Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Note

Social Media, e-ID and Privacy

Background for the e-Me project

Note no DART/02/2011

Authors Lothar Fritsch

Date April 11, 2011

About the author

Dr. Lothar Fritsch is currently a research scientist at the Norwegian Computing Center (Norsk Regnesentral, NR) in Oslo. Lothar's work focuses on the analysis of security and privacy requirements in upcoming application areas. Particularly he worked on the deployment of privacy functionality into new systems with respect to requirements engineering, business models, and verification. Lothar is an experienced researcher who worked on international projects such as the EU-funded IST PRIME, FIDIS and WiTness projects, the Norwegian PETweb and PETweb II projects as well as the CEN-ESSI workshops on signature standardization. He has published in various international proceedings, writing with international co-authors.

Norsk Regnesentral

Norsk Regnesentral (Norwegian Computing Center, NR) is a private, independent, non-profit foundation established in 1952. NR carries out contract research and development projects in the areas of information and communication technology and applied statistical modelling. The clients are a broad range of industrial, commercial and public service organizations in the national as well as the international market. Our scientific and technical capabilities are further developed in co-operation with The Research Council of Norway and key customers. The results of our projects may take the form of reports, software, prototypes, and short courses. A proof of the confidence and appreciation our clients have for us is given by the fact that most of our new contracts are signed with previous customers.

Norsk Regnesentral
Norwegian Computing Center
Postboks 114, Blindern
NO-0314 Oslo, Norway

Besøksadresse
Office address
Gaustadalléen 23
NO-0373 Oslo, Norway

Telefon · telephone
(+47) 22 85 25 00
Telefaks · telefax
(+47) 22 69 76 60

Internett · internet
www.nr.no
E-post · e-mail
nr@nr.no

Title	**Social Media, e-ID and Privacy**
Authors	**Lothar Fritsch**
Date	April 11, 2010
Year	2011
Publication number	DART/02/2011

Abstract

The e-Me project is a research project funded by the Research Council of Norway. It focuses on the improvement of identity management and accessibility in social networking and social communities. The latter have become a trend in internet applications in a fast-paced market that is at risk of leaving the weaker members of society behind. e-Me aims at the improvement especially of identity management and privacy functionalities from the e-inclusion perspective.

Keywords	Social networks, identity management, e-inclusion, privacy, universal design, usability
Target group	e-Me project participants, Provacy researchers, Social media users
Availability	Public
Project number	320457 (NR internal number)
Research field	Identity Management, Privacy enhancing technology, Information security
Number of pages	30

Contents

List of figures

List of tables

1 Social media and social networks

1.1 Definitions

Many parties have described the concept of social communities and social networks on information systems. I suggest to consider the following sections with some of the definitions.

1.1.1 Boyd / Ellisson definition

One of the most oft-quoted definitions of social network sites was developed by Boyd and Ellison, who write that these are "web- based services that allow individuals to (1) construct a public or semi- public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site" (Boyd and Ellison, 2007: 211)[1].

1.1.2 From the PrimeLife[2] EU project

"Social software: The term social software characterizes infrastructures, platforms and applications that enable users to communicate, collaborate and coordinate themselves via networks, to establish and maintain relationships and thus in some way map social aspects of real life to an online environment. Schmidt defines social software as web-based applications that support management of information, relationships and representation of one's self to (a part of) the public in hypertextual and social networks. Therefore, three primary functions of social software can be identified (…):

- Information Management: finding, evaluating and administration of information
- Self-Management: present aspects of yourself on the Internet
- Relationship Management: represent and maintain contacts to others via Internet"

1.1.3 Mynatt et al, 1997

This definition is from the "Computer-supported Collaborative Work" community, which focused on the interaction processes mediated with collaborative information systems. Many articles and books have been published in this community, where work and interaction processes based on data are the main subject of study. However, it should be difficult to ignore the intersection of today's social network computing and the CSCW community's work:

" We introduced network communities as embodying a particular design direction in supporting collaborative activity. In this section, we attempt to characterize network communities in more detail. We have chosen to use the term community rather than collaboration to point toward a more long-term and multilayered relationality. Community has been defined variously as being based on geographic area, social norms, or types of social interaction. Without contesting the particularities of these differences, we would like to point to the loose consensus around community as referring to a multidimensional, cohesive social

[1] http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

[2] PrimeLife FP7 ICT EU research project, http://www.primelife-project.eu

grouping that includes, in varying degrees: shared spatial relations, social conventions, a sense of membership and boundaries, and an ongoing rhythm of social interaction."

