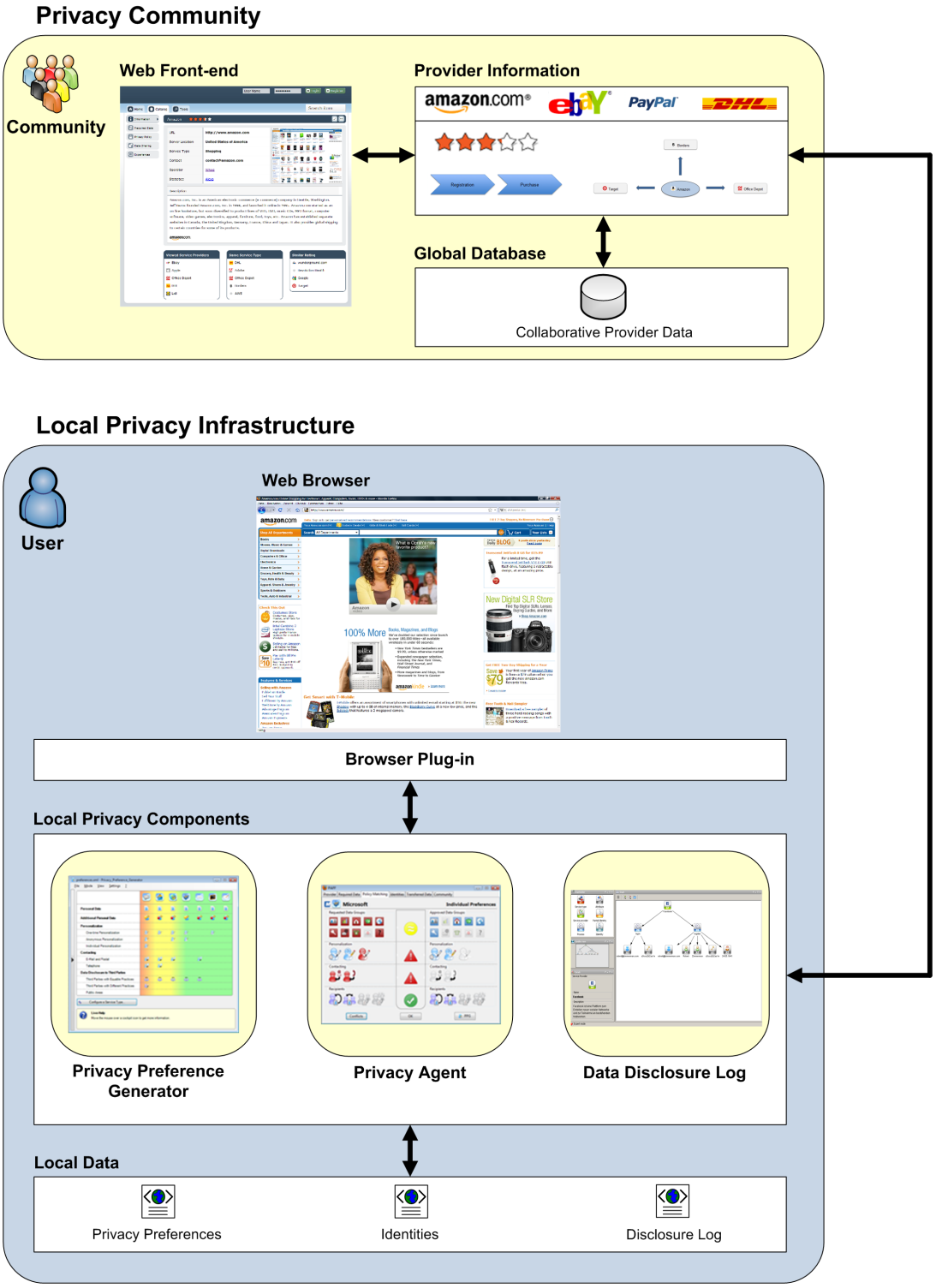


Figure 5.1: User-centric Privacy Architecture

Catering the needs of predominantly inexperienced users, our solution offers a tool that allows users to define privacy preferences in an understandable way. We facilitate the individual definition of privacy preferences for pre-defined Internet service types, guaranteeing more realistic and practical results. In addition to a user-friendly configuration wizard, our solution provides a clear configuration summary and evaluation.

In Chapter 7 we present the main challenges of a suitable Privacy Preference Generator component, before we introduce our proposed solution.

Privacy Agent

The Privacy Agent component assists users in making an informed decision about the actual disclosure of personal data. With regard to the visited Web site, this support involves the presentation of relevant information from the privacy community and the evaluation of a published privacy policy. Returning users benefit from information about linkable partial identities and already disclosed personal data.

In particular, our developed Web browser extension automatically looks-up and displays reputational community information for the visited Web site. Furthermore, user's privacy preferences, which are generated by the Privacy Preference Generator component, are matched with machine-readable privacy policies, out of which a disclosure recommendation is calculated and signaled to the user. Moreover, the use of partial identities and their potential linkabilities are examined. Finally, a summary of personal data that have been disclosed to the visited service provider is offered.

Chapter 8 thoroughly discusses requirements and usability deficits of existing privacy agents and introduces our proposed solution for the Privacy Agent component.

Data Disclosure Log

The Data Disclosure Log component records personal data transfers and provides a clear overview of past personal data flows. Such an overview enables users to know the recipients of past personal data transactions at any time. This knowledge represents a prerequisite for an ex post revocation of personal data. The Data Disclosure Log component requires both a tracking tool that monitors personal data disclosures in the Web browser and usable interfaces that illustrate logged data transactions in a comprehensible way.

Implementing the required functionality, our solution provides a browser plug-in that detects and stores data transactions in the Web browser. A corresponding application visualizes these logged disclosures and relates them to used partial identities. The visualization tool offers multiple views, including a dynamically generated graph.

After an evaluation of requirements and related work, Chapter 9 introduces our proposed solutions of a logging tool and a visualization tool.

5.3 Conducted User Studies

In the following chapters we elaborate usable solutions that provide intuitive user interfaces and understandable explanations. In order to meet the needs of heterogeneous user groups, we conducted four user surveys, which served as valuable input for the development as well as the evaluation of the proposed components.

This section describes the general conditions of each survey. The outcomes are described within the requirements and evaluation sections of the following chapters.

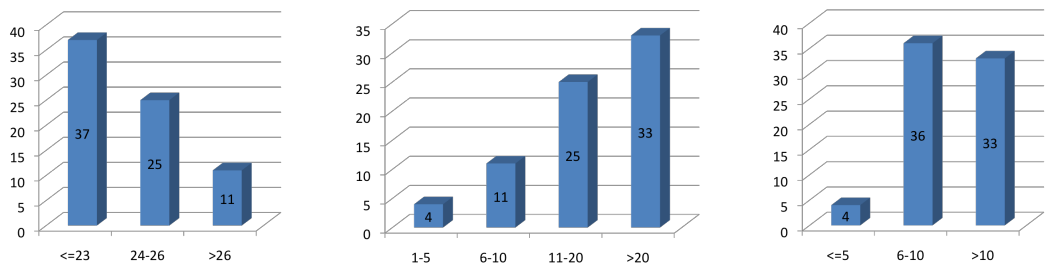
5.3.1 Questionnaire Survey: Privacy Attitude and Privacy Behavior

At the beginning of our efforts we conducted a user study that captured the general privacy knowledge and the privacy behavior of Internet users.

We created a questionnaire, which was divided into three parts. The first part was devoted to users' awareness of privacy threats in the World Wide Web. Specifically, this part focused on the perceived personal data requests of service providers, the sensitivity of certain personal data types and the knowledge about privacy rights. The second part covered the data disclosure behavior and the importance of a privacy policy. Finally, the last part analyzed the knowledge about existing privacy-enhancing technologies and their deficits.

The questionnaire was handed to 73 information systems students at the University of Regensburg. The portion of undergraduate and graduate students was nearly equal. 87% of the test persons were male. The low fraction of female participants reflects the proportion within the information systems programs. Also related to their field of study, the vast majority of test persons showed frequent Internet usage as well as a long Internet experience. Figure 5.2 illustrates the distribution of returned questionnaires with regard to age, Internet usage and Internet experience.

Even though the test sample exclusively consisted of information systems students, the results of the questionnaire study revealed the low awareness of privacy and the poor knowledge of privacy-enhancing technologies. These findings evolved into the primary design constraints of the initial versions of the privacy components.



(a) Age (b) Internet Usage (hours/week) (c) Internet Experience (years)

Figure 5.2: Questionnaire Survey – Distribution of Test Persons

5.3.2 Online Survey: Disclosure Behavior and Records

Within the design process of a suitable logging tool of the Data Disclosure Log component we deployed an online survey that captured opinions and attitudes of users towards such an application. In order to project the accruing amount of personal data transactions, we also analyzed the browsing and data transaction behavior of users. An additional goal of this survey was the identification of an adequate host browser for the targeted logging application. For the implementation and evaluation we used the online survey tool LimeSurvey[1].

The online questionnaire was filled by 351 participants, which allowed us to reliably identify the required environment of a suitable tool as well as the dimensions of a resulting log file.

5.3.3 User Test 1: Initial Privacy Components

Our first user test evaluated the existing Privacy Bird (see Chapter 4.2.2.4), a draft version of the Privacy Preference Generator component (see Chapter 7.4) and a prototypical implementation of a 3D graph-based Data Disclosure Log (see Chapter 9.5.1).

For the user experiment we recruited twenty test persons, acknowledging frequent recommendations that a single-digit sample is insufficient [COB06, Fau03, PL01, SS01]. In addition, we arranged a rehearsal test, which allowed the improvement and adjustment of the test procedure. Aiming at a heterogeneous test sample, the invited test persons showed a diverse academic and professional background. However, basic knowledge of Microsoft Windows as well as the occasional use of the World Wide Web were prerequisites for participating candidates. In order to avoid biased results, persons with close relationships to the interviewers were not considered.

Seeking to put test outcomes into the right perspective, we grouped participants with regard to their age and their privacy attitude. For the attribute age we defined the groups teens (0 - 19 years), no-teens (20 - 49 years) and elderly (50+ years). Utilizing studies of Alan Westin (see Chapter 3.5.2), we classified each test person as fundamentalist, pragmatist or unconcerned. The result is depicted in Table 5.1.

$[TestPersons]$	Teens	No Teens	Elderly	
Fundamentalists	2	2	1	5
Pragmatists	5	6	1	12
Unconcerned	2	0	1	3
	9	8	3	20

Table 5.1: User Test 1 – Distribution of User Groups

At the beginning, we provided test persons with short tutorials of each tested component. Subsequently, the test persons were confronted with tasks that involved the tested privacy tools. Table 5.2 shows an average time schedule of a conducted test. Avoiding sequence effects and

[1] http://www.limesurvey.org/

interaction effects, all six sequential permutations of the three privacy tools were alternately applied. We encouraged test persons to think aloud, in order to gain as many immediate impressions and as much feedback as possible. If necessary, the interviewer initiated ad-hoc interviews at critical moments. After the completion of a task, additional interview questions followed, which covered the understandability and the perceived user-friendliness of the tested tool.

$[minutes]$	Tutorial	Tasks	Interview	Σ
Privacy Bird	3	8	4	15
Privacy Preference Generator	1	8	4	13
Data Disclosure Log	3	7	4	14
Final Interview			3	3
Σ	7	23	15	45

Table 5.2: User Test 1 – Test Outline

All tests were conducted on a pre-configured machine. Only one person was tested at a time. The test was moderated by interviewers familiar with privacy and usability. Facilitating a high-quality examination, we recorded all on-screen activities along with a voice stream, using Camtasia Studio[2].

The outcome of the first user test allowed us to evaluate initial versions of the local privacy tools. In particular, the test results helped us make significant improvements of the Privacy Preference Generator component and the Data Disclosure Log component.

5.3.4 User Test 2: Final Privacy Components

Finally, we conducted a second user test that evaluated advanced versions of the privacy community, the Privacy Preference Generator component, the Privacy Agent component and the Data Disclosure Log component, including the 2D graph-based visualization tool as well as the underlying logging tool.

We scheduled individual tests with 26 participants. As we sought an unbiased feedback, the criteria for the selection of test persons were equal to our first user test (see above). In an effort to compensate and motivate participants of this test, we rewarded each test candidate with a 5 Euro coupon redeemable at a local consumer electronic store. The coupon was handed out before the start of each test.

The test sample consisted of 17 university students, while nine test persons were graduated professionals. Hence, 15 out of the 26 test persons were 25 years old or younger, seven between 26 and 30, and four between 30 and 45. 22 of all test persons were male. Out of the 17 students nine were enrolled in a technical program and five in a business program. From the remaining students two were pursuing a teaching degree and one a diploma in mathematics.

On average each test cycle lasted 75 minutes. A short tutorial opened the test of each privacy

[2]http://www.techsmith.com/camtasia.asp

$[minutes]$	Tutorial	Tasks	Interview	Σ
Introduction	2			2
Privacy Community	3	5	3	11
Privacy Preference Generator	3	14	2	19
Privacy Agent	2	12	6	20
Data Disclosure Log	5	13	5	23
Σ	15	44	16	75

Table 5.3: User Test 2 – Test Outline

tool, followed by several tasks and a closing interview section. The test outline is illustrated in Table 5.3. Again, we conducted two rehearsal tests and used a preconfigured notebook with on-screen and voice recording.

The second user test served as final evaluation of the privacy community and the local privacy components. The test results allowed us to make final adjustments to the designed user interfaces and textual explanations.

5.4 Summary

Considering the findings of the first part of this work, this chapter identifies the dependence on service providers as well as usability deficits as main obstacles of today's solutions aiming for the protection of personal user data.

Addressing these shortcomings, we introduce a collaborative privacy architecture that provides an open privacy community for the provider-independent maintenance of service provider information. Moreover, three local privacy components support users in controlling potential, actual and past personal data disclosures with a special emphasis on usability and user acceptance. Finally, we present the conditions of four conducted user studies that played an essential role for the design and the development of the proposed solutions.

Chapter 6

Privacy Community

One of the main features of our proposed privacy architecture is service provider independence. This goal is achieved by setting up an open privacy community, which maintains service provider information in a collaborative fashion.

After a summary of related work, this chapter highlights the privacy community's allotted role in our privacy architecture. Subsequently, the content and the functionality of the targeted privacy community are discussed, before the design of a Web service-based solution is introduced. Finally, the chapter sketches an agenda for the launch and the operation of a privacy community.

6.1 Related Work

Aiming at protecting Internet users from fraud, SPAM, viruses and spyware, the Web of Trust provides a global database that allows users to exchange their impressions about Web sites (see Chapter 4.2.2.3).

Compared to the proposed privacy community, Web of Trust does not focus on the protection of personal user data. Instead, the tool offers the exchange of four general ratings, including a broadly specified privacy category. No fine-grained information about data handling practices is captured. Conclusively, Web of Trust does not have the scope to cater the needs of an open, provider-independent privacy architecture.

Other instruments designed to offer users reputational information include privacy seals, which certify privacy methods of service providers (see Chapter 4.2.2.2). Unlike the proposed privacy community, privacy seals do not follow a collaborative approach. In contrast, a certifying third party attests that a service provider's privacy practices are in-line with the privacy standards of the issuing constitution. As the exact standards are unknown to most users, the presence of a privacy seal frequently misleads users in their judgment of a Web site's trustworthiness.

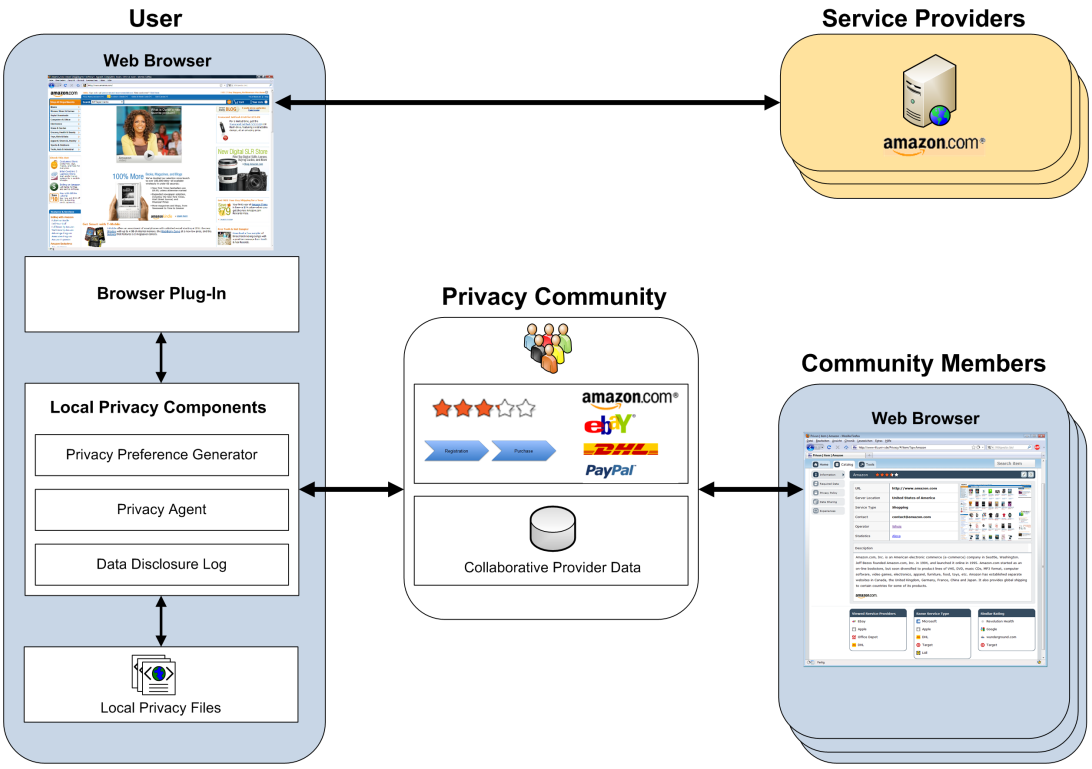


Figure 6.1: Collaborative Privacy Community

6.2 Outline

The goal of the introduced privacy community is the provision of a platform that gives Internet users the opportunity to share various privacy-related information about service providers, such as experiences, evaluations and data handling practices [KKP09]. Serving as an open information source in the proposed privacy architecture, the privacy community is utilized by local privacy components, which are presented in Chapters 7, 8 and 9. The integration of collaborative, provider-independent information significantly increases the potential as well as the applicability of these components that – in combination – build a powerful, user-centric privacy infrastructure with no footprint on the service provider side.

Figure 6.1 depicts the privacy community in the center of our proposed privacy architecture. All service provider information is maintained and organized in a Wikipedia-like [LC01] Web front-end. For each service provider privacy-related information is grouped into an article, allowing all Internet users (Community Members) to view and edit articles in the Web browser. In addition, open Web service interfaces allow the flexible integration of privacy-related service provider information into the local privacy components.

6.3 Content and Functions

This section defines the provided content and functions of our proposed privacy community [KKP09]. The selection was primarily driven by the needs of the local privacy components, but also by the potential of a provider-independent privacy environment.

6.3.1 Static Provider Information

When an unknown Web site is visited, users generally have the option to trust a service provider at face value or to look for information about its reputation and data handling practices. A survey [FML+01] 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 seems understandable. Addressing this fact, the privacy community provides users with an overview of static service provider data.

This general information is primarily utilized by the local Privacy Agent component and is displayed on demand (see Chapter 8.3.2), enabling users to easily get a first impression of a service provider.

In particular, the privacy community offers the following static service provider information:

- the service provider's URL
- the physical location of the server
- the offered service type
- information about the revocation of already transferred personal data
- contact information
- a short textual description of the service provider

A URL is required to uniquely identify service providers. The server location clarifies legal matters, as different privacy laws apply in different countries. The service provider's service type is queried by the Privacy Agent component. This capability facilitates the individual generation of privacy preferences for each service type (see Chapter 7.4) and their application during a policy matching (see Chapter 8.3.3). Aggregating service providers for visualization purposes, the Data Disclosure Log component also benefits from an automatic service type look-up (see Chapter 9.4).

In addition, the privacy community offers information (e.g. a link or an e-mail address) about the removal of disclosed personal data that have been transferred to a service provider (see Chapter 8.3.2), which helps users exercise their rights to control already transferred data (see

Chapter 3.4). Exact contact information facilitates prosecution, if personal information is misused, or if users want to enforce their rights to revoke their personal data. Furthermore, a short textual description specifies the main characteristics of a service provider.

6.3.2 Required Amount of Personal Data

In addition to static provider information the privacy community enables users to know in advance what personal data are requested by a certain service in the World Wide Web.

Users generally understand the necessity to disclose, for example, 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 these excessive data. The online survey we conducted (see Chapter 5.3.2) revealed that 77% of all test persons cancel registration and purchasing processes, if too much personal information is requested. Unfortunately, with today's technical means users are unable to determine in advance what personal information is necessary to use a specific service. In an effort to find out, users have to start the process of filling a set of Web forms. In many cases the most privacy-sensitive information is requested on the last form page. If the user decides not to proceed, he/she wasted valuable time and disclosed the already transferred information with no use.

The introduced privacy community spares users from this negative experience and enables them to exchange the amount and kind of personal data required for each process a service provider offers. In this context, a process refers to each separate service offer, such as *Purchase* or *Newsletter Subscription*. In addition, the community stores the reliance of a process on the completion of a different one. The process *Purchase* could, for instance, require the completion of the process *Registration*. The privacy community also performs an automatic evaluation of the required personal data with regard to the offered process and service type, which further assists users in estimating personal data requests of service providers. The amount of required personal data as well as the results of the automatic evaluation is employed and displayed to the user by the local Privacy Agent component (see Chapter 8.3.2).

The amount of required personal data represents a fundamental element of privacy policies. Its online availability in a privacy community facilitates the local Privacy Agent component to retrieve this information and match it with individual privacy preferences, if no sufficient machine-readable privacy policy is offered by a service provider (see Chapter 8.3.3).

6.3.3 Third Party Recipients

The release of personal user data to third parties represents a considerably privacy-sensitive factor in the context of service providers' data handling practices.

While the P3P specification only defines third party categories, the proposed privacy community allows the exchange of individual third parties the service provider shares personal data with. These parties can include affiliated companies and other business partnerships. The presence

of those corporate networks serves as valuable input for the Data Disclosure Log component that visualizes potential threats caused by linkable partial identities transferred to cooperating service providers (see Chapter 9.4.3.3).

6.3.4 Collecting and Explaining Privacy Policies

In general, a service provider's textual privacy policy is the only available information source about its data handling practices. Studies as well as a conducted questionnaire survey (Chapter 5.3.1) show, however, that privacy policies are not regarded as understandable and are read by only a small fraction of Internet users [JPJ05, Pol07].

Addressing the needs of the majority of Internet users, the privacy community allows experienced users to write and share an understandable explanation of a provider's privacy policy. As privacy experts comprehend all aspects of a policy, they have the ability to paraphrase important elements in a form that – compared to a published policy as well as automatic privacy policy summaries (see Chapter 4.2.2.1) – is easy to understand.

Furthermore, as privacy policies change over time, the privacy community maintains a history of privacy policies, containing both textual policies as well as machine-readable P3P policies. Such a policy history enables users to determine the policy that has been valid, when personal data have been disclosed. The privacy community also allows users to rate current and past privacy policies of service providers with regard to the stated data handling practices.

6.3.5 Adherence to Privacy Policies

As the presence of a privacy-friendly privacy policy is no guarantee that a service provider follows that expressed policy, the presented privacy community enables users to rate the policy adherence of service providers. Based on their individual experiences users evaluate, whether or not a service provider processes personal data as stated in the privacy policy. For example, if not expressed in the privacy policy, a received e-mail that promotes a product would justify a negative policy adherence rating of that service provider. Displayed by the local Privacy Agent component (see Chapter 8.3.2), this information is of considerable importance for a disclosure decision of the user.

6.3.6 Individual Experiences

Finally, the offered service provider information is complemented by individual user experiences. These open postings can contain any privacy-related positive or negative experiences and are not related to a specific aspect of the provider's data handling practices.

Integrated into the Privacy Agent component, the individual experiences are utilized for the presentation of reputational information about a service provider (see Chapter 8.3.2).

6.3.7 Sharing Privacy Preferences with Community Members

The Privacy Preference Generator component captures individual privacy preferences, which are in turn processed by the Privacy Agent component to calculate disclosure recommendations (see Chapters 7 and 8.3.3). The quality of these recommendations strongly relies on the accuracy of privacy preferences. Even though the Privacy Preference Generator component alleviates this challenge by offering a user-friendly and understandable user interface, building accurate privacy preferences is a critical task, especially for inexperienced users.

For this reason, the privacy community facilitates the exchange of privacy preferences among users. Using an integrated social networking component [BE07], users have the option to upload privacy preferences and share them with selected members.

Imported privacy preferences of a trusted privacy expert or organization represent valuable assistance for inexperienced users, resulting in improved disclosure recommendations of the local Privacy Agent component.

6.4 Design

Considering the privacy community's content and functions as well as its position within the privacy architecture, this section introduces design aspects of our proposed solution. We start with a description of the community's service architecture and user management, before the graphical user interface of the community is presented.

6.4.1 Service Architecture

In Section 6.2 we demonstrate that both the Web front-end and the local privacy components on the user side simultaneously access the community. The integration of heterogeneous client applications requires the specification of standardized interfaces, which is ideally realized by a Service-oriented Architecture (SOA) [MLM+06]. Implementing the concept of a SOA, the privacy community encapsulates the offered information pieces and actions into fine-grained Web services. Each Web service provides a machine-readable WSDL [CMRW07, CHL+07] service definition, which clearly specifies its interface. The communication of the privacy community with its clients via SOAP messages [GHM+07a, GHM+07b] guarantees a consistent data exchange format. Figure 6.2 shows the privacy community's interaction with the remaining components of the privacy architecture.

For the community's Web front-end we utilize an Ajax [Gar05] Web architecture, allowing asynchronous, interactive communications between the Web front-end and the community server. The Ajax engine transforms JavaScript [Fla06] requests of the user into SOAP requests, which are forwarded to the community back-end on the server side. The Web service server receives and processes requests querying the provider database, before requested data are sent back to the client via SOAP. The Ajax engine of the Web front-end transforms these SOAP

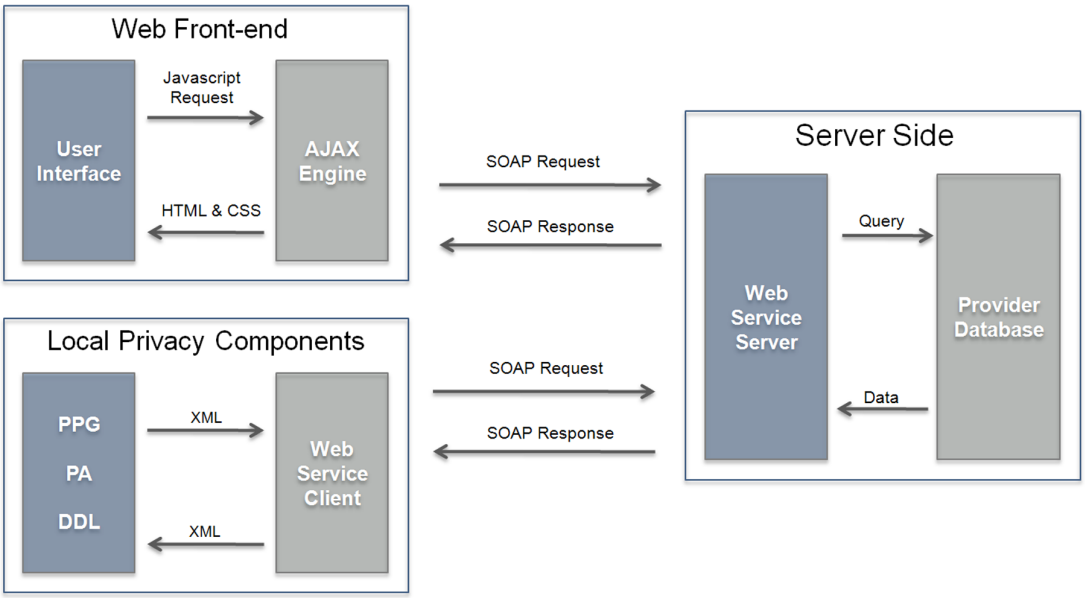


Figure 6.2: Interaction of the Privacy Community

messages to a user-friendly GUI using HTML [RLHJ99] and CSS [BCHL09]. Chapter 10.3.1 provides implementational details about the Web front-end and the Web service infrastructure.

The local privacy components – the Privacy Preference Generator (PPG), the Privacy Agent (PA) and the Data Disclosure Log (DDL) components – directly access the Web service server via SOAP messages.

6.4.2 User Management

The internal user management of the privacy community administers three user roles. Offering an open information source, the basic user role is assigned to every unregistered user and allows access to all available information about service providers. Furthermore, it permits users to edit articles collaboratively and to create new service provider articles. In order to prevent vandalism, the privacy community provides adequate backup and versioning functionality.

If users want to directly exchange information with connected members, a simple registration is necessary. Registration only requires a username and a password. The community does not request any additional personal user information. Unlike basic users, registered users have the option to upload and share generated privacy preferences with connected members. Likewise, privacy preferences of connected members can be downloaded and imported into the Privacy Preference Generator. We point out that the involved 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, helping inexperienced users to estimate the quality of advises and preferences.

Finally, users holding the administrator role define vocabularies of personal data types and

service types (see Chapter 10.2). In addition, administrators specify available processes as well as the appropriate amount of personal data for each process (see Section 6.3.2). If necessary, administrators are able to block users.

6.4.3 Graphical User Interface

The designed Web front-end aims for a clear layout and a high degree of user-friendliness. With regard to the assigned role users are able to look up, view and edit service provider articles, register, login and share privacy preferences with selected members, and administer users and structural data of the community. For the sake of brevity, we focus on the service provider catalog and the layout of a service provider article. The interested reader is referred to the prototypical implementation of the privacy community, which is accessible at the following link[1].

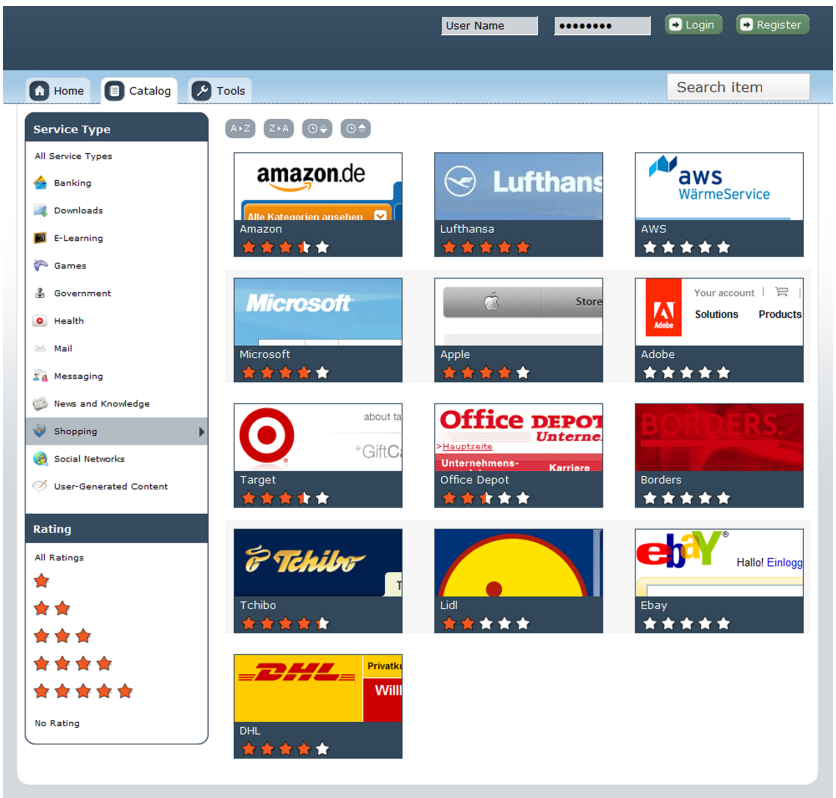
The welcome page shortly explains the purpose and the content of the privacy community and its related local privacy components. From this starting page, the user has the option to enter the catalog page, which lists all service provider maintained in the privacy community (see Figure 6.3(a)). Service providers are represented by tiles that contain a large provider logo, contributing to an easy association of the underlying article. In addition to the provider logo, a calculated average privacy rating is shown at the bottom of each tile. The rating is presented as star rating whose interpretation and usage is familiar to most users. Selections at the left side allow users to filter service providers based on their service type and their average privacy rating. Alternatively, users can type a service provider name in the search field at the top right of the page. Here, an auto-complete function eases the correct article selection. A third page provides a detailed presentation of the local privacy components and offers the download of an installer (see Chapter 10.7).

If a service provider is not listed in the privacy community, users can create a new service provider article at any time. The optional registration and login functionality is offered at the header of each page.

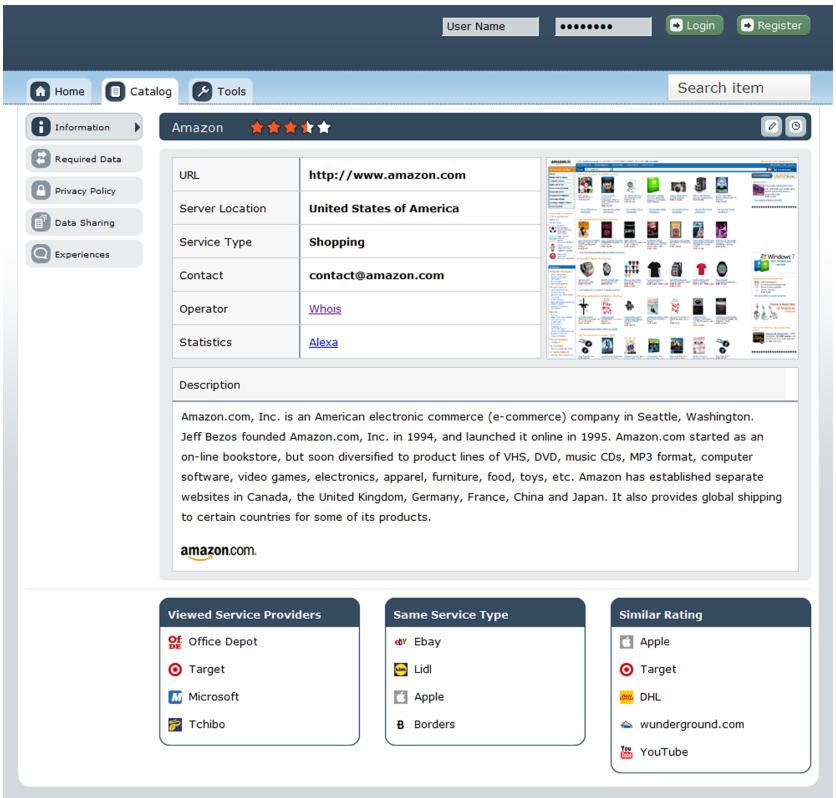
The privacy-related content of service provider articles is divided into five tabs. In the following, we present the community article of the eCommerce provider Amazon[2].

The initial tab offers a quick overview of the service provider (see Figure 6.3(b)), including the service provider's name, its average rating, a dynamically generated screenshot of its current Web site as well as privacy contact information (see Section 6.3.1). In addition, a general textual description of the service provider is presented. Moving the mouse over the average rating in the tab header triggers an overlay that lists the individual star ratings of each subcategory. At the bottom of the page three lists contain the most recently visited service providers, randomly selected service providers of the same service type and service providers with the most similar average privacy ratings. These article links allow for a quick comparison of similar service

[1] http://www-ifs.uni-regensburg.de/Privacy/
[2] http://www.amazon.com/

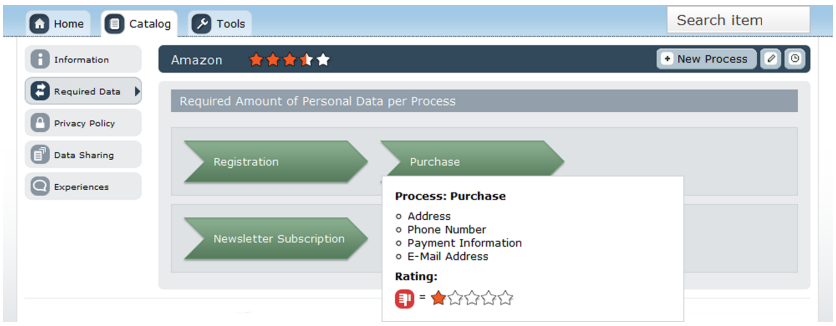


(a) Service Provider Catalog

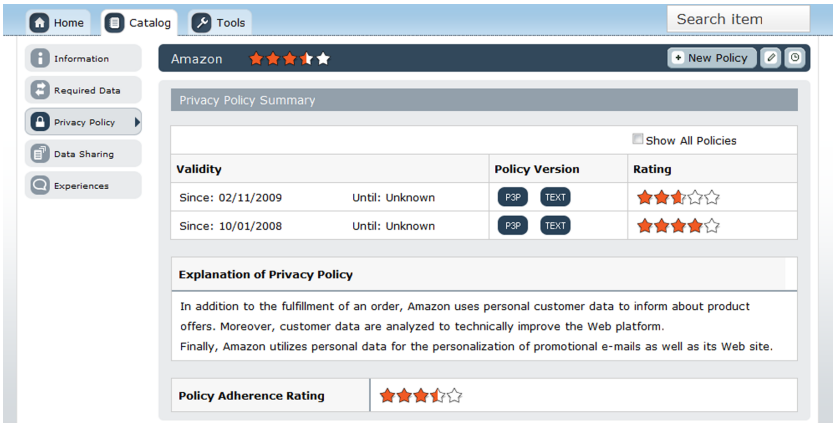


(b) Article – Static Provider Information

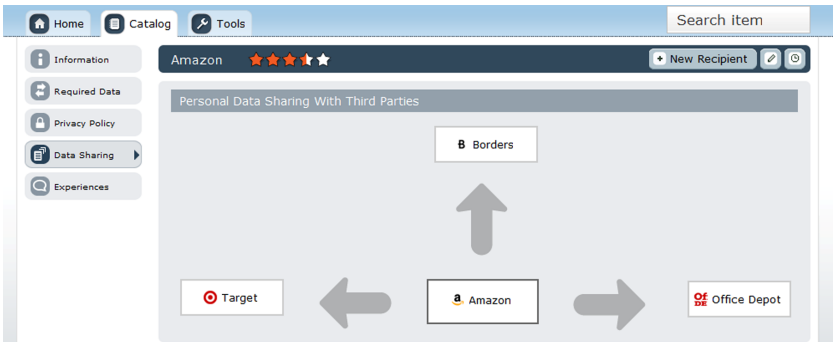
Figure 6.3: Design of the Privacy Community



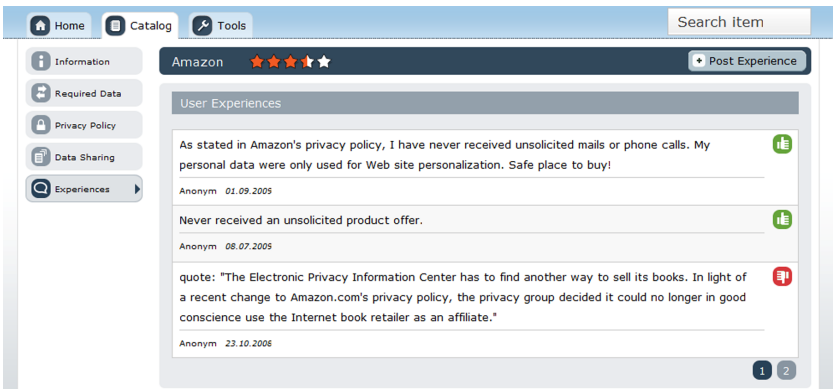
(c) Article – Required Amount of Personal Data



(d) Article – Privacy Policy Summary



(e) Article – Third Party Recipients



(f) Article – User Experiences

Figure 6.3: Design of the Privacy Community (cont.)

providers, facilitating an effective evaluation of service providers' privacy practices. The tab header and the three lists at the bottom of the page are presented in all remaining article tabs.

The second tab shows the required amount of personal data for all offered processes, which are represented by green arrows (see Figure 6.3(c)). If the user clicks an arrow, a text box unfolds containing the required personal data elements. In our example, the process *Purchase* is selected, which requires an *Address*, a *Phone Number*, *Payment Information* and an *E-mail Address*. The arrangement of the process arrows indicates that the *Purchase* process requires the completion of the *Registration* process that may require additional personal data. Based on comparable *Purchase* processes, the community evaluates the amount of required personal data as negative.

Concentrating on the data handling practices of the service provider, the fourth tab lists a privacy policy history – including the currently valid policy – along with a star rating of each policy (see Figure 6.3(d)). The history is capable of storing textual and P3P versions of privacy policies. Below the list of privacy policies, a short explanation of the effective privacy policy is offered to inexperienced users. At the bottom of the tab, users can rate the policy adherence of the service provider.

Focusing on personal data sharing, the fourth tab employs a directed graph to visualize third parties the service provider shares personal user data with (see Figure 6.3(e)). Originating from the examined service provider, arrows point to additional data recipients, which are represented by white boxes that contain their names and Favicons. If the user clicks a data recipient, the article of the respective provider is loaded.

Finally, the fifth tab lists posted user comments about privacy-related experiences with the service provider (see Figure 6.3(f)).

Edit buttons in the tab headers facilitate the addition and the revision of the collaboratively maintained information. If clicked, overlays capture the revised user input, using text boxes and drag and drop selections. Also placed in the tab header, a Versions button allows the recovery of older revisions.

6.4.4 Evaluation

Within the scope of our second user test (Chapter 5.3.4) we evaluated the design and the structure of an early version of the community Web front-end. We also assessed the user acceptance of and the potential user participation in a collaborative privacy community.

In order to measure the unbiased understandability of the page layout, test persons were only informed about the general purpose of the community. No detailed explanation of the Web page was provided. Before the first assignment, test persons had the opportunity to get familiar with the structure and the tabs of the Web page.

The first task targeted the submission of a provided privacy posting about the eCommerce shop Tchibo[3]. As the tested version of the community initially loaded the article of Amazon, this

[3] http://www.tchibo.com

task required the search for the proper service provider in the community using the offered navigation elements. Doing that, more than half of the test persons used the search field, which was placed in the page header, while the remaining test persons clicked on the index tab, the initial version of the catalog page, which presented a textual listing of all available service providers. The outcome and the user feedback led us to offer a general welcome page as well as a more meaningful catalog page, which both contribute to a more intuitive page navigation. Nevertheless, the test results underscored the necessity of offering both a service provider catalog and a search field. Once the proper article was loaded, all test persons clicked the correct tab and posted a comment with no difficulties.

Within the following task we asked test persons to post a particular rating for the policy adherence of Tchibo. Again, most test persons found the relevant tab and succeeded smoothly. Four out of 26 test persons, however, were not familiar with the handling of star ratings and needed guidance of the interviewers. Being asked after the completion of the task, 17 test persons were able to explain the difference between the ranking of the privacy policy itself and the policy adherence ranking, while nine persons could not differentiate both ratings. This fact highlighted the need of explaining labels, which were added in the final revision of the community.

Subsequently, test persons were asked to publish a set of data types required by a *Purchase* process at Tchibo. Only nine test persons solved that task problem-free. The majority of test persons did not intuitively find the position of the Edit and Save buttons, which led us to reallocate their positions.

In the interview section 22 out of 26 test persons agreed that the information of a service provider article was structured and designed in an understandable way. 24 out of 26 test persons would consult and use the privacy community in real-life scenarios. Being asked about the reliance of the community data, 18 test persons stated they would trust the privacy community, once the community gained enough members. The remaining eight test persons voiced their concern about the openness of the community, which – in theory – allows service providers to manipulate community data to their favor. One of these test persons admitted that a rising level of popularity and submitted data would alleviate that threat, as proven by well-known open reviewing systems of Ebay and Amazon.

17 test persons agreed that they would actively participate in editing articles of the privacy community. The remaining nine test persons would not add or change community data, hinting at their general reluctance to post content in public forums and other open content management systems.

A few test persons suggested the incorporation of other non-privacy-related information like delivery time or shipping costs. In our opinion, however, these data would contradict to the goals of the privacy community and would not contribute to its reputation as a provider-independent information source.

Finally, the user test revealed that 15 out of 26 test persons would upload and share their individual privacy preferences with selected users of the privacy community.

6.5 Agenda

This section outlines criteria for the successful launch as well as the long-term financing of an online privacy community.

As mentioned earlier in this chapter, the idea of a collaborative privacy community is primarily based on the successful Wikipedia concept [LC01]. Consequently, the launch of a privacy community can be related to success factors for the launch of a general ”Wiki”. For an enterprise Wiki, an insufficient structure and lacking user participation have been identified as the most serious obstacles [Bar06]. Usability aspects of the privacy community have been focused from the beginning of the design process. Our conducted user test proved the understandability of the community’s user interface, leaving the involvement of users as the remaining success factor.

In order to achieve the critical number of participating users, the successful launch of a privacy community requires the full exploitation of the Wiki effect. The Wiki effect is defined as a large number of Internet users visiting a Web site on a regular basis and voluntarily contributing to the structure, shape and quality of its content [EGH07]. In order to benefit from the Wiki effect, Davies recommends selectively seeding initial content that introduces the Wiki’s goal to new participants [Dav04]. The character of this initial content should not be final or complete, as this could prevent users from editing that content.

Hence, before the launch of a privacy community, articles of the best-known service providers should be created and filled with sparse information. In particular, we recommend entering the offered service type, the amount of required personal data and the release of personal data to third parties, as this information represents essential input for the local privacy components.

Addressing the promotion of a privacy community, comments in public forums and other public communication channels contribute to a Wiki’s level of popularity [Par06]. In addition, a press release is recommended, which should be forwarded to authors of blogs and other topic-related Web sites. Furthermore, the creation of a clear tutorial is suggested that outlines the community’s main goals and functionality. A tutorial enables new users to gain a quick overview and increases the chances that users contribute actively.

Finally, crucial success factors of the privacy community are security and trust considerations. In theory, the offered platform is capable of tracking the information requests of users. As local privacy components query data from the privacy community on a regular basis, a malicious community provider could create and misuse detailed navigation logs. Our proposed privacy community acknowledges this potential privacy threat and allows the anonymous accessing and editing of all collaboratively maintained data. When users have built the necessary level of trust in the community, they can opt to register, which allows them to connect to friends and to exchange experiences and privacy preferences.

After a successful launch and a risen number of users the long-term maintenance of a privacy community inevitably involves expenses, e.g. for the operation of the server infrastructure and, possibly, the entailing maintenance personnel. In the context of Web 2.0 applications,

Alby discusses the potentials of both advertising and a fee-based membership [Alb07]. These sources of financing can also be applied to the maintenance of a privacy community.

Technically there are multiple ways to place advertisements into a privacy community. Considering our proposed solution, advertisement should not be used at the expense of usability and should not affect the clear structure of an article. For the same reason, intrusive advertisements like pop-up windows should be avoided. While dynamic advertisement applications such as Google AdSense provide attractive models for the generation of revenues, we do not recommend the integration of contextual advertisements into a privacy community, as this could result in the placement of a service provider's ad banner in the community article of that provider. Such a behavior does not underscore provider-independence and could weaken users' trust in the privacy community.

A fee-based membership represents a further business model for a privacy community. If this financing source is chosen, the privacy community should provide a basic service offer for free. Special, value-added functions could require a membership involving monthly or annual fees. In particular, the social networking component that facilitates the direct exchange of privacy preferences and experiences could be defined as a premium function only available for paying members. Premium functions could also involve certain features of the local privacy components.

Finally, a privacy community could raise donations to cover operating costs, highlighting its service provider independence. This option would, however, require the abandonment of advertisement. The prominent example of Wikipedia shows that donations can cover operating costs of a large community. Donations are made by satisfied users who are convinced of Wikipedia's goal that knowledge should be accessible to anyone.

The goal of the introduced privacy community is the provider-independent enhancement of privacy. If the increased level of privacy and the improved privacy awareness of both users and providers are recognized, the privacy community could equally convince users of its higher goals and motivate to make donations.

6.6 Summary

This chapter starts with a definition of the privacy community's central role within our proposed privacy architecture and outlines its conceptual architecture.

We continue with a presentation of the content and functions of the collaborative privacy community. Targeting a provider-independent source of privacy-relevant service provider information, the privacy community collects static provider information and individual experiences. Moreover, our solution enables users to know in advance, what personal data are required for a specific service offer and which third parties the service provider shares this information with. Furthermore, the community maintains explanations and ratings of privacy policies along with a policy adherence rating. Finally, our presented solution facilitates the exchange of uploaded privacy preferences among registered users.

Considering the characteristics of the accessing clients, we employ a Service-oriented Architecture for the communication within the proposed privacy architecture. Web services serve as exclusive interfaces for both the Web front-end and the local privacy components. The Ajax Web front-end provides intuitive navigation and clearly arranged service provider articles, which can be viewed and edited in an understandable way. A conducted user test proved the usability of the Web front-end as well as the user acceptance of an open privacy community.

Finally, we provide an agenda for the successful deployment of a privacy community including its long-term maintenance. The selective seed of content and a widespread Internet promotion are necessary to quickly gain a sufficient number of users and to take advantage of the Wiki effect. Furthermore, we identify moderate advertisements in the Web front-end, a fee-based membership and donations as promising ways to cover operating costs.

The introduction of the local privacy components in the following chapters will further underscore the advantages and the potential of the presented privacy community within our proposed privacy infrastructure.

Chapter 7

Privacy Preference Generator

Focusing on the definition of potential information flows, the Privacy Preference Generator component builds the first local privacy component in our proposed privacy architecture. Its main purpose is the usable and understandable definition of disclosure rules, so called privacy preferences, that define the circumstances and conditions under which users are willing to disclose personal data.

These individual disclosure rules are employed by the Privacy Agent (see Chapter 8.3.3) to detect potential conflicts with machine-readable privacy policies of service providers. As a consequence, the resulting recommendations of the Privacy Agent require privacy preferences of considerably high quality and accuracy.

After a definition of the term privacy preferences, this chapter describes design challenges and requirements of a user-friendly tool for the capture of privacy preferences. We continue with a discussion of related work, before we introduce the concept and the design of a solution that features the privacy preference generation for individual Internet service types and the consideration of privacy experiences during the configuration process. Finally, we lay out further approaches that are capable of capturing privacy preferences.

7.1 Privacy Preferences

According to the P3P Privacy Preference Exchange Language specification privacy preferences are defined as

> ”the user’s desires regarding the collection and treatment of [personal] information” [CLM02a].

In Chapter 3.5.4 we analyze the main drivers that impact the disclosure of personal data. A comprehensive user study of Cranor et al. reveals that the disclosure behavior of users primarily depends on the *requested personal data types* as well as the *purpose of collecting personal information*, including the *disclosure to third parties* [CRA99]. These factors dictate the main dimensions of a suitable tool that captures privacy preferences of the user.

7.2 Design Challenges

The subjective disclosure behavior of users (see Chapter 3.5) requires user-friendly and intuitive user interfaces of any privacy preference generation tool that aims for a high degree of user acceptance. Analyzing the development of user interfaces within the context of P3P, Cranor et al. [CGA06] define the following design challenges:

- **Complexity of Users' Privacy Preferences**
  In theory, the diversity of possible privacy preferences is infinite. The development of a tool should find the balance between providing simple interfaces that allow the definition of understandable preferences and the possibility of fine-tuning preferences to individual personal needs. In this context, the observed mismatch of pre-defined preferences and the actual behavior of users marks a special challenge (see Chapter 3.5.3). In many cases users' behavior differs from pre-defined preferences, because an advantage is expected for releasing additional personal data.

- **Inexperienced Users**
  Users' lack of experience in the area of privacy represents a further design challenge. Even though users are increasingly concerned about their privacy (see Chapter 3.5.1), their knowledge about privacy threats and technologies that could help protect their privacy is considerably low [JPJ05]. Furthermore, users are not familiar with technical and legal terms related to privacy [Pol07, FHP04]. As a consequence, basic contents and dialogues should explain all privacy terms sufficiently. A simple and intuitive user interface will allow the vast majority of inexperienced Internet users to use and understand a privacy preference generation tool.

- **Complexity of the P3P Specification**
  The first version of P3P provides eight element types, the latest version 1.1 introduces additional types. Disregarding optional attributes, 36.000 different policies can potentially be modeled with P3P [CGA06]. A user interface trying to capture privacy preferences at that fine-grained level is obviously not usable. This is especially evident, if one considers the previously discussed low level of user experience.

The described design challenges underscore the prevalent contrast between users' low familiarity with privacy matters and the potential complexity of privacy preferences and corresponding privacy policies. The design of a suitable tool that captures privacy preferences must find the balance between these crucial factors.

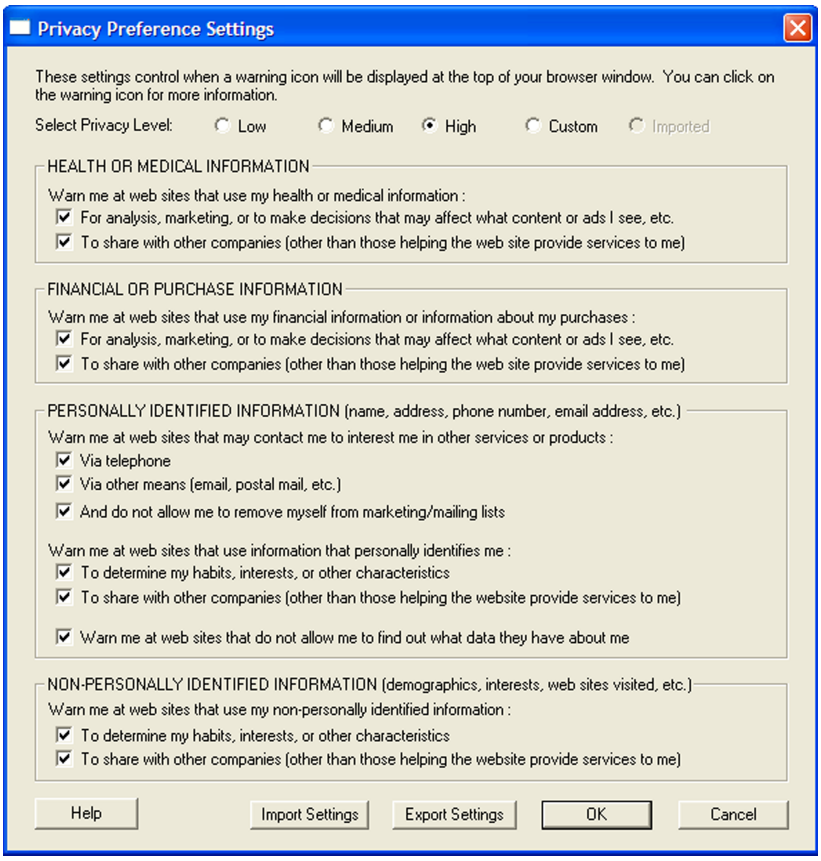


Figure 7.1: Privacy Preference Settings of the Privacy Bird

7.3 Related Work

The Privacy Bird[1] implements a complete P3P-based privacy agent that focuses on usability aspects. As described in Chapter 4.2.2.4, the Privacy Bird automatically retrieves P3P policies of service providers, which are interpreted and matched with privacy preferences of the user. A bird in the browser header is used to signal the result of the matching process to the user.

As the Privacy Bird calculates tailored disclosure recommendations, the tool requires privacy preferences of the user as input. After the initial start individual privacy preferences are prompted. These preferences are captured offering a dedicated window, which is shown in Figure 7.1.

The simplicity of the interface is motivated by the design goals of the Privacy Bird, which aims for a low complexity of the privacy topic and, as a consequence, a high user acceptance [CGA06]. For this reason, only fundamental configuration options are available.

The configuration page of the Privacy Bird divides privacy preferences into four groups that represent personal data categories. According to a survey [CRA99] the first two groups (*health or medical information* and *financial or purchase information*) contain particularly sensitive

[1] http://www.privacybird.org/

user data, while the last two groups (*personally identified information* and *non-personally identified information*) account for user concerns of being identified when releasing personal data. The options of each group address data handling practices of service providers, such as contacting or third party recipients. Alternatively, users can select one out of three pre-defined profiles (*low*, *medium* or *high*).

Analyzing the usability of the Privacy Bird's configuration page, we examined the privacy preference generation process within our first user test (see Chapter 5.3.3). Most test persons had difficulties understanding the texts of the sparsely designed window. The interpretation of the four data category groups caused further problems, as the groups did not show the involved data types. The radio buttons at the top of the page switch to *custom* as soon as an option is changed. This confuses users, as no evaluation of their configuration remains. Furthermore, users do not completely comprehend the privacy-related language used in the explanations of the configuration options and do not feel confident about the selected options.

The outcome of the conducted user test revealed that the available privacy preference settings of the Privacy Bird do not provide a satisfactory user experience. The dedicated window does not offer the necessary explanation and graphical support for privacy novices. Moreover, as one single page captures all preferences, the interface does not provide detailed configuration options for more experienced users and relies on the aggregation of four broad data categories. Addressing these usability shortcomings, the following section introduces a dedicated tool that generates privacy preferences in a user-understandable way.

7.4 Service Type-based Privacy Preference Generation

Considering the design challenges and deficits of existing solutions, this section introduces a user-friendly, P3P-based Privacy Preference Generator that provides a configuration wizard as well as a clear preference summary [KP09]. In the following, we present the main capabilities of our solution. Exemplifying the design features of our presented application, we illustrate the privacy preference generation process with screenshots. Chapter 10.3.2 lays out additional implementation details of the introduced tool.

7.4.1 Multi-Level User Interface

As described earlier, the potential complexity of individual privacy preferences and service providers' privacy policies represent significant obstacles for the development of a suitable solution. A tool that generates fine-grained privacy preferences inevitably requires intense user interaction as well as a considerable amount of time to complete the definition process. Such an approach, however, is contrary to the prevalent inexperience of most users. Meeting the needs of this majority of Internet users calls for a tool that is understandable and logically structured. The tool should be restricted to a manageable but still expressive amount of configuration options, in order to facilitate a quick generation of privacy preferences. An elementary interface

enables inexperienced users to fully utilize the application and, as a consequence, encourages users to understand and get informed about privacy-related topics and threats [Tid05, PK03]. When users' level of experience gets more advanced, they will be able to define more detailed privacy preferences.

Our proposed solution acknowledges the prevalent heterogeneity of users. For this reason, we present a multi-level user interface, splitting the user interface into two complexity modes with differing amount of configuration options. Offering modes of diverse complexity, our application targets both inexperienced privacy novices and privacy experts. As inexperienced users may advance over time, modes can be switched at any time during the privacy preference generation process.

The basic mode is designed for the majority of Internet users with little knowledge about privacy threats and data handling practices of service providers. This mode accounts for most users' low level of experience in this area. Only fundamental and most necessary settings are available. All options are explained in an understandable way using clear examples. Choosing the basic mode, the user generates a complete set of individual privacy preferences in about five minutes.

The advanced mode is designed for experienced users. In this mode a more detailed configuration environment is offered, allowing the Privacy Agent to make more individualized recommendations. This mode takes advantage of most elements of the P3P vocabulary, giving privacy experts the chance to fine-tune their settings at a high degree.

After the Privacy Preference Generator is installed and started the first time, the user is offered a detailed tutorial explaining all settings of the tool. Subsequently, the user is asked to choose the preferred mode used for the configuration process. As mentioned above, the selected mode can be changed at any time enabling users to switch and reassess their level of experience. The content and the layout of each mode are described in the following sections.

7.4.2 Service Type-based Approach

In addition to the complexity of preferences and the inexperience of most Internet users, one of the main challenges is the irrational and unpredictable disclosing behavior of users (see Chapter 3.5.3). Even though the compensation users receive for disclosing comprehensive personal data is generally small, the perceived value for users is estimated much higher [Tre07]. As a consequence users tend to disclose more personal data than required to service providers, not questioning the purpose of the additional data requests.

These facts lead us to conclude that users transmit personal data goal-oriented, i.e. personal data are disclosed focusing on a specific user goal, such as purchasing a product at an online store. Furthermore, trying to achieve this goal, the decision of users which personal data to disclose is a subjective matter. For example, some users might disclose their phone number and their date-of-birth to an online store, even if these data are not necessary for the fulfillment of an online shopping service.

Addressing this goal-oriented behavior of users, our solution provides the definition of privacy preferences for each of twelve pre-defined service types. Each service type represents a World Wide Web service category and corresponds to an abstract user goal. The definition of privacy preferences for each Internet service type allows for more practical and accurate results, as it permits e.g. to define a user's willingness to disclose his/her credit card information to an online shop, and not to release this information to any other service type, such as an Internet forum or a Web mail service. Additionally, individual privacy preferences for service types enable our solution to provide an adequate set of personal data for each service type (see Chapter 7.4.3), which substantially assists users in the configuration step of specifying personal data they are willing to disclose.

For the identification of meaningful and realistic service types we first looked at the P3P specification. With the release of version 1.1 the Primary Purpose element (`PPurpose`) allows service providers to specify the primary purpose for collecting personal data [CDE+06]. The 23 `PPurpose` elements were designed to group service offers and served as basis for our definition. Aiming at a more user-friendly and condensed service type set, we analyzed surveys aiming at the categorization of Internet services. A survey conducted within the PRIME project identifies ten service categories out of 46 most frequently used online services [Ber05]. SevenOne Interactive [Sev07] published a survey about Internet user groups. The survey is based on users' online interests, Internet usage, eCommerce behavior as well as demographic characteristics. Altogether, the survey analyzed 52 Internet content types, out of which thirteen homogeneous clusters were built.

Based on the P3P `PPurpose` elements and the described surveys, we define the following twelve service types for our solution:

- Banking
- Downloads
- E-Government
- E-Learning
- Games
- Health
- Mail
- Messaging
- News and Knowledge
- Shopping
- Social Networks
- User-generated Content

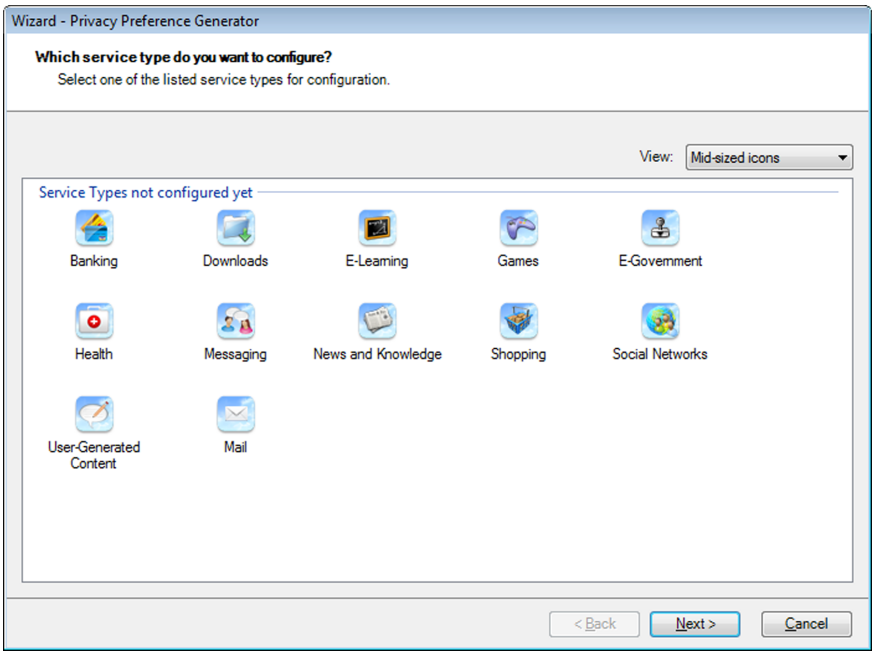


Figure 7.2: Service Type Selection

After the initial start of the tool and the selection of a complexity mode, the user chooses the first out of twelve service types (see Figure 7.2). Subsequently, a configuration wizard collects user's privacy preferences for the selected service type. After the completion of the wizard the user can choose to configure the next service type or to directly proceed to the privacy cockpit (see Chapter 7.4.6). A non-configured service type indicates that the user is not willing to interact with and disclose personal data to service providers of that service type. In order to facilitate easy recognition, we designed icons that allow for an easy association with respective service types.

At this point we would like to point out that the service type-based approach entails the reliable identification of a Web site's service type by the privacy agent. In Chapter 8.3.3 we present a solution that employs a Web service offered by the privacy community.

7.4.3 Recommended Data Sets

A central component of privacy preferences is the amount of personal data a user is willing to disclose. For this purpose, the P3P specification offers a basic set of data types and data categories that primarily focuses on online and demographic data [CDE+06]. In order to represent all commonly used personal data transmitted in the World Wide Web, we extended the P3P vocabulary with additionally required data types. Appendix A.1 provides a complete list of the data type vocabulary used within our privacy infrastructure. The data type definitions are centrally administered in the privacy community (see Chapter 10.2.2).

As our solution captures privacy preferences for twelve service types, we enable users to

individually define a set of data types they are willing to disclose to service providers of the respective service type.

In general, a given service in the World Wide Web should not require more than a certain, adequate amount of personal user data. This set of appropriate personal data generally depends on the service type a certain service belongs to. For example, an instant messaging service should not require more than a nickname and, possibly, an e-mail address, while online shopping services additionally need users' payment information and delivery address to fulfill their service offers.

In an effort to address the needs of especially inexperienced users, our solution provides a recommended adequate set of personal data for each service type. A survey conducted within the scope of a user test (see Chapter 5.3.4) helped us identify data sets each service type generally requires. Furthermore, our results incorporate the findings of a PRIME survey [Ber05]. Appendix A.2 provides a list of all service types with their recommended data sets. Both service type definitions and their recommended data sets are centrally maintained in the privacy community (see Chapter 10.2.3).

As we defined the appropriate amount of personal data requests for each service type, our application is able to provide this information during the privacy preference generation process. In the basic mode the application recommends assuming this data set and not releasing additional data to a service provider. In order to offer a well-arranged user interface, we aggregated personal data types into data groups and designed individual icons for each data group.

In our example, Figure 7.3 shows that the selected service type *Shopping* requires data from the data groups *Financial Data*, *Login Data*, *Name Data*, *Online Data* and *Postal Address Data*. Moving the mouse over a data group triggers a tool tip that shows the individual data types of that group.

In addition, the user is asked to specify the behavior of the Privacy Agent, if a Web site requests more data than the recommended data set. The user can choose to:

- Not accept the release of additional data
- Accept the release of personal data, but to be specifically warned and informed about this practice
- Accept the release of additional personal data without warning (not recommended)

This configuration is offered to the user in accordance with the `behavior` element of APPEL [CLM02a]. The `behavior` element can take on the values `block`, `limited` and `request`. The values are applied to the three options respectively. Additionally, the second option employs the APPEL `prompt` element to trigger a user notification.

If the advanced mode is chosen, the wizard allows direct adjustment of the data set the user is willing to release (Figure 7.4). The left area lists available personal data types a user possesses. All data types are arranged in their corresponding data groups. The right box shows data types selected for disclosure. By default, the right box contains the recommended data set.

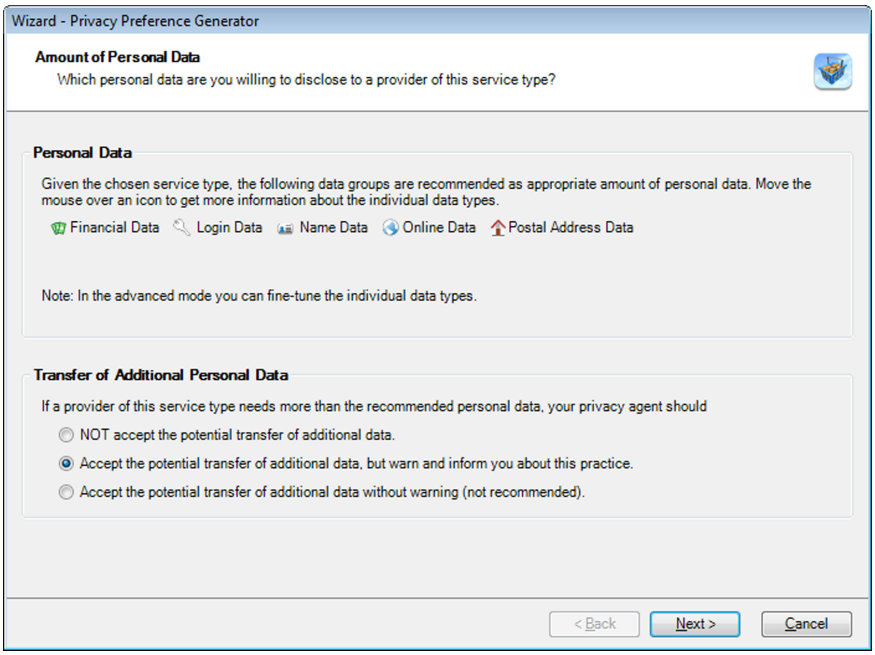


Figure 7.3: Personal Data Page of the Basic Mode

If an element is double-clicked, it is moved from one box to the other. The same functionality is provided by arrow buttons placed between both boxes. An additional button resets the elements to the recommended data set.

7.4.4 Purpose of Collecting Data

Apart from the amount of personal data, controlling and restricting the usage of these data is a primary factor for the disclosure of personal data (see Chapter 3.5.4). In the context of our proposed tool, a configuration wizard captures purposes a service provider – in addition to the fulfillment of the original service – may use transferred personal data for. The most common and privacy-relevant additional purposes are personalization and contacting.

Personalization techniques enable service providers to tailor the content of their Web sites to users' needs and interests, by processing users' disclosed personal data and their surfing behavior [Sea03]. A well-known example of personalization measures is the individual product offering a returning customer obtains at an online shopping Web site. A further, more intrusive case is an individually tailored advertisement.

Our proposed application enables users to specify the preferred personalization activities of service providers. The P3P specification defines five personalization elements [CDE+06]. Catering usability needs, we aggregated the elements into three distinctive groups:

- **One-time Personalization**
  A service provider collecting personal data for the purpose of one-time personalization adjusts content and design of a Web site using transferred personal user information

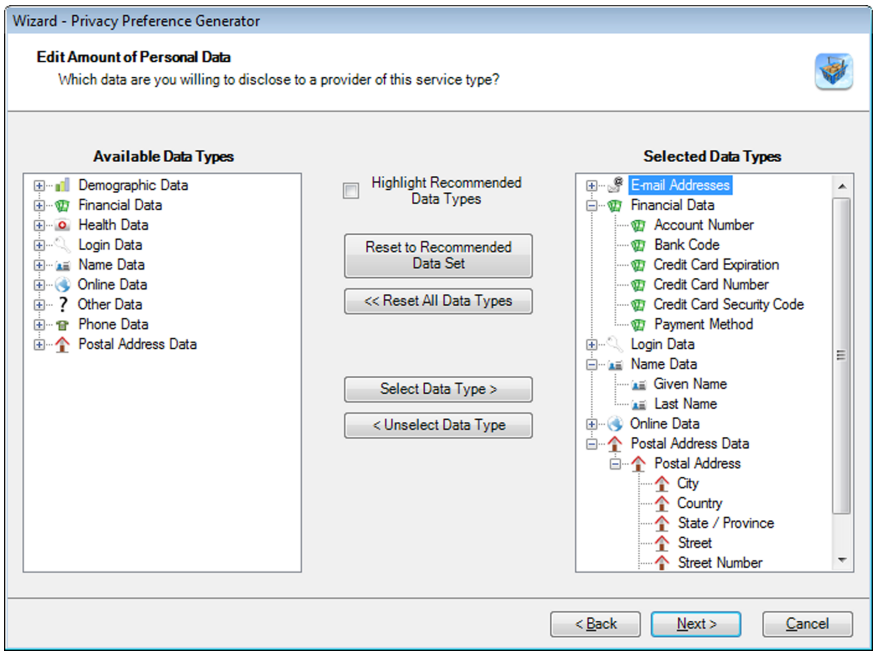


Figure 7.4: Personal Data Page of the Advanced Mode

as well as clickstream information. Personalization is only conducted for the current visit. Personal data collected for one-time personalization are not permanently stored by the service provider. Hence, personalization is not applied at any future visit. The P3P element indicating one-time personalization is `tailoring`.

- **Pseudonymous Personalization**
  Unlike one-time personalization a service provider collecting personal user information for pseudonymous personalization stores personal user data for future personalization measures. Additionally, the service provider uses personal user data to analyze Web offers. This form of personalization only uses non-identifiable user data, such as gender and year-of-birth. The corresponding P3P elements for pseudonymous personalization are `pseudo-analysis` and `pseudo-decision`.

- **Individual Personalization**
  In addition to pseudonymous personalization service providers collecting personal user data for individual personalization also use identifying user information (e.g. date-of-birth, postal address and e-mail address) for personalization measures. The P3P elements representing this group are `individual-analysis` and `individual-decision`.

A further prominent and valuable purpose of collecting personal data is direct contacting. In many cases this activity is triggered by a service provider's marketing department that promotes a product or offers an individualized service. The communication channels users can be contacted are e-mail, postal or telephone.

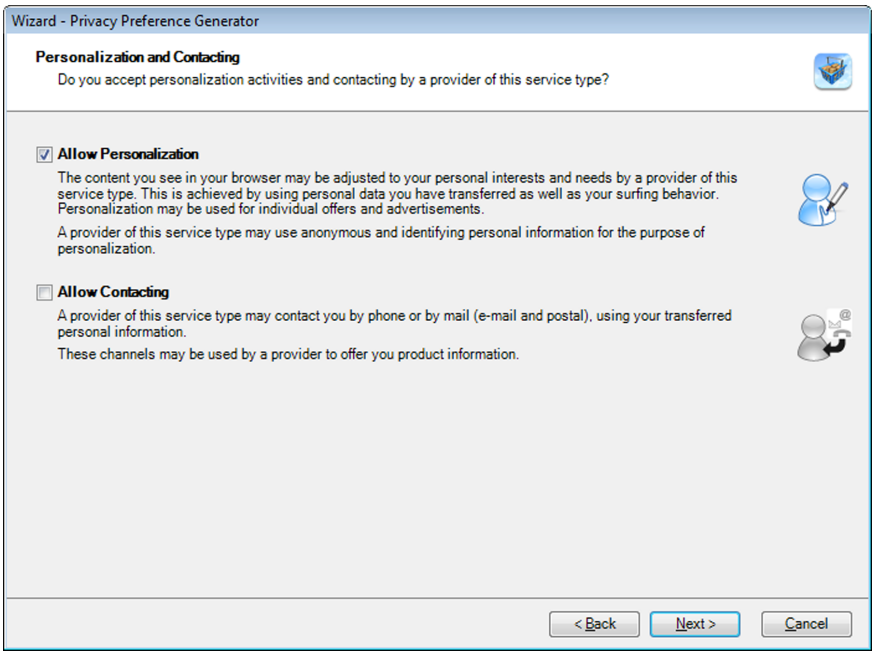


Figure 7.5: Personalization and Contacting Page of the Basic Mode

As many users dislike being directly contacted, our presented configuration wizard provides options to adjust privacy preferences accordingly. Following the P3P specification we adopt the two contacting categories mail (e-mail and postal) and telephone. The respective P3P elements are `contact` and `telemarketing`.

Figure 7.5 depicts the personalization and contacting page for the basic mode. As users in this mode tend to be inexperienced and value user-friendliness over rich configuration settings, we limited available options to allowing or disallowing personalization and contacting as a whole and left out further options of fine-tuning. All options are explained thoroughly. In our example the user chooses to allow personalization and not to allow contacting for the selected service type *Shopping*. Serving understandability and user acceptance of the application, we designed supportive icons, which are placed at the right of each option. Underscoring the selection of the user, the icons are grayed out for unselected options.

In the advanced mode we allow more fine-grained configuration of each purpose. For this purpose, we split the page of the basic mode into a personalization page and a contacting page. The personalization page presents all three identified personalization types, the contacting page the two available contacting categories. Figure 7.6 shows the personalization page in the advanced mode. In our example the user opts to allow one-time personalization as well as pseudonymous personalization and does not allow individual personalization. Again, the user selection is supported by individually designed icons for each option.

After a change in the tool settings, an extended advanced mode enables users to further refine the selection of each option. Employing the `required` attribute of the P3P specification, users can choose to:

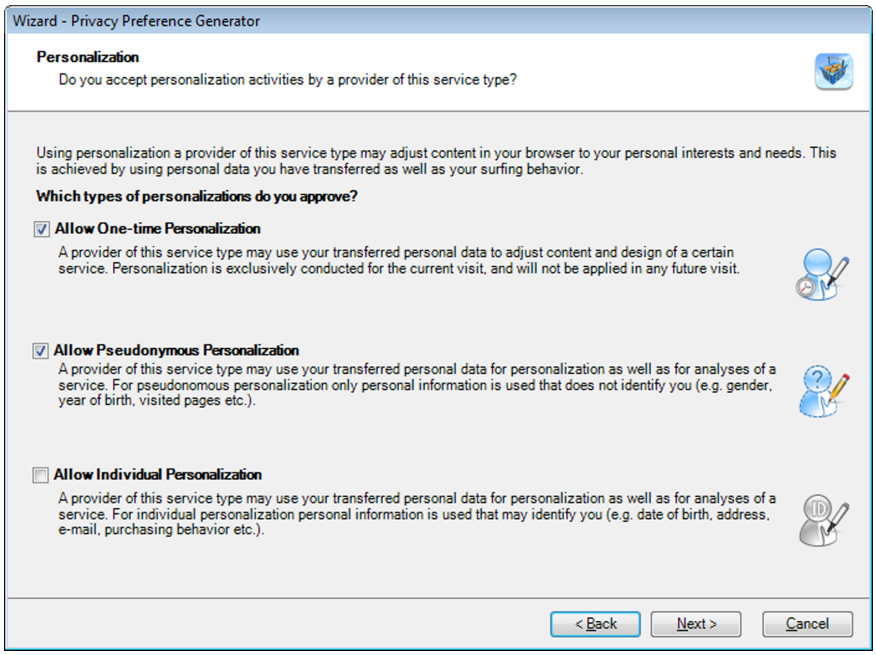


Figure 7.6: Personalization Page of the Advanced Mode

- Not allow a configuration option (equivalent to an unchecked box)
- Only allow a configuration option, if an explicit user consent is required (opt-in)
- Only allow a configuration option, if an opt out option is provided (opt-out)
- Always allow a configuration option (equivalent to a checked box)

The extended personalization page of the advanced mode in Figure 7.7 shows that the preferences of the user concerning one-time personalization and pseudonymous personalization are bound to conditions. The user only accepts the purpose one-time personalization, if the service provider offers an option that disables this practice (opt-out). For pseudonymous personalization, the user even requires his/her explicit consent (opt-in) during or after the disclosure of personal data. Again, the user chooses not to allow individual personalization.

7.4.5 Disclosure to Third Parties

The final page of the configuration wizard is dedicated to recipients of personal user data. As service providers pass personal user data to third parties for various reasons, our application gives users means to customize this privacy-sensitive practice in their privacy preferences.

Like for personalization, we looked at the P3P specification and aggregated third party groups based on the recipient elements [CDE+06]. The P3P element `ours` represents the service provider itself, or agents that act on the behalf of the service provider. It is set by default in users' privacy preferences. The remaining P3P recipient elements specify third party entities, which are grouped as follows:

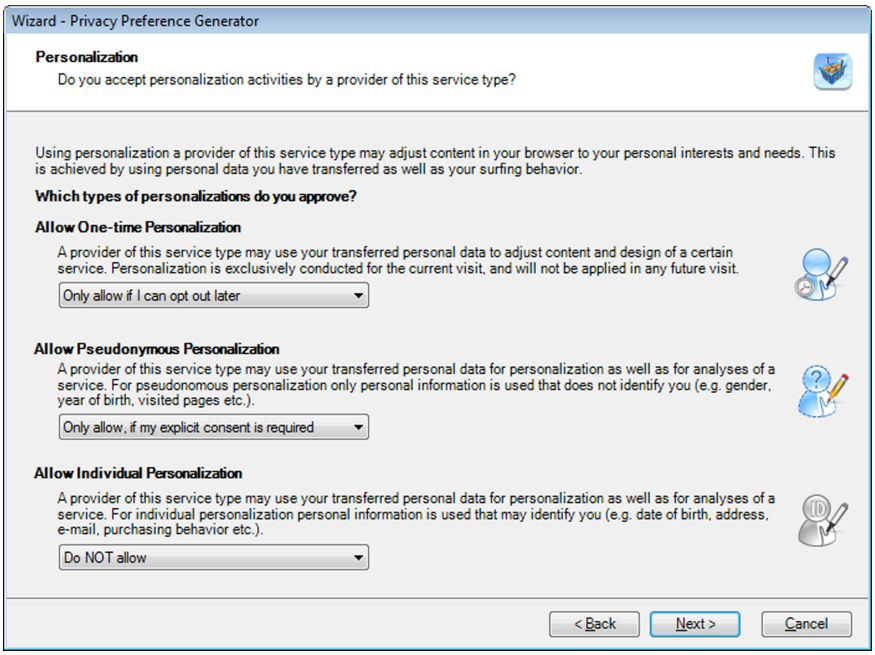


Figure 7.7: Personalization Page of the Extended Advanced Mode

- **Third Parties Following Equable Practices**
  This group comprises third parties with equable privacy and data handling practices. These recipients use personal user data on their own behalf. However, all parties are required to follow the same restrictions concerning user data as the service provider. The corresponding P3P element for this group is `same`.

- **Third Parties Following Different Practices**
  The privacy and data handling practices of this group differ from those of the service provider. These third parties are not affiliated with the service provider and use personal data on their own behalf. The P3P elements for this group are `delivery` and `other-recipient`.

- **Public Areas**
  A final third party group represents all publicly available areas like fora and public directories. It also includes third parties with unknown privacy and data handling practices. The associated P3P recipient elements are `unrelated` and `public`.

The third party page design, including the composition of each mode, resembles the personalization and contacting pages. In the basic mode, the user can solely choose whether or not to accept the disclosure to third parties following equable practices (the first defined third party group) and the disclosure to other third parties (the combination of the latter third party groups). In the advanced mode all three groups are available as individual options.

Figure 7.8 presents the third party configuration page of the basic mode. The user only accepts

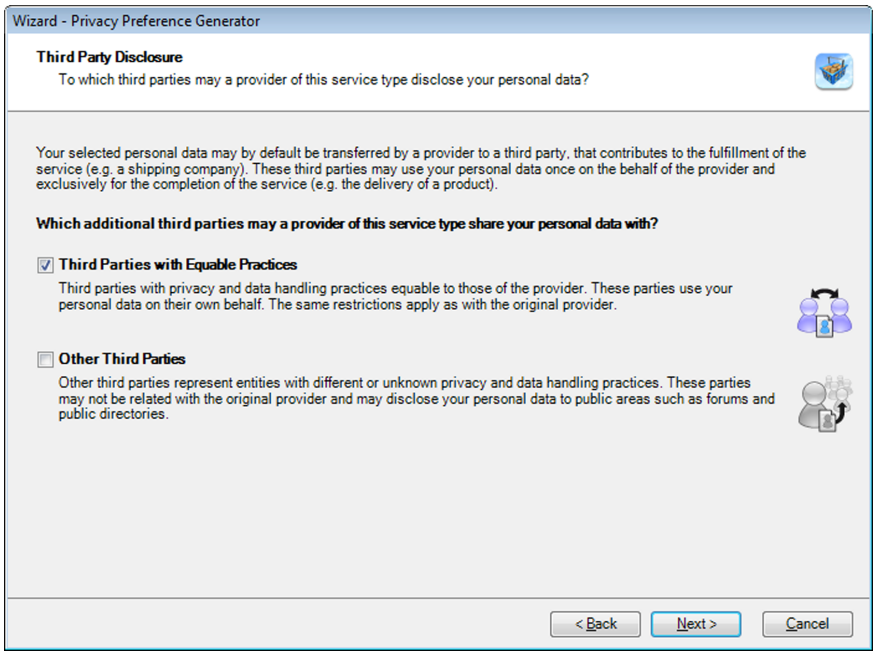


Figure 7.8: Third Party Page of the Basic Mode

disclosure of personal data to third parties following the same privacy and data handling practices as the service provider.

7.4.6 Cockpit

After the completion of service type configurations, our solution offers a clear overview of users' privacy preferences. For this purpose, the tool switches to the privacy cockpit, the main page of the Privacy Preference Generator. This page provides a quick and comprehensible summary of all configured service types, enabling users to oversee and double-check all settings and options. In addition, the cockpit page presents an evaluation of each configured service type with regard to the privacy impact of the configuration.

The columns of the dynamically generated table show all configured service types. The table rows list all available configuration options of the selected mode. As users value consistent interfaces [Tid05], the selected mode – just like in the configuration wizard – does not change the design of the cockpit page, but only impacts the number of rows, as the number of available options depends on the selected mode.

Figure 7.9 shows a screenshot of the privacy cockpit in the advanced mode including seven configured service types. Icons in the table mark the selection of a particular option for a specific service type. Conversely, a missing icon in a table cell indicates that the configuration option for a service type was not approved by the user. Targeting user acceptance, we reused the supportive icons of the configuration wizard. If the user moves the mouse over an icon,

the Live-Help area shows additional explanations of that configuration option, applying the concept of a Two Panel Selector [Tid05].

In order to give users a meaningful and understandable feedback of their configuration, service types are sorted horizontally by their impact on users' privacy. Doing this, we calculate a privacy level for each service type, taking equally into account the configurations for personalization, contacting and disclosures to third parties. Enabling users to switch modes at any time while maintaining the privacy level scores of each service type, the options of the basic mode have a higher weight in each aggregated configuration category than the more fine-grained options of the advanced mode. The resulting score for each service type ranges from 0 (all options selected; lowest privacy level) to 1 (no options selected; highest privacy level).

Applying the calculated privacy levels, service types are arranged in ascending order. Fostering user acceptance and providing a quick recognition and evaluation of privacy levels, we align all service types on a dynamic color scale. The color scale underscores the calculated privacy level, starting with red and fading to yellow and green. With regard to the calculated privacy level score, we allocate service types as follows:

- Privacy Level Score between 0 and 0.33: Red
- Privacy Level Score between 0.34 and 0.66: Yellow
- Privacy Level Score between 0.67 and 1: Green

In the exemplary configuration of Figure 7.9 the service type *News and Knowledge* shows the lowest privacy level score and, as a result, is listed first. As its calculated privacy level score is 0.28, the service type is accompanied with a red background color. The adjacent service types *Banking* and *Social Networks* show a score of 0.44 and 0.5, respectively, resulting in a yellow background color. The remaining service types *Shopping*, *Mail*, *E-Learning* and *User-generated Content* are presented with a green background color with privacy level scores of 0.67, 0.72, 0.89 and 1, respectively. We point out that the background color scale shifts dynamically, as the configurations and the privacy level scores change.

A click on the "Configure a Service Type" button at the bottom of the page opens the service type selection page (Figure 7.2) that allows the configuration of an additional service type. If the user chooses to change the configuration, a click on a service type icon in the table restarts the configuration wizard for that service type.

7.4.7 Quick Mode

While the configuration of a single service type can be completed in a manageable amount of time, the full configuration of twelve service types can turn into a lengthy task, especially for service types with repetitive preferences. Seeking a method that enables the quick set-up of multiple service types, we added a quick mode functionality to the privacy cockpit, which is displayed in Figure 7.10.

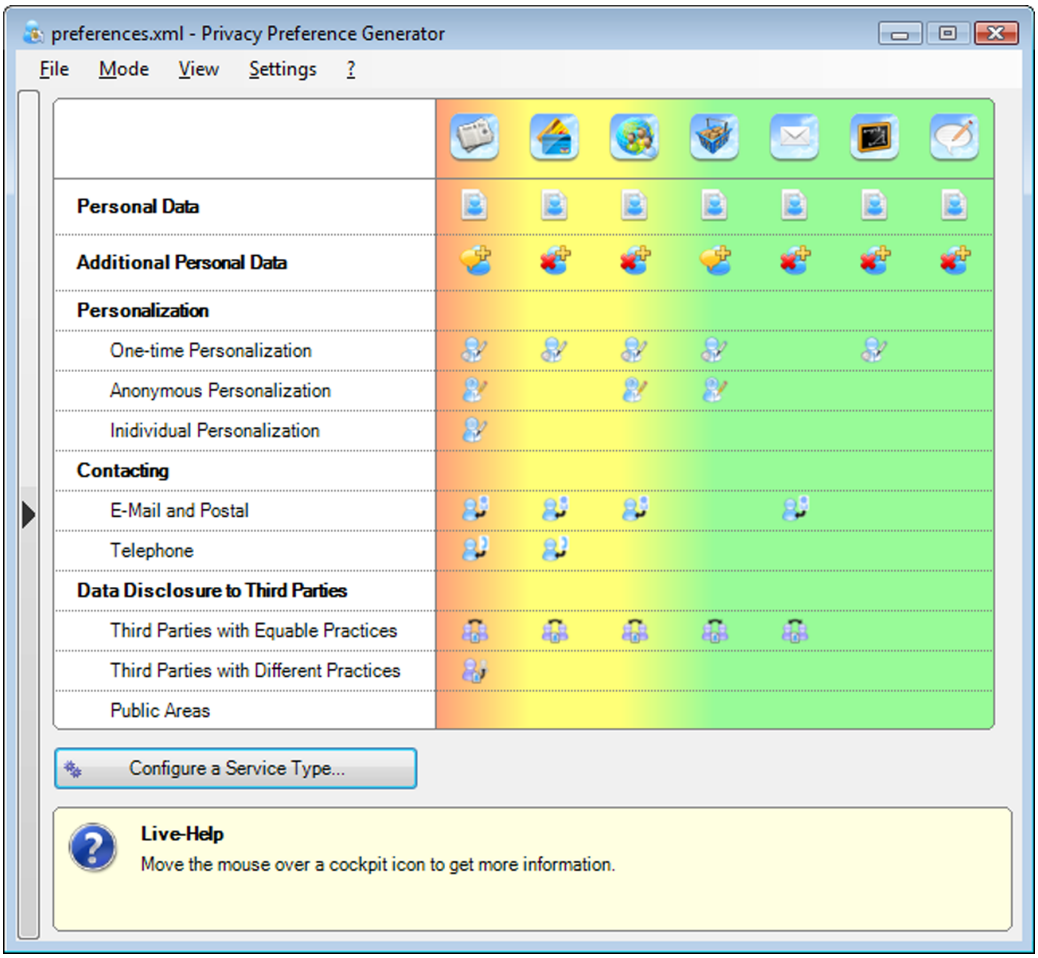


Figure 7.9: Privacy Cockpit – Overview of Configured Privacy Preferences

Expanding the privacy cockpit page, the quick mode places a box at the left side of the cockpit, which lists all unconfigured service types. A link below that box leads to the quick mode section of the tutorial. The quick mode is activated and deactivated using the stretched vertical button at the left of the dynamic table, guaranteeing easy accessibility to the quick mode functionality [Tid05].

Once the quick mode is activated, unconfigured service types can be dragged and dropped from the left box to the cockpit table. Likewise, users can drag already configured service types to the list of unconfigured service types. Being used in all modern operating systems [Spe05], we assume that all users are familiar with drag and drop actions. Signaling users the consequences of drag and drop activities [Tid05], the target area is highlighted with a blue frame, once the user started dragging a service type.

If the user drags and drops an unconfigured service type on a particular service type in the dynamic table, the dropped service type assumes all settings from the targeted service type and is aligned and sorted along with all configured service types. The user can also drop a service

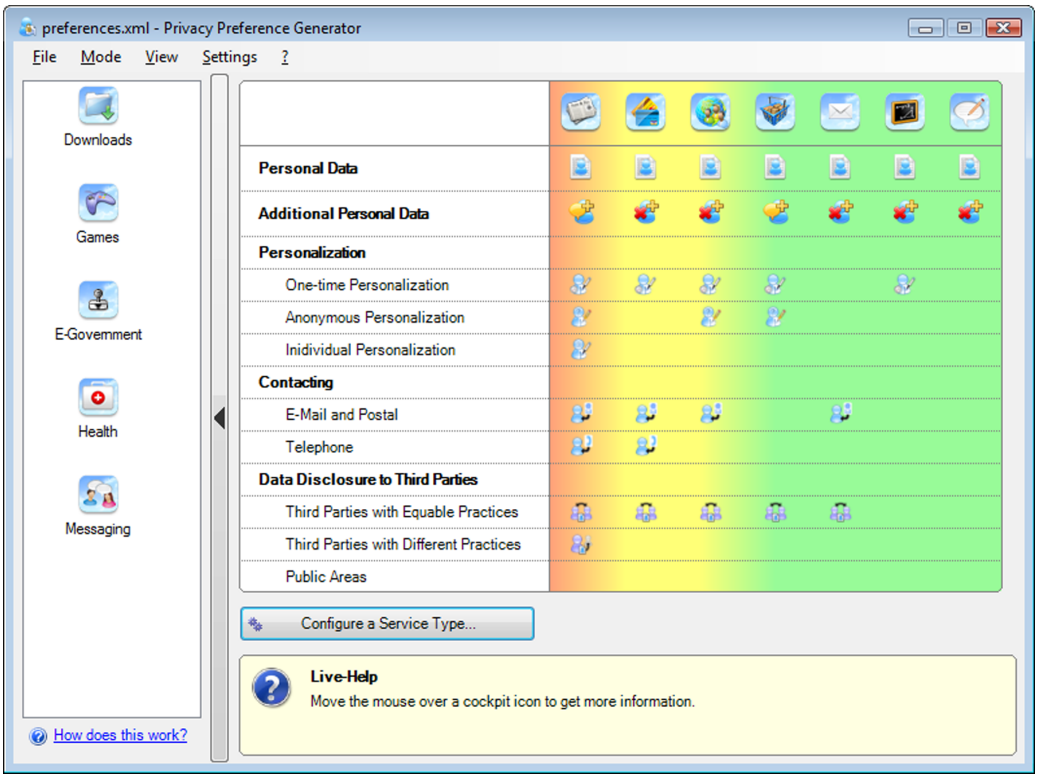


Figure 7.10: Privacy Cockpit with Activated Quick Mode

type on the right border of the table. This action assigns the highest privacy settings to that service type with no configuration options selected.

An activated quick mode also enables users to drag and drop already configured service types on each other. Again, the dropped service type assumes all configuration options.