1.1.4 Stanoevska-Slabeva, 2002

A more sociologically oriented definition separates social networking into the social aspect, and the supporting platform:

"For the purposes of this paper, on-line communities are defined through their features as associations of participants who share a common language, world, values, and interests, obey a commonly defined organizational structure, and communicate and cooperate ubiquitously connected by electronic media and possibly represented by avatars. In accordance with this definition, online communities exist at the intersection of complex technical and social systems. “Neither technology nor sociality can supplant the need for the other, and the two are conceptually inseparable”. Therefore, on-line communities have two interrelated constitutional elements: the association of community participants, and the enabling digital platform."

1.2 Example platforms

In this section, a few of the many example platforms are listed into one of three categories identified in e-Me in the discussions on the kick-off meeting.

Social network sites divide into three main purposes. Each purpose is illustrated with one or more examples:

1. Stages for socializing with various private circles

   Facebook.com
   Nettby.no

2. Platforms that focus on common interests

   Aktivitetesvenner.no
   InclusivePlanet.com
   Møteplassen.no

3. Professional networking platforms

   LinkedIn.com
   XING.com

Many other categorizations exist. One of the most prominent classifications was done in the EU project PICOS in its deliverable D2.2: Categorization of Communities.

It must be noted that many social networks are reactive to the market, aiming at the inclusion of more members to their networks. New communication functions, group functionalities, or community purposes are added as soon as they show a potential for better user binding. Some communities change their purpose, or might get absorbed by others.

1.3 Taxonomization of social networks

The PICOS project has identified 10 different dimensions that can be used to analyze social networks. Figure 1 shows the PICOS dimensions:

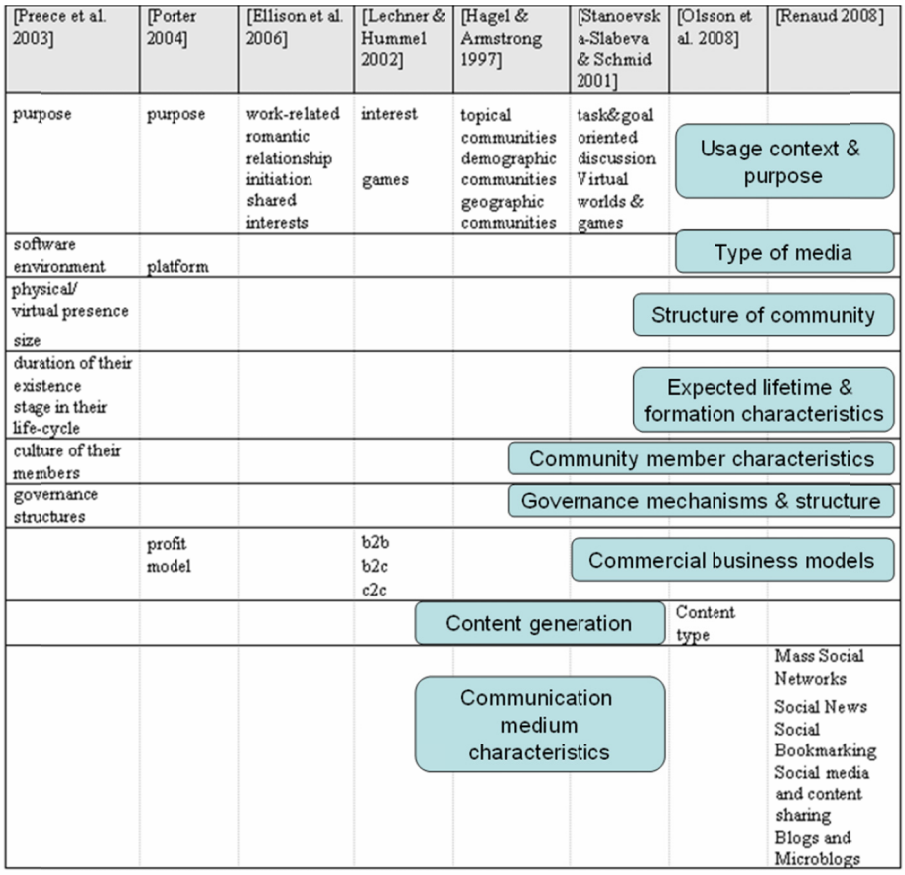


Figure 1: Overview and summarization of influencing approaches on the PICOS categorization model in (PICOS D2.2).

The dimensions in the diagram can be used to taxonomize social networks. For example, user geography, sociography and common interests are part of the "usage context & purpose" dimension. An application of Figure 1 to a - hypothetical - Norwegian angling community based on mobile phone applications with GPS and photo use would result in the taxononmization shown in Table 1.