7.4.8 Evaluation

After the initial identification of design challenges we developed an early prototype of the Privacy Preference Generator. The application included most features of the current version and underwent a user test (Chapter 5.3.3) that especially emphasized usability.

After the initial start of the tool and the selection of a mode, test persons were directly guided to the cockpit (Figure 7.11). Similar to the final version of the privacy cockpit, the cockpit page of the early prototype evaluated the generated privacy preferences and grouped configured service types into the privacy columns *Very Good*, *Good*, *Average* and *Poor*. This terminology confused users, as it was unclear, whether the column headings referred to the level of privacy or the disclosure of personal data itself. Furthermore, the use of color did not support the meaning of each group. We solved that issue by discontinuing the use of static groups in favor of a dynamic color scale in the background.

Furthermore, the red check marks at the bottom of each group were criticized, as they failed to

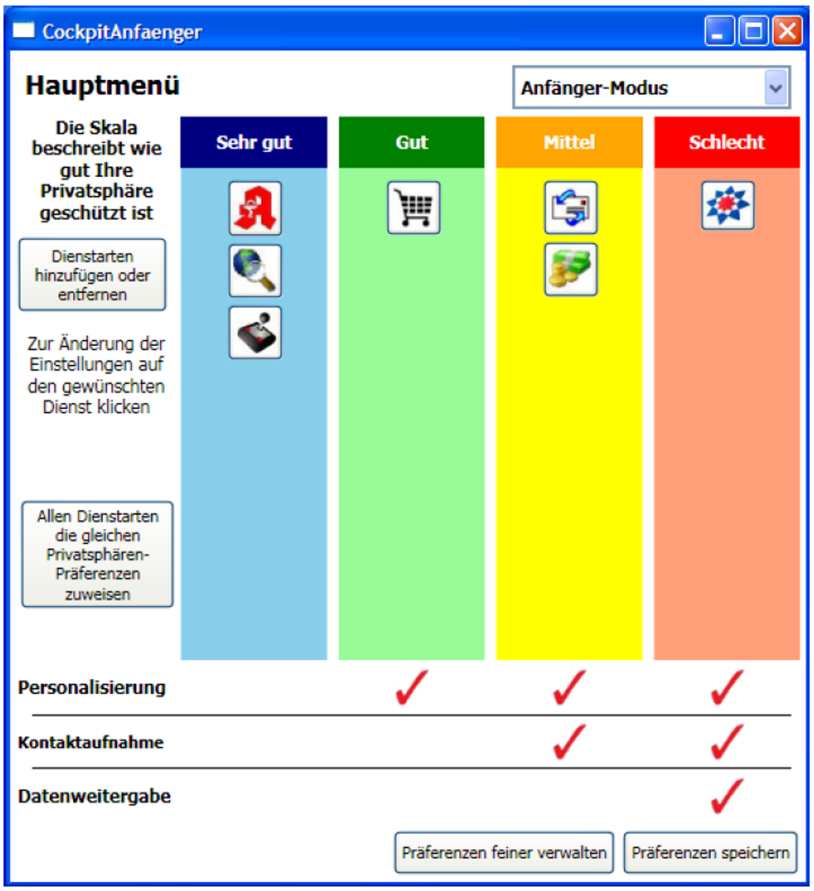


Figure 7.11: Privacy Cockpit of the Early Prototype (German User Interface)

reflect the chosen settings in an understandable way. Additional deficits of the privacy cockpit included unrecognizable buttons, missing drag and drop functionality of service types as well as a non-intuitive mode change.

The test outcome of the configuration wizard triggered improvements of the explaining texts as well as the supportive icons. We learned that test persons primarily focused on the text, when they were about to make a selection. The icons only had supportive character, but were highly appreciated by all users. Furthermore, we discovered that simple, little space consuming icons are favored over detailed, more complex pictures.

Moreover, we recognized the need of a recommended data set for each service type, as inexperienced users were unable to determine an appropriate amount of personal data types in an acceptable amount of time.

Furthermore, a fundamental finding of the first user test was the need of a uniform design for all complexity modes. The early prototype applied different design patterns for each mode, which caused confusion among test persons when the chosen mode was changed.

In spite of the existing deficits, however, all test persons agreed that the definition of privacy preferences with the tested tool offered significantly more user-friendliness than the priorly

tested Privacy Bird. Also, the concept of a configuration wizard along with a privacy cockpit was widely appreciated. Our initial idea to determine the appropriate mode for a user by using Westin's privacy attitude categorization (see Chapter 3.5.2) proved to be not applicable.

The findings of the first test served as valuable input for the development of the final solution, which is described and illustrated in the previous sections. In an effort to unveil final weaknesses, we evaluated our proposed tool within the scope of our second user test (Chapter 5.3.4).

After an initial tutorial, the first task included the definition of privacy preferences for the service type *Shopping*. As described earlier, test persons had to choose a complexity mode when the application initially started. 19 out of 26 test persons chose the basic mode, while six decided to start with the advanced mode. One person chose the at that time available expert mode. Many test persons expressed their confusion about three different complexity modes. This user feedback led us to merge the configuration settings of the advanced mode and the expert mode. Offering a basic mode and an advanced mode, we still provide a multi-level user interface that caters the needs of inexperienced users and privacy experts. In addition, we provide an opt-in/opt-out menu option for experts who want to fine-tune their advanced settings.

Most test persons praised the layout and the explanations of the configuration wizard. Some test persons had difficulties in understanding the personalization settings. For this reason, we prepared clearer explanations and redesigned the supportive icons. The help texts were broadly recognized and utilized for the understanding of the configuration options. Consequently, all test persons successfully completed the configuration with no difficulties.

For the second task, test persons were asked to make particular changes to the pre-configured service type *Health*. In order to solve this task, we asked test persons to switch into the advanced mode. Most test persons changed the mode smoothly. Only two persons needed time to find the proper menu item, but did not need any additional hints.

The basic principle to move data types on the disclosed data page was instantly recognized and posed no observable hurdle. Only a few test persons had difficulties finding the right data types in the tree, which is mainly caused by the partially assumed P3P data type definition. The adjustment of the remaining settings in the configuration wizard, again, was completed smoothly.

The third task focused on the privacy cockpit. Test persons were asked to interpret the pre-configured service type *Banking*. The findings of this task underscored the understandability of the privacy cockpit. All test persons correctly interpreted the settings. Only a few asked for a clearer highlighting of the live help area. Addressing voiced hints of some test persons, we added grid lines to the privacy cockpit to further improve the quick comprehension of the service type configurations.

In the interview section we obtained valuable feedback that allowed us to improve the design of the icons. Furthermore, we learned that most users agreed that the tool as a whole was understandable and easy-to-use. During the discussions we recognized that both privacy experts and privacy novices realized the purpose of the configuration they made, even though many test

persons have not been confronted with privacy-enhancing technologies before the test. 21 test persons agreed that the tested tool facilitates them to exactly define their privacy preferences. Only a few persons expressed their concern, whether machine-readable preferences could exactly reflect their disclosure behavior. Finally, all test persons unanimously believed that the Privacy Preference Generator along with the Privacy Agent has the potential to become a valuable tool for the protection of personal data.

7.5 Additional Approaches

The previous section introduces our proposed solution for a user-friendly Privacy Preference Generator component. During the course of development, we considered the potential of additional approaches for the usable generation of privacy preferences, which are theoretically discussed in the following.

7.5.1 Creating Privacy Preference Templates

Seeking means to provide users with useful privacy preferences in an acceptable amount of time, we also analyzed the potential of offering privacy preference templates to users.

Within the scope of our first user test (Chapter 5.3.3) each test person was asked to configure privacy preferences as accurately as possible. Using Westin's classification (Chapter 3.5.2), we also determined the privacy attitude of test persons, looking for similarities of privacy preferences within each user group.

Analyzing the results of the test, we learned that the preferences of each user group did not match significantly and did not permit the forging of reliable sets of privacy preferences for each group. The test sample as a whole, however, showed that test persons almost unanimously did not accept to be contacted via phone. Furthermore, we observed that users generally disapproved the disclosure of their personal data to third parties.

Apart from the test outcome, a disadvantage of privacy preference templates is the low level of observed user acceptance. A template of privacy preferences does not contribute to understandability and, as a consequence, does not gain users' trust, which is supported by the second user test we conducted (Chapter 5.3.4). Being asked about their opinion, the majority of test persons did not trust a preference template and favored a user-friendly configuration tool over a timesaving template. A few users, however, would use a template as a starting point for the definition of privacy preferences.

Considering these findings, we conclude that privacy preferences of users are too heterogeneous and too individual to build significant clusters. Templates could – at the most – serve as basic preferences at the beginning of the configuration process. In our proposed solution we offer recommended data sets and a quick mode that enable users to define privacy preferences in a timesaving and accurate manner.

The proposed privacy community takes up the potential of shared privacy preferences and

enables inexperienced users to adopt privacy preferences of more experienced friends (see Chapter 6.3.7). Our second user test proved that provided preferences that originated from a familiar person substantially boosted the acceptance of these preferences. The vast majority of test persons agreed that they would consider using privacy preferences specified by a known privacy expert.

7.5.2 Process-based Privacy Preference Generation

Instead of following a service-based definition, we also analyzed the potential of configuring privacy preferences for individual Internet processes, such as *Registration* or *Purchase* (see also Chapter 6.3.2).

This approach enables users to define privacy preferences explicitly for each process. Like its service type-based counterpart, a corresponding privacy agent identifies the Internet process being used and matches it with the privacy policy of the visited service provider.

However, a severe shortcoming of the process-based approach is the lacking comparability of equal processes. Using a configuration tool as introduced in this chapter, it is not possible to define general privacy preferences for a given Internet process. The process *Registration* of a *Shopping* service type, for example, requires a lot more personal data than the same process of a *Messaging* service type. This fact already represented the driving force of our proposed service type-based approach that takes into account the goal-oriented actions of Internet users.

A further difficulty of the process-based approach is the fact that a service provider's service offer generally consists of several processes, while in most cases only one privacy policy is published. Even though the P3P standard facilitates separate policies for each Web page of the provider, it seems barely possible to divide information of one or many privacy policies into individual processes without extensive support of a privacy community.

7.5.3 Building Privacy Preferences from a Transaction History

Within the local Data Disclosure Log component a transaction logging application records all personal data transfers in the Web browser (see Chapter 9.3). This transaction history facilitates the identification of patterns in the disclosure behavior of the user.

In an effort to gain privacy preferences for each service type, the implementation of this approach required the addition of the Internet service type to each logged transaction of the parsed transaction log file. In order to generate meaningful disclosure rules, our analysis also required the aggregation of transactions belonging to the same service provider.

We merely focused on disclosure patterns of personal data types. Personalization and contacting configurations as well as disclosures to third parties could equally be determined, if a sufficient amount of P3P privacy policies is logged. In the following we analyze three identification approaches.

An algebraic approach allows the identification of set unions and intersections of past data

disclosures [FKPT07]. Built set unions collect all personal data types that have been transferred to service providers of a particular service type. As expected, set unions result in extremely permissive disclosure rules, as the transfer of a particular data type to a single service provider leads to an inclusion in the resulting disclosure rule. On the contrary, intersections represent the opposite extreme, as an inclusion of a data type into the resulting rule requires the transfer to each service provider of a given service type. The built disclosure rules show a high level of confidence. However, in real life scenarios it seems unlikely that the analysis of a transaction log file results in useful intersections, as the number of transactions rises. Hence, the calculated disclosure rules of both procedures show little practical value. Both the set union and the intersection are considerably susceptible to outliers. The elimination of outliers inevitably requires substantial user interactions, which outweighs the advantages of this automatic approach.

Targeting a solution that reduces the impact of outliers, we employed a threshold-based approach, which required the generation of personal data disclosure vectors for each service provider. Each dimension represents the disclosure (1) or non-disclosure (0) of a particular data type to a service provider. The following vectors list personal data types for three online shopping providers:

$$Amazon = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}; Ebay = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}; BestBuy = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

In order to account for the relevance of each service provider for the considered service type, we incorporate the transaction frequency by weighing each vector with the number of transactions. For our presented example, we assume the user executed six transactions with Amazon, four with Ebay and two with Best Buy, resulting in a relative weight of 0.5, 0.33 and 0.17, respectively. The following vectors show the service provider vectors multiplied with the relative weights:

$$Amazon = \begin{pmatrix} 0.5 \\ 0 \\ 0.5 \\ 0.5 \\ 0 \end{pmatrix}; Ebay = \begin{pmatrix} 0.33 \\ 0 \\ 0.33 \\ 0 \\ 0 \end{pmatrix}; BestBuy = \begin{pmatrix} 0.17 \\ 0.17 \\ 0 \\ 0.17 \\ 0 \end{pmatrix}$$

In a final step all vectors are summed up and compared to pre-defined thresholds. If a threshold value is reached or exceeded the particular data type is included in the disclosure rule. For our example, the first and third data type exceeds the defined thresholds.

$$Threshold = \begin{pmatrix} 0.5 \\ 0.4 \\ 0.5 \\ 0.8 \\ 0.5 \end{pmatrix}; \Sigma = \begin{pmatrix} 1 \\ 0.17 \\ 0.83 \\ 0.67 \\ 0 \end{pmatrix}$$

The defined thresholds reflect the sensitivity of a particular data type and express the required level of confidence for a data type to get included in the disclosure rules. The high threshold of the fourth data type, for instance, indicates a sensitive data type like a credit card number.

The presented threshold-based approach represents a suitable solution for the automatic creation of privacy preferences. Unlike set unions and intersections, outliers do not bias the outcome noticeably. Furthermore, the incorporated transaction frequency of service provider results in meaningful and practical disclosure rules. An implemented Java-based demonstration proved the applicability of this approach. However, the determination of appropriate threshold values remains a challenging task and requires careful considerations. Existing user surveys as well as individual user preferences could help determine suitable values.

Finally, we analyzed the potential of data mining techniques for the automatic creation of privacy preferences [WF05]. The given task requires unsupervised learning algorithms for the discovery of patterns within the user's transaction history. Existing clustering algorithms are capable of identifying similar disclosures within a particular service type. However, users need to select the most appropriate clusters for the generation of disclosure rules. From a usability perspective, inexperienced users are not able to understand the clustering procedure, resulting in a low acceptance and adoption of the built preferences.

On balance, compared to algebraic and data mining approaches the threshold-based approach represents the most promising way to automatically create privacy preferences from a transaction log. For the definition of threshold values intense interpretation of user surveys is required. A weakness all considered approaches have in common is the required accuracy of the transaction log file. This remaining uncertainty and the inferior user acceptance of automatic approaches led us to offer a user-friendly manual configuration tool for our proposed privacy infrastructure.

7.6 Summary

The Privacy Preference Generator component is responsible for the understandable capture of disclosure rules, which are required for an automatic evaluation of privacy policies.

The chapter starts with a definition of privacy preferences and highlights decision criteria for the disclosure of personal data. We discuss design challenges for the development of a suitable tool that generates privacy preferences, before we identify usability deficits of related work.

We introduce a user-friendly tool that captures privacy preferences for up to twelve Internet service types, allowing for more precise and practical disclosure rules. Two complexity modes

tailor the tool for inexperienced users and privacy experts. Building on the vocabulary of the P3P specification, a wizard guides users through a set of configuration options. The wizard shows clear explanations and recommends an adequate personal data set for each service type. The privacy cockpit provides an understandable overview and evaluation of the generated privacy preferences. Two conducted user tests helped us assess usability and user acceptance.

Finally, we discuss further approaches we considered for the definition of privacy preferences. Privacy preference templates can serve as basis for the configuration process, but cannot replace the manual configuration. We showed that the generation of privacy preferences for Internet processes is not applicable and inferior to the presented service type-based solution. Analyzing automatic generation procedures from a transaction log file, only the threshold-based approach delivers reasonable disclosure rules. Due to the fuzzy results and the low user acceptance, automatic approaches do not meet the requirements of a usable configuration tool.

Chapter 8

Privacy Agent

The Privacy Agent component is designed to assist users in protecting actual data disclosures. Specifically, the component provides information that supports users in making a more informed disclosure decision.

This chapter starts with a requirements analysis of a suitable privacy agent in the Web browser. After the identification of usability deficits of existing solutions, we introduce an integrating browser extension that provides contextual privacy-related service provider information, combining resources of the proposed privacy architecture. In particular, this information involves collaborative community data that are relevant for actual data disclosures. The browser extension also matches service providers' privacy policies with privacy preferences of the user and signals the resulting disclosure recommendation to the user. Finally, information about linkable partial identities and disclosed personal data is offered.

8.1 Requirements

One of the biggest privacy-related challenges users are facing is the decision whether or not to release personal data to an unknown service provider. In order to support the predominantly inexperienced users, the primary goal of the targeted Privacy Agent should be the provision of an improved basis for decision-making. Due to the characteristics of the Privacy Agent, it should be designed and implemented as a browser extension and present service provider information in the context of the visited Web site.

In Chapter 4 we highlighted that the acceptance of privacy-enhancing technologies requires the design of usable interfaces. Patrick et al. define 14 design features for the development of a privacy agent, which are partially applicable to the context of the targeted Privacy Agent [PKHvB03]. Specifically, the authors recommend a consistent visual layout, a functional arrangement of the used elements and a structured design that provides both a clear overview and the ability to focus individual components. Furthermore, a usable solution should protect personal data non-intrusively [BS93]. Also, an agent should not burden the user with additional tasks [CG05].

The proposed privacy architecture features a privacy community that stores collaboratively maintained service provider information (see Chapter 6). These data allow users to get an overview as well as comprehensive reputational information about a visited service provider. Hence, the context-based Privacy Agent should utilize this open source of privacy-related information and display static provider information as well as privacy ratings and individual user postings. A contextual browser agent should also allow users to rate the privacy policy as well as the policy adherence of a visited service provider, enabling users to quickly edit collaborative content.

The privacy community also maintains the amount of personal data required for the fulfillment of a process. Knowing this information in advance represents a valuable information source that impacts the disclosure decision and saves users valuable time. Hence, the Privacy Agent within our proposed privacy infrastructure should look up and present this information.

As the Privacy Agent is intended to provide a quick access to the privacy community, a login option should be available for registered community members. In this context, the Privacy Agent should facilitate the upload of privacy preferences to the community. Likewise, the import of privacy preferences of connected users into the local Privacy Preference Generator should be offered without the need of accessing the community's Web front-end.

Known as the basic functionality of a privacy agent, the matching of machine-readable privacy policies with privacy preferences should be performed. As we generate separate preference sets for each pre-defined Internet service type (see Chapter 7.4), the Privacy Agent is required to determine the service type of a visited Web site. In order to foster understandability, the results of the matching process should be highlighted and presented clearly, including the source and a detailed explanation of detected conflicts [LG05]. As the adoption rate of P3P policies among service providers is relatively low (see Chapter 4.1.2), the Privacy Agent within our collaborative privacy architecture should resort to available community information, if no P3P policy is available.

An interview section of our second user test (see Chapter 5.3.4) revealed that many test persons were not aware of privacy threats that originate from potential linkabilities of partial identities (see Chapter 2.3). Hence, the Privacy Agent should visualize used partial identities of the user, including detected linkabilities.

With regard to the visited Web site, already disclosed personal data should be highlighted, utilizing the personal transaction log file generated by the Data Disclosure Log component (see Chapter 9.3). Such a disclosure history allows users to see at a glance which personal data were transferred to a visited service provider in the past.

As the Privacy Agent is placed at a central position on the user side, it should provide direct access to the remaining local privacy components as well as the community article of the visited Web site.

8.2 Usability Deficits of Existing Privacy Agents

In order to derive design goals for a suitable user interface, we analyzed and evaluated existing browser-based privacy agents. Similar to the previous chapter we concentrated on the well-known Privacy Bird (see Chapter 4.2.2.4), which represents the only available privacy agent that focuses on usability aspects.

We tested the Privacy Bird within the scope of our first user test (see Chapter 5.3.3). During the preparation of the test, we noticed the insufficient integration of the application in the Web browser. A new bird icon window is opened for each browser tab, which quickly fills the task bar of the operating system. Furthermore, it complicates the association of Privacy Bird instances with opened browser windows.

Unlike the user test conducted by Cranor et al. [CGA06], our test did not target a usability comparison of the Privacy Bird with the obviously inferior built-in privacy policy presentation of the Microsoft Internet Explorer or a full-text privacy policy. Focusing on understandability of the user interface, the outcome of our test showed that users tend to misinterpret the yellow bird icon, which indicates a Web site that does not provide a P3P policy. In comparison to the red icon, which signals a policy violation, the yellow icon pretends higher privacy protection, even though the data collection and data handling practices of the Web site are completely unknown and not evaluated by the tool. Cranor et al. mention this issue, but stick to the yellow bird icon, stating that a red icon for an unknown policy would discourage users, as the overwhelming majority of Web sites does not provide P3P policies [CGA06].

Moreover, the user test revealed usability weaknesses of the policy summary. Most test candidates criticized the design and the confusing structure of the window and had difficulties finding the proper information. As the policy summary mainly focuses on detected conflicts as well as a textual description of the provider's privacy policy, a relation to the individually defined privacy preferences is not evident to the user. Consequently, most test persons did not succeed in finding the sources of detected conflicts. Even though the design of the tool targeted inexperienced users, we observed that most candidates had difficulties understanding the used privacy language. A reason for that outcome might be that most test candidates' native language was not English (see also [Pet04]).

In addition to the evaluation of the Privacy Bird, we also considered design weaknesses of existing policy viewers that are seeking to translate P3P policies in an explaining textual form (see Chapter 4.2.2.1).

8.3 Design of An Integrating Privacy Agent

In the following we introduce the design of the browser-based Privacy Agent within the collaborative privacy infrastructure. The section starts with a technical overview of the presented tool, before we describe the design and the functionality of the provided tabs. Chapter 10.3.3 lays out additional details of the prototypical implementation.

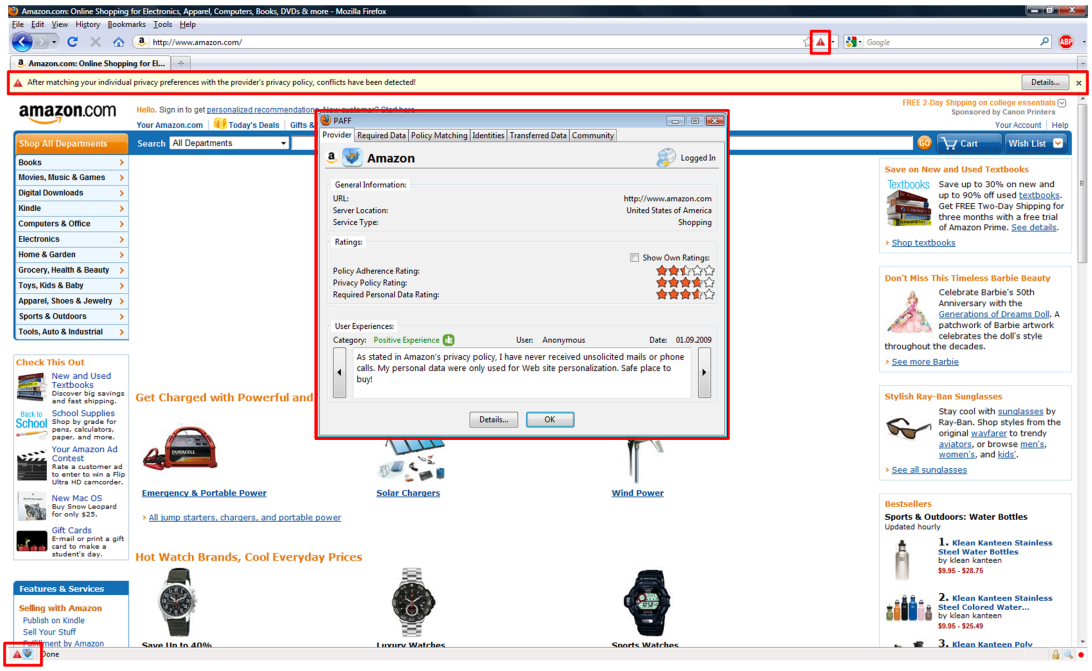


Figure 8.1: Embedded Privacy Agent in the Web Browser

8.3.1 Overview

As defined in our proposed privacy architecture, the Privacy Agent should provide users with information, which is relevant for the disclosure of personal data. Hence, the tool should utilize all available resources of the underlying privacy infrastructure. As mentioned earlier, a suitable solution needs to be integrated in the Web browser.

Recent Web browser usage statistics[1] as well as the powerful environment for the development of extensions led us to choose Mozilla Firefox[2] as host browser for the presented Privacy Agent.

The Firefox framework provides a comprehensive tool set for the development of complex browser-based applications. The integration of the XML User Interface Language (XUL)[3] facilitates the simple development of application interfaces. As our efforts emphasize usability aspects, this characteristic is of primary importance. The application logic is implemented using the script language JavaScript [Fla06]. The Firefox framework closely connects XUL and JavaScript, which results in a clear separation of user interface design and application logic [Pro07].

For the realization of the application design we followed the XUL Accessibility Guidelines[4] recommended by the Mozilla Developer Center.

Figure 8.1 depicts a screenshot of the proposed Privacy Agent, embedded in the Firefox browser.

[1] http://marketshare.hitslink.com/browser-market-share.aspx
[2] http://www.mozilla.com/firefox/
[3] http://planetxul.com/
[4] https://developer.mozilla.org/En/XUL_accessibility_guidelines

By default, the Privacy Agent operates in the background and does not directly affect the surfing behavior of the user. In this state only three small icons indicate the presence of the application.

The left side of the status bar contains a status icon that reflects the recommendation of the matched privacy policy. The same icon is placed at the right side of the address line. A conducted user test did not reveal a preferred position of the status icon (see Section 8.3.7), which led us to offer both alternative icons. An additional icon in the status bar presents the detected service type of the visited service provider.

A left-click on either status icon opens the main window of the introduced Privacy Agent. We point out that the user needs to actively open the main window. A more intrusive automatically triggered pop-up window is avoided.

Considering the collected requirements, we carefully selected and aggregated relevant information and functionality of the main window. The resulting six tabs are presented and discussed in the following sections.

8.3.2 Collaborative Service Provider Information

The initial two tabs of the designed Privacy Agent show selected privacy community data of the visited Web site. The displayed information was chosen with regard to their relevance to actual data disclosures. The service provider information is exclusively queried from the privacy community, utilizing the offered Web service interfaces.

When the user opens the Privacy Agent, the first tab lists general service provider information of the community (see Figure 8.2). The main purpose of this tab is to convey a general impression of an unknown service provider.

The tab header shows the Favicon, the name and the applicable service type icon of the service provider. The right side of the header indicates the login state of the privacy community. The tab header remains unchanged in the following tabs of the Privacy Agent, allowing users to relate the remaining information to the visited service provider.

The top segment of the first tab lists the provider's URL, the physical location of the server as well as a textual presentation of the service type. The presented information is part of the static provider information maintained by the privacy community (see Chapter 6.3.1) and displays users an approximate service provider classification.

The following segment is dedicated to service provider ratings. Resembling the design of the community's Web front-end, each rating is displayed in the form of an orange star rating. In particular, the policy adherence rating, the policy rating of the effective privacy policy and the data usage rating (automated evaluation of the required amount of personal data) are presented to the user (see Chapters 6.3.5, 6.3.4 and 6.3.2, respectively).

Easing the active participation in the privacy community, we offer users the possibility to rate the policy adherence as well as the effective privacy policy by clicking the respective amount of stars. With respect to the user's login state the rating is submitted to the community anonymously or is associated to the logged-in user. Furthermore, logged-in users have the

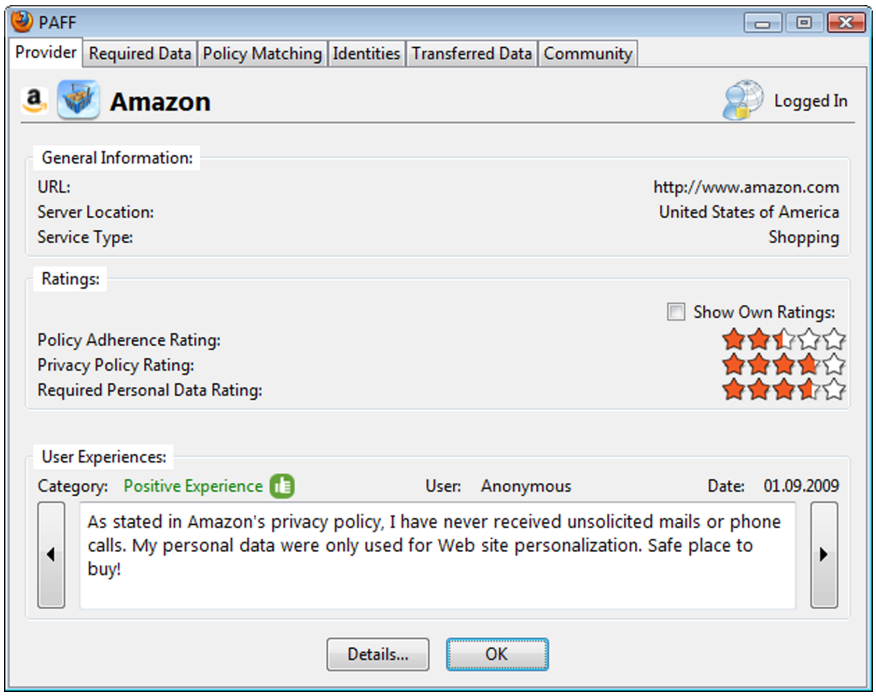


Figure 8.2: Privacy Agent – Service Provider Information, Ratings and Postings

option to see their own rating, if the respective box is checked. Own ratings are represented by yellow stars.

The last segment presents community postings of individual user experiences (see Chapter 6.3.6). While the most recent entry is presented first, the user can browse the chronologically sorted postings using the arrow buttons at the left and right side of the text box. Additionally, the segment highlights the category (positive or negative) as well as the user name (if not anonymously submitted) and the date of each posting.

As proved by numerous reviewing systems, ratings along with individual user experiences significantly impact the behavior of users and, hence, provide valuable reputational information about service providers in the context of the Privacy Agent.

If the user wishes to know additional privacy-related information about a presented service provider, the Details button at the bottom of the tab opens the community article of the visited service provider in a new browser tab.

The second tab of the designed Privacy Agent depicts the service provider's offered processes along with the required amount of personal data per process, which is also maintained by the privacy community (see Chapter 6.3.2). The design of the second tab is depicted in Figure 8.3.

Available processes are represented by green arrows, corresponding to the design of the respective article tab in the privacy community. When the user clicks a process, a box unfolds that shows the required amount of personal data. For the sake of clearness, we aggregated certain groups of personal data types, like *Address* and *Payment Information* (see Chapter 10.2.2). If the mouse is moved over a data element, a tool tip visualizes the individual data types

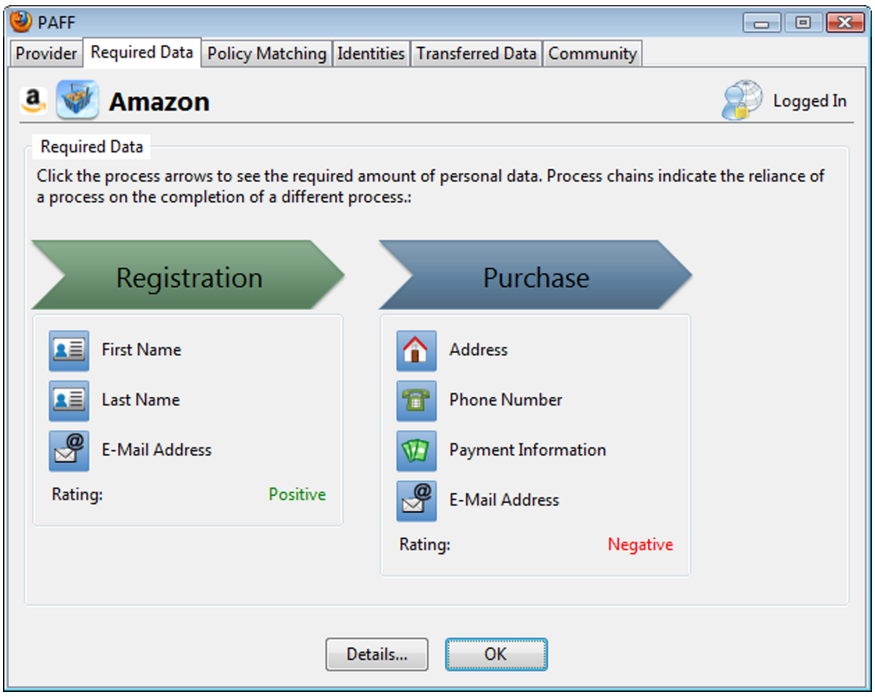


Figure 8.3: Privacy Agent – Required Amount of Personal Data

of an aggregated element. Furthermore, each personal data element belongs to a data group, which is indicated by the corresponding icon. The icons allow for a quick comprehension of the required types of personal data. In addition to the amount of personal data, the tab also presents the community evaluation of the personal data amount with regard to the offered process. This automated rating is shown below the list of personal data elements.

Similar to the privacy community, the arrangement of the process arrows resembles process chains, which indicate the required completion of a process before the execution of a depending process. In Figure 8.3, for instance, the process *Purchase* is selected that requires the completion of the process *Registration*. A selected process is colored blue, while processes that are not part of the selected process chain are faded out.

Again, a click on the Details button opens the community article of the visited service provider and allows users to directly edit service provider information.

The second tab of the Privacy Agent proves the potential of open information about required personal data. Integrated into a browser extension, this information facilitates users to quickly see and weigh the demanded personal data against the desired purpose. An automatic evaluation further extends the basis for decision-making.

8.3.3 Matching Privacy Policies With Privacy Preferences

The third tab of the introduced Privacy Agent is devoted to the evaluation of published privacy policies. If a P3P policy of a service provider is available, it is matched with individual privacy

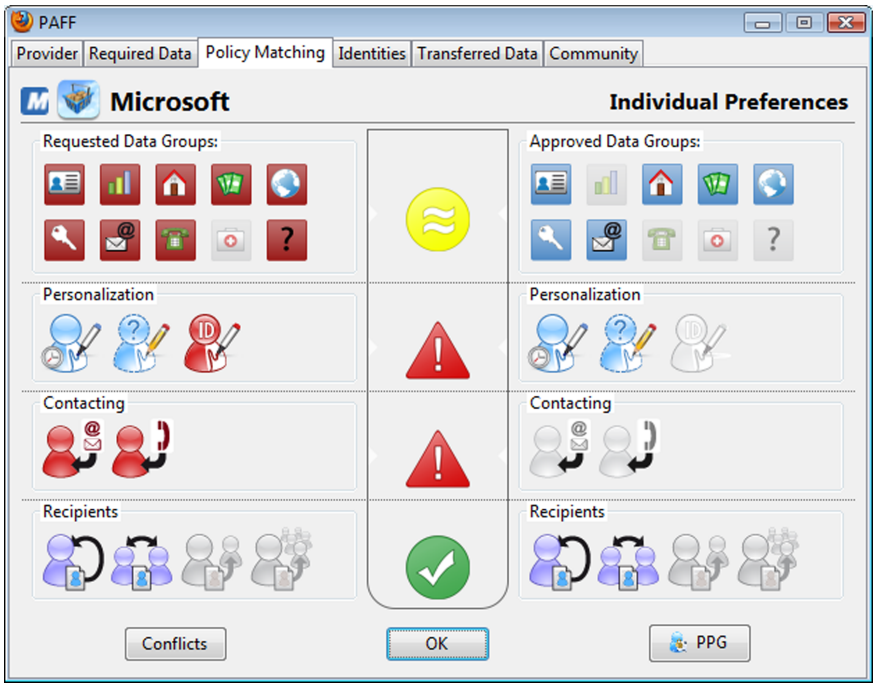


Figure 8.4: Privacy Agent – P3P Policy Matching

preferences of the user (see Chapter 7.4). Based on the service type of the visited service provider, the respective set of privacy preferences is automatically applied. The presented tab calculates a resulting disclosure recommendation, which is signaled to the user.

The design process of the presented tab targeted a user interface that confronts contents of the service provider's privacy policy with configured privacy preferences of the user. Such an at-a-glance overview facilitates users to quickly understand the outcome of the matching process as well as the sources of detected conflicts. Figure 8.4 shows the result of our design efforts.

The left side of the tab visualizes the content of the processed P3P policy, which is contrasted with the privacy preferences of the user on the right side. Horizontally, the tab is divided into four segments that correspond to the configuration categories of the Privacy Preference Generator. A column in the middle highlights the results of the matching process for each category.

The first category focuses on personal data the service provider requests in its privacy policy. Guaranteeing a consistent user experience, we utilized the aggregated data groups along with their icons, which are centrally administered in the privacy community (see Chapter 10.2.2). A blue colored data group icon on the provider side indicates that at least one data type of that data group is requested by the service provider. Consequently, grayed-out icons symbolize that no data types of that data group are listed in the provider's privacy policy. A tool tip triggered by a mouse-over event shows the full name of an icon group.

Representing the privacy preference configuration for the respective service type, the state of

the data group icons on the right side signals the user's willingness to disclose data types of that group to a service provider. In particular, a blue data group icon indicates that at least one data type of that group is approved for disclosure.

If the privacy policy of the service provider asks for personal data that are not approved within the privacy preferences of the user, the affected data group icons on the provider side are replaced by red icons that allow users to instantly recognize the sources of identified conflicts. As a blue data group icon within the privacy preference section does not necessarily represent the approval of all data types of that group, it does not prevent the appearance of a conflict in the same data group of the left section.

The second and third horizontal categories visualize the service provider's personalization and contacting purposes for collecting the specified personal data (see Chapter 7.4.4). The final category is devoted to the personal data disclosure to third parties (see Chapter 7.4.5). Fostering user acceptance, we adopted the purpose and third party aggregations as well as their supportive icons from the corresponding configuration stages of the Privacy Preference Generator. Again, requested/approved elements are represented by blue icons, while non-requested/approved elements are grayed-out. Conflicts are highlighted by red icons on the provider side. Optional `required` conditions are considered for the matching of each element (see Chapter 7.4.4).

The middle column underscores the result of the matching process in each category. A check box in a green circle symbolizes that no conflict is detected in the respective category. In contrast, an exclamation point in a red triangle indicates the presence of a conflict. Alternatively, the first category shows wave lines in a yellow circle, if a conflict was detected and if the user chose to accept the release of additional personal data, but to be specifically warned and informed about that practice (see Chapter 7.4.3). Finally, a gray question mark is presented, if either no privacy policy information is found for the category or the user did not configure privacy preferences for the applicable service type.

The Conflicts button at the bottom of the tab opens a pop-up window that presents a detailed list of detected conflicts, including a detailed textual explanation for each category. Clicking the PPG button opens the Privacy Preference Generator and allows users to review and change privacy preferences on the spot.

Providing users with a visible notification about detected conflicts, an aggregated conflict status is calculated and presented to the user in both the status bar and the address bar (see Section 8.3.1). The status icon utilizes the icons of the middle column and adopts the most severe conflict status of the four categories. Furthermore, a notification box at the top of the browser window signals detected conflicts and offers users to open the main window (see Figure 8.1). Again, we point out that by default, the user is not interrupted in his browsing behavior and is only discreetly but noticeably hinted at potential conflicts. If the user opts to receive more information, the main window of the Privacy Agent can be opened at any time.

While the presented tab offers a user-friendly tool for the utilization of privacy preferences, the lagging acceptance as well as the predominantly low quality of P3P policies limits the practical use of the presented functionality. For this reason, our solution captures the required amount of

Figure 8.5: Privacy Agent – Matching Privacy Preferences with Community Data

personal data from the privacy community (see Chapter 6.3.2), if no P3P policy of the service provider is available or if the offered P3P policy does not provide sufficient information. In doing so, the presented Privacy Agent accumulates the amount of personal data for all offered processes and utilizes that information for the evaluation of the data group category. Figure 8.5 shows a screenshot of that scenario, which further proves the potential of a collaborative privacy infrastructure.

In the future, the privacy community could also maintain purpose and third party information of privacy policies, allowing privacy experts to capture that information from the textual privacy policies of service providers and enter it in the privacy community. Such an extended privacy community would result in a totally provider-independent evaluation of privacy policies.

8.3.4 Usage and Linkability of Partial Identities

In Chapter 2.3 we discuss the possibility of linkable partial identities in the World Wide Web. Addressing this privacy threat, the fourth tab of the Privacy Agent provides users with an illustration of used partial identities and calculates their potential linkabilities.

In order to detect used partial identities, the individual disclosure history of the user is processed, which is built by the Data Disclosure Log component (see Chapter 9.3). A partial identity is regarded as a set of data types that involves a group of transactions. If transactions with a single service provider show contradicting data types, multiple partial identities are built. Identified partial identities are displayed to the user along with their linkabilities. In addition, partial identities are aggregated into non-linkable identity classes.

The direct linkability between two partial identities is determined by comparing non-recurring data type values. These unique values occur with certain data types. If two partial identities possess identical values of such a data type, the linkability between those partial identities can be identified reliably. Applied to the k-unlinkability model [Mal08], k equals 1, which clearly indicates a linkability.

Unique data type values occur, for instance, for the data type *E-mail Address*, as a corresponding value originates from a single person. A value of the data type *Zip Code*, on the other hand, can be led back to many persons and cannot be applied to link partial identities. In some cases, the classification of recurring and non-recurring data types is not clear. For instance, a phone number, which per se is regarded as a unique value, could be shared by several persons in a household. Furthermore, certain data types are only unique in combination with another data type. An example for a non-recurring data type pair is a bank account number in combination with a routing number. Table 8.1 shows a list of unique data types and data type pairs the presented Privacy Agent considers for the calculation of direct linkability between two partial identities.

Data Types with Non-recurring Values
E-mail Address
Phone Number (landline)
Phone Number (mobile)
Credit Card Number
Instant Messaging ID
Domain
Social Security Number
Data Type Pairs with Non-recurring Values
Routing Number AND Account Number
Full Name AND Address

Table 8.1: Data Types and Data Type Pairs with Unique Values

After the calculation of direct linkabilities, the presented solution aggregates partial identities by computing transitive linkabilities. In this process we assume that the knowledge of an attacker is unknown. For this reason, all disclosed partial identities are considered for the identification of transitive linkabilities.

Transitivity can lead to the linkability of two unrelated partial identities, if a third partial identity is considered. We follow the model of Steinbrecher and Köpsell [SK03] and group transitively linkable partial identities into equivalence classes, which are defined by a transitive and reflexive relation (see Chapter 2.3). In the context of this work, we refer to these linkable identity groups as identity classes.

The calculated identity classes along with their corresponding partial identities are presented to the user and allow the recognition of transitively linkable partial identities as well as the limits of linkability contexts (see Figure 8.6). In the presented tab the user has the option to choose individual names for identity classes and partial identities by clicking the respective text boxes.

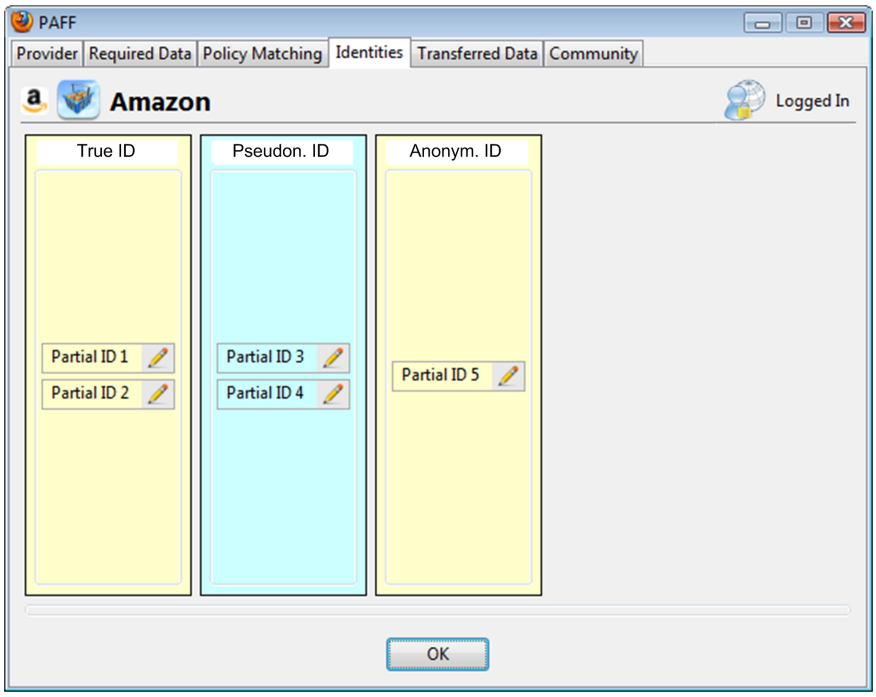


Figure 8.6: Privacy Agent – Used Partial Identities

Furthermore, each displayed partial identity provides a button, which opens a window with information about the involved data types and direct linkabilities to other partial identities.

The tab creates and maintains a personal identity file, which includes information about the detected identity classes, partial identities and their linkabilities. The file is used by the Data Disclosure Log component for the visualization of past transactions (see Chapter 9.4).

Building on the potential of the collaborative privacy community, a future extension of the presented functionality could revise the assumption that the knowledge of an attacker is unknown. Instead, the tab could access the privacy community and query third parties a service provider cooperates with (see Chapter 6.3.3). This knowledge would allow the generation of more precise identity classes that account for the potential information sources of service providers.

8.3.5 Information about Already Disclosed Personal Data

The fifth tab offers partial functionality of the Data Disclosure Log component by providing a short overview of personal data already disclosed to a visited service provider. For this purpose, the tab reads the transaction log file of the user (see Chapter 9.3) and lists all transferred data types for each used partial identity.

Figure 8.7 shows the design of the developed tab. The left box at the top of the tab presents the number of personal data transactions with the visited service provider as well as date and process type of the most recent transaction. Below, a box visualizes direct linkabilities between

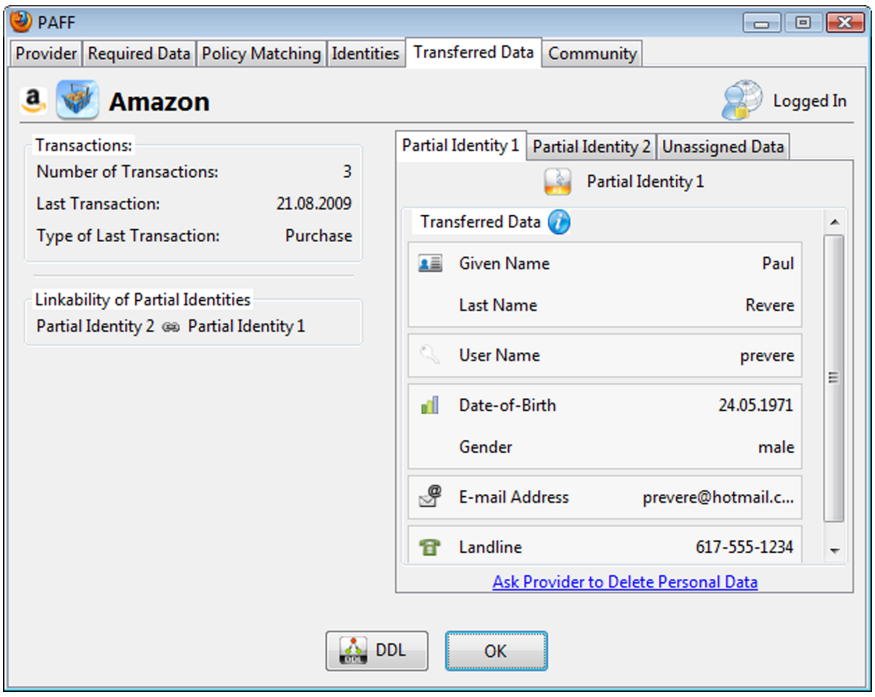


Figure 8.7: Privacy Agent – Transferred Personal Data

used partial identities. Tabs on the right side show disclosed data types of each partial identity used for the interaction with the visited service provider. Similar to the policy matching tab, data types are aggregated into data groups, which are supported by their respective icons (see Chapter 10.2.2). Each partial identity tab offers a link that facilitates the removal of disclosed personal data. The link opens a new message in the standard e-mail client and provides a text that asks the service provider to delete the disclosed personal data of the used partial identity. The service provider's e-mail address for privacy issues is automatically requested from the privacy community (see Chapter 6.3.1) and placed in the address line of the generated message. This functionality should encourage users to more frequently take advantage of their right to request a removal of already disclosed personal data by alleviating the burden of such a request. Finally, the DDL button at the bottom of the tab opens the Data Disclosure Log visualization tool that offers a detailed presentation of past transactions (see Chapter 9.4).

This tab in combination with the identity tab primarily addresses returning users and provides compact information about prior personal data transactions, allowing users to quickly estimate the amount and sensitivity of already disclosed data.

8.3.6 Privacy Community Login and Privacy Preferences Exchange

Finally, the last tab of the Privacy Agent allows registered privacy community members to login. The login state of the user is relevant for the submitted privacy ratings (see Section 8.3.2). After a successful login, the final tab presents a list of all connected members in the privacy community and enables users to download privacy preferences of a selected user

(see Chapter 6.3.7). The downloaded privacy preferences are automatically imported into the Privacy Preference Generator and applied for the following policy evaluations (see Section 8.3.3). Likewise, a button allows users to upload self-generated privacy preferences. Similar to prior tabs, a button at the bottom of the page opens the privacy community Web front-end in a new browser tab.

8.3.7 Evaluation

A user test (see Chapter 5.3.4) evaluated the policy matching functionality in the third tab of the Privacy Agent (see Section 8.3.3).

At the beginning, all test persons completed the test of the Privacy Preference Generator and became aware of the purpose and the content of privacy preferences. We provided a Privacy Agent installation with pre-loaded privacy preferences for the service type *Shopping*. Also, a short tutorial about the features of the Privacy Agent was presented.

In the first task, test candidates were asked to visit the eCommerce shop Tchibo[5] and to determine the result of the policy matching process. The vast majority of test persons found the green icon in the status bar or the address bar and interpreted it correctly. Only five of the 26 test persons had difficulties identifying the icons in the browser, but succeeded after approximately thirty seconds. As mentioned earlier, the test did not expose a favored position of the conflict status icon, which is why we place it at either position.

The following task involved a visit of a mock-up Web shop[6], whose privacy policy was not in accordance with the provided privacy preferences. After the completion of the first task, all test persons were able to find and interpret the red conflict status icon. Being asked about the sources of the conflict, 24 out of 26 persons instantly clicked the icon and were able to name the categories and elements of conflicts, reading the red-colored icons in the policy matching tab.

Evaluating the understandability of the user interface, we confronted test persons with the following questions:

- Does the service provider collect login data?
- Do the requested personal data include financial data?
- Does the service provider use your personal data for individual personalization?
- Who does the service provider share your personal data with?

Answering the questions, all test persons unanimously interpreted the icons and provided tool tips for each data group and purpose element correctly. Moreover, we observed that the consistent use of icons and terminology in the Privacy Agent and the Privacy Preference Generator contributed to the usability and intuitiveness of the user interface.

[5] http://www.tchibo.com/
[6] http://www-ifs.uni-regensburg.de/students/ps08/

In the following, we presented a compact view of the policy matching tab that visualized the privacy policy as well as the privacy preferences using only a single icon for each category. All but three users disliked the condensed representation of the tab and favored the more detailed view that provided a lot more information. These findings led us to discontinue the development of a compact view and to only offer a single, detailed view for the policy matching tab.

The interview section focused on the amount, form and content of the presented information. Again, most test persons agreed that the used icons and help texts are understandable and easy-to-read. Some test persons suggested additional help texts, which made us provide a conflict window with supplemental explanations about detected conflicts. On balance, 23 out of 26 persons agreed that the tab in combination with the Privacy Preference Generator facilitates a more informed disclosure decision and would use it in real-life scenarios. The main criticism of the remaining candidates involved the lacking enforcement of privacy policies, which cannot be directly related to the Privacy Agent itself.

A few test persons admitted that in conflict scenarios they might adjust their privacy preferences, in order to achieve a positive evaluation, which would ease their minds. These statements prove the discrepancy of pre-defined privacy preferences and observed user behavior (see Chapter 3.5.3) and underscore the need for tools that not only protect users from uncontrolled data disclosures, but also inform about the consequences of their actions.

8.4 Summary

The Privacy Agent component represents the central tool of the local privacy components and is built to support users in deciding about actual disclosures of personal data in the Web browser.

The chapter starts with a broad requirements analysis, which focuses on functionality and information a suitable implementation should provide. We continue with an identification of usability deficits of existing solutions and derive design goals. In the remainder of the chapter, we present our proposed solution for a browser-based Privacy Agent. The initial tabs provide reputational information about the visited service provider as well as the amount of required personal data per process, which is queried from the privacy community. The following tab presents a clear overview of the service provider's P3P privacy policy and the privacy preferences of the user. An understandable policy matching is performed, out of which a disclosure recommendation is calculated and signaled to the user. An additional tab identifies used partial identities, which are grouped with regard to potential transitive linkabilities. The remaining tabs are devoted to a quick overview of already disclosed personal data and the exchange of privacy preferences with connected users of the privacy community.

A conducted user test of the privacy policy matching proves the understandability and consistency of the designed user interface.

Chapter 9

Data Disclosure Log

Finally, the proposed privacy architecture offers a Data Disclosure Log component that enables users to review past information flows of personal data. The component gives users a clear overview of service providers that possess their personal data. This overview is a prerequisite for the enforcement of users' right to edit and revoke personal data once they have been transferred to a service provider.

The introduced component requires a logging tool that records data disclosures in the Web browser. A corresponding visualization tool illustrates these personal data transactions offering multiple views.

After an analysis of requirements of a suitable solution, this chapter evaluates the characteristics of existing approaches. In the remainder of the chapter, we introduce a tool that logs personal data disclosures in the Web browser as well as alternative solutions for the visualization of logged transactions.

9.1 Requirements

The online survey (see Chapter 5.3.4) we conducted confirms users' need for a powerful Data Disclosure Log component that analyzes and visualizes personal data disclosures in an understandable way, making the individual disclosure behavior of users more transparent. Dividing the component into a logging tool and a visualization tool, the following sections collect requirements that result from shortcomings of existing privacy-enhancing technologies as well as the outcome of the mentioned online survey.

9.1.1 Requirements of a Logging Tool

The primary goal of a logging tool is the complete recording of personal data the user transfers to service providers in the Web browser. A logging tool should seamlessly store these transactions securely and provide it to the corresponding visualization tool. Security threats require the

application to exclusively operate on the local system and not to share or offer any collected data online.

As HTML Web form fields [RLHJ99] are the primary container for the collection and transfer of personal data to service providers, an application that records personal data disclosures is required to analyze and identify these Web form fields. While irrelevant form fields should be disregarded, the application should relate form fields that contain personal data to the respective data type. Apart from personal data, a logging tool should also detect and store dynamic data and bond them to the respective transactions. In the context of the logging tool, dynamic data represent, for example, the user's IP address, cookies and a timestamp. In order to determine the conditions under which personal data were transferred to a service provider, a logging tool should find and store a machine-readable privacy policy of the respective data recipient and store it along with the transmitted personal data.

In addition to disclosed personal data, the service context of a personal data transaction should be captured as well. This requires the proper identification of the used process (like *Registration* or *Purchase*), contributing to more meaningful disclosure loggings (see also Chapter 6.3.2).

Our online survey (see Chapter 5.3.4) revealed that the majority of users favors the ability to edit identified transactions over a fully automatic process. These findings require the development of a semi-automatic identification approach.

Apart from the recording of personal data, the conducted online survey revealed the need for encrypted storage of elements that include sensitive data like a credit card number or a password. If an attacker gains access to the local transaction log file, no information can be obtained from encrypted data types, which significantly lowers the potential damage originating from a lost or stolen log file.

After the identification process is completed, the identified personal data transactions need to be stored in an open, standardized format, allowing the visualization tool to seamlessly read logged transactions.

Like for all solutions proposed in this work, usability and user-friendliness must be a primary design goal. As the transaction log file is the main input of the underlying visualization tool, user acceptance of the logging tool is of special importance. This implies a clear design of user interfaces and a simple user control.

9.1.2 Requirements of a Visualization Tool

A corresponding visualization tool is required to read and interpret recorded transactions of the logging tool. These raw data need to be converted and aggregated in a way that meaningful visualizations can be displayed.

A comprehensive visualization tool should also utilize information about used identity classes, partial identities and linkabilities captured and calculated by the Privacy Agent component (see Chapter 8.3.4). Ideally, this information is connected and related with information about logged transactions.

Exploiting the full potential of the collaborative privacy infrastructure, available information of the privacy community (see Chapter 6) should be incorporated. This facilitates the request of additional information about service providers, which allows for an aggregation and filtering of providers. Information from the privacy community also allows the presentation of more detailed service provider information.

Specifically, a visualization tool should enable users to analyze their disclosure behavior from different perspectives. The tool should provide understandable filtering functionality with regard to service providers and data types. Dedicated views should enable users to check, which personal data were transferred to a certain service provider as well as which service provider possesses a specific personal data type. In addition, a chronological analysis is required, which allows a temporal overview of data disclosures. Finally, the targeted tool should present selected relations between service providers, disclosed personal data types, partial identities and potential linkabilities, offering a clear, comprehensible overview of all involved information types.

In order to gain a high degree of user acceptance, special considerations should be taken for the design of suitable user interfaces [Bro98]. An intuitive user interface allows users to derive a mental model of the visualization tool and to estimate its behavior, when unfamiliar functions are used. A cohesive design will contribute to orientation and is of paramount importance.

In particular, Patrick et al. suggest the use of intuitive control elements, like familiar menus, lists and buttons [PKHvB03]. Furthermore, the application of common interaction patterns like drag-and-drop is recommended. A user-friendly tool should also employ familiar functions like zooming, selecting and moving objects. Finally, the differing level of user experience should be considered by adopting the user interface accordingly (see also Chapter 7.4.1).

9.2 Evaluation of Existing Approaches

This section compares the requirements of a Data Disclosure Log component with existing solutions. As we experienced the Data Track to be the best-known solution for the recording and displaying of past transaction, we lay out its goals and analyze its weaknesses. Furthermore, we evaluate additional applications that log personal user data in the Web browser. The evaluation of existing approaches helped us design and implement a suitable logging application as well as a corresponding visualization tool.

9.2.1 Data Track

Developed within the PRIME project, the goal of Data Track is to record personal data transactions and to offer users a transparent view of disclosed personal data [PFHB06] (see Chapter 4.2.2.5). In addition, the component provides supplemental functions that allow users to interact with service providers, facilitating an ex post adjustment of transferred personal data.

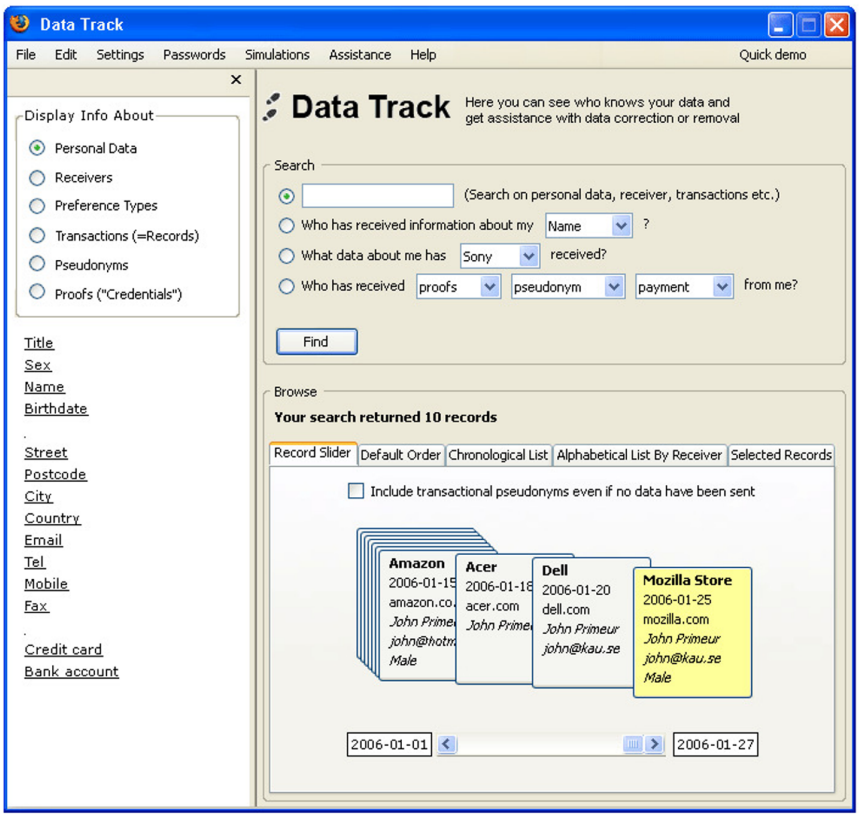


Figure 9.1: Data Track User Interface [Pet08]

The design of the user interface addresses the intuitive capture of user input as well as the understandable representation of disclosed personal data. Figure 9.1 shows a screenshot of the Data Track user interface.

Data Track offers several representations of data transactions. Addressing the component's understandability, a conducted user test showed that a chronological stack of data transactions earned the highest user acceptance [FHPB+07]. The test also underscored the importance of intuitive data filters. Data Track selects transactions by offering questions that alternatively filter data transactions according to service providers and personal data types.

The presented Data Track provides widespread functionality for the management of already disclosed personal data. The component, however, relies on the powerful PRIME architecture, which builds on a complex privacy infrastructure on the client side and the provider side. In this work, we propose and prove the advantages of a provider-independent privacy infrastructure (see Chapter 5) and show the benefits for local privacy components like the Data Disclosure Log component. Moreover, compared to Data Track this chapter evaluates and employs the application of more sophisticated visualization schemes for the user-friendly presentation of personal data transactions.

9.2.2 Transaction Logging

Apart from Data Track there are additional browser-based tools that log personal user data for various purposes.

The most essential applications of that kind are password managers that carry out a simple form of identity management. They support users in memorizing passwords of numerous accounts and generally are integrated in the Web browser. Password managers store login information of service providers and refill login forms, when the user returns to a provider. In some implementations, the password manager automatically submits the login form and logs the user in. Examples of password managers are the built-in Mozilla Firefox password manager[1] and the extension Secure Login[2].

While password managers facilitate the use of numerous complex passwords, they are obviously not suitable for the purposes of the Data Disclosure Log component. Password managers merely focus on the storage of user names and passwords and are not built to log personal data disclosures.

Another group of browser tools targets the automatic filling of Web forms. These form fillers memorize user data that are submitted in a Web form and automatically fill future Web forms with these personal user data. This functionality involves the automatic identification of Web form fields.

Roboform[3] is a form filler that combines the functionality of a password manager. The proprietary tool allows the definition of identities including personal data and Passcards, which are used to automatically fill Web forms and login forms. According to the project Web site Roboform employs artificial intelligence for the automatic detection of Web forms. A detailed analysis of Roboform is not possible, as the source code of the application is not publicly available.

Sxipper[4] provides similar identity management features. Unlike Roboform, Sxipper to some degree allows the analysis of its functionality. For the storage of identity information and filled Web forms a JSON file[5] is used. Like Roboform, Sxipper allows users to select one out of multiple pre-defined identities before the automatic filling of Web forms. For the protection of personal data Sxipper uses a symmetric encryption function. The automatic identification of Web forms is performed using a collaborative database, which is trained by all participating users.

Other form fillers we analyzed are InFormEnter[6] and Autofill Forms[7].

Even though the presented form fillers follow a different goal compared to our sketched transaction logging component, the concepts as well as the shortcomings of the presented

[1] http://www.mozilla.com/en-US/firefox/firefox.html
[2] http://securelogin.mozdev.org/
[3] http://www.roboform.com/
[4] http://www.sxipper.com/
[5] http://json.org/
[6] http://informenter.mozdev.org/
[7] http://autofillforms.mozdev.org/

tools served as input for the design of our proposed logging solution that is introduced in the following section. The main reason for the implementation of a self-made tool was the need for a customizable solution that provides an open interface for the corresponding visualization tool. While Roboform chooses a proprietary approach, Sxipper is mainly focused on identity management and does not allow the definition of a vocabulary, which prevents its integration into our privacy infrastructure. Moreover, Sxipper's approach of a global database raises privacy concerns, as users can potentially be tracked, when they fill multiple forms with the identical IP address.

An additional shortcoming all form fillers have in common is the lack of monitoring functionality. Applications like Tamper Data[8] intercept HTTP requests and present personal data and additional attributes to be transferred. Making data disclosures and the logging process more transparent, a similar but user-friendlier monitoring is required by our targeted logging tool.

9.3 Browser-based Transaction Logging

The accurate logging of transactions in the Web browser is a prerequisite for the corresponding visualization tool and represents a fundamental element of the Data Disclosure Log component. This section introduces the concept and the design of a browser-based tool that logs personal data disclosures. Identifying and storing these transactions, our proposed solution offers an open interface for the accessing visualization tool. Chapter 10.3.4 provides further prototypical details, including the defined data exchange format. The used data type vocabulary is centrally administered in the privacy community and distributed to all local privacy components (see Chapter 10.2.2).

9.3.1 Identifying Personal Data Transactions

As Web forms represent the main container for the transfer of personal data in the World Wide Web, the development of a browser-based logging tool requires a thorough analysis of HTML Web form usage. Listing 9.1 shows an example of a Web form, including text fields for the submission of personal data. For a more detailed definition of Web forms, the interested reader is referred to [RLHJ99].

```
1  <form action="form_handler.php" method="post">
2     <fieldset>
3        <label for="fname">First Name:</label>
4        <input type="text" id="fname" name="firstname" value="Your First Name">
5        <label for="lname">Last Name:</label>
6        <input type="text" id="lname" name="lastname" value="Your Last Name">
7        <label for="email">E-Mail Address:</label>
8        <input type="text" id="email" name="e-mail" value="Your E-Mail Address">
9        ...
10    </fieldset>
```

[8] http://tamperdata.mozdev.org/

```
11 </form>
```

Listing 9.1: Example HTML Web Form

Our analysis covered a wide selection of service providers from various areas, such as customizable search engines, Web shops that maintain a profile with diverse user data as well as online communities that offer user profiles containing a pool of various personal information. Our sample also included AJAX Web sites [Gar05], used by an increasing number of service providers. The outcome of our analysis allowed us to develop algorithms and rules for the identification of personal data type fields and the used process, which are introduced in the following.

The source code analysis revealed that service providers tend to use distinctive keywords within the definition of Web form fields. In many cases these keywords hint at the personal data type being captured. For the identification we developed an algorithm that checks all attribute values of a form field. The algorithm compares these attributes with pre-defined keywords of personal data types. The set of distinctive keywords for the data type *E-mail Address*, for instance, equals {`e mail, e-mail, email`}. Other keyword sets were built accordingly. For the generation of suitable keyword sets we considered Web sites of English and German service providers. Web sites using other languages require the adjustment of the generated keyword sets.

The data type *password* marks an exception of the keyword analysis, as it can easily be identified searching the Web form for `type=password`.

The check of Web form attributes like `name`, `id`, `title` and `alt`, however, does not result in a reliable identification of form fields. These attributes are invisible on the user side and can be assigned with random values by the service provider. For this reason, the identification algorithm focuses on the visible `label` and `value` attributes.

As the described parsing of Web forms alone is not sufficient for the accurate detection of form fields, a supplemental algorithm utilizes the user input in Web form fields. Employing regular expressions, the algorithm scans text strings prompted by the user and compares them to patterns that indicate a specific data type.

The identification capabilities of the user input analysis highly depend on the data type. Text strings representing *User Name*, *Password*, *Given Name*, *Last Name* or *Account Number* show little potential for the recognition of text patterns. These data types typically are of variable length and possess little distinctive characteristics. *User Name* and *Password*, for example, are strings of random letters and digits, which show no identification pattern. *Given Name* and *Last Name* generally start with a capital letter, while *Account Number* consists of a number of digits, also providing no reliable identification patterns.

On the other hand, *ZIP Code*, *Routing Number* and *Credit Card Security Code* show a higher identification rate. These data types have a fixed number of digits, allowing the algorithms to differentiate these values from other Web form input. This group of data types includes all instances of phone numbers in countries that issue phone numbers of fixed length.

The highest identification potentials bear data types like *E-mail Address* and *Credit Card Number*. An e-mail address can easily be detected due to the distinctive `@` in the middle of a text string followed by a valid domain. A credit card number has a fixed length and can be checked using the Luhn algorithm[9]. Additionally, in most countries the address line, which consists of the data types *Street* and *Street Number*, follows a standardized pattern, which can easily be identified.

Further improving the identification rate of form fields, users can optionally enter their user profile (see also Section 9.2), which is constantly compared with user inputs. The captured user profile is exclusively used for the identification of form fields. If a form field input of the user matches an element of the user profile, the data type of that form element is reliably identified. The user profile is accessible and changeable at any time. Addressing security, sensitive data of the user profile are encrypted. As the storage of a user profile requires trust of the user in the tool, the dedicated purpose of the user profile and the applied security precautions are highlighted in the user interface.

Once the tool detects a selected Web form, the described identification algorithms are executed sequentially. First, the Web form is scanned for keywords, which allow a preliminary identification of data types. As the user fills form fields with personal data, the input is compared with the entered user profile. If no user profile was created or values of the profile do not match, the user input is analyzed with regular expressions. The combination of these procedures facilitates a reliable, automatic identification of personal data types in Web forms, which represent the most essential elements of a logged transaction.

Apart from the identification of personal data disclosures, the proposed logging tool also detects the context of personal data submissions. In particular, the tool identifies *Login*, *Registration* and *Purchase* processes. The detection of processes assumes the prior identification of Web form fields.

The most distinctive characteristic of a *Login* process is the presence of a single `password` form element. A Web form belonging to a *Registration* process, on the other hand, generally provides two password fields and contains several additional fields. Furthermore, our Web site analysis revealed the general presence of the substring "`regist`" in the Web site's URL. If a registration involves several consecutive Web forms, the presented tool employs session identifiers and timestamps to merge the Web forms to a single transaction. *Purchase* processes are detected by searching the submit buttons for distinctive keywords. Our study identified the following keyword set: {`Buy, Check out, Purchase, Confirm`}.

As the presented automatic detection procedures cannot guarantee correct identifications in any possible scenario, we followed a semi-automatic approach and enable users to check and correct identified transactions during the interaction with service providers. The user interface also provides a history that allows the ex post selection and editing of all logged transactions, facilitating users to both control and improve the identification results.

[9] Identification cards – Identification of issuers – Part 1: Numbering system (ISO/IEC 7812-1:2006)

In the future, the automatic identification of form fields and processes could benefit from collaborative input. Following the Sxipper approach, the collaborative privacy community could maintain Web form information and store it along with the respective service provider information. This approach, however, requires a critical separation of privacy-sensitive transaction records and Web form information.

In addition to submitted personal data types and process information, the presented tool logs the used IP address, cookies and a timestamp. This dynamic information provides supplemental privacy-related information about transactions for the corresponding visualization tool. Furthermore, the machine-readable P3P privacy policy – if available – is stored and linked to a transaction.

After all transaction details have been analyzed and identified, the logging tool stores an entry to the user's transaction log file, including the personal data types, the transmitted values, dynamic information and the used process. As demanded in the requirement analysis, sensitive user data like *Credit Card Number* and *Password* are not stored as plain text. Instead, we employ hash functions [Eck07] for the protection of these personal data.

9.3.2 Design of the FireLox Extension

As discussed earlier in this work, browser usage statistics as well as the presence of a powerful extension development environment led us to choose Mozilla Firefox as host browser for the development of browser extensions (see also Chapter 8.3.1).

Designing our proposed solution, we chose *FireLox* as memorable name for a Firefox logging tool. Similar to the development of the introduced Privacy Agent, we employed the XML User Interface Language (XUL)[10] for the design of FireLox. The logic described in the previous section is implemented using JavaScript [Fla06].

For the user-friendly activation and management of FireLox we added two buttons at the right side of the browser's status bar (see Figure 9.2). A Record button allows users to activate and deactivate the recording of personal data transactions. A red spot on that button indicates that personal data disclosures are recorded, while a blue spot signals that logging is paused. An accompanying magnifying glass button opens and closes the FireLox sidebar, which displays the identified transaction information, the transaction history and the user profile.

As depicted in Figure 9.2, the FireLox sidebar is divided into three tabs. The initial Form Data tab shows dynamic user information as well as Web form data of the visited Web page, including the identified data types with user inputs and the detected process. If the Web page contains more than one Web form, the sidebar displays information of the Web form the user selected. After a Web page is loaded, the data type and process identifications are only based on the results of the keyword analysis. As the user fills the selected Web form, the identifications are automatically fine-tuned using inputs from the user profile as well as the regular expression analysis. Once the user completed filling the Web form, the data type and process identification

[10] http://planetxul.com/

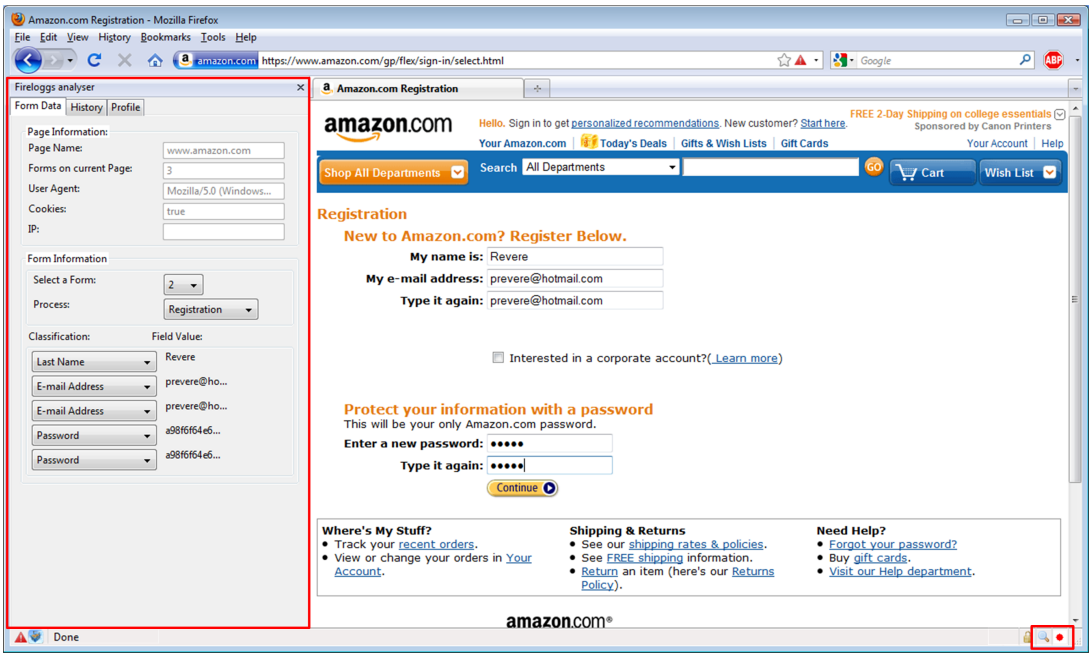


Figure 9.2: FireLox User Interface

can be adjusted using the respective drop down menus in the sidebar. After the submit button of the Web form is clicked, the complete transaction is added to the transaction log file.

The History tab provides an overview of all past transactions stored in the transaction log file. A drop down menu filters transaction according to service providers. If the user chooses to edit a selected transaction, the transaction log file is updated accordingly.

The Profile tab lists a user profile with personal data. As described in the previous section, this information is utilized for a more precise identification of Web form fields, if the user opts to provide this additional information.

9.3.3 Evaluation

We evaluated the introduced FireLox logging tool within the scope of our second user test (see Chapter 5.3.4).

After a short introduction of the tool the participating test persons showed no difficulties in interpreting the toolbar and activating the Record button. When test persons were asked to fill a registration form of a Web mail provider, we did not observe any usability issues in the Form data tab. Comments of test persons merely led us to change the use of color and the arrangement of the drop down menus. Test persons generally praised the possibility to change personal data type and process identifications during the filling of a Web form as well as after the submission of personal data in the History tab.

Only seven out of 26 test persons voiced their general concern about a tool that protocols personal data transactions. However, the outcome of the test underscored the need to clearly

demonstrate the dedicated purpose and the advantages of such a tool, as the presence of a personal data log per se represents a potential privacy threat and not a privacy-preserving measure. The secure storage of sensitive user data was acknowledged and considered as prerequisite for the widespread use of a logging tool.

In addition, the test revealed a clear rejection of an integrated automatic Web form filler, which 17 out of 26 test persons did not favor.

9.4 Visualizing Logged Personal Data Transactions

After the presentation of a suitable logging component, this section introduces the concept, design and implementation of an application for the user-friendly visualization of personal data transactions. The proposed solution along with its technical implementation employs basic concepts of graph theory and graph drawing. For more detailed theoretical foundations of these topics the interested reader is referred to [CH91, DBETT99]. After a conceptual overview and a description of the used entities, we present the design of the proposed visualization tool, followed by an evaluation.

9.4.1 Conceptual Overview

This section describes the architecture as well as the interfaces of our proposed visualization tool. The design followed the Information Visualization Data State Reference Model [CR98, Chi00], which defines the visualization process from raw data to visualization views.

The proposed tool is divided into several visualization components (see Figure 9.3). One of the two import interfaces parses local XML files containing logged personal data transactions as well as information about used identity classes and partial identities. A second import interface requests service provider information using the privacy community's Web services. All captured information is loaded into a local database that integrates imported data. Each provided visualization component requests and processes information from the local database and displays a specific view of personal data transactions to the user. Two export interfaces enable users to print and store generated visualizations. In the following, considerations and design factors of the presented application architecture are discussed.

XML interfaces facilitate the exchange of information between local privacy components. The transaction log file represents the most essential data source, containing information about personal data disclosures to service providers (see Section 9.3). The identity file provides information about identity classes and partial identities (see Section 8.3.4) and facilitates the presentation of partial identity usage as well as the visualization of potential linkability threats.

The privacy community (see Chapter 6) serves as additional input source for the proposed application. Service provider information of the community is queried using the published Web service interfaces. In order to integrate Web services into the visualization tool, a Web

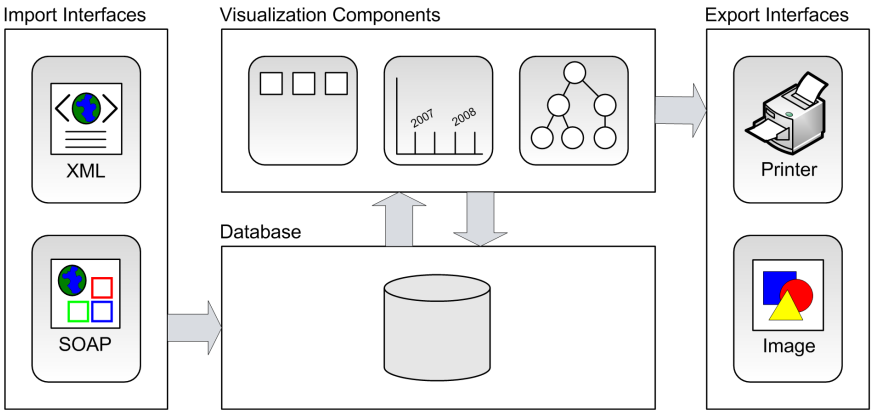


Figure 9.3: Architectural Overview of the Visualization Tool

service client performs a Remote Procedure Call (RPC) employing the SOAP communication protocol [GHM+07a, GHM+07b].

In particular, Web services of the privacy community are utilized to request static information about service providers, which is displayed as supplemental information in the prepared visualizations. The application also accesses the service type of service providers, which allows for a graphical categorization of service providers. For the visualization of linkabilities in provider networks, the developed tool requests information about personal data sharing with third parties.

In order to present understandable translations and explanations of the displayed data types and service types, the tool also loads vocabulary information from the privacy community (see Chapter 10.2).

The visualization of interconnections between various input sources requires the integration of all available data into a homogeneous data pool. Its ability to handle complex queries is a vital factor for the corresponding visualization capabilities. A conducted online survey (see Chapter 5.3.2) indicates that the transaction log file for an average user grows by 1480 entries per year. The number of data sets and the necessity of complex queries entail the need for low data access time, which is intimately connected to the application's overall performance. These constraints represent the main criteria for an appropriate data layer.

As fast data access is inevitable for complex visualizations based on large amounts of raw data, native parsing, accessing and querying of XML files via XQuery is not a suitable solution given the described requirements. Even though XQuery allows complex data queries at runtime, this approach does not provide the required data access performance. Furthermore, such an approach would require a complex integration of all data sources. A suitable solution, however, should perform an integration before the visualization.

For this reason, we employ a relational database to integrate and store available input data. The advantages of a tailored data model that caters the needs of the application outweigh the additional effort. Furthermore, a database provides a powerful query language that allows

(a) Identity (b) Partial Identity (c) Service Type (d) Service Provider (e) Process (f) Data Type

Figure 9.4: Icons of Defined Entity Types

complex queries over several data sources. As the database is initialized during the start of the application and loaded into the system memory, the performance is additionally increased.

Chapter 10.3.5 provides further implementational details, including the data model designed for the presented visualization tool.

9.4.2 Definition of Entity Types

In the previous section we describe the involved input data sources and stress the need for a consistent integration. The common goal of the corresponding visualization components is the understandable and intuitive presentation of relations between various types of data.

Ware classifies data into entities and their relations to each other [War04]. An entity represents an object of interest, while relations define the structure that interconnects entities. Both entities and relations possess additional attributes.

Identifying objects of interest for our scenario, we define six entity types out of all information available in the generated database. These entity types represent aggregated information units and serve as foundation for all developed visualization components. The ability to relate entity types allows for a hierarchical display of dependencies. An entity describes a concrete instance of an entity type.

In the following we describe the identified entity types along with the relation to each other. The entity types were defined considering the relevance to the proposed tool and the resulting information value to the user. For this reason, the entity type `Transaction` – although a central entity type in the data model – was not defined, as by itself it provides no integral information to the user. Figure 9.4 displays icons of each defined entity type, guaranteeing a consistent and comprehensible user experience.

- `Identity`
  The entity type `Identity` refers to identity classes that group partial identities with regard to their transitive linkabilities (see Chapter 8.3.4). `Partial Identity` represents an inferior data type.

- `Partial Identity`
  A `Partial Identity` is regarded as a data type subset of a user's complete identity (see Chapter 2.2). Partial identities are used for interaction with service providers.

- `Service Type`
  The entity type `Service Type` consists of twelve pre-defined service types (see Chapter 7.4.2) and assigns entities of `Service Provider` to one of these twelve instances. For example, the service provider Amazon is subordinated to the service type *Shopping*.

- `Service Provider`
  This entity type represents an abstract Web site a user interacts with and discloses personal data to. An instance of this entity type is, for example, Amazon. `Service Provider` marks a central entity type, as it maintains several relations to other entity types. `Service Type` can be interpreted as superior entity type, as it aggregates service providers. Inferior entity types are `Process` and `Data Type`.

- `Process`
  `Process` groups transactions into processes like *Registration* or *Purchase*. The goal of this entity type is to allocate data types to one of these categories and – as a consequence – to make data type disclosures to service providers more transparent. This allows, for example, to understand that credit card information has been transferred to Amazon in the context of a *Purchase* process.

- `Data Type`
  This entity type contains all personal data types transferred to service providers. Entities of this entity type are, for example, *User Name*, *Date-of-Birth* or *E-mail Address*.

9.4.3 Application Design

This section introduces visualization components, which are designed in accordance with the requirements identified in Section 9.1. Considering recommendations of the Visual Information Seeking Mantra [Shn96], the given scenario requires the development of multiple views, which reduce the complexity of the presented visualization. Wang Baldonado et al. define guidelines for the use of multiple views [WBWK00]. Following these guidelines the presented application provides five individual views to the user. Each view presents disclosed personal data and identity information from a different perspective.

The goal of all visualization views is the abstraction of the textual representation. A clear graphical representation and its accompanying complexity reduction facilitate a quick comprehension of all essential information.

Figure 9.5 shows the application's navigation bar with buttons that lead to the respective views. In the following sections, the design and the realization of each view are discussed. We start with the chronological view (Timeline), before we introduce the graph view (Graph) and the linkability view (Linkabilities), the most complex data visualizations of this tool. Finally, we present supplemental views, which include the service provider view (Overview) and the record slider view (Slider).

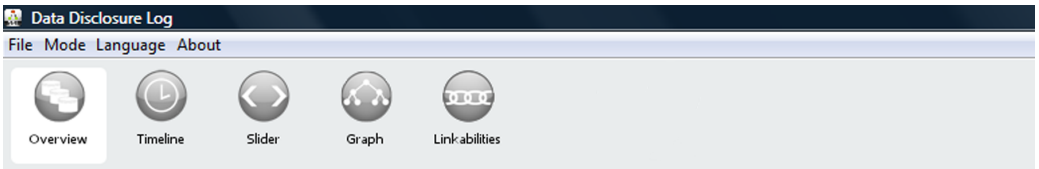


Figure 9.5: Navigation Bar

9.4.3.1 Chronological View

The primary goal of the introduced application is the presentation of relations between entities. A comprehensive visualization tool, however, should also be capable of displaying a timely sequence of data disclosures. In the following, we describe the design of such a view.

The chronological view performs a temporal analysis of all communications with service providers, allowing users, for instance, to identify frequently interacted service providers. For the reason of clarity an adjustment of the observation period is available. As opposed to other views, the chronological view focuses on a single entity type (`Service Provider`) and does not show relations to other types.

Accounting for the needs of a usable interface, we employ a horizontal timeline for the chronological presentation of service provider interactions. As users are already familiar with such an element, the application benefits from a short introduction phase and a higher user acceptance.

The structure and design of the chronological view is depicted in Figure 9.6. The view is divided into four parts. Most space is dedicated to the scrollable timeline, as proposed by the Center Stage pattern [Tid05]. The horizontal axis is labeled with date values. The column above each date stacks all service providers the user transferred personal data to at that certain date.

Service providers are represented by a standardized icon frame introduced in Section 9.4.2. If possible, an individual service provider icon is dynamically generated, downloading the respective Favicon of the service provider and dynamically setting it into the green icon frame. Such an individual service provider icon contributes to a quick comprehension of the shown information.

The top side window provides an interface for the selection of a time interval, enabling users to define the observation period. Facilitating an intuitive user control, the active interval selection resembles a calendar, if the start or the end date field is clicked.

Below that window, a filter enables users to limit the displayed service provider interactions with regard to the checked data types. This dynamic filter, for instance, allows users to specify the exclusive consideration of service providers an *E-mail Address* or a *Date-of-Birth* has been disclosed to, which is known as Dynamic Queries [Tid05].

If a service provider is selected by clicking a respective icon in the timeline, a detail window provides supplemental information about that service provider using static information queried

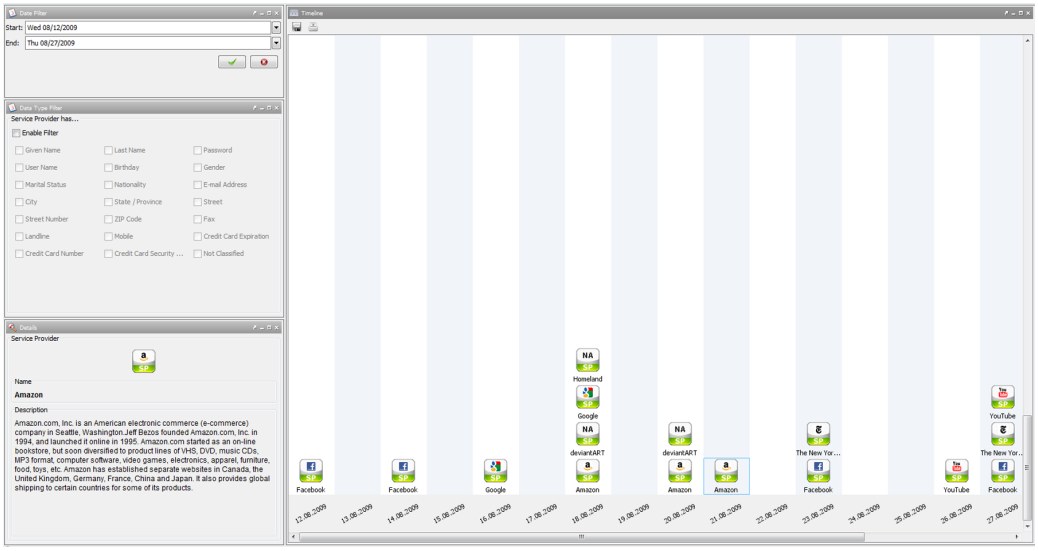


Figure 9.6: Chronological View

from the privacy community. The interplay of the main window and the detail window implements the concept of a Two-Panel Selector [Tid05].

9.4.3.2 Graph View

The graph view aims for a generic presentation of diverse relations between entities, using a graph-based visualization scheme. As the proposed application integrates various data sources and defines entity types for the better abstraction of the textual representation, the comprehension of a large amount of information is possible, which facilitates the visualization of relations that are not evident to the user.

In the graph view, the selection of entities and the display of their relations are individually controlled by the user, who is enabled to drag and drop selective entity nodes to the graph area. If an entity is added to the graph, relations to existing entities in the graph are visualized as edges, allowing users to dynamically build a tree of any available combination of entities.

The described procedure allows a maximum number of degrees of freedom. In order to lower the complexity for inexperienced users, this user group is limited to adding a whole entity *type* to the graph, which includes all available entity nodes of that type. Accounting for their low level of experience, this less complex graph generation simplifies user control. More experienced users can add and relate individual entities to the graph. The user experience is captured at the initial start of the application, offering two complexity modes to the user (see also Chapter 7.4.1).

In order to facilitate a generic, implementable solution of the graph view, we specify relations that can be built and visualized from a particular entity type. This prevents the visualization of entity combinations that cannot be related. In particular, we define two directions of relations between entity types. These hierarchies alleviate the burden of developing generic algorithms.

At the same time, pre-defined hierarchies do not limit users from selecting any available combination of entities.

The Top-Down hierarchy defines `Identity` as the most superior entity type, followed by `Partial Identity` and `Service Type` that categorizes `Service Providers`. Furthermore, `Service Provider` is superior to the entity type `Process`, which is used to aggregate `Data Types`. Applying this hierarchy of entity types, a user could, for example, select a particular service provider and visualize its related processes (child nodes of the selected service provider) and their involved data types (child nodes of each process) in the graph.

The Bottom-Up hierarchy consists of three entity types and defines `Data Type` as top element and `Process` and `Service Provider` as child elements. This hierarchy allows, for instance, the selection of a particular data type and the presentation of service providers (child nodes) it was transferred to.

As mentioned above, the definition of entity type hierarchies limits the complexity of the view generation. Furthermore, a hierarchy facilitates the representation of relations between entities as a directional graph. The applicable hierarchy is implicitly chosen by the user, when the first element of the graph is added. If a `Data Type` or `Process` entity is initially selected, the Bottom-Up hierarchy applies. For all remaining entity types, the Top-Down hierarchy is chosen.

Figure 9.7 shows the design of the graph view user interface, which is divided into four windows. Similar to the chronological view, the main window focuses on the display of entities. Due to the hierarchical relations, we choose a hierarchical layout for the presentation of the directional graph [STT81], allowing a clear and intuitive representation of relations between entities. Alternatively, we provide a radial layout that aligns child nodes on outer circles originating from a root node in the center.

Available entity types the user can add to the graph are presented in a side window and are represented by the designed entity type icons (see Section 9.4.2). The first graph node is added by dragging and dropping an entity type to the main window. A pop-up window captures the selection of the single entity of that type, which serves as root element of the graph. The procedure for the addition of further entity types depends on the chosen complexity mode.

In the basic mode the user drags and drops whole entity types in the main window. An iterative algorithm scans all available instances of that entity type and adds entities to the graph that are related to the already displayed entities in the graph. Based on a single entity this procedure allows for a quick arrangement of the chosen elements and their relations to each other.

The advanced mode facilitates a more customizable visualization of relations between entities. Again, a single entity serves as root element of the graph. In the consequence, all possible relations based on this entity are calculated in advance. Subsequently, the user is able to add individual entities using menus that are offered by right-clicking a graph node.

The algorithm of the advanced mode is shown in Figure 9.8. For each identity class of the user

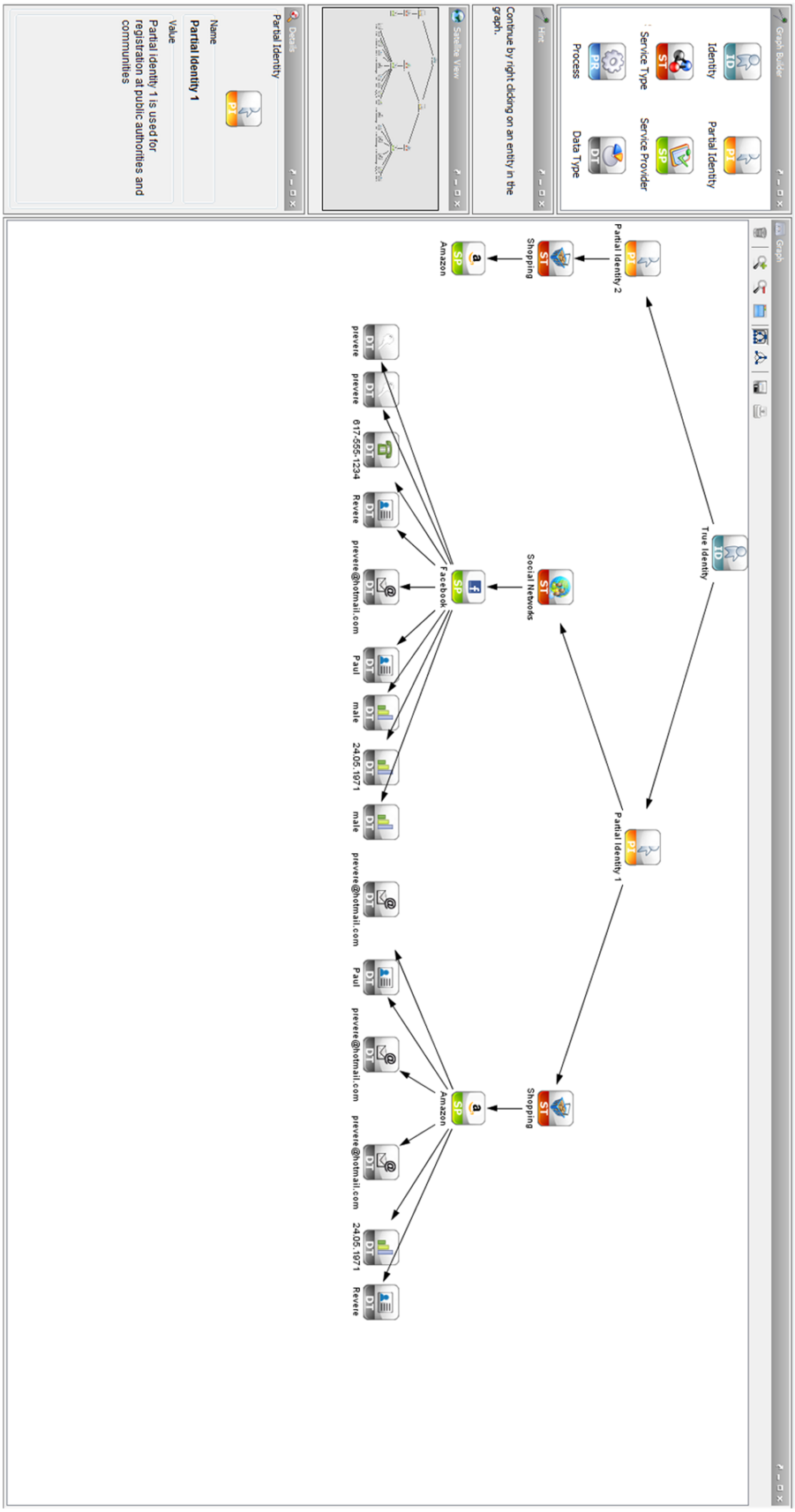


Figure 9.7: Graph View

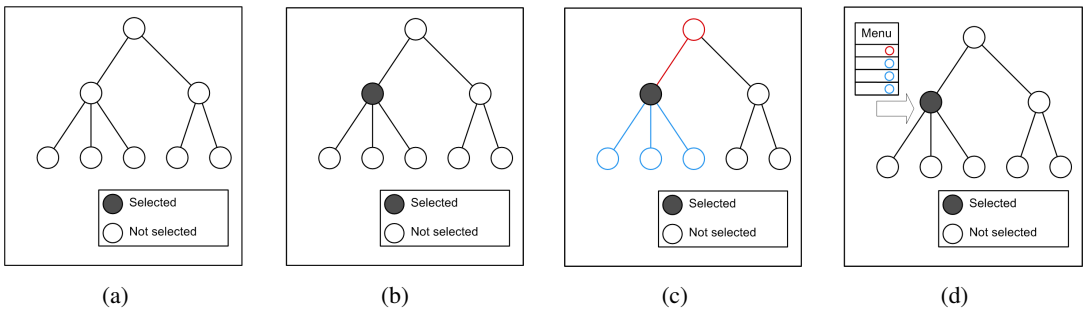


Figure 9.8: Generation of the Graph View in the Advanced Mode

a complete graph including all related entities of the Top-Down hierarchy is pre-calculated (a). Once the user chooses a root entity, the identity graph is implicitly chosen. The chosen root node is marked as selected in the graph (b). In the consequence, all preceding and following entities are determined (c). These nodes represent those entities the selected entity is related to. These relations are embedded in a selection menu, which is added to the root node (d). For the visualization the whole tree is scanned and all nodes marked as selected are visualized. If the user selects an additional node in the menu, step (c) and (d) is executed for that added node.

By default, all graph nodes are represented by the designed entity type icons. Like the chronological view the graph view dynamically generate individual `Service Provider` icons downloading the repective Favicons. In addition, individual `Service Type` and `Data Type` icons are generated by downloading the corresponding icons from the privacy community (see also Chapter 10.2). Again, the downloaded icons are dynamically set in the icon frame of the respective entity type.

Below the entity selection window a tool tip window provides hints that guide users through the graph generation process. In order to keep an overview of larger graphs, a satellite window displays a tiny view of the graph. Similar to the chronological view, the graph view provides an additional window that presents detailed information about entities selected in the graph.

The usability of the interface is realized by a clear separation of information. The windows integrate elements used in other local privacy components, contributing to understandability and fostering user acceptance.

9.4.3.3 Linkability View

An additional goal of the proposed visualization tool is to show the potential spread of personal data originating from linkable partial identities (see Chapter 2.3). As cooperating service providers – in theory – are able to exchange collected personal user data, linkable partial identities can potentially be aggregated to a larger partial identity known to all recipients. The presented linkability view addresses this threat and visualizes service provider networks the user disclosed linkable partial identities to.

In order to display potential linkability threats, the linkability view processes identity infor-

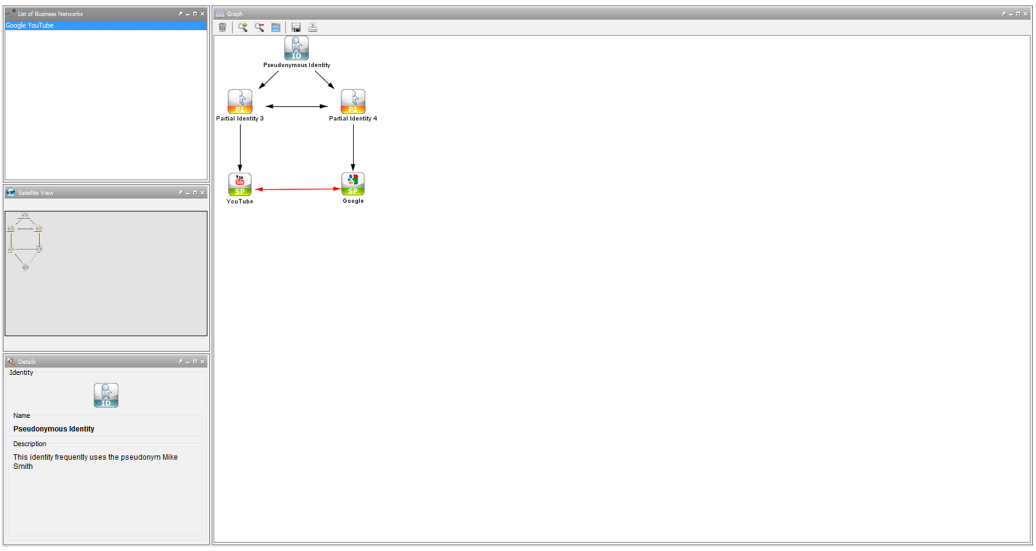


Figure 9.9: Linkability View

mation from the Privacy Agent component (see Chapter 8.3.4) as well as service provider information about data sharing with third parties, which can be requested from the privacy community (see Chapter 6.3.3).

The resulting user interface consists of four windows. The provider network along with the involved identity classes and the linkable partial identities used for the interaction with service providers represents the central element of the view. Similar to the graph view, the relations between all elements are displayed using a hierarchical graph. Identity classes serve as root elements of the graph. Their linkable partial identities are represented as child nodes. The involved service providers of the provider network are child nodes of the partial identities.

Figure 9.9 shows the resulting graph visualization, which underscores the connection of linkable partial identities and cooperating service providers. A highlighted horizontal, bi-directional relation (red edge) between two service providers emphasizes the potential aggregation of linkable partial identities among the involved service providers.

For each service provider network, separate graphs are generated, which are listed in the top side window and can individually be selected by the user. A satellite window provides an overview for the navigation in large graphs. As already described in the previous sections, a detail window provides supplemental information about selected graph nodes.

The internal generation of the linkability view requires the determination of all service provider networks the user interacted with, using the respective Web service of the privacy community. Figure 9.10 sketches the generation process that is executed for each identified provider network. The process is divided into three steps. First, all identity classes linkable partial identities have been disclosed to are determined. Subsequently, linkable partial identities used for the interaction with the provider network are identified and added to the graph as child nodes. The involved service providers are added as child nodes of the partial identities. After the

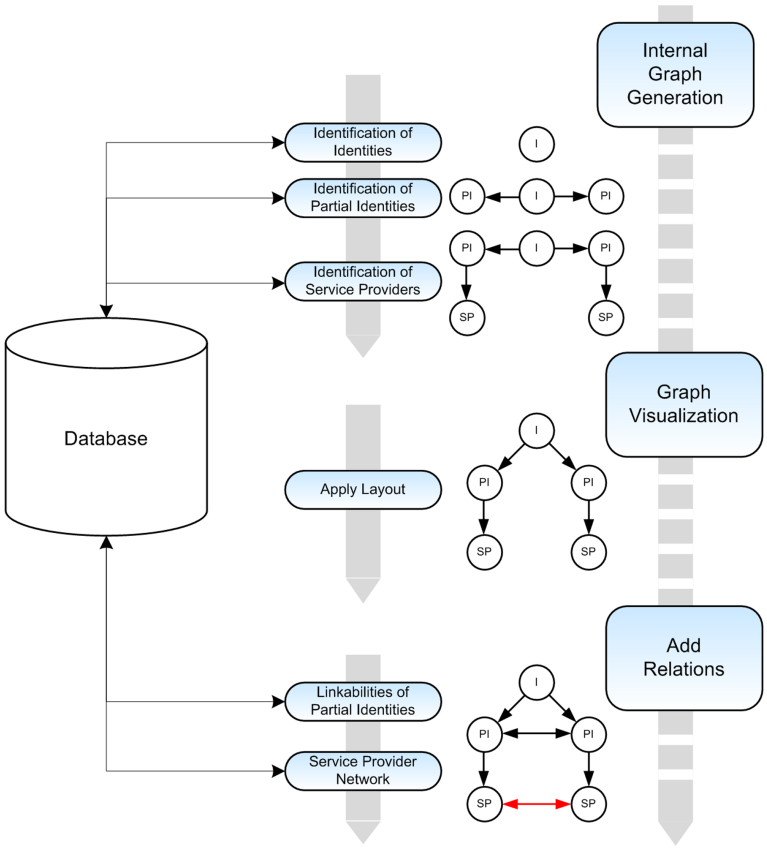


Figure 9.10: Generation of the Linkability View

generation of a hierarchical graph visualization, linkabilities between partial identities as well as cooperations of service providers are added as edges. As described earlier, relations between service providers are highlighted in red.

The presentation of provider networks along with disclosed linkable partial identities makes a valuable contribution to the transparency of past personal data disclosures, as it allows the recognition of linkability threats that could lead to an unintentional spread of users' personal data.

9.4.3.4 Supplemental Views

In addition to the main views presented in the previous sections, the presented visualization tool provides simple views that focus on personal data disclosures to service providers, employing intuitive forms of presentation. Accordingly, these views concentrate on the entity types `Service Provider` and `Data Type`.

Figure 9.11 shows the service provider view, which represents the first view in the navigation bar. The service provider view displays all service providers the user disclosed personal data to. Service providers are visualized as tiles in the main window. Apart from the service provider,

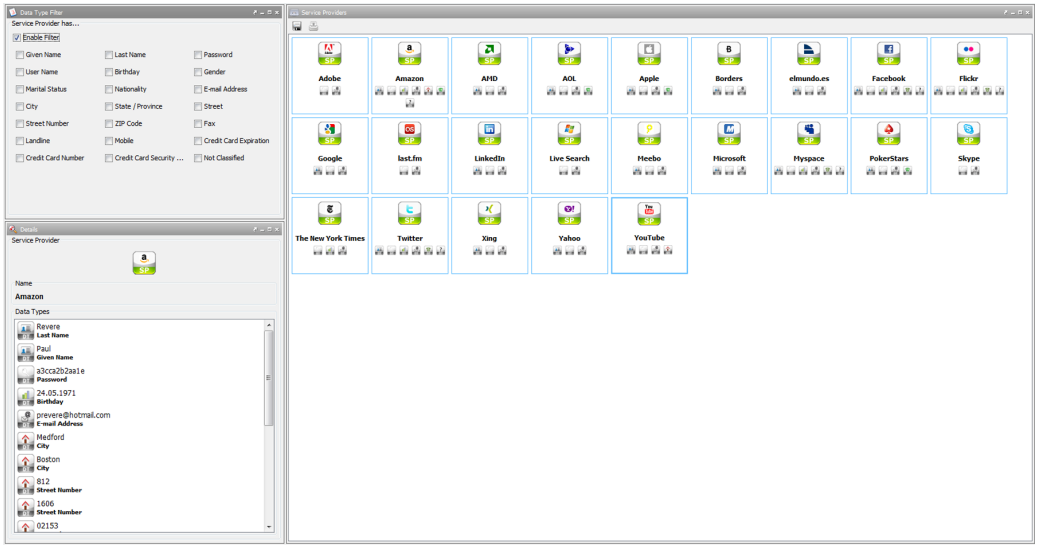


Figure 9.11: Service Provider View

icon groups that represent certain data type groups (see Chapters 7.4.3 and 8.3.3) are fit into the tile, allowing users to instantly estimate the kind of personal data disclosed to a provider.

As already utilized in the chronological view, a filter window enables users to fade out service providers with regard to combinations of transferred personal data types. If the user selects a service provider tile, a list of disclosed individual data types is provided in a detail window.

Also highlighting service providers the user interacted with, the record slider view aligns service providers in a Cover Flow-like fashion, which is depicted in Figure 9.12.

The record slider view provides two spacious windows for the display of static service provider information from the privacy community and a list of disclosed data types. Reusing an element of the chronological view, a final window allows for the filtering of service providers based on the time of interaction.

9.4.4 Evaluation

Analyzing the usability of our solution, the presented visualization tool was tested within the scope of our second user test (see Chapter 5.3.4).

Each test person was provided with a mock-up transaction log file, which was loaded into the tool. After the initial start of the tool, the test candidates were asked to read and understand a prepared tutorial, which took about five minutes on average.

The first task involved the service provider view. Test persons were asked about data type disclosures to a specific service provider. The solution of this question posed no further difficulties. All test persons unanimously associated the tiles with service providers, intuitively clicked the correct tile and read the list of disclosed data types from the detail window.

Affecting the same view, test persons were requested to identify service providers their *Date-*

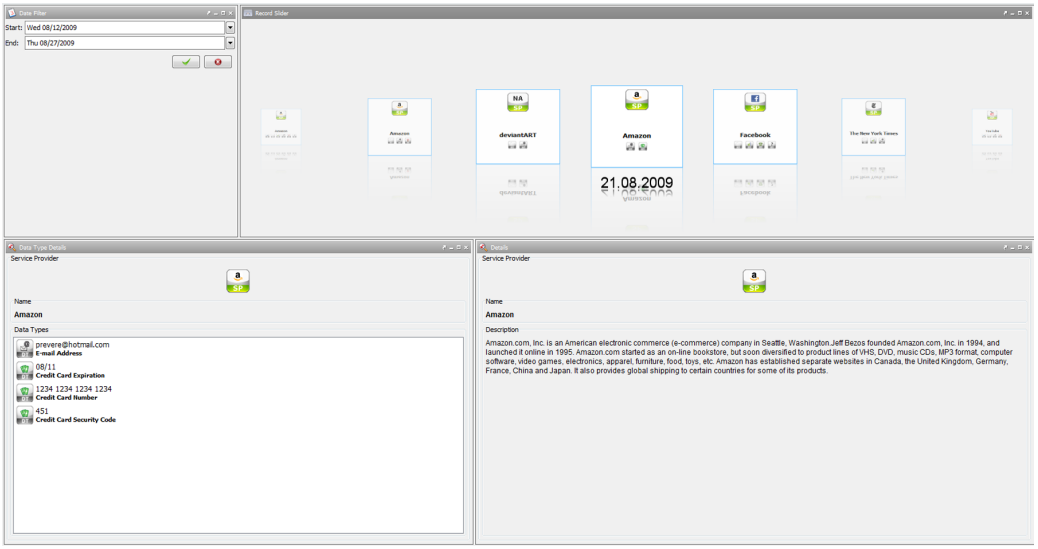


Figure 9.12: Record Slider View

of-Birth was transferred to. 21 out of 26 test persons instantly activated the filter window and checked the correct box. The remaining five persons tried to solve the question by going through all service providers in the main window. Acknowledging the comments of these users, we slightly changed the wording of the filter window.

The next question asked about disclosures of the *Credit Card Number* in a certain time period. 18 out of 26 test persons correctly identified and clicked the Timeline button in the navigation bar, which led to the chronological view. Six test persons needed guidance from the interviewers how to find the button of the respective view. After reaching the chronological view, only two test persons were given hints how to use the time period filter.

The following questions targeted the graph view and represented the most challenging tasks. In particular, test persons were asked to display disclosed data types for each process executed at the service provider Amazon. 22 test persons correctly navigated to the correct graph view. Building the correct graph, most test persons had difficulties applying the offered entity types correctly. The majority of these persons also did not choose Amazon as root element, which was required for the proper solution of the task. The observed issues during the graph generation led us to create an animated built-in tutorial that describes the purpose and demonstrates the graph generation process of the graph view. Once the graph was generated, all users correctly interpreted the visualized relations and confirmed the intuitive presentation. Concerning the layout, 16 test persons preferred the hierarchical layout, while seven persons opted for the radial layout.

A similar task tested the graph view in the advanced mode. The change of the graph generation from a simple drag and drop in the basic mode to a menu-controlled node addition in the advanced mode caused further difficulties that served as additional input for the now available tutorial.

In a final task test persons were prompted to identify linkability threats originating from a certain service provider. Navigating to the linkability view, 21 out of 26 test persons selected the correct graph and recognized prevalent linkability threats. The remaining five test persons needed a short hint to find the proper view.

In the interview section all test persons attested a clear presentation of the displayed information and relations. The test candidates also praised the intuitive user control. Asked about scenarios they would use the presented visualization tool for, test persons named the misuse of their personal data, the analysis of their disclosing behavior and the identification of service provider networks. The answers prove that test persons understood and valued the advantages of the developed visualization tool.

9.5 Alternative Visualization Solutions

The solution introduced in the previous section was preceded by additional visualization approaches we developed and considered for the proposed privacy infrastructure. The following sections present a 3D graph and an Adobe AIR solution.

9.5.1 3D Graph Approach

This section introduces a visualization tool that displays personal data disclosures employing a 3D graph. After the introduction of a 3D visualization scheme, we present details of the designed user interface, followed by an evaluation.

9.5.1.1 3D Visualization

In addition to the presented visualization tool in Section 9.4, we weighed the potential of 3D graph representations. Targeting an intuitive graph, we employed concepts of the geon theory of Biederman [Bie87]. The author proves that human perception is based on the recognition of 36 3D primitives, so called geons. Biederman refers to analogies to the human languages that are composed of phonemes. In the context of UML diagrams [Obj07b, Obj07a] user experiments of Irani and Ware proved a superior understandability of geons compared to regular diagrams [IW03].

As the graph-based visualization of personal data transactions shares many constructs of UML class diagrams, such as generalizations, aggregations and associations, we apply the technique of geon UML diagrams to the generation of graphs that show relations between our defined entity types (see Section 9.4.2). Compared to 2D graphs, a 3D representation of relations offers additional visualization opportunities, especially for the presentation of large amounts of data. On the other hand, a more complex visualization poses usability challenges, which affect the user acceptance of the application.

For the visual translation into geon-based diagrams we followed the rules of Irani and Ware

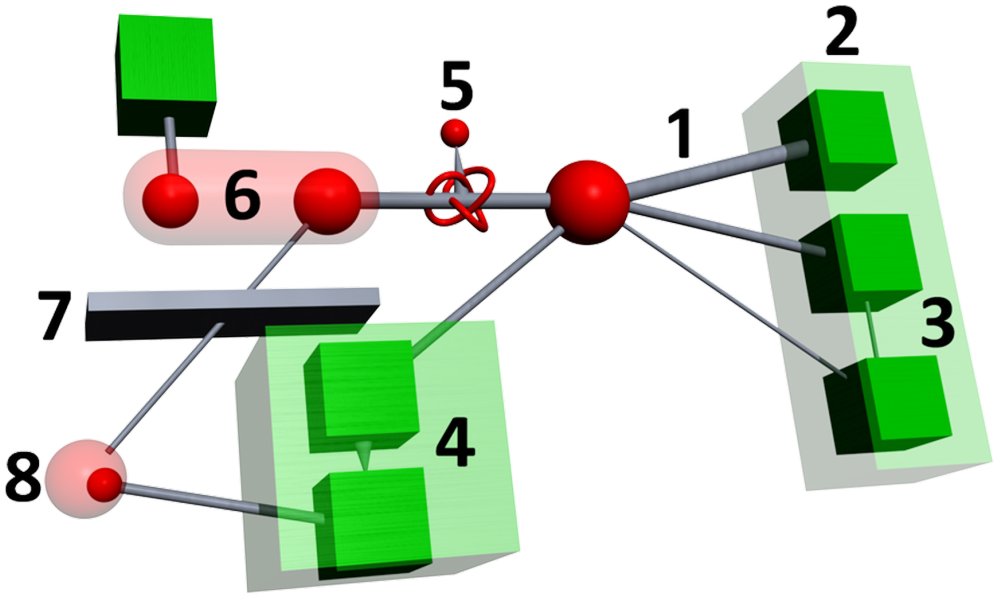


Figure 9.13: 3D Geon Visualization

[IW03]. Fundamental entities are visualized as simple 3D primitives, relations between entities by the geon skeleton. Less significant entities are shown as geon appendices. The size of entities is chosen proportionally to their importance. Emphasizing a 3D shape, geons are shadowed. Secondary attributes play a less significant role. A view presents all geons, which are aligned orthogonally to the viewing perspective.

Figure 9.13 shows an exemplary 3D visualization of personal data transactions, following the above-described guidelines. Marking the central elements of the visualization, partial identities are represented as red spheres, while service providers are shown as green cuboids. This definition allows for a clear identification of these fundamental entity types. The chosen shapes facilitate the projection of textures like icons on the shapes, supporting the perception and the spatial layout of a 3D shape [War04]. The diameter of a partial identity sphere is chosen with regard to the amount and the privacy sensitivity of its containing data types.

The relations between all entities are represented as cylinders. For relations between a partial identity and a service provider the thickness of the cylinders reflects the interaction frequency of the user (1), which corresponds to the number of transactions. Service providers that offer the same service type are grouped using transparent green cuboids that cover a set of service providers (2). Cylinders between service providers highlight service provider cooperations (3). Pyramids indicate hierarchical connections in corporate networks (4). Linkabilities of partial identities are highlighted using a red knot (5). In addition, the intercept set of linkable data types is shown above the knot. Partial identities can be grouped into identity classes using long-shaped transparent red bubbles (6). A border between partial identities allows for the

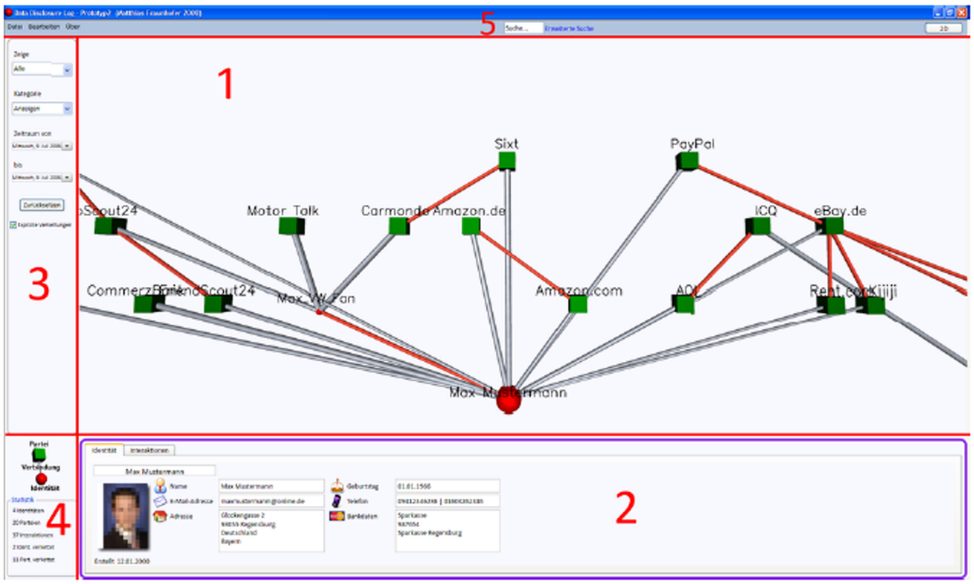


Figure 9.14: 3D User Interface

separation of anonymous and identifiable partial identities (7). Finally, a red sphere surrounded by a transparent red sphere indicates a subset of a partial identity (8).

9.5.1.2 Application Design

The development of the presented 3D visualization scheme resulted in a .NET 3.5 prototype that is depicted in Figure 9.14.

Implementing the concept of Smooth Transition [SIG07], the application initially presents a 3D graph containing all available entities (1). This graph only shows the most essential details about entities and relations. If the user zooms objects of interest, the application gradually increases the amount of entity details, like data types and additional textures. The Smooth Transition enables users both to maintain an overview of the graph and to receive supplemental entity information on demand. A double-click on an entity triggers a pre-defined animated zoom of that entity.

The detail window displays additional information about a selected entity (2). Applying a Cardstack pattern [Tid05], the first tab contains static entity information like data types of a partial identity. The second tab shows information about the involved transactions. For selected service providers a third tab lists partners the provider shares personal user data with. The border color of the detail window is matched with the color of the selected entity. This interplay maintains the connection of both elements, applying the principles of Color-Coded Sections [Tid05] as well as the Rule of Self-Evidence [WBWK00].

Offering further means to reduce the complexity of the user interface, a dynamic filter is provided that fades out elements of the graph based on particular entity type criteria (3). A final segment presents a short legend and statistics about the total number of entities and the amount

of displayed entities (4). Finally, a text box allows users to directly search for elements in the graph (5). The search results are presented both in the main window and in the detail window.

9.5.1.3 Evaluation

The described 3D visualization tool was tested within the first conducted user test (see Chapter 5.3.3).

In general, the test candidates' feedback of a 3D visualization that presents personal data disclosures was positive. The test revealed no severe shortcomings of the application's usability and navigation. The majority of test persons interpreted the displayed graph correctly. Most test persons, however, did not intuitively click and select entities of the graph, which revealed the need for an introducing tutorial and explaining tool tips. As already observed in other user tests, the recognition of linkabilities was primarily misinterpreted due to the lack of user knowledge in that area.

Even though the presented approach provides an innovative opportunity to visualize disclosed personal data, we chose to offer and integrate a 2D-based solution into our privacy infrastructure (see Section 9.4). Main reasons for our choice were user studies that identify deteriorating user performances and an inferior user preference of 3D interfaces [RCMC00, CM02]. Even though 3D visualizations offer impressive presentation schemes, the user interfaces inevitably become complex and less intuitive.

Nevertheless, this section proved the applicability of a 3D application in the context of the Data Disclosure Log component. Implementing the required interfaces, the presented application can be integrated into our privacy infrastructure at any time.

9.5.2 Adobe AIR Approach

Finally, we weighed the potential of a visualization employing Adobe AIR[11]. This section sketches the characteristics of Adobe AIR and presents a corresponding visualization application.

Recent trends show the emergence of Rich Internet Applications (RIAs), which run in the Web browser and offer powerful desktop application functionality [Sch08]. Built on common Internet technologies, like JavaScript [Fla06] and Adobe Flash[12], RIAs facilitate the development of platform-independent online applications. Famous representatives of RIAs are the mail client Google Mail[13] and the instant messenger Meebo[14].

Exploiting the advantages of platform-independent application development, Adobe offers the opportunity to run Flash-based RIAs outside the browser. The underlying Adobe Flex framework[15] combined with Adobe AIR facilitates the generation of powerful RIAs that can be

[11] http://www.adobe.com/products/air/
[12] http://www.adobe.com/products/flash/
[13] http://mail.google.com/
[14] http://www.meebo.com/
[15] http://www.adobe.com/products/flex/

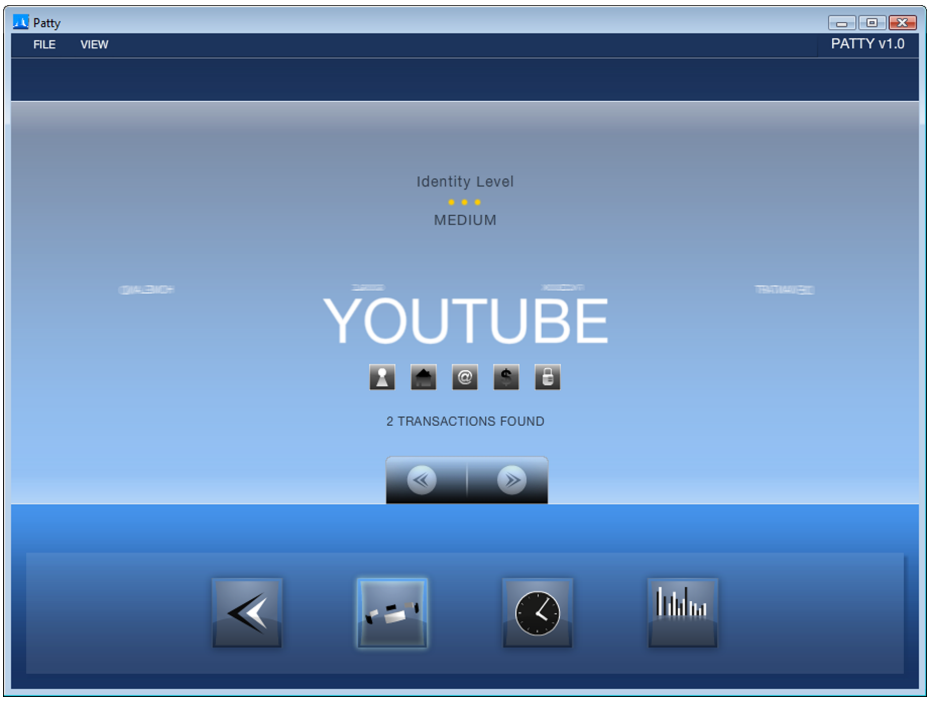


Figure 9.15: Adobe AIR – Carousel View

deployed as desktop applications [Wid08]. In order to run these applications, the installation of the Adobe AIR runtime environment is required, which is available for Microsoft Windows, Mac OS X and Linux. Offline RIAs facilitate the design of innovative graphical user interfaces that differ from standardized elements of operating systems.

In our effort to identify the most suitable visualization of logged personal data transactions, we developed a corresponding Adobe AIR application. The program consists of four parts including a built-in tutorial. After the start four buttons lead to the respective parts of the program. Mouse-over events trigger animated help texts in the main window. After a button is clicked, an animated transition leads to the selected view.

The first view focuses on service providers and is shown in Figure 9.15. The view enables users to get a quick overview of providers the user interacted with. Service providers are lined up in an animated 3D carousel that can be rotated by using the arrow buttons. While unselected service providers in the background are still visible to the user, the carousel view displays detailed information about the selected service provider. This information includes an indicator about the size of the used partial identity with regard to the user's full identity. Below the service provider name five icons indicate the data type disclosure from particular data groups. Furthermore, the total number of transactions is presented.

Figure 9.16 shows a chronological view of data disclosures, which aggregates transactions by calendar month or day. Similar to the chronological view in Section 9.4.3.1, the user is able to define the observation period of the timeline. The main window allows users to expand individual days or months. The expansion triggers an animation that unwinds the stacked

Figure 9.16: Adobe AIR – Chronological View

personal data transactions and shows additional information like the disclosure time and the used process.

A final view focuses on disclosed personal data. A pie diagram presents the aggregated amount and sensitivity of personal data disclosures for each service provider. If the user clicks a slice of the diagram, more detailed information is displayed. All transitions of the user interface are animated.

The presented Adobe AIR application shows a promising way of visualizing personal data transactions. The underlying Flash framework allows the design of intuitive user controls that feature impressive animations. However, Adobe Flex and Adobe AIR are considered as relatively young technologies. The application development proved the early maturity stage of Adobe Flex 4.0 and Adobe AIR, which both show a small development community at this point. These caveats as well as the positive user feedback on graphs led us to choose a conventional visualization approach for the Data Disclosure Log component.

9.6 Summary

The Data Disclosure Log component of the proposed user-centric privacy infrastructure logs and visualizes past personal data disclosures. Dividing the component into a logging tool and a visualization tool, this chapter first collects requirements of both tools, before existing approaches in these areas are evaluated.

We introduce a transaction logging tool that identifies personal data types and used processes

in the Web browser. The data type identification process follows a hybrid approach. First, the tool analyzes attributes of Web form fields using pre-defined lists of keywords. In an additional step, the user input is compared with regular expressions as well as a user profile. After the submission of personal data, the developed browser extension stores the detected information in a transaction log file, which can be checked and edited at any time.

The next section introduces a versatile tool for the visualization of recorded personal data disclosures. The tool offers multiple views that process and display entity types of personal data transactions from many perspectives. A chronological view aligns service provider transactions on a timeline and allows for a temporal analysis of data disclosures. A graph view facilitates the visualization of various relations between entities, enabling users to dynamically generate a self-defined graph. The tool also displays the presence of linkability threats in a dedicated view. Supplemental views complement the visualization needs of users.

Conducted user tests of the logging tool as well as the visualization tool proved the usability and user acceptance of our design efforts.

In the remainder of the chapter we discuss two alternative solutions for a suitable visualization tool. We present and evaluate the potentials of a 3D graph-based implementation as well as an Adobe AIR application. While a 3D graph allows for richer, more complex representations, studies reveal that users prefer intuitive 2D interfaces to complex 3D visualizations. We also proved that an Adobe AIR application has the potential to provide a portable solution offering innovative non-standardized user interfaces.

Part III

Implementation

Chapter 10

Prototype

In this chapter we lay out prototypical details of the privacy infrastructure introduced in the previous part of this work. Based on the user-centric privacy architecture proposed in Chapter 5, we present the system architecture including the designed components. After describing the definition and the central administration of the global privacy vocabularies, we provide implementational details of each component. We continue with an overview of the component interaction. Finally, we present the installer of the local privacy infrastructure, which assists users in setting up the developed local privacy components on the user side. The installer is available for download on the privacy community Web site[1].

10.1 System Architecture

While Part II of this work discussed the purposes, the technical characteristics and the user interfaces of each component, this section concentrates on the component distribution on system level, which is depicted in Figure 10.1.

Similar to the conceptual architecture in Chapter 5.1 (see Figure 5.1), the online privacy community marks the central component of the system architecture, representing an open information source of privacy-relevant service provider information, which is maintained in a dynamic Web front-end (see Chapter 6). Web service interfaces facilitate the flexible integration of these valuable data into the local privacy components. Likewise, the Privacy Preference Generator component, built for the user-understandable capture of individual privacy preferences (see Chapter 7), and the Privacy Agent component, a browser extension that presents information relevant for a personal data disclosure decision (see Chapter 8), form respective components on system level.

The Data Disclosure Log component, which is responsible for the logging and visualization of past personal data disclosures, is divided into two individual tools. While the FireLox browser extension detects and records personal data transactions in the Web browser (see Chapter

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

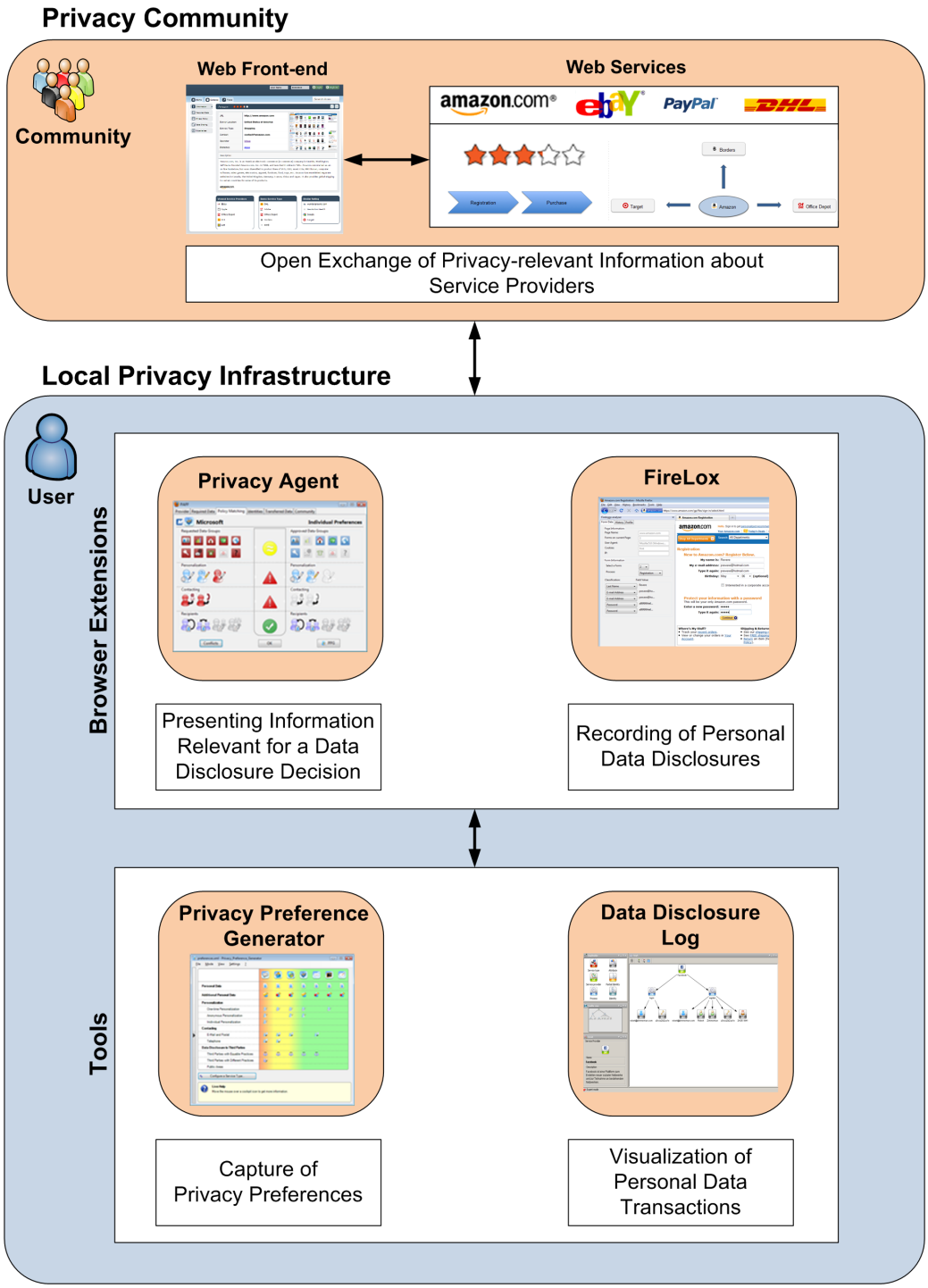


Figure 10.1: System Component Distribution of the Privacy Infrastructure

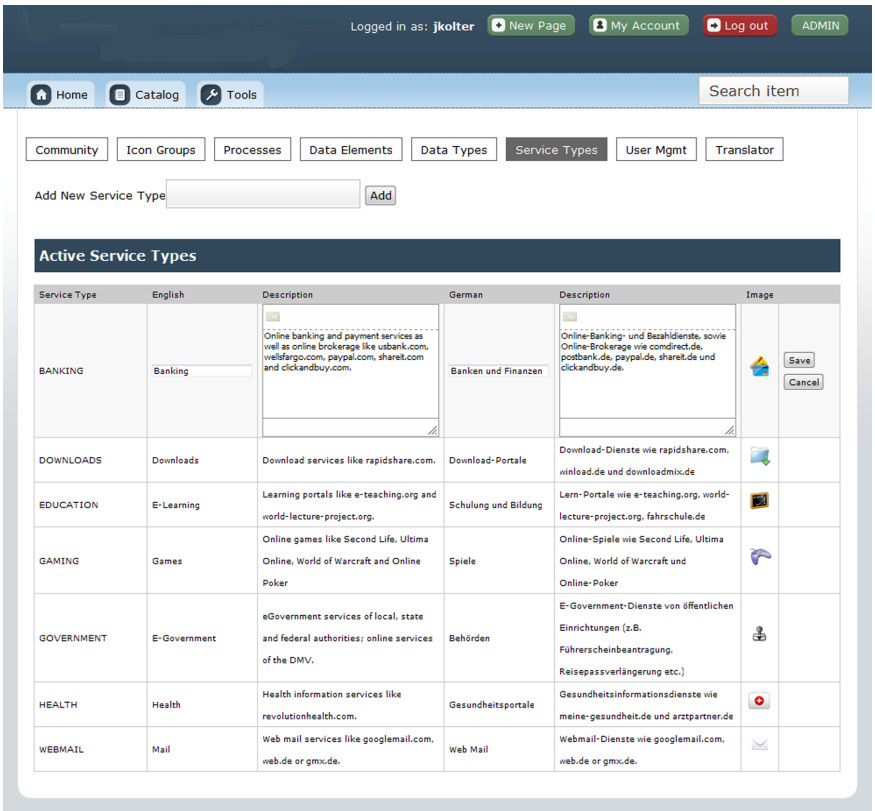


Figure 10.2: Vocabulary Maintenance in the Privacy Community

9.3), the Data Disclosure Log visualization tool offers multiple views on these logged data disclosures (see Chapter 9.4).

10.2 Privacy Vocabularies

As the components of the introduced privacy infrastructure exchange information that refers to specific data types and service types, global vocabularies used by all five system components are required.

10.2.1 Administration

Analyzing the architecture of the presented privacy infrastructure, we chose the central privacy community as suitable place for the administration of the privacy vocabularies. The characteristics of the privacy community allow for the central maintenance and storage of this information, which can be added, changed and removed by administrators of the privacy community. Figure 10.2 shows a screenshot that depicts the vocabulary maintenance in the community.

In the following, we sketch the content of the defined data type and service type vocabularies and discuss their configuration and provision within the privacy infrastructure.

10.2.2 Data Type Definitions

As personal data mark the protection goal of this work, the vocabulary defining individual data types represents the foundation of all information exchange within the privacy infrastructure. Catering the needs of all components, we specified a three-level hierarchy of personal data.

The most fine-grained units of personal data are individual *data types*. These fundamental information pieces refer to individual personal data types of the user and are used by all five privacy components. An example of such a data type is *Zip Code*, which refers to the five digit postal code of the user's address.

Aggregating certain groups of data types, we define *data elements*, which represent the atomic personal data pieces in the privacy community, when the required amount of personal user data for a specific process is maintained (see Chapter 6.3.2). A typical example of a data element is *Address*, including the data types *Street*, *Street Number*, *City*, *State*, *Zip Code* and *Country*.

Finally, we group all data types into ten distinct *data groups* and provide a characteristic icon for each data group, which represents all included data types. Data groups are used for the graphical presentation of data type sets in the Privacy Preference Generator (see Chapter 7.4). The Privacy Agent uses data groups to signal requested and approved data types in the policy matching tab (see Chapter 8.3.3). The Privacy Agent also utilizes data groups for the presentation of the required amount of personal data and the list of already disclosed data (see Chapters 8.3.2 and 8.3.5, respectively). The Data Disclosure Log resorts to data groups for the dynamic generation of data type nodes in the graph view (see Chapter 9.4.3.2) and in the details window of the service provider view and the record slider view (see Chapter 9.4.3.4). An exemplary data group is *Phone Data*, which includes data types that refer to the user's *Landline* phone number, *Mobile* phone number and *Fax* number.

In addition to identifiers of data types, data elements and data groups, we maintain the following attributes and classifications:

- the English and German name of each data type
- the English and German name of each data element
- the English and German name of each data group
- the physical icon location of each data group
- data types belonging to each data element
- data types belonging to each data group

After the definition in the privacy community, the vocabulary data are transformed into an XML file, which is distributed to the privacy components on the user side. As data elements are exclusively used by the privacy community, the built XML file is limited to data types and icon groups along with their attributes and relations.

Listing 10.1 shows an extract of the resulting `datagroups.xml` file. The listing presents the definition of the data group `Name Data`, including its ID, the physical location of its icon and the German and English translation. Subelements of the data group (`icon_group`) consist of data type definitions, which contain their IDs and German and English translations.

```
<?xml version="1.0" encoding="UTF-8"?>
<icon_groups xmlns="http://www-ifs.uni-regensburg.de/Privacy">
  <icon_group id="NameData" icon="images/data/NameInfo.png">
    <locale id="de">Namensdaten</locale>
    <locale id="en">Name Data</locale>
    <DataType>
      <Family>
        <locale id="de">Nachname</locale>
        <locale id="en">Last Name</locale>
      </Family>
      <Given>
        <locale id="de">Vorname</locale>
        <locale id="en">Given Name</locale>
      </Given>
      <Maidenname>
        <locale id="de">Mädchenname</locale>
        <locale id="en">Maiden Name</locale>
      </Maidenname>
      <Middle>
        <locale id="de">Zweiter Vorname</locale>
        <locale id="en">Middle Name</locale>
      </Middle>
      <Nickname>
        <locale id="de">Spitzname</locale>
        <locale id="en">Nickname</locale>
      </Nickname>
      <Prefix>
        <locale id="de">Vorsatz</locale>
        <locale id="en">Prefix</locale>
      </Prefix>
      <Suffix>
        <locale id="de">Suffix</locale>
        <locale id="en">Suffix</locale>
      </Suffix>
    </DataType>
  </icon_group>
  ...
</icon_groups>
```

Listing 10.1: Extract of the file `datagroups.xml`

In the process of defining the presented data type vocabulary, we initially resorted to the P3P basic data set [CDE+06], which primarily focuses on online data and demographic data. In an effort to offer attributes of a complete user model, we added elements of the UbisWorld ontology[2] that aims for a comprehensive real world simulation, focusing on ubiquitous computing and user modeling.

Appendix A.1 lists the complete data type vocabulary. For data groups like *Demographic Data*

[2] http://www.ubisworld.org/

that contain numerous data types, we facilitate the definition of an additional hierarchy level, which is defined by the `hid` attribute. This additional hierarchy level is utilized by the Privacy Preference Generator for the individual selection of data type amounts in the advanced mode (see Chapter 7.4.3).

10.2.3 Service Type Definitions

In the context of this work, we divide service providers into pre-defined service types, which correspond to their offered service category. The according service provider classification is collaboratively maintained in the privacy community (see Chapter 6.3.1).

This service provider classification enables the Privacy Preference Generator to capture individual privacy preferences for each service type (see Chapter 7.4.2). The Privacy Agent looks up the service type of a service provider to select the applicable set of privacy preferences (see Chapter 8.3.3). In addition, service types are used for the aggregation of service providers in the graph view of the Data Disclosure Log (see Chapter 9.4.3.2).

Similar to the data type vocabulary in the previous section, we maintain supplemental information for each service type:

- the English and German name
- the English and German description
- the physical location of the service type icon
- the recommended data set (amount of data types appropriate for disclosure)
- the corresponding primary purpose elements of the P3P specification

Each service type definition contains a set of personal user data that is appropriate for disclosure. This data set is recommended by the Privacy Preference Generator when the amount of personal data types is captured (see Chapter 7.4.3). If no service type information about a service provider is available in the privacy community, the Privacy Agent identifies the service type by matching detected primary purpose elements in a published P3P policy with the maintained primary purpose elements of the service type definitions.

The resulting XML file `services.xml` contains the described information for all service types. The file extract shown in Listing 10.2 demonstrates the file structure by listing the properties of the service type *Shopping*.

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <service id="ONLINESHOPPING">
3     <icon filename="images/services/shopping.png"/>
4     <label>
5       <lang id="de">
6         <name>Einkauf</name>
7         <description>Diverse Einkaufsdienste wie amazon.de, bol.de, musicload.de,
```

```
8                       ebay.de und itunes.de.</description>
9       </lang>
10      <lang id="en">
11        <name>Shopping</name>
12        <description>All kinds of online shopping services like amazon.com,
13                      ebay.com and itunes.com</description>
14      </lang>
15    </label>
16    <mindata>  <!-- Data types appropriate for disclosure -->
17      <Accountnumber/>
18      <Bankcode/>
19      <Bytes/>
20      <City/>
21      <Clickstream/>
22      <Clientevents/>
23      <Clientip/>
24      <Cookies/>
25      <Country/>
26      <Creditcardexpiration/>
27      <Creditcardnumber/>
28      <Creditcardsecuritycode/>
29      <Email/>
30      <Family/>
31      <Given/>
32      <Hostname/>
33      <Housenumber/>
34      <Http/>
35      <Httpmethod/>
36      <Id/>
37      <Interactionrecord/>
38      <Login/>
39      <Password/>
40      <Paymentmethod/>
41      <Postalcode/>
42      <Referer/>
43      <Searchtext/>
44      <Stateprov/>
45      <Statuscode/>
46      <Street/>
47      <Timestamp/>
48      <Uri/>
49      <Useragent/>
50    </mindata>
51    <ppurpose>
52      <survey/>
53      <sales/>
54      <payment/>
55      <marketing/>
56      <delivery/>
57      <charity/>
58      <arts/>
59      <downloads/>
60    </ppurpose>
61  </service>
62  ...
63 </services>
```

Listing 10.2: Extract of the file `services.xml`

Altogether, we defined twelve service types by assuming and aggregating the primary purpose elements of the P3P specification [CDE+06]. Furthermore, we considered the results of selected surveys (see Chapter 7.4.2). Appendix A.2 provides the complete service type definitions.

10.2.4 Configuration

As laid out in the previous sections, the data type and service type vocabularies are centrally administered in the privacy community. For the distribution to the local privacy components, the vocabulary data are transformed into the XML vocabulary files `datagroups.xml` and `services.xml`.

The XML vocabulary files are generated whenever vocabulary information is updated in the administration area of the community. The XML files are stored in the download directory of the privacy community and can be accessed by the local components of the privacy infrastructure. In order to avoid redundant vocabulary downloads, the Privacy Agent downloads the vocabulary files on a regular basis and shares them with the remaining local privacy components, by placing the vocabulary files at a pre-defined location on the user side (see Section 10.5). While the `datagroups.xml` file is utilized by all local privacy components, the `services.xml` file is required by all components but the FireLox browser extension.

Figure 10.3 illustrates the central vocabulary configuration process of the privacy infrastructure.

10.3 Implementation Details

This section provides technical details of the individual component implementations, concentrating on the chosen programming environments, the designed data models as well as the input and output interface definitions.

10.3.1 Privacy Community

The online privacy community offers a central information pool that contains privacy-related information about service providers. These data are collaboratively maintained in a Wikipedia-like Web front-end and can be accessed by the local privacy components via Web service interfaces. Chapter 6 lays out the content and the design of the proposed privacy community.

The back-end of the privacy community employs NuSOAP[3], a PHP-based SOAP toolkit that provides the required Web service functionality for our proposed privacy infrastructure. The Web service interface definitions can be accessed following this link[4].

The SOAP server accesses a MySQL database[5], which stores all service provider information as well as the presented data type and service type vocabularies. Figure 10.4 depicts the data model of the privacy community.

[3] http://sourceforge.net/projects/nusoap/
[4] http://www-ifs.uni-regensburg.de/Privacy/soap_ws/
[5] http://www.mysql.com/

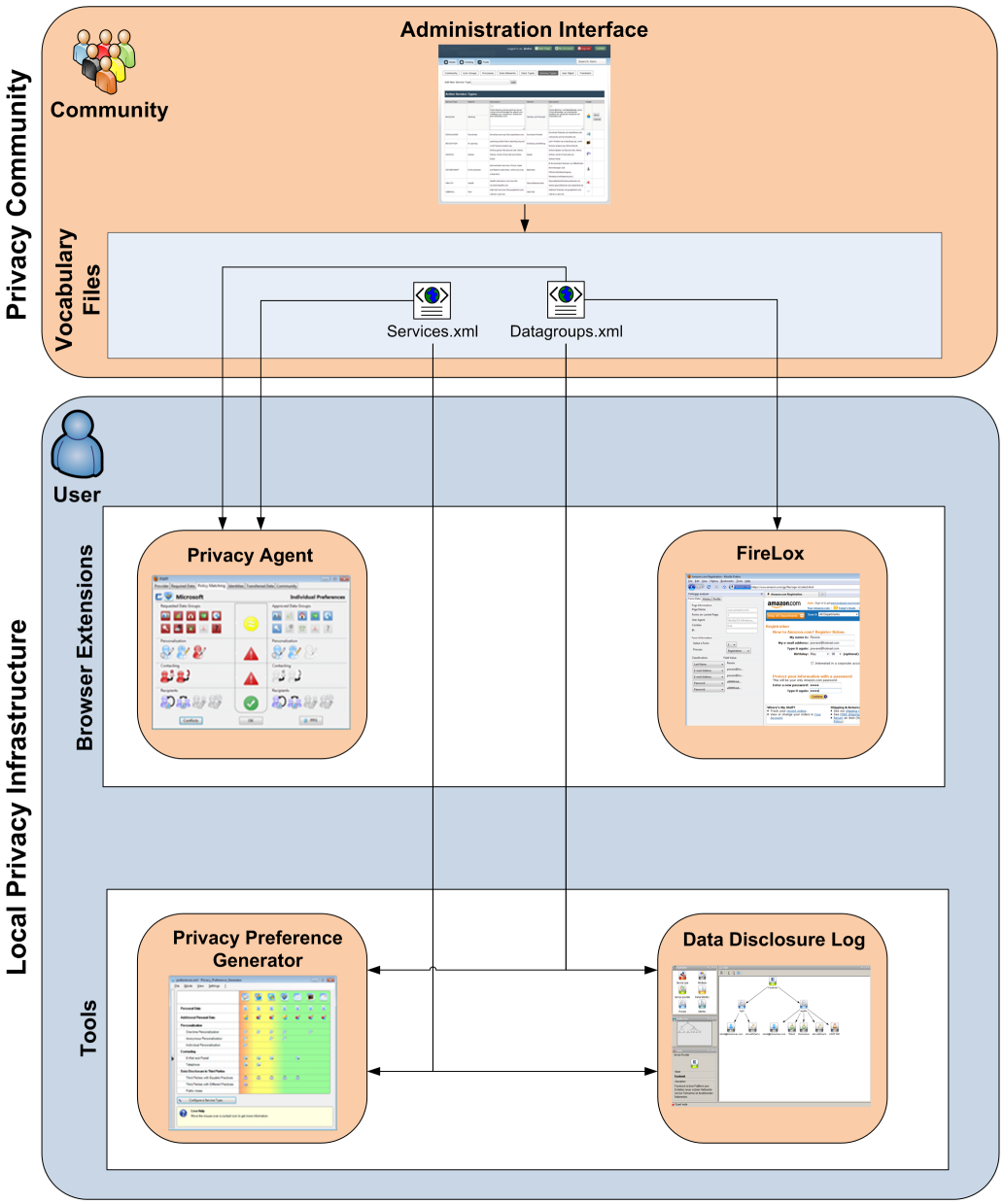


Figure 10.3: Configuration of Global Vocabularies within the Privacy Infrastructure

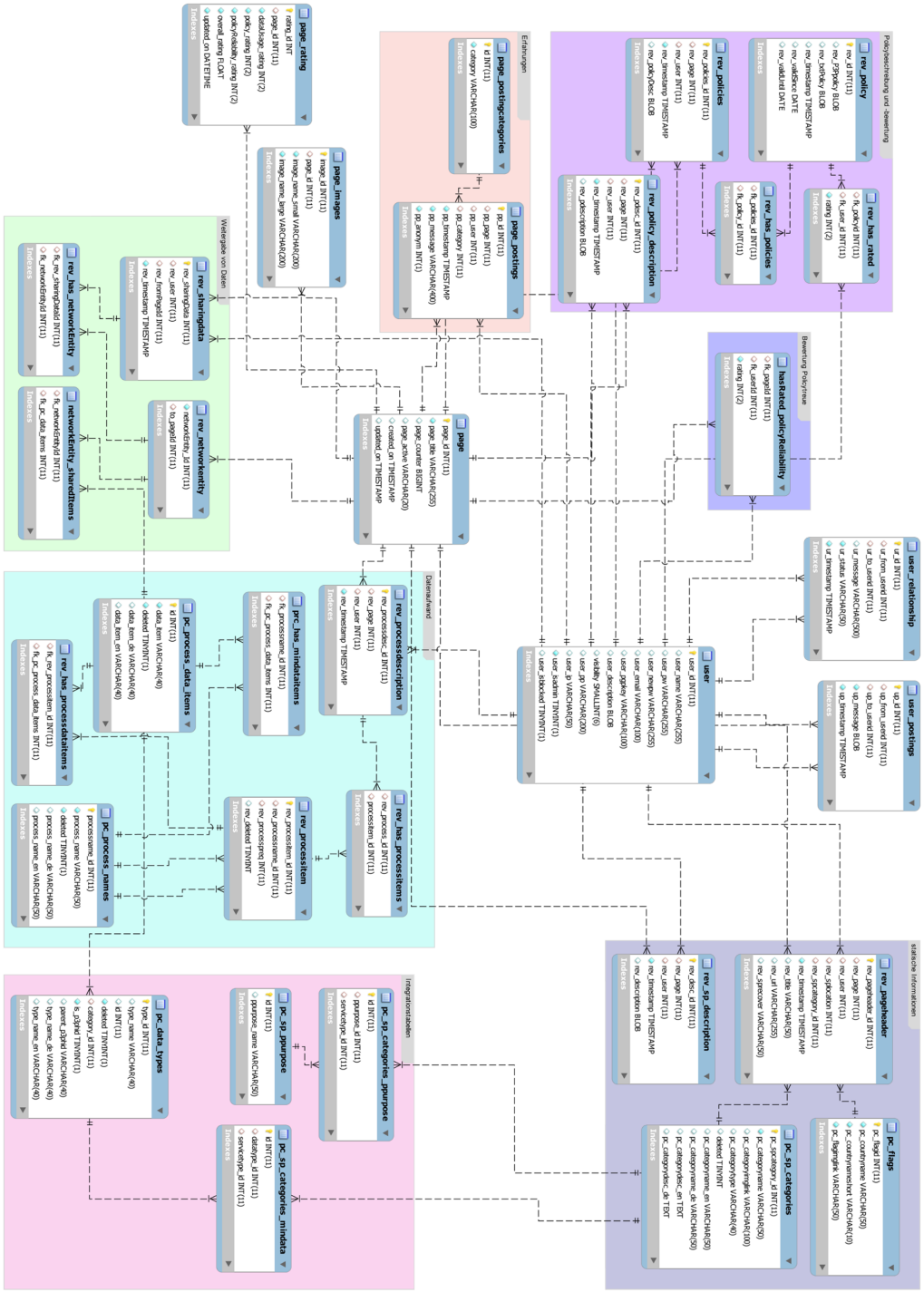


Figure 10.4: Data Model of the Privacy Community

For the Web front-end we utilize the Web application framework CodeIgniter[6]. The PHP-based framework facilitates the MVC-compliant development of dynamic Web applications and allows for the smooth integration of the back-end Web services. On the client side, the JavaScript framework jQuery[7] offers AJAX and dynamic HTML technologies, which provide necessary drag & drop and auto-complete functions, overlays, as well as high performance DOM parsing.

10.3.2 Privacy Preference Generator

The Privacy Preference Generator is designed as stand-alone tool that captures individual privacy preferences for up to twelve pre-defined service types. These privacy preferences represent rules under which users are willing to disclose personal data. The concept and design of the proposed Privacy Preference Generator is presented in Chapter 7.

The tool was implemented in Visual Basic using Visual Studio 2008[8]. Hence, the Privacy Preference Generator requires an installed .NET framework[9] version 3.5 at runtime.

After the completion of the privacy preference configuration, the resulting settings are saved in an XML file that contains all captured information. Listing 10.3 shows an extract of an exemplary privacy preference file.

```
1  <?xml version="1.0" encoding="utf-8"?>
2  <privacypreferences xmlns="http://www-ifs.uni-regensburg.de/Privacy">
3    <info>
4      <mode>advanced</mode>
5    </info>
6      <service id="ONLINESHOPPING">
7        <retention>
8          <stated-purpose />
9        </retention>
10       <additionaldata>request-prompt</additionaldata>
11       <data>
12         <Accountnumber />
13         <Bankcode />
14         <Bytes />
15         <City />
16         <Clickstream />
17         <Clientevents />
18         <Clientip />
19         <Cookies />
20         <Country />
21         <Creditcardexpiration />
22         <Creditcardnumber />
23         <Creditcardsecuritycode />
24         <Email />
25         <Family />
26         <Given />
27         <Hostname />
```

[6] http://codeigniter.com/
[7] http://jquery.com/
[8] http://www.microsoft.com/visualstudio/en-us/default.mspx
[9] http://msdn.microsoft.com/en-us/netframework/default.aspx

```
28          <Housenumber />
29          <Http />
30          <Httpmethod />
31          <Id />
32          <Interactionrecord />
33          <Login />
34          <Password />
35          <Paymentmethod />
36          <Postalcode />
37          <Referer />
38          <Searchtext />
39          <Stateprov />
40          <Statuscode />
41          <Street />
42          <Timestamp />
43          <Uri />
44          <Useragent />
45        </data>
46        <purpose>
47          <current />
48          <personalization>
49            <tailoring required="opt-out" />
50            <pseudo-analysis required="opt-in" />
51            <pseudo-decision required="opt-in" />
52          </personalization>
53          <contacting />
54        </purpose>
55        <recipient>
56          <ours required="always" />
57          <same required="always" />
58        </recipient>
59      </service>
60      ...
61    </services>
62  </privacypreferences>
```

Listing 10.3: Extract of an Exemplary Privacy Preference File

The listing demonstrates the configuration of the service type *Shopping*. In particular, the file defines personal data types the user is willing to disclose to a provider of this service type. In addition, approved data usages and third party recipients are listed. The first subelement of the file specifies the chosen complexity mode, which is read and applied when the application is started. Appendix B.1 presents a complete exemplary privacy preference file including seven configured service types.

A configuration file of the Privacy Preference Generator sets the local directory of the vocabulary files and the privacy preference file. The file also specifies the chosen language as well as the URL of the privacy community, which is required for the download of data group and service type icons.

10.3.3 Privacy Agent

The Privacy Agent is implemented as an extension for the Mozilla Firefox browser and provides service provider information that facilitates a more informed decision about personal data disclosures. This information includes service provider data queried from the privacy community, a P3P privacy policy matching with captured privacy preferences, a linkability detection of used partial identities and personal data already disclosed to the visited service provider. Chapter 8 discusses the functionality and the graphical user interface of the Privacy Agent.

As laid out in Chapter 8.3.1, the application logic is implemented in JavaScript [Fla06]. User interfaces are defined with the XML User Interface Language (XUL)[10], which facilitates the dynamic generation of graphical elements.

The identity tab of the Privacy Agent (see Chapter 8.3.4) analyzes the transaction log file of the user (see Section 10.3.4), in order to identify partial identity usages and their linkabilities. The resulting identity class and partial identity definitions are stored in an identity file. An extract of an exemplary identity file is shown in Listing 10.4.

```
1  <?xml version="1.0" encoding="utf-8" ?>
2  <identities xmlns="http://www-ifs.uni-regensburg.de/Privacy">
3    <identity id="default">
4      <label>
5        <lang id="de">
6          <name>Echte Identität</name>
7          <description>Die echte Identität von Paul Revere</description>
8        </lang>
9        <lang id="en">
10         <name>True Identity</name>
11         <description>The true identity of Paul Revere</description>
12       </lang>
13     </label>
14     <partialIdentity id="1">
15       <label>
16         <lang id="de">
17           <name>Teilidentität 1</name>
18           <description>Die Teilidentität 1 wird für die Registrierung bei
19                        offiziellen Stellen und Communities verwendet</description>
20         </lang>
21         <lang id="en">
22           <name>Partial Identity 1</name>
23           <description>Partial identity 1 is used for registration at public
24                        authorities and communities</description>
25         </lang>
26       </label>
27       <partialIdentityInformation>
28         <linkable>
29           <partialIdentity id="2"/>
30         </linkable>
31       </partialIdentityInformation>
32       <serviceProviderInformation>
33         <usedBy>
```

[10] http://planetxul.com/

```
34            <hostname name="http://www.facebook.com" title="Facebook"/>
35            <hostname name="http://www.amazon.com" title="Amazon"/>
36          </usedBy>
37        </serviceProviderInformation>
38        <data>
39          <Given>Paul</Given>
40          <Family>Revere</Family>
41          <Email>prevere@hotmail.com</Email>
42          <Id>prevere</Id>
43          <Bdate>24.05.1971</Bdate>
44          <Gender>male</Gender>
45          <Telephone>617-555-1234</Telephone>
46        </data>
47      </partialIdentity>
48      ...
49    </identity>
50    ...
51  </identities>
```

Listing 10.4: Extract of an Exemplary Identity File

The listed extract defines the first partial identity of the true identity class of the fictional person Paul Revere. The file contains German and English names and descriptions for both the identity class and the partial identity. The partial identity contains information about the linkability to other partial identities, the disclosure to service providers and the involved personal data types. Appendix B.2 presents a complete identity file including five defined partial identities.

The Privacy Agent accesses Web services of the privacy community, in order to display privacy-relevant information of the visited service provider. In addition, the Privacy Agent downloads actual data type and service type icons from the privacy community on a regular basis.

A configuration file defines the locations of the vocabulary files and the local privacy files, the URL of the privacy community as well as the chosen language of the user interface.

10.3.4 FireLox

The Mozilla Firefox extension FireLox enables the recording of personal data disclosures in the Web browser. The logging tool analyzes Web form fields and identifies data types as well as the involved process. The detected information and the user-filled input are stored in an XML log file. The identification algorithm and the user interface of FireLox are presented in Chapter 9.3.

Similar to the Privacy Agent, the application logic is implemented in JavaScript, while the graphical user interface is specified employing XUL.

The resulting transaction log file groups all personal data transactions by service providers, which simplifies the processing for the accessing Privacy Agent and Data Disclosure Log. Listing 10.5 presents an extract of an exemplary transaction log file.

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <xml xmlns="http://www-ifs.uni-regensburg.de/Privacy">
3    <hostname name="http://www.amazon.com" title="Amazon">
```

```
4     <transaction process="Registration" id="1">
5       <dynamic>
6         <Uri>http://www.amazon.com/gp/css/homepage.html</Uri>
7         <Timestamp>1250592961</Timestamp>
8         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9)
9                    Gecko/2008052906 Firefox/3.0</Useragent>
10        <Cookies>true</Cookies>
11      </dynamic>
12      <data>
13        <Miscfields>
14          <Notclassified>lorem</Notclassified>
15          <Notclassified>ipsum</Notclassified>
16        </Miscfields>
17        <Given>Paul</Given>
18        <Family>Revere</Family>
19        <Email>prevere@hotmail.com</Email>
20        <Password>a3cca2b2aa1e</Password>
21        <Bdate>24.05.1971</Bdate>
22        <Addresses>
23          <Postal>
24            <Street>Arlington Road</Street>
25            <Housenumber>1606</Housenumber>
26            <City>Boston</City>
27            <Stateprov>MA</Stateprov>
28            <Postalcode>02101</Postalcode>
29          </Postal>
30        </Addresses>
31      </data>
32    </transaction>
33    ...
34    <transaction process="Purchase" id="3">
35      <dynamic>
36        <Uri>http://www.amazon.com/gp/css/homepage.html</Uri>
37        <Timestamp>1250852161</Timestamp>
38        <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9)
39                   Gecko/2008052906 Firefox/3.0</Useragent>
40        <Cookies>true</Cookies>
41      </dynamic>
42      <data>
43        <Miscfields>
44        </Miscfields>
45        <Email>prevere@hotmail.com</Email>
46        <Password>a3cca2b2aa1e</Password>
47        <Financialdata>
48          <Account type="Credit Card">
49            <Creditcardnumber>1376151d963cad089697872af7dcec48bbd3ae29
50             </Creditcardnumber>
51            <Creditcardexpiration>08/11</Creditcardexpiration>
52            <Creditcardsecuritycode>451</Creditcardsecuritycode>
53          </Account>
54        </Financialdata>
55      </data>
56    </transaction>
57  </hostname>
58  ...
59 </xml>
```

Listing 10.5: Extract of an Exemplary Transaction Log File

The listing demonstrates the logging of two timely separated personal data transactions of the fictional person Paul Revere. Both transactions disclosed personal data to the service provider Amazon. The first transaction was identified as *Registration* process and included the personal data types *Given Name*, *Last Name*, *E-mail Address*, *Password*, *Date-of-Birth* and a postal address. FireLox identified the following transaction as *Purchase* process. During that transaction *E-mail Address* – presumably for identification – *Password* and credit card information have been disclosed to the service provider.

A FireLox configuration file stores the locations of the transaction log file and the data type vocabulary file as well as the preferred language.

10.3.5 Data Disclosure Log

The Data Disclosure Log displays logged personal data transactions employing intuitive visualization techniques. The tool offers multiple views that present past personal data disclosures from different perspectives. Chapter 9.4 introduces the concept and the designed views of the visualization tool.

The stand-alone tool is implemented as Java application and runs platform-independently with every Java 6 Runtime Environment[11]. For the design of the graphical user interfaces we utilize the Java Swing toolkit. The dynamic graph generation is enabled by Visual Library[12].

As discussed in Chapter 9.4.1, we employ a relational database for the integration and storage of available input data that include the transaction log file, the identity file and both vocabulary files. In particular, we utilize the H2 database[13]. All required tables are generated at the start of the program and filled with data of the XML input files. Figure 10.5 shows the designed data model, which caters the needs of the developed visualization views.

In addition to the local privacy and vocabulary files, the database is loaded with selected service provider information queried from the privacy community. The privacy community is also accessed for the download of data type and service type icons, which are used for the dynamic generation of graph node icons.

A configuration file of the Data Disclosure Log specifies the local directory of the transaction log file, the identity file and the vocabulary files. Furthermore, the file sets the URL of the privacy community and the chosen language.

10.4 Component Interaction

After the presentation of the components' implementational details this section summarizes the relations and dependencies of the developed tools by highlighting the information exchange

[11] http://java.sun.com/javase/6/
[12] http://graph.netbeans.org/
[13] http://www.h2database.com/

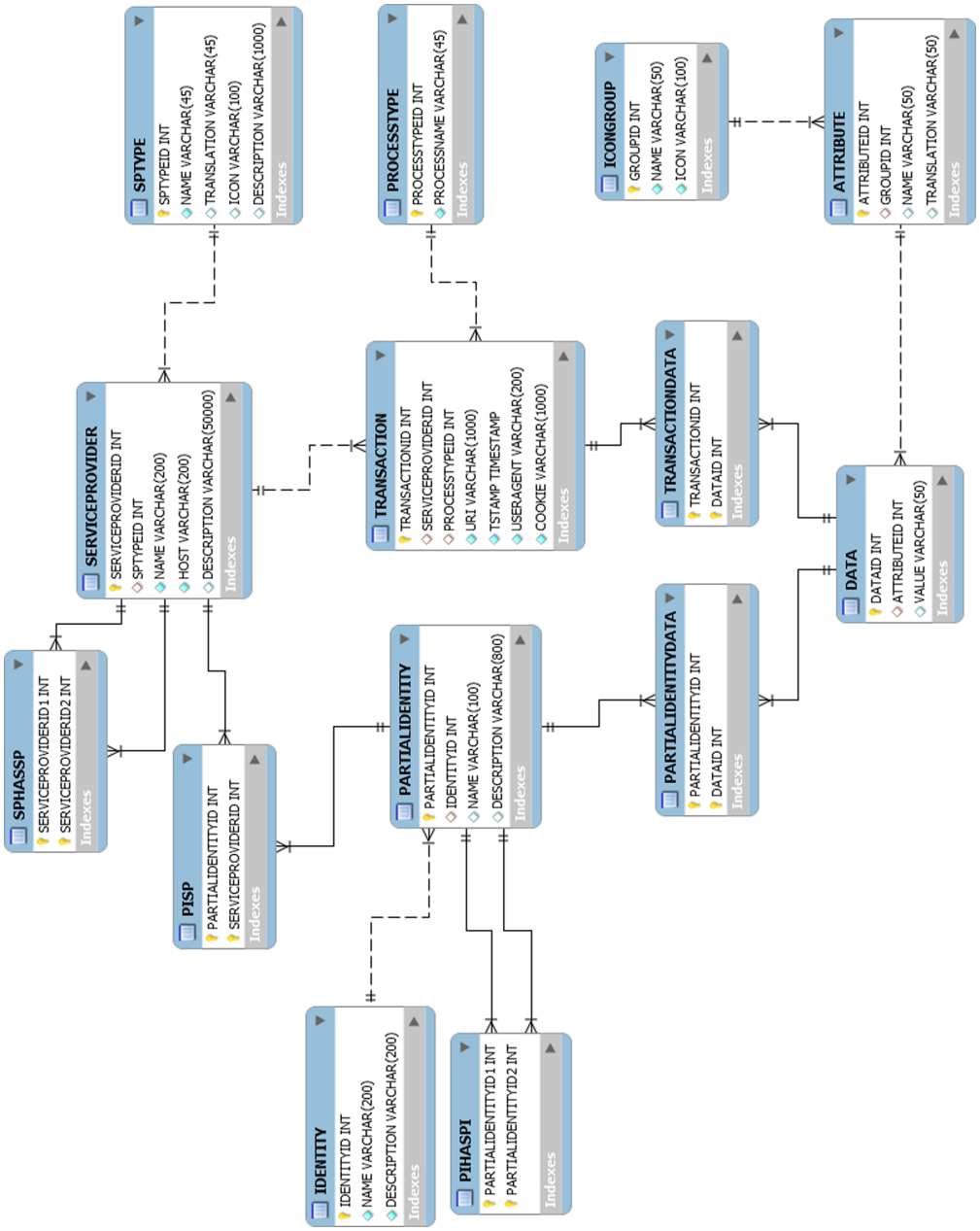


Figure 10.5: Data Model of the Data Disclosure Log

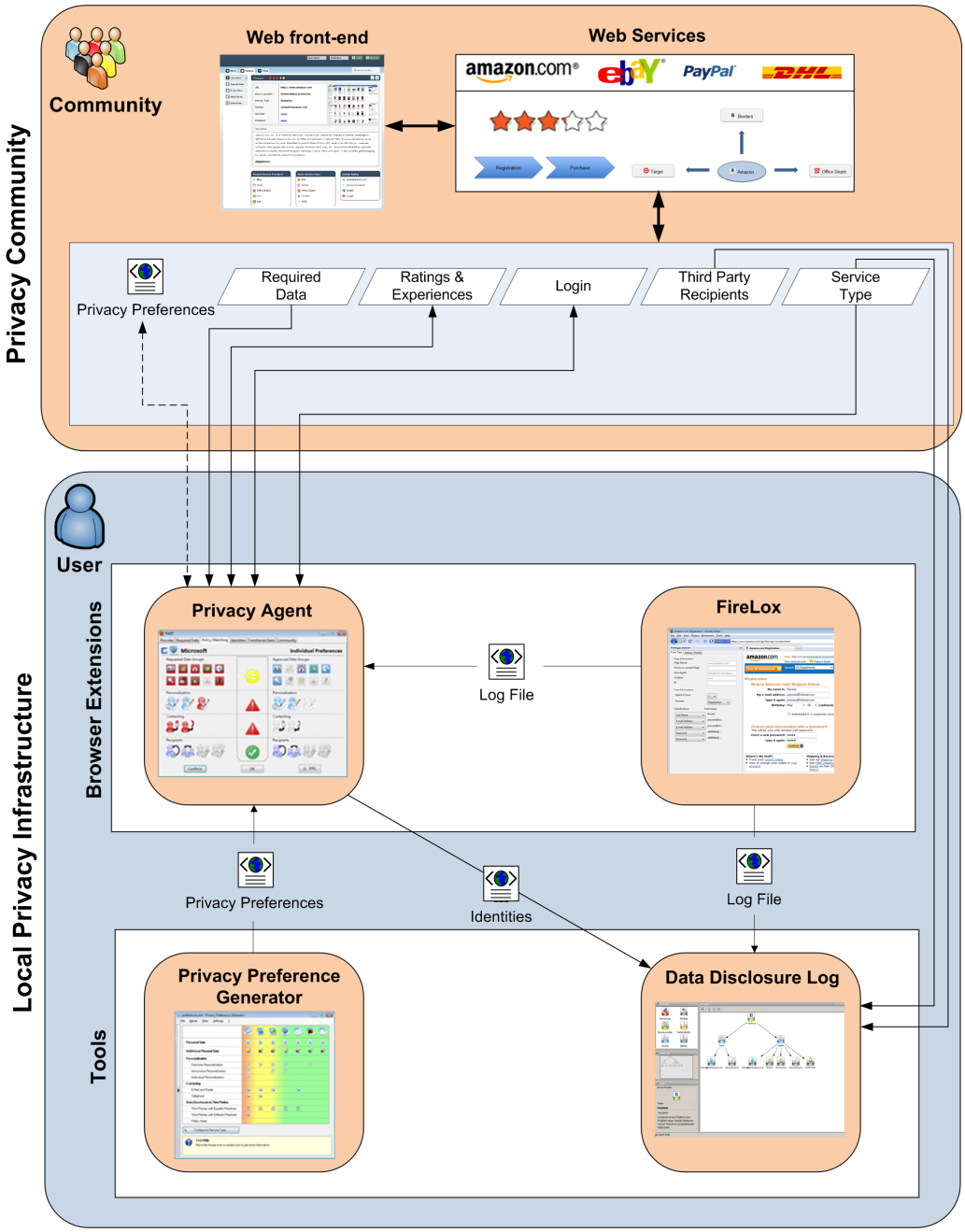


Figure 10.6: Interaction of System Components

within the privacy infrastructure. Figure 10.6 depicts the information flows between the privacy components at runtime.

Privacy preferences are captured by the Privacy Preference Generator and utilized by the Privacy Agent for the privacy policy matching. Registered community users have the option to upload and share privacy preferences with selected users. Likewise, privacy preferences of connected community members can be downloaded. The optional privacy preferences upload and download processes are managed by the Privacy Agent.

The transaction log file generated by the FireLox browser extension is made available to the Data Disclosure Log and the Privacy Agent. The log file represents the essential information source for the visualizations of the Data Disclosure Log. The Privacy Agent processes the transaction log file to display personal data already transferred to a visited service provider.

Finally, the identity file created by the Privacy Agent is utilized by the Data Disclosure Log, in order to relate personal data transactions to used partial identities.

Apart from the data exchange of the four local privacy components, the Privacy Agent and the Data Disclosure Log access and integrate Web services of the privacy community. In addition to the retrieval of static service provider and service type information, the Privacy Agent queries the required amount of personal data as well as available user ratings and experiences. Moreover, the Privacy Agent facilitates the direct submittance of ratings. The login service is a prerequisite for the controlled exchange of privacy preferences. The Data Disclosure Log requests service type information as well as third party recipients, which are required by the graph view and the linkability view, respectively.

10.5 Installer

In an effort to offer the developed privacy infrastructure to a maximum number of users, we created an installer that sets up the four local privacy components and writes the configuration files of the tools.

The installer is offered for download[14] on the privacy community Web site (see footnote) and can be executed on every Microsoft Windows-based machine. Prerequisites of the installation are the Mozilla Firefox 3.5 browser, the .NET 3.5 framework and the Java 6 Runtime Environment. The installer checks the existence of these components, before the installation routine is started.

A wizard guides the user through the installation process. At the beginning, the wizard captures the preset default language of all tools (English or German). In a next step, the user is prompted to select the local privacy components to be installed (see Figure 10.7). The wizard recommends the selection of all four components. At this point the user can opt to copy sample privacy files that allow the instant demonstration of the tool's capabilities. Subsequently, the user is requested to choose the installation directory of the privacy components. Simplifying the

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

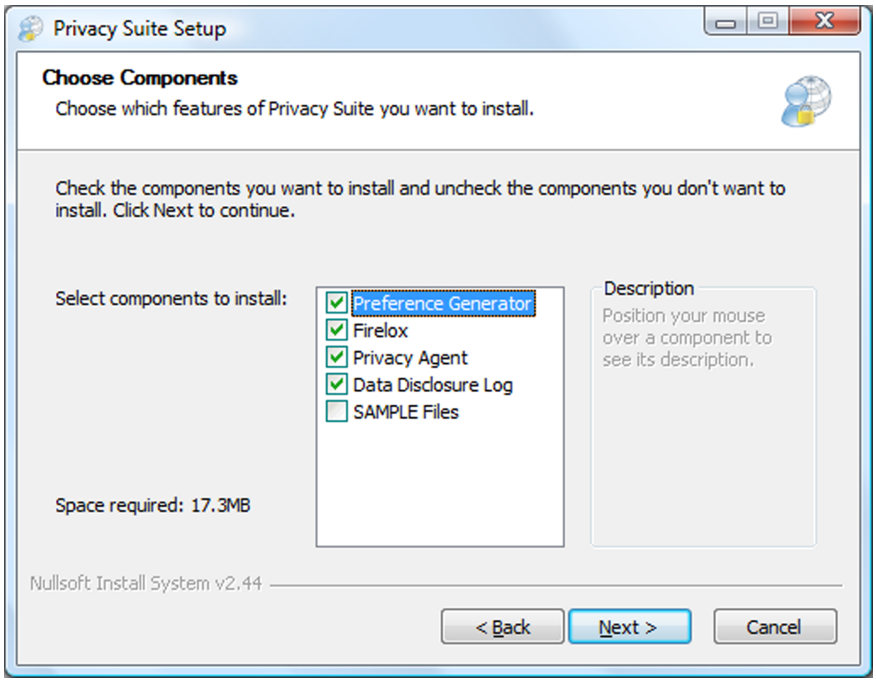


Figure 10.7: Installation Wizard – Component Selection

exchange of local privacy information, the user is also asked to select a personal privacy directory for the storage of the privacy preference file, identity file and transaction log file. The same directory is used to place the downloaded data type and service type vocabulary files.

After the completion of the wizard the selected browser extensions and stand-alone tools are installed. Depending on the chosen settings, the configuration files of the components are generated. If no sample privacy files are copied, an empty privacy preference file, identity file and transaction log file are created in the personal privacy directory. The installer adds desktop icons for the Privacy Preference Generator and the Data Disclosure Log. After the first start of the Firefox browser, the user is informed about the added extensions. The first start of the Web browser also triggers the Privacy Agent to download and update the vocabulary files in the personal privacy directory.

As a result, a fully operational local privacy infrastructure is set up, preconfigured and ready for usage.

Chapter 11

Conclusions

The landscape of the World Wide Web with all its versatile services heavily relies on the disclosure of private user attributes. Unfortunately, the growing amount of personal data collected by service providers poses a significant privacy threat for Internet users. Targeting privacy in the context of personal data disclosures, this work introduced and developed solutions for the provider-independent protection of personal data, putting a special focus on usability and user-friendliness. In the following we recapitulate the contributions of this work by summarizing the results of the research questions we phrased in Chapter 1.2.

- **What are digital identities and how are they managed in the World Wide Web?**
  In Chapter 2 we lay out definitions and the use of identities and partial identities in the World Wide Web. While an identity includes all properties and characteristics of the user, a partial identity contains a subset of personal user data and represents the user on the provider side. Partial identities are used intuitively, when the user interacts with service providers. The rising number of cooperating service providers increases the possibility that linkable partial identities are exchanged among connected service providers and merged to a larger partial identity of the user. Identity management systems enable users to control the disclosure of partial identities more consciously. While the current trend leads to user-centric identity management, according systems do not offer sufficient mechanisms for privacy protection.

- **To what extent do service providers collect personal data and what is the reaction of users?**
  After the definition of the protection goal, Chapter 3 shifts the focus towards privacy aspects in the context of personal data disclosures. We highlight the incentives of service providers to collect personal user data and point out the limited user protection provided by effective privacy legislations. We show the growing privacy concerns of Internet users and present the shift towards more sensitive privacy attitudes. Furthermore, we acknowledge mismatches of self-expressed privacy preferences and the observed user behavior. Finally, we identify the requested kind of personal data, the intended use and third party recipients as most important factors that impact a disclosure decision of users.

- **How can existing privacy-enhancing technologies protect personal user data?**
  Chapter 4 provides a detailed overview of privacy policy languages and existing privacy solutions that are designed to protect personal user data. In spite of its frequently cited deficiencies, the P3P specification in combination with APPEL offers the only complete framework capable of defining a machine-readable privacy policy and matching it with privacy preferences of the user. However, more flexible policy frameworks like XACML and EPAL bear the potential to offer superior expressiveness and evaluation capabilities in the future. A wide range of existing privacy tools like policy viewers, privacy seals and privacy agents have been developed to protect potential, actual and past personal data disclosures. While privacy agents and logging tools offer the highest value for users, we found that the described tools are barely used by the vast majority of Internet users.

- **What are the usability requirements of privacy-enhancing technologies?**
  In Chapter 4.3 we conclude that usability deficits are a major reason for the lacking acceptance of the discussed privacy tools. In an effort to identify usability requirements of users, we conducted several user studies, which are outlined in Chapter 5.3. Initial studies involved questionnaire and online surveys that delivered valuable input for our early prototypes. These prototypes underwent two extensive user tests, which involved the completion of tasks as well as face-to-face interviews. The results of the timely separated user tests revealed remaining usability weaknesses of our intermediate versions and guaranteed the quality of the final solutions.

- **How can privacy management be executed more provider-independently?**
  In addition to usability deficits, the dependency on service providers prevents a widespread usage of existing privacy tools that protect personal user data. Considering the apparent conflict of interests between users who are seeking to protect their privacy and service providers that rely on valuable user data, Chapter 5 introduces a user-centric privacy architecture, which facilitates the provider-independent exchange of privacy-related information about service providers. The privacy architecture includes a collaborative privacy community as well as three local privacy tools that cater for the usability requirements of average Internet users.

- **What are the content and design requirements of a collaborative privacy community?**
  Introduced in Chapter 6, the privacy community represents the central element of our proposed privacy architecture. In order to facilitate a provider-independent privacy management, the privacy community provides a platform for the open exchange of privacy-related information relevant for the protection of personal data. In addition to static service provider information, a privacy community should maintain supplemental provider information that is valuable for Internet users. This information includes the amount of personal data required for the execution of a particular service of a provider. Moreover, the community should store third parties the service provider shares collected

personal data with as well as historic and effective privacy policies. In this context, the privacy community should enable users to rate a privacy policy and the adherence to that policy. Finally, an open privacy community should facilitate the submittance of individual user experiences and the selective exchange of privacy preferences.

We developed a user-friendly Web front-end that enables a Wikipedia-like maintenance of service provider articles that contain the proposed privacy-relevant data. In addition to the front-end, the privacy community offers all stored information pieces via Web service interfaces, allowing local privacy components to access and integrate required information on demand.

- **How can privacy preferences of the user be captured and fine-tuned in a user-friendly way?**
  In Chapter 7 we introduce a usable tool that captures conditions under which users are willing to disclose personal data. Considering the varying level of privacy experience among Internet users, the Privacy Preference Generator offers two different complexity modes. Moreover, we acknowledge the goal-oriented disclosure behavior of users and facilitate the definition of individual privacy preferences for up to twelve pre-defined service types. The designed configuration wizard offers clear user interfaces using understandable language and supportive icons. A privacy cockpit presents an overview of the configured settings as well as intuitive feedback. We show that automatic approaches are inferior to our developed user-understandable configuration tool.

- **How can a privacy agent in the Web browser improve the user's basis for decision-making, when personal data are about to be disclosed?**
  Chapter 8 presents a developed browser extension that provides valuable information, when the user is about to disclose personal data to a visited service provider. This information includes selected data queried from the privacy community as well as the results from matching the service provider's privacy policy with user's previously defined privacy preferences. The offered information is complemented by a mechanism that observes the usage of partial identities as well as a list of already disclosed personal data. Embedded into the Web browser, the introduced Privacy Agent displays the composed information in a clear and understandable way. Contributing to consistency, we reused graphical elements of the Privacy Preference Generator.

- **How can historic transactions of personal data be stored and visualized in a user-comprehensible manner?**
  Chapter 9 proposes a usable solution for the recording and visualization of past personal data disclosures. We introduce a browser extension that detects and stores personal data submitted to service providers in the Web browser. In order to identify the transfer of specific personal data types, a hybrid approach employs a keyword analysis of Web form fields and an evaluation of the user input using regular expressions and a pre-defined user profile. An additional tool displays the resulting transaction log using intuitive

visualization techniques. A chronological view facilitates the temporal analysis of past personal data disclosures. A flexible graph-based view enables the user-defined visualization of relations between involved entity types. Moreover, a linkability view visualizes disclosures of linkable partial identities in corporate networks. We prove that a 3D graph approach and an Adobe AIR approach show innovative visualization schemes, but do not provide the user acceptance and maturity of our proposed solution.

All introduced components of the proposed privacy architecture have been prototypically implemented (see Chapter 10 for implementation details) and are available for download[1]. The privacy community facilitates the central administration of data type and service type vocabularies, which are distributed to the local privacy components.

The elaborated solutions realize *provider independence* and provide *usable user interfaces*, which both have been identified as main shortcomings of existing privacy-enhancing technologies. Hence, the developed privacy infrastructure, that is fully operational, makes a significant contribution towards the broad usage and acceptance of tools that protect personal user data.

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

Appendix A

Vocabularies

A.1 Data Type Vocabulary

```
<?xml version="1.0" encoding="UTF-8"?>
<icon_groups xmlns="http://www-ifs.uni-regensburg.de/Privacy">
  <icon_group id="NameData" icon="images/data/NameInfo.png">
    <locale id="de">Namensdaten</locale>
    <locale id="en">Name Data</locale>
    <DataType>
      <Family>
        <locale id="de">Nachname</locale>
        <locale id="en">Last Name</locale>
      </Family>
      <Given>
        <locale id="de">Vorname</locale>
        <locale id="en">Given Name</locale>
      </Given>
      <Maidenname>
        <locale id="de">Mädchenname</locale>
        <locale id="en">Maiden Name</locale>
      </Maidenname>
      <Middle>
        <locale id="de">Zweiter Vorname</locale>
        <locale id="en">Middle Name</locale>
      </Middle>
      <Nickname>
        <locale id="de">Spitzname</locale>
        <locale id="en">Nickname</locale>
      </Nickname>
      <Prefix>
        <locale id="de">Vorsatz</locale>
        <locale id="en">Prefix</locale>
      </Prefix>
      <Suffix>
        <locale id="de">Suffix</locale>
        <locale id="en">Suffix</locale>
      </Suffix>
    </DataType>
  </icon_group>
  <icon_group id="LoginData" icon="images/data/LoginInfo.png">
    <locale id="de">Zugangsdaten</locale>
    <locale id="en">Login Data</locale>
    <DataType>
      <Biometricinformation>
        <locale id="de">Biometrische Informationen</locale>
        <locale id="en">Biometric Information</locale>
      </Biometricinformation>
      <Cert>
        <locale id="de">Zertifikat</locale>
        <locale id="en">Certificate</locale>
      </Cert>
      <Dna hid="Biometricinformation">
        <locale id="de">DNS</locale>
        <locale id="en">DNA</locale>
      </Dna>
      <Facegeometry hid="Biometricinformation">
        <locale id="de">Gesichtsgeometrie</locale>
        <locale id="en">Face Geometry</locale>
      </Facegeometry>
      <Fingerprint hid="Biometricinformation">
        <locale id="de">Fingerabdruck</locale>
        <locale id="en">Fingerprint</locale>
      </Fingerprint>
      <Handgeometry hid="Biometricinformation">
        <locale id="de">Handgeometrie</locale>
        <locale id="en">Hand Geometry</locale>
      </Handgeometry>
      <Id>
        <locale id="de">Benutzername</locale>
        <locale id="en">User Name</locale>
      </Id>
      <Iris hid="Biometricinformation">
```

```
70          <locale id="de">Iris-Scan</locale>
71          <locale id="en">Iris Scan</locale>
72        </Iris>
73        <Key>
74          <locale id="de">Schlüssel</locale>
75          <locale id="en">Key</locale>
76        </Key>
77        <Keystrokedynamics hid="Biometricinformation">
78          <locale id="de">Tippverhalten</locale>
79          <locale id="en">Keystroke Dynamics</locale>
80        </Keystrokedynamics>
81        <Login>
82          <locale id="de">Login</locale>
83          <locale id="en">Login</locale>
84        </Login>
85        <Password>
86          <locale id="de">Passwort</locale>
87          <locale id="en">Password</locale>
88        </Password>
89        <Signature>
90          <locale id="de">Unterschrift</locale>
91          <locale id="en">Signature</locale>
92        </Signature>
93        <Vein hid="Biometricinformation">
94          <locale id="de">Venen-Scan</locale>
95          <locale id="en">Vein Scan</locale>
96        </Vein>
97        <Voice hid="Biometricinformation">
98          <locale id="de">Stimme</locale>
99          <locale id="en">Voice</locale>
100       </Voice>
101     </DataType>
102   </icon_group>
103   <icon_group id="DemographicData" icon="images/data/DemoInfo.png">
104     <locale id="de">Demographische Daten</locale>
105     <locale id="en">Demographic Data</locale>
106     <DataType>
107       <Age hid="Catpersonal">
108         <locale id="de">Alter</locale>
109         <locale id="en">Age</locale>
110       </Age>
111       <Agegroup hid="Catpersonal">
112         <locale id="de">Altersgruppe</locale>
113         <locale id="en">Age Group</locale>
114       </Agegroup>
115       <Ancestry hid="Catpersonal">
116         <locale id="de">Abstammung</locale>
117         <locale id="en">Ancestry</locale>
118       </Ancestry>
119       <Bdate hid="Catpersonal">
120         <locale id="de">Geburtstag</locale>
121         <locale id="en">Date-of-Birth</locale>
122       </Bdate>
123       <Birthplace hid="Catpersonal">
124         <locale id="de">Geburtsort</locale>
125         <locale id="en">Place of Birth</locale>
126       </Birthplace>
127       <Bodyheight hid="Catappearance">
128         <locale id="de">Körpergröße</locale>
129         <locale id="en">Body Height</locale>
130       </Bodyheight>
131       <Bodyweight hid="Catappearance">
132         <locale id="de">Körpergewicht</locale>
133         <locale id="en">Body Weight</locale>
134       </Bodyweight>
135       <Catappearance>
136         <locale id="de">Erscheinungsbild</locale>
137         <locale id="en">Appearance</locale>
138       </Catappearance>
139       <Catavocationandeducation>
140         <locale id="de">Beruf und Bildung</locale>
141         <locale id="en">Profession and Education</locale>
142       </Catavocationandeducation>
143       <Catfamily>
144         <locale id="de">Familie</locale>
145         <locale id="en">Family</locale>
146       </Catfamily>
147       <Catlanguage>
148         <locale id="de">Sprachen</locale>
149         <locale id="en">Languages</locale>
150       </Catlanguage>
151       <Catpersonal>
152         <locale id="de">Persönliches</locale>
153         <locale id="en">Personal</locale>
154       </Catpersonal>
155       <Childrenstatus hid="Catfamily">
156         <locale id="de">Kinder</locale>
157         <locale id="en">Children Status</locale>
158       </Childrenstatus>
159       <Clothingsize hid="Catappearance">
160         <locale id="de">Kleidergröße</locale>
161         <locale id="en">Clothing Size</locale>
162       </Clothingsize>
163       <Department hid="Catavocationandeducation">
164         <locale id="de">Abteilung</locale>
165         <locale id="en">Department</locale>
166       </Department>
167       <Educationlevel hid="Catavocationandeducation">
168         <locale id="de">Bildungsgrad</locale>
169         <locale id="en">Level of Education</locale>
170       </Educationlevel>
171       <Employer hid="Catavocationandeducation">
172         <locale id="de">Arbeitgeber</locale>
173         <locale id="en">Employer</locale>
174       </Employer>
175       <Employment hid="Catavocationandeducation">
176         <locale id="de">Erwerbstätigkeit</locale>
177         <locale id="en">Employment</locale>
178       </Employment>
179       <Eyecolor hid="Catappearance">
180         <locale id="de">Augenfarbe</locale>
```

```
181         <locale id="en">Eye Color</locale>
182       </Eyecolor>
183       <Familystatus hid="Catfamily">
184         <locale id="de">Familienstand</locale>
185         <locale id="en">Marital Status</locale>
186       </Familystatus>
187       <Firstlanguage hid="Catlanguage">
188         <locale id="de">Muttersprache</locale>
189         <locale id="en">Native Language</locale>
190       </Firstlanguage>
191       <Gender hid="Catpersonal">
192         <locale id="de">Geschlecht</locale>
193         <locale id="en">Gender</locale>
194       </Gender>
195       <Haircolor hid="Catappearance">
196         <locale id="de">Haarfarbe</locale>
197         <locale id="en">Hair Color</locale>
198       </Haircolor>
199       <Jobtitle hid="Catavocationandeducation">
200         <locale id="de">Berufsbezeichnung</locale>
201         <locale id="en">Job Title</locale>
202       </Jobtitle>
203       <Nationality hid="Catpersonal">
204         <locale id="de">Nationalität</locale>
205         <locale id="en">Nationality</locale>
206       </Nationality>
207       <Politicalaffiliation hid="Catpersonal">
208         <locale id="de">Politische Zugehörigkeit</locale>
209         <locale id="en">Political Affiliation</locale>
210       </Politicalaffiliation>
211       <Religiousdenomination hid="Catpersonal">
212         <locale id="de">Konfession</locale>
213         <locale id="en">Religious Denomination</locale>
214       </Religiousdenomination>
215       <Salary hid="Catavocationandeducation">
216         <locale id="de">Einkommen</locale>
217         <locale id="en">Salary</locale>
218       </Salary>
219       <Secondlanguage hid="Catlanguage">
220         <locale id="de">Fremdsprache</locale>
221         <locale id="en">Foreign Language</locale>
222       </Secondlanguage>
223       <Shoesize hid="Catappearance">
224         <locale id="de">Schuhgröße</locale>
225         <locale id="en">Shoe Size</locale>
226       </Shoesize>
227       <Siblingsstatus hid="Catfamily">
228         <locale id="de">Geschwister</locale>
229         <locale id="en">Siblings Status</locale>
230       </Siblingsstatus>
231       <Socialsecuritynumber hid="Catpersonal">
232         <locale id="de">Sozialversicherungsnummer</locale>
233         <locale id="en">Social Security Number</locale>
234       </Socialsecuritynumber>
235       <Unionaffiliation hid="Catpersonal">
236         <locale id="de">Gewerkschaftszugehörigkeit</locale>
237         <locale id="en">Union Affiliation</locale>
238       </Unionaffiliation>
239       <Wealth hid="Catavocationandeducation">
240         <locale id="de">Vermögen</locale>
241         <locale id="en">Wealth</locale>
242       </Wealth>
243     </DataType>
244   </icon_group>
245   <icon_group id="EMail" icon="images/data/EMailInfo.png">
246     <locale id="de">E-Mail-Adressen</locale>
247     <locale id="en">E-mail Addresses</locale>
248     <DataType>
249       <Email>
250         <locale id="de">E-Mail-Adresse</locale>
251         <locale id="en">E-mail Address</locale>
252       </Email>
253     </DataType>
254   </icon_group>
255   <icon_group id="PostalData" icon="images/data/PostalInfo.png">
256     <locale id="de">Adressdaten</locale>
257     <locale id="en">Postal Address Data</locale>
258     <DataType>
259       <City hid="Postal">
260         <locale id="de">Stadt</locale>
261         <locale id="en">City</locale>
262       </City>
263       <Coordinate hid="Geographiclocation">
264         <locale id="de">Koordinaten</locale>
265         <locale id="en">Coordinates</locale>
266       </Coordinate>
267       <Country hid="Postal">
268         <locale id="de">Land</locale>
269         <locale id="en">Country</locale>
270       </Country>
271       <Geographiclocation>
272         <locale id="de">Geographische Position</locale>
273         <locale id="en">Geographic Location</locale>
274       </Geographiclocation>
275       <Housenumber hid="Postal">
276         <locale id="de">Hausnummer</locale>
277         <locale id="en">Street Number</locale>
278       </Housenumber>
279       <Organization hid="Postal">
280         <locale id="de">Organisation</locale>
281         <locale id="en">Organization</locale>
282       </Organization>
283       <Postal>
284         <locale id="de">Postanschrift</locale>
285         <locale id="en">Postal Address</locale>
286       </Postal>
287       <Postalcode hid="Postal">
288         <locale id="de">Postleitzahl</locale>
289         <locale id="en">ZIP Code</locale>
290       </Postalcode>
291       <Stateprov hid="Postal">
```

```
292         <locale id="de">Bundesland</locale>
293         <locale id="en">State / Province</locale>
294       </Stateprov>
295       <Street hid="Postal">
296         <locale id="de">Straße</locale>
297         <locale id="en">Street</locale>
298       </Street>
299     </DataType>
300   </icon_group>
301   <icon_group id="PhoneData" icon="images/data/PhoneInfo.png">
302     <locale id="de">Telefonnummern</locale>
303     <locale id="en">Phone Data</locale>
304     <DataType>
305       <Fax>
306         <locale id="de">Fax</locale>
307         <locale id="en">Fax</locale>
308       </Fax>
309       <Mobile>
310         <locale id="de">Mobil</locale>
311         <locale id="en">Mobile</locale>
312       </Mobile>
313       <Pager>
314         <locale id="de">Pager</locale>
315         <locale id="en">Pager</locale>
316       </Pager>
317       <Telephone>
318         <locale id="de">Festnetz</locale>
319         <locale id="en">Landline</locale>
320       </Telephone>
321     </DataType>
322   </icon_group>
323   <icon_group id="FinancialData" icon="images/data/FinancialInfo.png">
324     <locale id="de">Finanzdaten</locale>
325     <locale id="en">Financial Data</locale>
326     <DataType>
327       <Accountbalance>
328         <locale id="de">Kontostand</locale>
329         <locale id="en">Account Balance</locale>
330       </Accountbalance>
331       <Accountnumber>
332         <locale id="de">Kontonummer</locale>
333         <locale id="en">Account Number</locale>
334       </Accountnumber>
335       <Bankcode>
336         <locale id="de">Bankleitzahl</locale>
337         <locale id="en">Routing Number</locale>
338       </Bankcode>
339       <Creditcardexpiration>
340         <locale id="de">Kreditkartengültigkeit</locale>
341         <locale id="en">Credit Card Expiration</locale>
342       </Creditcardexpiration>
343       <Creditcardnumber>
344         <locale id="de">Kreditkartennummer</locale>
345         <locale id="en">Credit Card Number</locale>
346       </Creditcardnumber>
347       <Creditcardsecuritycode>
348         <locale id="de">Kreditkarten-Sicherheitscode</locale>
349         <locale id="en">Credit Card Security Code</locale>
350       </Creditcardsecuritycode>
351       <Depositnumber>
352         <locale id="de">Depotnummer</locale>
353         <locale id="en">Deposit Number</locale>
354       </Depositnumber>
355       <Paymentmethod>
356         <locale id="de">Zahlungsmethode</locale>
357         <locale id="en">Payment Method</locale>
358       </Paymentmethod>
359     </DataType>
360   </icon_group>
361   <icon_group id="HealthData" icon="images/data/HealthState.png">
362     <locale id="de">Gesundheitsdaten</locale>
363     <locale id="en">Health Data</locale>
364     <DataType>
365       <Arousal hid="Physiologicalstate">
366         <locale id="de">Erregung</locale>
367         <locale id="en">Arousal</locale>
368       </Arousal>
369       <Bloodpressure hid="Physiologicalstate">
370         <locale id="de">Blutdruck</locale>
371         <locale id="en">Blood Pressure</locale>
372       </Bloodpressure>
373       <Disability>
374         <locale id="de">Behinderung</locale>
375         <locale id="en">Disability</locale>
376       </Disability>
377       <Disease>
378         <locale id="de">Krankheit</locale>
379         <locale id="en">Disease</locale>
380       </Disease>
381       <Drugusage>
382         <locale id="de">Drogenkonsum</locale>
383         <locale id="en">Drug Usage</locale>
384       </Drugusage>
385       <Fatigue hid="Physiologicalstate">
386         <locale id="de">Erschöpfung</locale>
387         <locale id="en">Fatigue</locale>
388       </Fatigue>
389       <Heartbeat hid="Physiologicalstate">
390         <locale id="de">Herzfrequenz</locale>
391         <locale id="en">Heart Rate</locale>
392       </Heartbeat>
393       <Injury hid="Physiologicalstate">
394         <locale id="de">Verletzung</locale>
395         <locale id="en">Injury</locale>
396       </Injury>
397       <Medication>
398         <locale id="de">Medikation</locale>
399         <locale id="en">Medication</locale>
400       </Medication>
401       <Perspiration hid="Physiologicalstate">
402         <locale id="de">Perspiration</locale>
```

```
403         <locale id="en">Perspiration</locale>
404       </Perspiration>
405       <Physiologicalstate>
406         <locale id="de">Physiologischer Zustand</locale>
407         <locale id="en">Physiological State</locale>
408       </Physiologicalstate>
409       <Pupilsdilation hid="Physiologicalstate">
410         <locale id="de">Pupillenerweiterung</locale>
411         <locale id="en">Pupils Dilation</locale>
412       </Pupilsdilation>
413       <Respiration hid="Physiologicalstate">
414         <locale id="de">Atmung</locale>
415         <locale id="en">Respiration</locale>
416       </Respiration>
417       <Smokinghabits>
418         <locale id="de">Rauchgewohnheiten</locale>
419         <locale id="en">Smoking Habits</locale>
420       </Smokinghabits>
421       <Temperature hid="Physiologicalstate">
422         <locale id="de">Temperatur</locale>
423         <locale id="en">Temperature</locale>
424       </Temperature>
425     </DataType>
426   </icon_group>
427   <icon_group id="OnlineData" icon="images/data/OnlineInfo.png">
428     <locale id="de">Online-Daten</locale>
429     <locale id="en">Online Data</locale>
430     <DataType>
431       <Bytes hid="Clickstream">
432         <locale id="de">Bytes</locale>
433         <locale id="en">Bytes</locale>
434       </Bytes>
435       <Clickstream>
436         <locale id="de">Clickstream</locale>
437         <locale id="en">Clickstream</locale>
438       </Clickstream>
439       <Clientevents>
440         <locale id="de">Client-Ereignisse</locale>
441         <locale id="en">Client Events</locale>
442       </Clientevents>
443       <Clientip hid="Clickstream">
444         <locale id="de">Client-IP</locale>
445         <locale id="en">Client IP</locale>
446       </Clientip>
447       <Cookies>
448         <locale id="de">Cookies</locale>
449         <locale id="en">Cookies</locale>
450       </Cookies>
451       <Homepage>
452         <locale id="de">Homepage</locale>
453         <locale id="en">Homepage</locale>
454       </Homepage>
455       <Hostname hid="Clickstream">
456         <locale id="de">Hostname</locale>
457         <locale id="en">Hostname</locale>
458       </Hostname>
459       <Http hid="Clickstream">
460         <locale id="de">HTTP</locale>
461         <locale id="en">HTTP</locale>
462       </Http>
463       <Httpmethod hid="Clickstream">
464         <locale id="de">HTTP-Methode</locale>
465         <locale id="en">HTTP Method</locale>
466       </Httpmethod>
467       <Instantmessagingid>
468         <locale id="de">Instant Messaging ID</locale>
469         <locale id="en">Instant Messaging ID</locale>
470       </Instantmessagingid>
471       <Interactionrecord>
472         <locale id="de">Interaktionsaufzeichnung</locale>
473         <locale id="en">Interaction Record</locale>
474       </Interactionrecord>
475       <Referer>
476         <locale id="de">Referer</locale>
477         <locale id="en">Referer</locale>
478       </Referer>
479       <Searchtext>
480         <locale id="de">Suchtext</locale>
481         <locale id="en">Search Text</locale>
482       </Searchtext>
483       <Statuscode hid="Clickstream">
484         <locale id="de">Statuscode</locale>
485         <locale id="en">Status Code</locale>
486       </Statuscode>
487       <Timestamp hid="Clickstream">
488         <locale id="de">Zeitstempel</locale>
489         <locale id="en">Timestamp</locale>
490       </Timestamp>
491       <Uri hid="Clickstream">
492         <locale id="de">URL</locale>
493         <locale id="en">URL</locale>
494       </Uri>
495       <Useragent>
496         <locale id="de">User Agent</locale>
497         <locale id="en">User Agent</locale>
498       </Useragent>
499     </DataType>
500   </icon_group>
501   <icon_group id="OtherData" icon="images/data/OtherInfo.png">
502     <locale id="de">Sonstige Daten</locale>
503     <locale id="en">Other Data</locale>
504     <DataType>
505       <Ability>
506         <locale id="de">Fähigkeit</locale>
507         <locale id="en">Ability</locale>
508       </Ability>
509       <Catinterests>
510         <locale id="de">Interessen</locale>
511         <locale id="en">Interests</locale>
512       </Catinterests>
513       <Catother>
```

```
514         <locale id="de">Sonstige</locale>
515         <locale id="en">Other</locale>
516       </Catother>
517       <Characteristic>
518         <locale id="de">Charaktereigenschaft</locale>
519         <locale id="en">Characteristic</locale>
520       </Characteristic>
521       <Cinematicinterest hid="Culturalinterest">
522         <locale id="de">Film und Kino</locale>
523         <locale id="en">Cinema and Movie</locale>
524       </Cinematicinterest>
525       <Criminalrecords hid="Catother">
526         <locale id="de">Strafregister</locale>
527         <locale id="en">Criminal Records</locale>
528       </Criminalrecords>
529       <Culturalinterest>
530         <locale id="de">Kultur</locale>
531         <locale id="en">Culture</locale>
532       </Culturalinterest>
533       <Economicinterest hid="Catinterests">
534         <locale id="de">Wirtschaft</locale>
535         <locale id="en">Economy</locale>
536       </Economicinterest>
537       <Environmentalinterest hid="Catinterests">
538         <locale id="de">Umwelt</locale>
539         <locale id="en">Environment</locale>
540       </Environmentalinterest>
541       <Fashioninterest hid="Catinterests">
542         <locale id="de">Mode</locale>
543         <locale id="en">Fashion</locale>
544       </Fashioninterest>
545       <Identificationnumber>
546         <locale id="de">Identifizierungsnummer</locale>
547         <locale id="en">Identification Number</locale>
548       </Identificationnumber>
549       <Literaryinterest hid="Culturalinterest">
550         <locale id="de">Literatur</locale>
551         <locale id="en">Literature</locale>
552       </Literaryinterest>
553       <Musicinterest hid="Culturalinterest">
554         <locale id="de">Musik</locale>
555         <locale id="en">Music</locale>
556       </Musicinterest>
557       <Natureinterest hid="Catinterests">
558         <locale id="de">Natur</locale>
559         <locale id="en">Nature</locale>
560       </Natureinterest>
561       <Notclassified>
562         <locale id="de">Nicht klassifiziert</locale>
563         <locale id="en">Not Classified</locale>
564       </Notclassified>
565       <Nutritioninterest hid="Catinterests">
566         <locale id="de">Ernährung</locale>
567         <locale id="en">Nutrition</locale>
568       </Nutritioninterest>
569       <Performingartsinterest hid="Culturalinterest">
570         <locale id="de">Darstellende Kunst</locale>
571         <locale id="en">Performing Arts</locale>
572       </Performingartsinterest>
573       <Philosophyinterest hid="Catinterests">
574         <locale id="de">Philosophie</locale>
575         <locale id="en">Philosophy</locale>
576       </Philosophyinterest>
577       <Politicalinterest hid="Catinterests">
578         <locale id="de">Politik</locale>
579         <locale id="en">Politics</locale>
580       </Politicalinterest>
581       <Scienceinterest hid="Catinterests">
582         <locale id="de">Wissenschaft</locale>
583         <locale id="en">Science</locale>
584       </Scienceinterest>
585       <Sportsinterest hid="Catinterests">
586         <locale id="de">Sport</locale>
587         <locale id="en">Sports</locale>
588       </Sportsinterest>
589       <Technologyinterest hid="Catinterests">
590         <locale id="de">Technologie</locale>
591         <locale id="en">Technology</locale>
592       </Technologyinterest>
593       <Travelinginterest hid="Catinterests">
594         <locale id="de">Reisen</locale>
595         <locale id="en">Traveling</locale>
596       </Travelinginterest>
597       <Visualartsinterest hid="Culturalinterest">
598         <locale id="de">Bildende Kunst</locale>
599         <locale id="en">Visual Arts</locale>
600       </Visualartsinterest>
601     </DataType>
602   </icon_group>
603 </icon_groups>
```

A.2 Service Type Vocabulary

```
<?xml version="1.0" encoding="UTF-8"?>
<services xmlns="http://www-ifs.uni-regensburg.de/Privacy">
  <service id="BANKING">
    <icon filename="images/services/banking.png"/>
    <label>
      <lang id="de">
        <name>Banken und Finanzen</name>
        <description>Online-Banking- und Bezahldienste, sowie Online-Brokerage wie comdirect.de, postbank.de,
                     paypal.de, shareit.de und clickandbuy.de.</description>
      </lang>
      <lang id="en">
        <name>Banking</name>
        <description>Online banking and payment services as well as online brokerage like usbank.com,
                     wellsfargo.com, paypal.com, shareit.com and clickandbuy.com.</description>
      </lang>
    </label>
    <mindata>
      <Accountnumber/>
      <Bankcode/>
      <Bdate/>
      <Birthplace/>
      <Bytes/>
      <City/>
      <Clickstream/>
      <Clientevents/>
      <Clientip/>
      <Cookies/>
      <Country/>
      <Creditcardexpiration/>
      <Creditcardnumber/>
      <Creditcardsecuritycode/>
      <Depositnumber/>
      <Email/>
      <Employment/>
      <Family/>
      <Familystatus/>
      <Given/>
      <Hostname/>
      <Housenumber/>
      <Http/>
      <Httpmethod/>
      <Id/>
      <Interactionrecord/>
      <Login/>
      <Nationality/>
      <Password/>
      <Postalcode/>
      <Referer/>
      <Searchtext/>
      <Stateprov/>
      <Statuscode/>
      <Street/>
      <Timestamp/>
      <Uri/>
      <Useragent/>
    </mindata>
    <ppurpose>
      <payment/>
      <finmgt/>
    </ppurpose>
  </service>
  <service id="DOWNLOADS">
    <icon filename="images/services/downloads.png"/>
    <label>
      <lang id="de">
        <name>Download-Portale</name>
        <description>Download-Dienste wie rapidshare.com, winload.de und downloadmix.de</description>
      </lang>
      <lang id="en">
        <name>Downloads</name>
        <description>Download services like rapidshare.com.</description>
      </lang>
    </label>
    <mindata>
      <Bytes/>
      <Clickstream/>
      <Clientevents/>
      <Clientip/>
      <Cookies/>
      <Hostname/>
      <Http/>
      <Httpmethod/>
      <Id/>
      <Interactionrecord/>
      <Login/>
      <Password/>
      <Referer/>
      <Searchtext/>
      <Statuscode/>
      <Timestamp/>
      <Uri/>
      <Useragent/>
    </mindata>
    <ppurpose>
      <downloads/>
      <survey/>
    </ppurpose>
  </service>
  <service id="EDUCATION">
    <icon filename="images/services/elearning.png"/>
    <label>
      <lang id="de">
        <name>Schulung und Bildung</name>
        <description>Lern-Portale wie e-teaching.org, world-lecture-project.org, fahrschule.de</description>
      </lang>
      <lang id="en">
```

```
107         <name>E-Learning</name>
108         <description>Learning portals like e-teaching.org and world-lecture-project.org.</description>
109       </lang>
110     </label>
111     <mindata>
112       <Bytes/>
113       <Clickstream/>
114       <Clientevents/>
115       <Clientip/>
116       <Cookies/>
117       <Email/>
118       <Hostname/>
119       <Http/>
120       <Httpmethod/>
121       <Id/>
122       <Interactionrecord/>
123       <Login/>
124       <Nickname/>
125       <Password/>
126       <Referer/>
127       <Searchtext/>
128       <Statuscode/>
129       <Timestamp/>
130       <Uri/>
131       <Useragent/>
132     </mindata>
133     <ppurpose>
134       <education/>
135       <survey/>
136     </ppurpose>
137   </service>
138   <service id="GAMING">
139     <icon filename="images/services/games.png"/>
140     <label>
141       <lang id="de">
142         <name>Spiele</name>
143         <description>Online-Spiele wie Second Life, Ultima Online, World of Warcraft und
144                      Online-Poker</description>
145       </lang>
146       <lang id="en">
147         <name>Games</name>
148         <description>Online games like Second Life, Ultima Online, World of Warcraft and
149                      Online Poker</description>
150       </lang>
151     </label>
152     <mindata>
153       <Accountnumber/>
154       <Bankcode/>
155       <Bytes/>
156       <Clickstream/>
157       <Clientevents/>
158       <Clientip/>
159       <Cookies/>
160       <Creditcardexpiration/>
161       <Creditcardnumber/>
162       <Creditcardsecuritycode/>
163       <Email/>
164       <Hostname/>
165       <Http/>
166       <Httpmethod/>
167       <Id/>
168       <Interactionrecord/>
169       <Login/>
170       <Nickname/>
171       <Password/>
172       <Paymentmethod/>
173       <Referer/>
174       <Searchtext/>
175       <Statuscode/>
176       <Timestamp/>
177       <Uri/>
178       <Useragent/>
179     </mindata>
180     <ppurpose>
181       <gaming/>
182       <gambling/>
183       <payment/>
184       <survey/>
185     </ppurpose>
186   </service>
187   <service id="GOVERNMENT">
188     <icon filename="images/services/government.png"/>
189     <label>
190       <lang id="de">
191         <name>Behörden</name>
192         <description>E-Government-Dienste von öffentlichen Einrichtungen (z.B. Führerscheinbeantragung,
193                      Reisepassverlängerung etc.)</description>
194       </lang>
195       <lang id="en">
196         <name>Government</name>
197         <description>eGovernment services of local, state and federal authorities; online services
198                      of the DMV.</description>
199       </lang>
200     </label>
201     <mindata>
202       <Accountnumber/>
203       <Bankcode/>
204       <Bdate/>
205       <Bytes/>
206       <City/>
207       <Clickstream/>
208       <Clientevents/>
209       <Clientip/>
210       <Cookies/>
211       <Country/>
212       <Creditcardexpiration/>
213       <Creditcardnumber/>
214       <Creditcardsecuritycode/>
215       <Email/>
216       <Family/>
217       <Given/>
```

```
218         <Hostname/>
219         <Housenumber/>
220         <Http/>
221         <Httpmethod/>
222         <Id/>
223         <Interactionrecord/>
224         <Login/>
225         <Password/>
226         <Paymentmethod/>
227         <Postalcode/>
228         <Referer/>
229         <Searchtext/>
230         <Stateprov/>
231         <Statuscode/>
232         <Street/>
233         <Timestamp/>
234         <Uri/>
235         <Useragent/>
236       </mindata>
237       <ppurpose>
238         <government/>
239       </ppurpose>
240     </service>
241     <service id="HEALTH">
242       <icon filename="images/services/health.png"/>
243       <label>
244         <lang id="de">
245           <name>Gesundheitsportale</name>
246           <description>Gesundheitsinformationsdienste wie meine-gesundheit.de und arztpartner.de</description>
247         </lang>
248         <lang id="en">
249           <name>Health</name>
250           <description>Health information services like revolutionhealth.com.</description>
251         </lang>
252       </label>
253       <mindata>
254         <Age/>
255         <Arousal/>
256         <Bloodpressure/>
257         <Bytes/>
258         <Clickstream/>
259         <Clientevents/>
260         <Clientip/>
261         <Cookies/>
262         <Disability/>
263         <Disease/>
264         <Drugusage/>
265         <Email/>
266         <Fatigue/>
267         <Gender/>
268         <Heartbeat/>
269         <Hostname/>
270         <Http/>
271         <Httpmethod/>
272         <Id/>
273         <Injury/>
274         <Interactionrecord/>
275         <Login/>
276         <Medication/>
277         <Nickname/>
278         <Nourishment/>
279         <Password/>
280         <Perspiration/>
281         <Physiologicalstate/>
282         <Pupilsdilation/>
283         <Referer/>
284         <Respiration/>
285         <Searchtext/>
286         <Smokinghabits/>
287         <Statuscode/>
288         <Temperature/>
289         <Timestamp/>
290         <Uri/>
291         <Useragent/>
292       </mindata>
293       <ppurpose>
294         <health/>
295       </ppurpose>
296     </service>
297     <service id="INSTANTMESSAGING">
298       <icon filename="images/services/messaging.png"/>
299       <label>
300         <lang id="de">
301           <name>Instant Messaging</name>
302           <description>Web-basierte Instant-Messaging-Dienste wie meebo.com, ICQ2Go und
303                        MSN Web Messenger</description>
304         </lang>
305         <lang id="en">
306           <name>Messaging</name>
307           <description>Web-based instant messaging services like meebo.com, ICQ2Go
308                        and MSN Web Messenger.</description>
309         </lang>
310       </label>
311       <mindata>
312         <Bytes/>
313         <Clickstream/>
314         <Clientevents/>
315         <Clientip/>
316         <Cookies/>
317         <Hostname/>
318         <Http/>
319         <Httpmethod/>
320         <Id/>
321         <Instantmessagingid/>
322         <Interactionrecord/>
323         <Login/>
324         <Password/>
325         <Referer/>
326         <Searchtext/>
327         <Statuscode/>
328         <Timestamp/>
```

```
329         <Uri/>
330         <Useragent/>
331       </mindata>
332       <ppurpose>
333         <communicate/>
334       </ppurpose>
335     </service>
336     <service id="NEWS">
337       <icon filename="images/services/news.png"/>
338       <label>
339         <lang id="de">
340           <name>News- und Wissensportale</name>
341           <description>Newsseiten wie stern.de und spiegel.de aber auch Wissensportale wie wikipedia.org
342                        und wissen.de, sowie Sportseiten wie kicker.de und sport1.de.</description>
343         </lang>
344         <lang id="en">
345           <name>News and Knowledge</name>
346           <description>News portals like cnn.com and msnbc.com; knowledge portals like wikipedia.org and
347                        answers.com; sports portals like espn.com and foxsports.com.</description>
348         </lang>
349       </label>
350       <mindata>
351         <Bytes/>
352         <Clickstream/>
353         <Clientevents/>
354         <Clientip/>
355         <Cookies/>
356         <Hostname/>
357         <Http/>
358         <Httpmethod/>
359         <Id/>
360         <Interactionrecord/>
361         <Login/>
362         <Password/>
363         <Referer/>
364         <Searchtext/>
365         <Statuscode/>
366         <Timestamp/>
367         <Uri/>
368         <Useragent/>
369       </mindata>
370       <ppurpose>
371         <news/>
372       </ppurpose>
373     </service>
374     <service id="ONLINESHOPPING">
375       <icon filename="images/services/shopping.png"/>
376       <label>
377         <lang id="de">
378           <name>Einkauf</name>
379           <description>Diverse Einkaufsdienste wie amazon.de, bol.de, musicload.de, ebay.de
380                        und itunes.de.</description>
381         </lang>
382         <lang id="en">
383           <name>Shopping</name>
384           <description>All kinds of online shopping services like amazon.com, ebay.com
385                        and itunes.com</description>
386         </lang>
387       </label>
388       <mindata>
389         <Accountnumber/>
390         <Bankcode/>
391         <Bytes/>
392         <City/>
393         <Clickstream/>
394         <Clientevents/>
395         <Clientip/>
396         <Cookies/>
397         <Country/>
398         <Creditcardexpiration/>
399         <Creditcardnumber/>
400         <Creditcardsecuritycode/>
401         <Email/>
402         <Family/>
403         <Given/>
404         <Hostname/>
405         <Housenumber/>
406         <Http/>
407         <Httpmethod/>
408         <Id/>
409         <Interactionrecord/>
410         <Login/>
411         <Password/>
412         <Paymentmethod/>
413         <Postalcode/>
414         <Referer/>
415         <Searchtext/>
416         <Stateprov/>
417         <Statuscode/>
418         <Street/>
419         <Timestamp/>
420         <Uri/>
421         <Useragent/>
422       </mindata>
423       <ppurpose>
424         <survey/>
425         <sales/>
426         <payment/>
427         <marketing/>
428         <delivery/>
429         <charity/>
430         <arts/>
431         <downloads/>
432       </ppurpose>
433     </service>
434     <service id="SOCIALNETWORKING">
435       <icon filename="images/services/socialnetworking.png"/>
436       <label>
437         <lang id="de">
438           <name>Social-Networking-Portale</name>
439           <description>Social-Networking-Plattformen wie xing.de, studivz.de, myspace.com
```

```
440                     und lokalisten.de.</description>
441       </lang>
442       <lang id="en">
443         <name>Social Networks</name>
444         <description>Social networking platforms like myspace.com, facebook.com
445                     and xing.com.</description>
446       </lang>
447     </label>
448     <mindata>
449       <Bdate/>
450       <Bytes/>
451       <Clickstream/>
452       <Clientevents/>
453       <Clientip/>
454       <Cookies/>
455       <Email/>
456       <Family/>
457       <Gender/>
458       <Given/>
459       <Hostname/>
460       <Http/>
461       <Httpmethod/>
462       <Id/>
463       <Interactionrecord/>
464       <Login/>
465       <Password/>
466       <Referer/>
467       <Searchtext/>
468       <Statuscode/>
469       <Timestamp/>
470       <Uri/>
471       <Useragent/>
472     </mindata>
473     <ppurpose>
474       <communicate/>
475       <survey/>
476     </ppurpose>
477   </service>
478   <service id="USERGENERATED">
479     <icon filename="images/services/usergenerated.png"/>
480     <label>
481       <lang id="de">
482         <name>Foren und User-Generated Content</name>
483         <description>Webforen und Weblogs, sowie User-Generated-Content-Portale wie youtube.com,
484                     flickr.com und myvideo.de.</description>
485       </lang>
486       <lang id="en">
487         <name>User-Generated Content</name>
488         <description>Forums and weblogs as well as user generated content portals like youtube.com
489                     and flickr.com.</description>
490       </lang>
491     </label>
492     <mindata>
493       <Bytes/>
494       <Clickstream/>
495       <Clientevents/>
496       <Clientip/>
497       <Cookies/>
498       <Email/>
499       <Hostname/>
500       <Http/>
501       <Httpmethod/>
502       <Id/>
503       <Interactionrecord/>
504       <Login/>
505       <Nickname/>
506       <Password/>
507       <Referer/>
508       <Searchtext/>
509       <Statuscode/>
510       <Timestamp/>
511       <Uri/>
512       <Useragent/>
513     </mindata>
514     <ppurpose>
515       <communicate/>
516       <survey/>
517     </ppurpose>
518   </service>
519   <service id="WEBMAIL">
520     <icon filename="images/services/mail.png"/>
521     <label>
522       <lang id="de">
523         <name>Web Mail</name>
524         <description>Webmail-Dienste wie googlemail.com, web.de or gmx.de.</description>
525       </lang>
526       <lang id="en">
527         <name>Mail</name>
528         <description>Web mail services like googlemail.com, web.de or gmx.de.</description>
529       </lang>
530     </label>
531     <mindata>
532       <Bytes/>
533       <Clickstream/>
534       <Clientevents/>
535       <Clientip/>
536       <Cookies/>
537       <Email/>
538       <Family/>
539       <Given/>
540       <Hostname/>
541       <Http/>
542       <Httpmethod/>
543       <Id/>
544       <Interactionrecord/>
545       <Login/>
546       <Password/>
547       <Referer/>
548       <Searchtext/>
549       <Statuscode/>
550       <Timestamp/>
```

```
      <Uri/>
      <Useragent/>
    </mindata>
    <ppurpose>
      <communicate/>
    </ppurpose>
  </service>
</services>
```

Appendix B

Exemplary Local Privacy Files

B.1 Privacy Preference File

```
<?xml version="1.0" encoding="utf-8"?>
<privacypreferences xmlns="http://www-ifs.uni-regensburg.de/Privacy">
  <info>
    <mode>advanced</mode>
  </info>
  <services>
    <service id="BANKING">
      <retention>
        <stated-purpose />
      </retention>
      <additionaldata>refuse</additionaldata>
      <data>
        <Accountnumber />
        <Bankcode />
        <Bdate />
        <Birthplace />
        <Bytes />
        <City />
        <Clickstream />
        <Clientevents />
        <Clientip />
        <Cookies />
        <Country />
        <Creditcardexpiration />
        <Creditcardnumber />
        <Creditcardsecuritycode />
        <Depositnumber />
        <Email />
        <Employment />
        <Family />
        <Familystatus />
        <Given />
        <Hostname />
        <Housenumber />
        <Http />
        <Httpmethod />
        <Id />
        <Interactionrecord />
        <Login />
        <Nationality />
        <Password />
        <Postalcode />
        <Referer />
        <Searchtext />
        <Stateprov />
        <Statuscode />
        <Street />
        <Timestamp />
        <Uri />
        <Useragent />
      </data>
      <purpose>
        <current />
        <personalization>
          <tailoring required="always" />
        </personalization>
        <contacting>
          <contact required="always" />
          <telemarketing required="always" />
        </contacting>
      </purpose>
      <recipient>
        <ours required="always" />
        <same required="always" />
      </recipient>
    </service>
    <service id="EDUCATION">
      <retention>
        <stated-purpose />
      </retention>
      <additionaldata>refuse</additionaldata>
      <data>
        <Bytes />
        <Clickstream />
```

```
75          <Clientevents />
76          <Clientip />
77          <Cookies />
78          <Email />
79          <Hostname />
80          <Http />
81          <Httpmethod />
82          <Id />
83          <Interactionrecord />
84          <Login />
85          <Nickname />
86          <Password />
87          <Referer />
88          <Searchtext />
89          <Statuscode />
90          <Timestamp />
91          <Uri />
92          <Useragent />
93        </data>
94        <purpose>
95          <current />
96          <personalization>
97            <tailoring required="always" />
98          </personalization>
99          <contacting />
100       </purpose>
101       <recipient>
102         <ours required="always" />
103       </recipient>
104     </service>
105     <service id="NEWS">
106       <retention>
107         <stated-purpose />
108       </retention>
109       <additionaldata>request-prompt</additionaldata>
110       <data>
111         <Bytes />
112         <Clickstream />
113         <Clientevents />
114         <Clientip />
115         <Cookies />
116         <Hostname />
117         <Http />
118         <Httpmethod />
119         <Id />
120         <Interactionrecord />
121         <Login />
122         <Password />
123         <Referer />
124         <Searchtext />
125         <Statuscode />
126         <Timestamp />
127         <Uri />
128         <Useragent />
129       </data>
130       <purpose>
131         <current />
132         <personalization>
133           <tailoring required="always" />
134           <pseudo-analysis required="always" />
135           <pseudo-decision required="always" />
136           <individual-analysis required="always" />
137           <individual-decision required="always" />
138         </personalization>
139         <contacting>
140           <contact required="always" />
141           <telemarketing required="always" />
142         </contacting>
143       </purpose>
144       <recipient>
145         <ours required="always" />
146         <same required="always" />
147         <delivery required="always" />
148         <other-recipient required="always" />
149       </recipient>
150     </service>
151     <service id="ONLINESHOPPING">
152       <retention>
153         <stated-purpose />
154       </retention>
155       <additionaldata>request-prompt</additionaldata>
156       <data>
157         <Accountnumber />
158         <Bankcode />
159         <Bytes />
160         <City />
161         <Clickstream />
162         <Clientevents />
163         <Clientip />
164         <Cookies />
165         <Country />
166         <Creditcardexpiration />
167         <Creditcardnumber />
168         <Creditcardsecuritycode />
169         <Email />
170         <Family />
171         <Given />
172         <Hostname />
173         <Housenumber />
174         <Http />
175         <Httpmethod />
176         <Id />
177         <Interactionrecord />
178         <Login />
179         <Password />
180         <Paymentmethod />
181         <Postalcode />
182         <Referer />
183         <Searchtext />
184         <Stateprov />
185         <Statuscode />
```

```
186         <Street />
187         <Timestamp />
188         <Uri />
189         <Useragent />
190       </data>
191       <purpose>
192         <current />
193         <personalization>
194           <tailoring required="opt-out" />
195           <pseudo-analysis required="opt-in" />
196           <pseudo-decision required="opt-in" />
197         </personalization>
198         <contacting />
199       </purpose>
200       <recipient>
201         <ours required="always" />
202         <same required="always" />
203       </recipient>
204     </service>
205     <service id="SOCIALNETWORKING">
206       <retention>
207         <stated-purpose />
208       </retention>
209       <additionaldata>refuse</additionaldata>
210       <data>
211         <Bdate />
212         <Bytes />
213         <Clickstream />
214         <Clientevents />
215         <Clientip />
216         <Cookies />
217         <Email />
218         <Family />
219         <Gender />
220         <Given />
221         <Hostname />
222         <Http />
223         <Httpmethod />
224         <Id />
225         <Interactionrecord />
226         <Login />
227         <Password />
228         <Referer />
229         <Searchtext />
230         <Statuscode />
231         <Timestamp />
232         <Uri />
233         <Useragent />
234       </data>
235       <purpose>
236         <current />
237         <personalization>
238           <tailoring required="always" />
239           <pseudo-analysis required="always" />
240           <pseudo-decision required="always" />
241         </personalization>
242         <contacting>
243           <contact required="always" />
244         </contacting>
245       </purpose>
246       <recipient>
247         <ours required="always" />
248         <same required="always" />
249       </recipient>
250     </service>
251     <service id="USERGENERATED">
252       <retention>
253         <stated-purpose />
254       </retention>
255       <additionaldata>refuse</additionaldata>
256       <data>
257         <Bytes />
258         <Clickstream />
259         <Clientevents />
260         <Clientip />
261         <Cookies />
262         <Email />
263         <Hostname />
264         <Http />
265         <Httpmethod />
266         <Id />
267         <Interactionrecord />
268         <Login />
269         <Nickname />
270         <Password />
271         <Referer />
272         <Searchtext />
273         <Statuscode />
274         <Timestamp />
275         <Uri />
276         <Useragent />
277       </data>
278       <purpose>
279         <current />
280         <personalization />
281         <contacting />
282       </purpose>
283       <recipient>
284         <ours required="always" />
285       </recipient>
286     </service>
287     <service id="WEBMAIL">
288       <retention>
289         <stated-purpose />
290       </retention>
291       <additionaldata>refuse</additionaldata>
292       <data>
293         <Bytes />
294         <Clickstream />
295         <Clientevents />
296         <Clientip />
```

```
297         <Cookies />
298         <Email />
299         <Family />
300         <Given />
301         <Hostname />
302         <Http />
303         <Httpmethod />
304         <Id />
305         <Interactionrecord />
306         <Login />
307         <Password />
308         <Referer />
309         <Searchtext />
310         <Statuscode />
311         <Timestamp />
312       <Uri/>
313       <Useragent/>
314       </data>
315       <purpose>
316         <current />
317         <personalization />
318         <contacting>
319           <contact required="always" />
320         </contacting>
321       </purpose>
322       <recipient>
323         <ours required="always" />
324         <same required="always" />
325       </recipient>
326     </service>
327   </services>
328 </privacypreferences>
```

B.2 Identity File

```
<?xml version="1.0" encoding="utf-8" ?>
<identities xmlns="http://www-ifs.uni-regensburg.de/Privacy">
  <identity id="default">
    <label>
      <lang id="de">
        <name>Echte Identität</name>
        <description>Die echte Identität von Paul Revere</description>
      </lang>
      <lang id="en">
        <name>True Identity</name>
        <description>The true identity of Paul Revere</description>
      </lang>
    </label>
    <partialIdentity id="1">
      <label>
        <lang id="de">
          <name>Teilidentität 1</name>
          <description>Die Teilidentität 1 wird für die Registrierung bei offiziellen Stellen und Communities
                       verwendet</description>
        </lang>
        <lang id="en">
          <name>Partial Identity 1</name>
          <description>Partial identity 1 is used for registration at public authorities and
                       communities</description>
        </lang>
      </label>
      <partialIdentityInformation>
        <linkable>
          <partialIdentity id="2"/>
        </linkable>
      </partialIdentityInformation>
      <serviceProviderInformation>
        <usedBy>
          <hostname name="http://www.facebook.com" title="Facebook"/>
          <hostname name="http://www.amazon.com" title="Amazon"/>
        </usedBy>
      </serviceProviderInformation>
      <data>
        <Given>Paul</Given>
        <Family>Revere</Family>
        <Email>prevere@hotmail.com</Email>
        <Id>prevere</Id>
        <Bdate>24.05.1971</Bdate>
        <Gender>male</Gender>
        <Telephone>617-555-1234</Telephone>
      </data>
    </partialIdentity>
    <partialIdentity id="2">
      <label>
        <lang id="de">
          <name>Teilidentität 2</name>
          <description>Die Teilidentität 2 wird zum Einkaufen verwendet und enthält
                       Kreditkarteninformationen</description>
        </lang>
        <lang id="en">
          <name>Partial Identity 2</name>
          <description>Partial identity 2 is used for online shopping and therefore contains
                       credit card information</description>
        </lang>
      </label>
      <partialIdentityInformation>
        <linkable>
          <partialIdentity id="1"/>
        </linkable>
      </partialIdentityInformation>
      <serviceProviderInformation>
        <usedBy>
          <hostname name="http://www.amazon.com" title="Amazon"/>
        </usedBy>
      </serviceProviderInformation>
      <data>
        <Given>Paul</Given>
        <Family>Revere</Family>
        <Email>prevere@hotmail.com</Email>
        <Bdate>24.05.1971</Bdate>
        <Addresses>
          <Postal>
            <Street>Arlington Road</Street>
            <Housenumber>1606</Housenumber>
            <City>Boston</City>
            <Stateprov>MA</Stateprov>
            <Postalcode>02101</Postalcode>
          </Postal>
        </Addresses>
        <Financialdata>
          <account type="Credit Card">
            <Creditcardnumber>1376151d963cad089697872af7dcec48bbd3ae29</Creditcardnumber>
            <Creditcardexpiration>08/11</Creditcardexpiration>
            <Creditcardsecuritycode>451</Creditcardsecuritycode>
          </account>
        </Financialdata>
      </data>
    </partialIdentity>
  </identity>
  <identity id="pseudonymous">
    <label>
      <lang id="de">
        <name>Pseudonyme Identität</name>
        <description>Die Identität verwendet u.a. das Pseudonym Mike Smith</description>
      </lang>
      <lang id="en">
        <name>Pseudonymous Identity</name>
        <description>This identity frequently uses the pseudonym Mike Smith</description>
      </lang>
    </label>
    <partialIdentity id="3">
```

```
107         <label>
108           <lang id="de">
109             <name>Teilidentität 3</name>
110             <description>Die Teilidentität 3 enthält persönliche Daten zur anonymen Interaktion
111                          mit Service Providern</description>
112           </lang>
113           <lang id="en">
114             <name>Partial Identity 3</name>
115             <description>The partial identity 3 contains personal data for the anonymous interaction
116                          with service providers</description>
117           </lang>
118         </label>
119         <partialIdentityInformation>
120           <linkable>
121             <partialIdentity id="4"/>
122           </linkable>
123         </partialIdentityInformation>
124         <serviceProviderInformation>
125           <usedBy>
126             <hostname name="http://www.youtube.com" title="YouTube"/>
127           </usedBy>
128         </serviceProviderInformation>
129         <data>
130           <Given>Mike</Given>
131           <Family>Smith</Family>
132           <Email>braveman@gmail.com</Email>
133           <Addresses>
134               <Postal>
135               <Street>Fake Street</Street>
136               <Housenumber>1201</Housenumber>
137               <City>Arlington</City>
138               <Stateprov>MD</Stateprov>
139               <Postalcode>35416</Postalcode>
140             </Postal>
141           </Addresses>
142         </data>
143       </partialIdentity>
144       <partialIdentity id="4">
145         <label>
146           <lang id="de">
147             <name>Teilidentität 4</name>
148             <description>Die Teilidentität 4 enthält persönliche Daten zur anonymen Interaktion
149                          mit Service Providern</description>
150           </lang>
151           <lang id="en">
152             <name>Partial Identity 4</name>
153             <description>The partial identity 4 contains personal data for the anonymous interaction
154                          with service providers</description>
155           </lang>
156         </label>
157         <partialIdentityInformation>
158           <linkable>
159             <partialIdentity id="3"/>
160           </linkable>
161         </partialIdentityInformation>
162         <serviceProviderInformation>
163           <usedBy>
164             <hostname name="http://www.deviantart.com" title="deviantART"/>
165             <hostname name="http://www.google.com" title="Google"/>
166           </usedBy>
167         </serviceProviderInformation>
168         <data>
169           <Email>braveman@gmail.com</Email>
170         </data>
171       </partialIdentity>
172     </identity>
173     <identity id="anonymous">
174       <label>
175         <lang id="de">
176           <name>Anonyme Identität</name>
177           <description>Anonyme Identität enthält die AOL E-Mail-Adresse</description>
178         </lang>
179         <lang id="en">
180           <name>Anonymous Identity</name>
181           <description>Anonymous identity contains AOL E-Mail address</description>
182         </lang>
183       </label>
184       <partialIdentity id="5">
185         <label>
186           <lang id="de">
187             <name>Teilidentität 5</name>
188             <description>Die Teilidentität 5 enthält persönliche Daten zur anonymen Interaktion
189                          mit Service Providern</description>
190           </lang>
191           <lang id="en">
192             <name>Partial Identity 5</name>
193             <description>The partial identity 5 contains personal data for the anonymous interaction
194                          with service providers</description>
195           </lang>
196         </label>
197         <partialIdentityInformation>
198           <linkable>
199           </linkable>
200         </partialIdentityInformation>
201         <serviceProviderInformation>
202           <usedBy>
203             <hostname name="http://www.nytimes.com" title="The New York Times"/>
204           </usedBy>
205         </serviceProviderInformation>
206         <data>
207           <Id>smith123</Id>
208           <Email>smith@aol.com</Email>
209         </data>
210       </partialIdentity>
211     </identity>
212   </identities>
```

B.3 Transaction Log File


```xml
<?xml version="1.0" encoding="UTF-8"?>
<xml xmlns="http://www-ifs.uni-regensburg.de/Privacy">
  <hostname name="http://www.facebook.com" title="Facebook">
    <transaction process="Registration" id="1">
      <dynamic>
        <Uri>http://de.facebook.com/register</Uri>
        <Timestamp>1250074561</Timestamp>
        <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
        <Cookies>true</Cookies>
      </dynamic>
      <data>
        <Miscfields>
          <Notclassified>lorem</Notclassified>
          <Notclassified>ipsum</Notclassified>
        </Miscfields>
        <Given>Paul</Given>
        <Family>Revere</Family>
        <Email>prevere@hotmail.com</Email>
        <Id>prevere</Id>
        <Password>a3cca2b2aa1e</Password>
        <Bdate>24.05.1971</Bdate>
        <Gender>male</Gender>
        <Familystatus>single</Familystatus>
        <Nationality>British</Nationality>
        <Telephone>617-555-4567</Telephone>
        <Fax>617-555-2345</Fax>
        <Mobile>617-555-3456</Mobile>
      </data>
    </transaction>
    <transaction process="Login" id="2">
      <dynamic>
        <Uri>http://de.facebook.com/login</Uri>
        <Timestamp>1250247361</Timestamp>
        <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
        <Cookies>true</Cookies>
      </dynamic>
      <data>
        <Miscfields>
        </Miscfields>
        <Id>prevere</Id>
        <Password>a3cca2b2aa1e</Password>
      </data>
    </transaction>
    <transaction process="Login" id="3">
      <dynamic>
        <Uri>http://de.facebook.com/login</Uri>
        <Timestamp>1251024961</Timestamp>
        <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
        <Cookies>true</Cookies>
      </dynamic>
      <data>
        <Miscfields>
          <Notclassified>lorem</Notclassified>
          <Notclassified>ipsum</Notclassified>
        </Miscfields>
        <Id>prevere</Id>
        <Password>a3cca2b2aa1e</Password>
        <Gender>male</Gender>
        <Familystatus>married</Familystatus>
        <Nationality>U.S.</Nationality>
        <Telephone>617-555-1234</Telephone>
      </data>
    </transaction>
    <transaction process="Login" id="4">
      <dynamic>
        <Uri>http://de.facebook.com/login</Uri>
        <Timestamp>1251370561</Timestamp>
        <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
        <Cookies>true</Cookies>
      </dynamic>
      <data>
        <Miscfields>
        </Miscfields>
        <Id>prevere</Id>
        <Password>a3cca2b2aa1e</Password>
      </data>
    </transaction>
  </hostname>
  <hostname name="http://www.amazon.com" title="Amazon">
    <transaction process="Registration" id="1">
      <dynamic>
        <Uri>http://www.amazon.com/gp/css/homepage.html</Uri>
        <Timestamp>1250592961</Timestamp>
        <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
        <Cookies>true</Cookies>
      </dynamic>
      <data>
        <Miscfields>
          <Notclassified>lorem</Notclassified>
          <Notclassified>ipsum</Notclassified>
        </Miscfields>
        <Given>Paul</Given>
        <Family>Revere</Family>
        <Email>prevere@hotmail.com</Email>
        <Password>a3cca2b2aa1e</Password>
        <Bdate>24.05.1971</Bdate>
        <Addresses>
          <Postal>
            <Street>Arlington Road</Street>
            <Housenumber>1606</Housenumber>
            <City>Boston</City>
            <Stateprov>MA</Stateprov>
            <Postalcode>02101</Postalcode>
          </Postal>
        </Addresses>
      </data>
```

```
107     </transaction>
108     <transaction process="Login" id="2">
109       <dynamic>
110         <Uri>http://www.amazon.com/gp/css/homepage.html</Uri>
111         <Timestamp>1250765761</Timestamp>
112         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
113         <Cookies>true</Cookies>
114       </dynamic>
115       <data>
116         <Miscfields>
117         </Miscfields>
118         <Email>prevere@hotmail.com</Email>
119         <Password>a3cca2b2aa1e</Password>
120         <Addresses>
121           <Postal>
122             <Street>Business Street</Street>
123             <Housenumber>812</Housenumber>
124             <City>Medford</City>
125             <Stateprov>MA</Stateprov>
126             <Postalcode>02153</Postalcode>
127           </Postal>
128         </Addresses>
129       </data>
130     </transaction>
131     <transaction process="Purchase" id="3">
132       <dynamic>
133         <Uri>http://www.amazon.com/gp/css/homepage.html</Uri>
134         <Timestamp>1250852161</Timestamp>
135         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
136         <Cookies>true</Cookies>
137       </dynamic>
138       <data>
139         <Miscfields>
140         </Miscfields>
141         <Email>prevere@hotmail.com</Email>
142         <Password>a3cca2b2aa1e</Password>
143         <Financialdata>
144           <Account type="Credit Card">
145             <Creditcardnumber>1376151d963cad089697872af7dcec48bbd3ae29</Creditcardnumber>
146             <Creditcardexpiration>08/11</Creditcardexpiration>
147             <Creditcardsecuritycode>451</Creditcardsecuritycode>
148           </Account>
149         </Financialdata>
150       </data>
151     </transaction>
152   </hostname>
153   <hostname name="http://www.deviantart.com" title="deviantART">
154     <transaction process="Registration" id="1">
155       <dynamic>
156         <Uri>http://www.deviantart.com/join/</Uri>
157         <Timestamp>1250592961</Timestamp>
158         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
159         <Cookies>true</Cookies>
160       </dynamic>
161       <data>
162         <Miscfields>
163         </Miscfields>
164         <Email>braveman@gmail.com</Email>
165         <Password>b257cf1g9</Password>
166       </data>
167     </transaction>
168     <transaction process="Login" id="2">
169       <dynamic>
170         <Uri>http://www.deviantart.com/join/</Uri>
171         <Timestamp>1250765761</Timestamp>
172         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
173         <Cookies>true</Cookies>
174       </dynamic>
175       <data>
176         <Miscfields>
177         </Miscfields>
178         <Email>braveman@gmail.com</Email>
179         <Password>b257cf1g9</Password>
180       </data>
181     </transaction>
182   </hostname>
183   <hostname name="http://www.nytimes.com" title="The New York Times">
184     <transaction process="Registration" id="1">
185       <dynamic>
186         <Uri>http://www.nytimes.com/gst/regi.html</Uri>
187         <Timestamp>1251024961</Timestamp>
188         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
189         <Cookies>true</Cookies>
190       </dynamic>
191       <data>
192         <Miscfields>
193         </Miscfields>
194         <Id>smith123</Id>
195         <Password>d469ez1s6</Password>
196         <Email>smith@aol.com</Email>
197         <Bdate>24.09.1976</Bdate>
198       </data>
199     </transaction>
200     <transaction process="Login" id="2">
201       <dynamic>
202         <Uri>http://www.nytimes.com/auth/login</Uri>
203         <Timestamp>1251370561</Timestamp>
204         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
205         <Cookies>true</Cookies>
206       </dynamic>
207       <data>
208         <Miscfields>
209         </Miscfields>
210         <Id>smith123</Id>
211         <Password>d469ez1s6</Password>
212       </data>
213     </transaction>
214   </hostname>
215   <hostname name="http://www.google.com" title="Google">
216     <transaction process="Registration" id="1">
217       <dynamic>
```

```
218         <Uri>https://www.google.com/accounts/NewAccount</Uri>
219         <Timestamp>1250420161</Timestamp>
220         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
221         <Cookies>true</Cookies>
222       </dynamic>
223       <data>
224         <Miscfields>
225         </Miscfields>
226         <Given>Mike</Given>
227         <Family>Smith</Family>
228         <Email>braveman@gmail.com</Email>
229         <Password>d257ca1g9</Password>
230       </data>
231     </transaction>
232     <transaction process="Login" id="2">
233       <dynamic>
234         <Uri>https://www.google.com/accounts/Login</Uri>
235         <Timestamp>1250592961</Timestamp>
236         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
237         <Cookies>true</Cookies>
238       </dynamic>
239       <data>
240         <Miscfields>
241         </Miscfields>
242         <Email>braveman@gmail.com</Email>
243         <Password>d257ca1g9</Password>
244       </data>
245     </transaction>
246   </hostname>
247   <hostname name="http://www.youtube.com" title="YouTube">
248     <transaction process="Registration" id="1">
249       <dynamic>
250         <Uri>http://youtube.com/signup</Uri>
251         <Timestamp>1251284161</Timestamp>
252         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
253         <Cookies>true</Cookies>
254       </dynamic>
255       <data>
256         <Miscfields>
257         </Miscfields>
258         <Given>Mike</Given>
259         <Family>Smith</Family>
260         <Email>braveman@gmail.com</Email>
261         <Password>d257ca1g9</Password>
262         <Addresses>
263           <Postal>
264             <Street>Fake Street</Street>
265             <Housenumber>1201</Housenumber>
266             <City>Arlington</City>
267             <Stateprov>MD</Stateprov>
268             <Postalcode>35416</Postalcode>
269           </Postal>
270         </Addresses>
271       </data>
272     </transaction>
273     <transaction process="Login" id="2">
274       <dynamic>
275         <Uri>http://youtube.com/login</Uri>
276         <Timestamp>1251370561</Timestamp>
277         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
278         <Cookies>true</Cookies>
279       </dynamic>
280       <data>
281         <Miscfields>
282         </Miscfields>
283         <Email>braveman@gmail.com</Email>
284         <Password>d257ca1g9</Password>
285       </data>
286     </transaction>
287   </hostname>
288   <hostname name="http://www.homeland.com" title="Homeland">
289     <transaction process="Registration" id="1">
290       <dynamic>
291         <Uri>http://youtube.com/signup</Uri>
292         <Timestamp>1250592961</Timestamp>
293         <Useragent>Mozilla/5.0 (Windows; U; Windows NT 5.1; de; rv:1.9) Gecko/2008052906 Firefox/3.0</Useragent>
294         <Cookies>true</Cookies>
295       </dynamic>
296       <data>
297         <Miscfields>
298         </Miscfields>
299         <Given>Joe</Given>
300         <Family>Biden</Family>
301         <Email>jbiden@googlemail.com</Email>
302         <Password>f369et1f4</Password>
303       </data>
304     </transaction>
305   </hostname>
306 </xml>
```

Bibliography

[ACR99] M. S. Ackerman, L. Cranor, and J. Reagle. Privacy in E-commerce: Examining User Scenarios and Privacy Preferences. In *Proceedings of the 1st ACM Conference on Electronic Commerce (EC '99)*, pages 1–8, New York, NY, USA, 1999. ACM.

[AG05] A. Acquisti and J. Grossklags. Privacy and Rationality in Individual Decision Making. *IEEE Security and Privacy*, 3(1):26–33, 2005.

[AHK+03] P. Ashley, S. Hada, G. Karjoth, C. Powers, and M. Schunter. Enterprise Privacy Authorization Language (EPAL 1.2). IBM Research Report, 2003.

[AHKS02] P. Ashley, S. Hada, G. Karjoth, and M. Schunter. E-P3P Privacy Policies and Privacy Authorization. In *Proceedings of the 2002 ACM Workshop on Privacy in the Electronic Society (WPES '02)*, pages 103–109, New York, NY, USA, 2002. ACM.

[AKSX03] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. An XPath-based Preference Language for P3P. In *Proceedings of the 12th International Conference on World Wide Web (WWW '03)*, pages 629–639, New York, NY, USA, 2003. ACM.

[Alb07] T. Alby. *Web 2.0 – Konzepte, Anwendungen, Technologien*. Carl Hanser Verlag, Munich, 2007.

[And04] A. H. Anderson. Privacy Policy Languages: XACML vs EPAL. In *5th Annual Privacy & Security Workshop*. Sun Microsystems Labs, October 2004.

[And06] A. H. Anderson. A Comparison of Two Privacy Policy Languages: EPAL and XACML. In *Proceedings of the 3rd ACM Workshop on Secure Web Services (SWS '06)*, pages 53–60, New York, NY, USA, 2006. ACM.

[Ars04] F. Arshad. Privacy Fox – A JavaScript-based P3P Agent for Mozilla Firefox. Technical report, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA, December 2004.

[Bar06] T. Bartels. Nutzung von Wikis als Wissensmanagement unterstützende Systeme in Unternehmen. Diplomarbeit, Universität Köln, 2006.

[BBC+07] A. Berglund, S. Boag, D. Chamberlin, M. Férnandez, M. Kay, J. Robie, and J. Siméon. XML Path Language (XPath) 2.0. W3C Recommendation, World Wide Web Consortium (W3C), January 2007.

[BCHL09] B. Bos, T. Celik, I. Hickson, and H. W. Lie. Cascading Style Sheets Level 2 Revision 1 (CSS 2.1) Specification. W3C Candidate Recommendation, World Wide Web Consortium (W3C), April 2009.

[BDS06] Bundesdatenschutzgesetz, August 2006.

[BE07] D. M. Boyd and N. B. Ellison. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1):210–230, 2007.

[Beh01] L. Behrens. Privacy and Security: The Hidden Growth Strategy. GartnerG2 Report, http://www.gartner.com/5_about/press_releases/2001/pr20010807d.html, August 2001.

[Ben99] Paola Benassi. TRUSTe: An Online Privacy Seal Program. *Communications of the ACM*, 42(2):56–59, 1999.

[Ber04] Stefan Berthold. XACML im Vergleich mit P3P und EPAL. Hauptseminar Technische Universität Dresden, July 2004.

[Ber05] M. Bergmann. PRIME Internal Privacy Preference Survey About Privacy Concerns and Conditions. Technical report, Technische Universität Dresden, TUD-FI07-04-Mai-2005, May 2005.

[BG05] J. Bhattacharya and S. K. Gupta. EPAL Based Privacy Enforcement Using ECA Rules. In S. Jajodia and C. Mazumdar, editors, *Proceedings of the 1st International Conference on Information Systems Security (ICISS 2005)*, pages 120–133, Berlin, Heidelberg, 2005. Springer, LNCS 3803.

[BGS05] B. Berendt, O. Günther, and S. Spiekermann. Privacy in E-commerce: Stated Preferences vs. Actual Behavior. *Communications of the ACM*, 48(4):101–106, 2005.

[Bie87] I. Biederman. Recognition-By-Components: A Theory of Human Image Understanding. *Psychological Review*, 94(2):115–147, 1987.

[BJW08] P. Bodorik, D. Jutla, and M. Xuehai Wang. Consistent Privacy Preferences (CPP): Model, Semantics, and Properties. In *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC '08)*, pages 2368–2375, New York, NY, USA, 2008. ACM.

[BK02] Ute Binder-Kissel. *Telefonakquisition – Ihr Draht zum Kunden*. Wissenschaft & Praxis, 2002.

[BMH05] M. Bauer, M. Meints, and M. Hansen. Structured Overview on Prototypes and Concepts of Identity Management Systems. FIDIS Deliverable D3.1, http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp3-del3.1.overview_on_IMS.final.pdf, September 2005.

[BPFP05] K. Borcea-Pfitzmann, E. Franz, and A. Pfitzmann. Usable Presentation of Secure Pseudonyms. In *Proceedings of the 2005 Workshop on Digital Identity Management (DIM '05)*, pages 70–76, New York, NY, USA, 2005. ACM.

[Bro98] C. M. Brown. *Human-Computer Interface Design Guidelines*. Intellect Books, 1998.

[BS93] V. Bellotti and A. Sellen. Design for Privacy in Ubiquitous Computing Environments. In *Proceedings of the 3rd European Conference on Computer-Supported Cooperative Work (ECSCW '93)*, pages 77–92, Norwell, MA, USA, 1993. Kluwer Academic Publishers.

[BSTW03] L. Brückner, J. Steffan, W. Terpstra, and U. Wilhelm. Active Data Protection with Data Journals. In *Proceedings of the GI Jahrestagung (Schwerpunkt "Sicherheit - Schutz und Zuverlässigkeit")*, pages 269–280, 2003.

[Bur82] J. K. Burgoon. Privacy and Communication. In M. Burgoon, editor, *Communication Yearbook 6*, pages 206–249. Sage Publications, Beverly Hills, CA, 1982.

[Bur97] H. Burkert. Privacy-enhancing Technologies: Typology, Critique, Vision. In P.E. Agre and M. Rotenberg, editors, *Technology and Privacy: The New Landscape*, pages 126–143, Boston, MA, USA, 1997. MIT Press.

[Cam05] K. Cameron. The Laws of Identity. http://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf, May 2005.

[CDE+06] L. Cranor, B. Dobbs, S. Egelman, G. Hogben, J. Humphrey, and al. The Platform for Privacy Preferences 1.1 Specification. W3C Working Group Note, World Wide Web Consortium (W3C), November 2006.

[CG05] L. Cranor and S. Garfinkel. *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly Media, Inc., 2005.

[CGA06] L. Cranor, P. Guduru, and M. Arjula. User Interfaces for Privacy Agents. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(2):135–178, 2006.

[CH91] J. Clark and Derek A. Holton. *A First Look at Graph Theory*. World Scientific, 1991.

[Chi00] E. H. Chi. A Taxonomy of Visualization Techniques Using the Data State Reference Model. In *Proceedings of the IEEE Symposium on Information Vizualization 2000 (INFOVIS '00)*, pages 69–75, Washington, DC, USA, 2000. IEEE Computer Society.

[CHL+07] R. Chinnici, H. Haas, A. A. Lewis, J.-J. Moreau, D. Orchard, and S. Weerawarana. Web Services Description Language (WSDL) version 2.0 Part 2: Adjuncts. W3C Recommendation, World Wide Web Consortium (W3C), June 2007.

[CJ06] K. Cameron and M. B. Jones. Design Rationale Behind the Identity Metasystem Architecture. http://www.identityblog.com/wp-content/resources/design_rationale.pdf, 2006.

[CK01] S. Clauß and M. Köhntopp. Identity Management and Its Support of Multilateral Security. *Computer Networks*, 37(2):205–219, 2001.

[CKK05] S. Clauß, D. Kesdogan, and T. Kölsch. Privacy Enhancing Identity Management: Protection Against Re-identification and Profiling. In *Proceedings of the 2005 Workshop on Digital Identity Management (DIM '05)*, pages 84–93, New York, NY, USA, 2005. ACM.

[Cla93] R. Clarke. Computer Matching and Digital Identity. In *Proceedings of the Computers, Freedom & Privacy Conference*, San Francisco, March 1993. http://cpsr.org/prevsite/conferences/cfp93/clarke.html.

[Cla06] S. Clauß. A Framework for Quantification of Linkability Within a Privacy-Enhancing Identity Management System. In *Emerging Trends in Information and Communication Security*, pages 191–205, Heidelberg, 2006. Springer, LNCS 3995.

[CLM02a] L. Cranor, M. Langheinrich, and M. Marchiori. A P3P Preference Exchange Language 1.0 (APPEL 1.0). W3C Working Draft, World Wide Web Consortium (W3C), April 2002.

[CLM+02b] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle. The Platform for Privacy Preferences 1.0 Specification. W3C Recommendation, World Wide Web Consortium (W3C), April 2002.

[CM02] A. Cockburn and B. McKenzie. Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*, pages 203–210, New York, NY, USA, 2002. ACM.

[CMRW07] R. Chinnici, J.-J. Moreau, A. Ryman, and S. Weerawarana. Web Services Description Language (WSDL) version 2.0 Part 1: Core Language. W3C Recommendation, World Wide Web Consortium (W3C), June 2007.

[COB06] S. Chiasson, P.C. van Oorschot, and R. Biddle. A Usability Study and Critique of Two Password Managers. In *Proceedings of the 15th conference on USENIX Security Symposium (USENIX-SS'06)*, pages 1–16, Berkeley, CA, USA, August 2006. USENIX Association.

[Com03] Commission of the European Communities. First Report on the Implementation of the Data Protection Directive (95/46/EC). http://ec.europa.eu/justice_home/fsj/privacy/lawreport/report_en.htm, Brussels, May 2003.

[Com07] Commission of the European Communities. Communication from the Commission to the European Parliament and the Council on the Follow-up of the Work Programme for Better Implementation of the Data Protection Directive. http://ec.europa.eu/justice_home/fsj/privacy/lawreport/index_en.htm, Brussels, March 2007.

[COP98] Children's Online Privacy Protection Act (COPPA), 1998.

[Coy99] K. Coyle. Pretty Poor Privacy: A Social Analysis of the Platform for Privacy Preferences. http://www.kcoyle.net/p3p.html, June 1999.

[CR98] E. H. Chi and J. Riedl. An Operator Interaction Framework for Visualization Systems. In *Proceedings of the 1998 IEEE Symposium on Information Visualization (INFOVIS '98)*, pages 63–70, Washington, DC, USA, 1998. IEEE Computer Society.

[CRA99] L. Cranor, J. Reagle, and M.S. Ackerman. Beyond Concern: Understanding Net Users' Attitudes About Online Privacy. Technical report, AT&T Labs-Research, TR 99.4.3, 1999.

[CS05] R. K. Chellappa and R. G. Sin. Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. *Information Technology and Management*, 6(2-3):181–202, 2005.

[CS07] G. Conti and E. Sobiesk. An Honest Man Has Nothing to Fear: User Perceptions on Web-based Information Disclosure. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*, pages 112–121, New York, NY, USA, 2007. ACM.

[CSS+05] J. Camenisch, A. Shelat, D. Sommer, S. Fischer-Hübner, M. Hansen, H. Krasemann, G. Lacoste, R. Leenes, and J. Tseng. Privacy and Identity Management for Everyone. In *Proceedings of the 2005 Workshop on Digital Identity Management (DIM '05)*, pages 20–27, New York, NY, USA, 2005. ACM.

[CSW97] S.-Y. Choi, D. O. Stahl, and A. B. Whinston. *The Economics of Electronic Commerce*. Pearson Education Ltd., 1997.

[CvOA07] J. Clark, P. C. van Oorschot, and C. Adams. Usability of Anonymous Web Browsing: An Examination of Tor Interfaces and Deployability. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*, pages 41–51, New York, NY, USA, 2007. ACM.

[Dav04] J. Davies. Wiki Brainstorming and Problems with Wiki Based Collaboration. Project Report Submitted for the Degree of Information Processing, Department of Computer Science, University of York, September 2004.

[DBETT99] G. Di Battista, P. Eades, R. Tamassia, and I. G. Tollis. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, 1999.

[ECC06] S. Egelman, L. Cranor, and A. Chowdhury. An Analysis of P3P-enabled Web Sites Among Top-20 Search Results. In *Proceedings of the 8th International Conference on Electronic Commerce (ICEC '06)*, pages 197–207, New York, NY, USA, 2006. ACM.

[Eck07] C. Eckert. *IT-Sicherheit: Konzepte - Verfahren - Protokolle*. Oldenbourg, 5th edition, 2007.

[EGH07] A. Ebersbach, M. Glaser, and R. Heigel. Gemeinsam weben. *c't*, (14):168–171, 2007.

[EHG07] T. M. Eap, M. Hatala, and D. Gasevic. Enabling User Control with Personal Identity Management. In *Proceedings of the IEEE International Conference on Services Computing (SCC 2007)*, pages 60–67. IEEE Computer Society, 2007.

[EMS07] T. El Maliki and J.-M. Seigneur. A Survey of User-centric Identity Management Technologies. In *Proceedings of the International Conference on Emerging Security Information, Systems, and Technologies (SECUREWARE '07)*, pages 12–17, Washington, DC, USA, 2007. IEEE Computer Society.

[Eur95] European Parliament. EU-Directive 95/46/EC. Official Journal of the European Communities No L 281 31, October 1995.

[Fau03] L. Faulkner. Beyond the Five-user Assumption: Benefits of Increased Sample Sizes in Usability Testing. *Behavior Research Methods, Instruments and Computers*, 35(3):379–383, 2003.

[Fed05] H. Federrath. Privacy Enhanced Technologies: Methods – Markets – Misuse. In *2nd International Conference on Trust, Privacy, and Security in Digital Business (TrustBus '05)*, pages 1–9, Heidelberg, 2005. Springer.

[FHP04] S. Fischer-Hübner and J.S. Pettersson. Evaluation of Early Prototypes. PRIME Deliverable D6.1.b, http://www.prime-project.eu/prime_products/reports/eval/, December 2004.

[FHPB+07] S. Fischer-Hübner, J. S. Pettersson, M. Bergmann, M. Hansen, S. Pearson, and M. Casassa Mont. HCI Designs for Privacy-Enhancing Identity Management. In A. Acquisti, S. De Capitani di Vimercati, S. Gritzalis, and C. Lambrinoudakis, editors, *Digital Privacy: Theory, Technologies and Practices*, chapter 11, pages 230–249. Auerbach Publications (Taylor and Francis Group), 2007.

[FKPT07] L. Fahrmeir, R. Künstler, I. Pigeot, and G. Tutz. *Statistik. Der Weg zur Datenanalyse*. Springer, Berlin, 6th edition, 2007.

[Fla06] D. Flanagan. *JavaScript: The Definitive Guide*. O'Reilly Media, Inc., 5th edition, 2006.

[FML+01] 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 *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '01)*, pages 61–68, New York, NY, USA, 2001. ACM.

[Gar05] J. J. Garrett. Ajax: A New Approach to Web Applications. Adaptive Path, http://www.adaptivepath.com/ideas/essays/archives/000385.php, February 2005.

[GHM+07a] M. Gudgin, M. Hadley, N. Mendelsohn, J.-J. Moreau, H. F. Nielsen, A. Karmarkar, and Y. Lafon. SOAP version 1.2 Part 1: Messaging Framework (Second Edition). W3C Recommendation, World Wide Web Consortium (W3C), April 2007.

[GHM+07b] M. Gudgin, M. Hadley, N. Mendelsohn, J.-J. Moreau, H. F. Nielsen, A. Karmarkar, and Y. Lafon. SOAP version 1.2 Part 2: Adjuncts (Second Edition). W3C Recommendation, World Wide Web Consortium (W3C), April 2007.

[GWB97] I. Goldberg, D. Wagner, and E. Brewer. Privacy-enhancing Technologies for the Internet. In *Proceedings of the 42nd IEEE International Computer Conference (COMPCON '97)*, pages 103–109, Washington, DC, USA, 1997. IEEE Computer Society.

[HAF05] B. A. Huberman, E. Adar, and L. R. Fine. Valuating Privacy. *IEEE Security and Privacy*, 3(5):22–25, 2005.

[HFHPB07] M. Hansen, S. Fischer-Hübner, J. S. Pettersson, and M. Bergmann. Transparency Tools for User-controlled Identity Management. In P. Cunningham and M. Cunningham, editors, *Expanding the Knowledge Economy: Issues, Applications, Case Studies*, pages 1360–1367. IOS Press, 2007.

[HI02] Harris-Interactive. Privacy On and Off the Internet: What Consumers Want. http://www.aicpa.org/download/webtrust/priv_rpt_21mar02.pdf, February 2002.

[HIP96] Health Insurance Portability and Accountability Act (HIPAA), 1996.

[HJW02] G. Hogben, T. Jackson, and M. Wilikens. A Fully Compliant Research Implementation of the P3P Standard for Privacy Protection: Experiences and Recommendations. In *Proceedings of the 7th European Symposium on Research in Computer Security (ESORICS '02)*, pages 104–125, London, UK, 2002. Springer.

[HM07] M. Hansen and S. Meissner, editors. *Verkettung digitaler Identitäten*. Independent Centre for Privacy Protection (ICPP), 2007.

[HMPR04] A. R. Hevner, S. T. March, J. Park, and S. Ram. Design Science in Information Systems Research. *MIS Quarterly*, 28(1):75–105, 2004.

[HW91] Harris Louis and Associates and A. F. Westin. Harris-Equifax Consumer Privacy Survey. Equifax Inc., Atlanta, Georgia, 1991.

[Ind03] Independent Centre for Privacy Protection (ICPP) and Studio Notarile Genghini. Identity Management Systems: Identification and Comparison Study. Contract N°19960-2002-10 F1ED SEV DE, September 2003.

[IP99] L. D. Introna and A. Pouloudi. Privacy in the Information Age: Stakeholders, Interests and Values. *Journal of Business Ethics*, 22(1):27–38, 1999.

[ITU95] ITU/ISO. ITU-T Rec. X.812, ISO/IEC 10181-3:1996, Information Technology - Open Systems Interconnection – Security Frameworks in Open Systems: Access Control Framework. http://www.itu.int/rec/T-REC-X.812-199511-I/en, 1995.

[IW03] P. Irani and C. Ware. Diagramming Information Structures Using 3D Perceptual Primitives. *ACM Transactions on Computuer-Human Interaction*, 10(1):1–19, 2003.

[JPJ05] 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.

[JZS07] A. Jøsang, M. Al Zomai, and S. Suriadi. Usability and Privacy in Identity Management Architectures. In *Proceedings of the 5th Australasian Symposium on ACSW Frontiers (ACSW '07)*, pages 143–152, Darlinghurst, Australia, 2007. Australian Computer Society Inc.

[KC05] P. Kumaraguru and L. Cranor. Privacy Indexes: A Survey of Westin's Studies. Technical report, Carnegie Mellon University, CMU-ISRI-5-138, http://reports-archive.adm.cs.cmu.edu/anon/isri2005/CMU-ISRI-05-138.pdf, December 2005.

[KCLC07] P. Kumaraguru, L. Cranor, J. Lobo, and S. Calo. A Survey of Privacy Policy Languages. In *Workshop on Usable IT Security Management (USM '07)*, July 2007.

[KD08] E. Kosta and J. Dumortier. Taxonomy. PICOS Deliverable D2.1, http://www.picos-project.eu/Public-Deliverables.29.0.html, July 2008. version 1.0.

[KDJ+05] P. Kolari, L. Ding, S. G. A. Joshi, T. Finin, and L. Kagal. Enhancing Web Privacy Protection through Declarative Policies. In *Proceedings of the 6th IEEE Workshop on Policy for Distributed Systems and Networks (POLICY 2005)*, pages 57–66, Washington, DC, USA, June 2005. IEEE Computer Society.

[KDK+04] P. Kolari, L. Ding, L. Kagal, S. Ganjugunte, A. Joshi, and T. Finin. Enhancing P3P Framework through Policies and Trust. Technical report, University of Maryland Baltimore County, September 2004.

[KFJ03] L. Kagal, T. Finin, and A. Joshi. A Policy Language for a Pervasive Computing Environment. In *Proceedings of the 4th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY '03)*, pages 63–74, Washington, DC, USA, 2003. IEEE Computer Society.

[KKP01] A. Kobsa, J. Koenemann, and W. Pohl. Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships. *The Knowledge Engineering Review*, 16(2):111–155, 2001.

[KKP09] J. Kolter, T. Kernchen, and G. Pernul. Collaborative Privacy – A Community-based Privacy Infrastructure. In *Proceedings of the 24th IFIP TC-11 International Information Security Conference (SEC 2009)*, pages 226–236, Berlin, May 2009. Springer.

[Kob04] A. Kobsa. Adaptive Methoden – Benutzermodellierung (Adaptive Methods – User Modeling). In R. Kuhlen, T. Seeger, and D. Strauch, editors, *Grundlagen der Information und Dokumentation (5th completely revised edition)*, pages 299–302. K. G. Saur, 2004.

[Kob07a] J. Kobielus. User-Centric Identity Still in Early Development. *Business Communications Review*, pages 50–55, May 2007.

[Kob07b] A. Kobsa. Privacy-Enhanced Personalization. *Communications of the ACM*, 50(8):24–33, 2007.

[Kob07c] A. Kobsa. Privacy-Enhanced Web Personalization. In P. Brusilovsky, A. Kobsa, and W. Nejdl, editors, *The Adaptive Web: Methods and Strategies of Web Personalization*, pages 628–670, Heidelberg, New York, 2007. Springer.

[KP01] M. Köhntopp and A. Pfitzmann. Informationelle Selbstbestimmung durch Identitätsmanagement. *it – Information Technology*, 43(5):227–236, 2001.

[KP09] J. Kolter and G. Pernul. Generating User-understandable Privacy Preferences. In *Proceedings of the 4th International Conference on Availability, Reliability and Security (ARES 2009)*, Fukuoka, Japan, March 2009. IEEE Computer Society.

[KS02] G. Karjoth and M. Schunter. A Privacy Policy Model for Enterprises. In *Proceedings of the 15th IEEE Workshop on Computer Security Foundations (CSFW '02)*, pages 271–281, Washington, DC, USA, June 2002. IEEE Computer Society.

[Lam06] E. Lammenett. Team Inmedias – Online Marketing Studie 2006. http://www.inmedias.de/internet-marketing/internet-marketing-email-marketing.html, February 2006.

[LC01] B. Leuf and W. Cunningham. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley Longman, Amsterdam, 2001.

[LG05] S. E. Levy and C. Gutwin. Improving Understanding of Website Privacy Policies with Fine-grained Policy Anchors. In *Proceedings of the 14th International*

Conference on World Wide Web (WWW '05), pages 480–488, New York, NY, USA, 2005. ACM.

[LHDL04] S. Lederer, I. Hong, K. Dey, and A. Landay. Personal Privacy through Understanding and Action: Five Pitfalls for Designers. *Personal Ubiquitous Computing*, 8(6):440–454, 2004.

[LLCL98] M. K. Landesberg, T. M. Levin, C. G. Curtin, and O. Lev. Privacy Online: A Report to Congress. Federal Trade Commission (FTC), http://www.ftc.gov/reports/privacy3/toc.shtm, June 1998.

[LSH08] R. Leenes, J. Schallaböck, and M. Hansen. PRIME White Paper version 3. https://www.prime-project.eu/prime_products/whitepaper/, 2008.

[LYA03] N. Li, T. Yu, and A. Antón. A Semantics Based Approach to Privacy Languages. Technical report, CERIAS, TR 2003-28, http://www.cs.purdue.edu/homes/ninghui/papers/p3p_csse.pdf, November 2003.

[Mal08] B. Malin. K-Unlinkability: A Privacy Protection Model for Distributed Data. *Data and Knowledge Engineering*, 64(1):294–311, 2008.

[May06] T. Mayer. *Personalisierungsstrategien im E-Commerce: Die Webloganalyse als Instrument der Personalisierung im Rahmen des eCRM*. Lang, Peter Frankfurt, 2006.

[MCMW06] P. Madsen, M. Casassa-Mont, and R. Wilton. A Privacy Policy Framework – A Position Paper for the W3C Workshop of Privacy Policy Negotiation. Position Paper, http://www.w3.org/2006/07/privacy-ws/papers/28-madsen-framework/, October 2006.

[Mey08] C. Meyer. Alter Raub, neuer Skandal: 17 Millionen Telekom-Nummern entwendet. *heise.de / c't*, October 2008. http://www.heise.de/security/Alter-Raub-neuer-Skandal-17-Millionen-Telekom-Nummern-entwendet--/news/meldung/116913.

[MLM+06] 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.

[Mos05a] T. Moses. eXtensible Access Control Markup Language version 2.0 (XACML). OASIS Standard, February 2005.

[Mos05b] T. Moses. Privacy Policy Profile of XACML v2.0. OASIS Standard, February 2005.

[MW93] J.-J. C. Meyer and R. J. Wieringa. Deontic Logic: A Concise Overview. In *Deontic Logic in Computer Science: Normative System Specification*, pages 3–16. John Wiley & Sons, Chichester, UK, 1993.

[NDR09] NDR Info. Handel mit persönlichen Verbraucherdaten blüht massenhaft weiter. News Aktuell, http://www.presseportal.de/pm/6561/1455558/ndr_norddeutscher_rundfunk, August 2009.

[Nel98] R. Nelson. Privacy Evaluator. W3C Note, World Wide Web Consortium (W3C), 1998.

[Obj07a] Object Management Group. OMG Unified Modeling Language, Infrastructure Specification, version 2.1.2, November 2007.

[Obj07b] Object Management Group. OMG Unified Modeling Language, Superstructure Specification, version 2.1.2, November 2007.

[Odl03] A. Odlyzko. Privacy, Economics, and Price Discrimination on the Internet. In *Proceedings of the 5th international Conference on Electronic Commerce (ICEC '03)*, pages 355–366, New York, NY, USA, 2003. ACM.

[Org80] Organisation for Economic Co-operation and Development (OECD). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, September 1980.

[Par06] L. Parry. A Tale of Two Wikis: Techniques for Building, Managing and Promoting Collaborative Communities. In *Proceedings of the International Wikimedia Conference (Wikimania 2006)*, Cambridge, MA, USA, August 2006.

[PBP06] A. Pfitzmann and K. Borcea-Pfitzmann. Identitätsmanagement und informationelle Selbstbestimmung. In A. Roßnagel, editor, *Allgegenwärtige Identifizierung? Neue Identitätsinfrastrukturen und ihre rechtliche Gestaltung*, number 33, pages 83–91. Nomos Verlagsgesellschaft - Schriftenreihe des Instituts für Europäisches Medienrecht, 2006.

[Pet04] J.S. Pettersson. P3P and Usability - The Mobile Case. In P. Duquennoy, S. Fischer-Hübner, J. Holvast, and A. Zuccato, editors, *Risk and Challenges of the Network Society*. Karlstad University Studies 2004:35, 2004.

[Pet08] J.S. Pettersson. HCI Guidelines. PRIME Deliverable D6.1, February 2008.

[PFHB06] J.S. 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.

[PH08] A. Pfitzmann and M. Hansen. Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology. Technical report, http://dud.inf.tu-dresden.de/Anon_Terminology.shtml, February 2008.

[PK99] T. Probst and M. Köhntopp. Datenschutzgerechter und datenschutzfördernder Einsatz von biometrischen Verfahren. In *Potential biometrischer Systeme als Privacy-Enhancing Technologies*, BSI Expert Panel, Bonn, November 1999.

[PK03] A. S. Patrick and S. Kenny. From Privacy Legislation to Interface Design: Implementing Information Privacy in Human-Computer Interactions. In R. Dingledine, editor, *Proceedings of the Privacy Enhancing Technologies Workshop (PET 2003)*, pages 107–124, Berlin, Heidelberg, 2003. Springer.

[PKHvB03] A. S. Patrick, S. Kenny, C. Holmes, and M. van Breukelen. Human Computer Interaction. In G. W. van Blarkom, J. J. Borking, and J. G. E. Olk, editors, *Handbook for Privacy and Privacy-Enhancing Technologies*, chapter 12, pages 249–290. College Bescherming Persoonsgegevens, The Hague, The Netherlands, 2003.

[PL01] C. Perfetti and L. Landesmann. Eight Is Not Enough. http://www.uie.com/articles/eight_is_not_enough/, June 2001.

[Pol07] I. Pollach. What's Wrong With Online Privacy Policies? *Communications of the ACM*, 50(9):103–108, 2007.

[Pro07] J. Protzenko. *XUL - Entwicklung von Rich Clients mit der Mozilla XML User Interface Language*. Open Source Press, Munich, 2007.

[PTG+06] K. Peffers, T. Tuunanen, C. E. Gengler, M. Rossi, W. Hui, V. Virtanen, and J. Bragge. The Design Science Research Process: A Model for Producing and Presenting Information Systems Research. In *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006)*, pages 84–147, Claremont, CA, USA, February 2006.

[RBDM07] I. K. Reay, P. Beatty, S. Dick, and J. Miller. A Survey and Analysis of the P3P Protocol's Agents, Adoption, Maintenance, and Future. *IEEE Transactions on Dependable Secure Computing*, 4(2):151–164, 2007.

[RC99] J. Reagle and L. Cranor. The Platform for Privacy Preferences. *Communications of the ACM*, 42(2):48–55, 1999.

[RCMC00] K. Risden, M.P. Czerwinski, T. Munzner, and D. Cook. An Initial Examination of Ease of Use for 2D and 3D Information Visualizations of Web Content. *International Journal of Human-Computer Studies*, 53(5):695–714, November 2000.

[RLHJ99] D. Raggett, A. Le Hors, and I. Jacobs. HTML 4.01 Specification. W3C Recommendation, World Wide Web Consortium (W3C), December 1999.

[RR06] D. Recordon and D. Reed. OpenID 2.0: A Platform for User-centric Identity Management. In *Proceedings of the 2nd ACM Workshop on Digital Identity Management (DIM '06)*, pages 11–16, New York, NY, USA, 2006. ACM.

[SAHJ04] W. H. Stufflebeam, A. I. Antón, Q. He, and N. Jain. Specifying Privacy Policies with P3P and EPAL: Lessons Learned. In *Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society (WPES '04)*, pages 35–35, New York, NY, USA, 2004. ACM.

[Sch08] A. Schulze. *Rich Internet Applikationen – Best Practices vom Core bis zum Desktop*. entwickler.press, 2008.

[Sea03] S. Searby. Personalisation – An Overview of Its Use and Potential. *BT Technology Journal*, 21(1):13–19, 2003.

[Sev07] SevenOne Interactive GmbH. @facts extra – Online Nutzertypen 2007. Unterföhring, 2007.

[SGB01] S. Spiekermann, J. Grossklags, and B. Berendt. E-privacy in 2nd Generation E-commerce: Privacy Preferences versus Actual Behavior. In *Proceedings of the 3rd ACM Conference on Electronic Commerce (EC '01)*, pages 38–47, New York, NY, USA, 2001. ACM.

[Shn96] B. Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages (VL '96)*, pages 336–343, Washington, DC, USA, 1996. IEEE Computer Society.

[SIG07] M. Shanmugasundaram, P. Irani, and C. Gutwin. Can Smooth View Transitions Facilitate Perceptual Constancy in Node-link Diagrams? In *Proceedings of Graphics Interface 2007 (GI '07)*, pages 71–78, New York, NY, USA, 2007. ACM.

[SK03] S. Steinbrecher and S. Köpsell. Modelling Unlinkability. In *Proceedings of the Privacy Enhancing Technologies Workshop (PET 2003)*, pages 32–47, Berlin, Heidelberg, 2003. Springer.

[Sol08] D. J. Solove. *Understanding Privacy*. Harvard University Press, Cambridge, MA, 2008.

[Spe05] B. Spell. *Pro Java Programming, Second Edition*. Apress, 2005.

[SS01] J. Spool and W. Schroeder. Testing Web Sites: Five Users Is Nowhere Near Enough. In *Extended Abstracts on Human Factors in Computing Systems (CHI '01)*, pages 285–286, New York, NY, USA, 2001. ACM.

[SSA06] S. Sackmann, J. Strüker, and R. Accorsi. Personalization in Privacy-aware Highly Dynamic Systems. *Communications of the ACM*, 49(9):32–38, 2006.

[Sta08] Statistisches Bundesamt Deutschland. Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien (IKT). Fachserie 15 - Reihe 4, Wiesbaden, April 2008.

[STT81] K. Sugiyama, S. Tagawa, and M. Toda. Methods for Visual Understanding of Hierarchical System Structures. *IEEE Transactions on Systems, Man and Cybernetics*, 11(2):109–125, February 1981.

[SVHW03] M. Schunter, E. Van Herreweghen, and M. Waidner. Translating EPAL to P3P – How to Keep Enterprise Privacy Promises in Sync with the Actual Practices. Position Paper, http://www.w3.org/2003/p3p-ws/pp/ibm2.html, March 2003.

[Swe02] L. Sweeney. K-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[Tay03] H. Taylor. Most People Are Privacy Pragmatists Who, While Concerned about Privacy, Will Sometimes Trade It Off for Other Benefits. http://www.harrisinteractive.com/harris_poll/index.asp?PID=365, March 2003.

[Tid05] J. Tidwell. *Designing Interfaces*. O'Reilly Media Inc., Sebastopol, 2005.

[TK04] M. Teltzrow and A. Kobsa. Impacts of User Privacy Preferences on Personalized Systems: A Comparative Study. In C.-M. Karat, J. Blom, and J. Karat, editors, *Designing Personalized User Experiences in eCommerce*, pages 315–332, Dordrecht, The Netherlands, 2004. Kluwer Academic Publishers.

[TMG08] Telemediengesetz, December 2008.

[Tre07] H. Treiblmaier. Beziehungsmarketing aus Kundensicht. *Wirtschaftsinformatik*, 49(1):42–48, 2007.

[U.S00] U.S. Departement of Commerce. Safe Harbor Privacy Principles. http://www.export.gov/safeHarbor/, July 2000.

[War04] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2nd edition, 2004.

[WB90] S. D. Warren and L. D. Brandeis. The Right to Privacy. *Harward Law Review*, 4(5):193–220, December 1890.

[WBWK00] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for Using Multiple Views in Information Visualization. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '00)*, pages 110–119, New York, NY, USA, 2000. ACM.

[Wes67] A. Westin. *Privacy and Freedom*. The Bodley Head, 1967.

[WF05] I. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems, 2nd edition, 2005.

[Wid08] S. Widjaja. *Rich Internet Applications mit Adobe Flex 3*. Hanser Fachbuch, 2008.

[WK07] Y. Wang and A. Kobsa. Respecting Users' Individual Privacy Constraints in Web Personalization. In *Proceedings of the 11th International Conference on User Modeling (UM '07)*, pages 157–166, Berlin, Heidelberg, 2007. Springer.

[WK08] Y. Wang and A. Kobsa. Privacy-Enhancing Technologies. In M. Gupta and R. Sharman, editors, *Handbook of Research on Social and Organizational Liabilities in Information Security*, pages 352–375, Hershey, PA, November 2008. IGI Global.

[WSS+01] A. Westerinen, J. Schnizlein, J. Strassner, M. Scherling, B. Quinn, S. Herzog, A. Huynh, M. Carlson, J. Perry, and S. Waldbusser. Terminology for Policy-Based Management. Internet Engineering Task Force, RFC 3198, http://www.ietf.org/rfc/rfc3198.txt, 2001.

[WT99] A. Whitten and J. D. Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. In *Proceedings of the 8th Conference on USENIX Security Symposium (SSYM'99)*, pages 14–14, Berkeley, CA, USA, 1999. USENIX Association.

[You78] J. B. Young. Introduction: A Look at Privacy. In J. B. Young, editor, *Privacy*. John Wiley & Sons, Chichester, New York, 1978.

[YPG00] R. Yavatkar, D. Pendarakis, and R. Guerin. A Framework for Policy-based Admission Control. Internet Engineering Task Force, RFC 2753, http://www.ietf.org/rfc/rfc2753.txt, 2000.

Acknowledgment

At this point, I would like to thank all people who supported me in creating this dissertation. Without their help, this work would not have been possible.

First of all, many thanks go to my advisor Günther Pernul who provided me with the necessary freedom of creativity and gave me the invaluable chance to publish the research results of this work at many scientific conferences. Also, I would like to thank my co-advisor Alfred Kobsa who provided me with stimulating input and valuable feedback.

Furthermore, I thank all my colleagues and friends at the Department of Information Systems who were open for fruitful discussions and who did a great job in proofreading this work.

Last but not least, I owe thanks to our students at the University of Regensburg who contributed to this work with their theses and projects. In particular, I would like to thank Michael Diepold, Matthias Fraunhofer, Pascal Jonietz, Stefan Kendlbacher, Thomas Kernchen, Bastian Kistner, Florian Müller, Michael Netter, Christopher Piosecny, Moritz Riesner, Achim Sautter, Matthias Ulmer and Thorsten Viehauser.