Norwegian Mobile Angling Community	Main characteristics
Usage context & purpose	Spare-time angling community with the purpose of socializing and mutual help among Norwegian angling enthusiasts
Type of media	Mobile phone, mobile network, central servers, optional web interface. Photo, video, text, and voice objects with tagging and geolocation. Supplementary fish species database and legal

	documents.
Structure of community	Virtual presence. 12.000 accounts, 3500 of them active every week.
Expected lifetime & formation characteristics	Open-ended lifetime. Came over "early-adopters"-phase. Well established. Due to language, growth prospectives are limited.
Community member characteristics	Independent, self-taught angling experts with affinity to hi-tech gadgets and open social discourse on internet platforms. Members of both "generation mobile" and "nature lovers".
Governance mechanisms and structure	Operator governs against problematic abuse and system errors. Self-moderation through trusted members for simple moderation.
Commercial business model	Driven by an angling shop with mail-order business. Targeted advertising and product placement is in use. No "premium accounts".
Content generation	Some limited editorial & blogging activity by the system owners. Most content created by members with blogs, photo albums, and link lists.
Communication medium characteristics	Blogs, personal messages, SMS, photo, video and sound collections, discussion forums, message distribution lists, link lists.

Table 1: Example of application of PICOS community characterization to a hypothetical angling enthusiast community using mobile devices.

Using this methodology, social networks can be analyzed for their purpose, their members, their business model, and their basic functionality.

2 Electronic identities and identity management

This section of the report will provide a high-level overview over the terms and concepts of electronic identities and identity management.

2.1 What are e-ID's and Identity Management?

Electronic identities, or e-ID's, are terms often used in discourse without clear definition. Generally, they can be characterized with the following statements:

- e-ID's are a portion of data used by algorithms in software or hardware that are meant to convince a computer that a particular person is using it - or that a known other computer system is sending messages.
- e-ID's can be attached to an "official" identity, for example a passport, an ID card, or social security numbers.
- Many e-ID's are based on "soft" identity, such as e-mail addresses, user pseudonyms, self-claimed names, names of computer game characters, or mobile phone numbers.
- e-ID's are used for many different purposes.
- Some e-ID's are attached to a communication channel, e.g. e-mail-addresses, and Skype names.
- e-ID's and their association with a person's identity is either under the control of the user (e.g. self-chosen password), or forced upon the user by the controller of an identity management system (governmental ID card, passport).
- e-ID's have a life cycle - they are created, maintained, deleted and archived.

From the above, we can conclude that e-ID's are controlled by various parties, are to varying degrees related to a real person, may be associated with a communication channel, and are generally used to manage a computer's relationship to a person or to another computer, or a program.

A decomposition of identity management and e-ID in use produced in the PETweb II[3] project is shown in Figure 2.

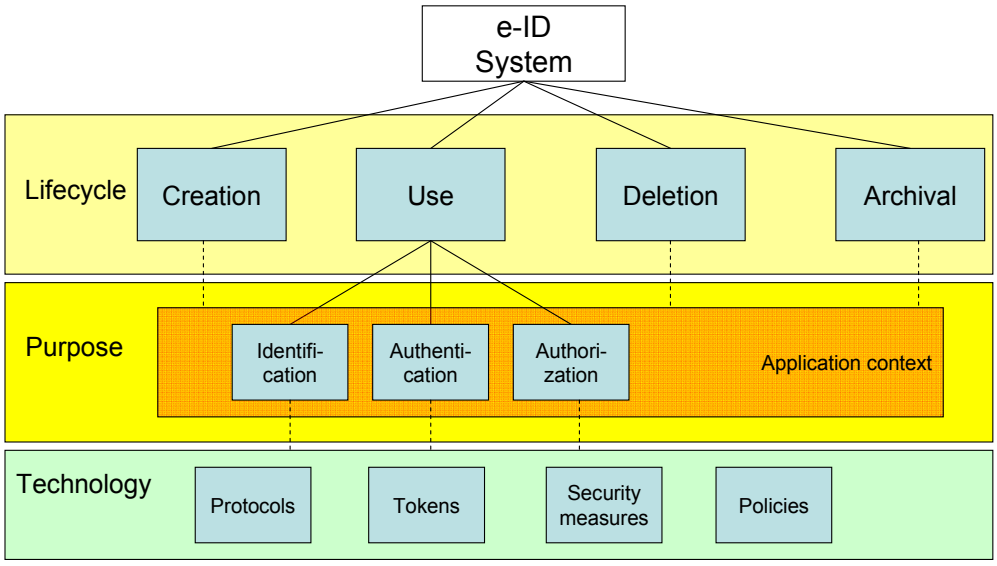


Figure 2: Decomposition of Identity Management Systems and e-ID [PETweb II project, Lothar Fritsch, internal meeting minutes 9/2010]

[3] PETweb II - Privacy respecting Identity Management for e-Norge, research project in the Norwegian VERDIKT program, 2009-2013. See http://petweb2.projects.nislab.no for more information.

2.2 The FIDIS typology for Identity Management

A closer analysis of identity management systems (IDMS) with respect to e-inclusion reveals a number of important research and development issues and challenges. The discussion of these will be structured by a classification of IMS systems [Fidis 2005a] where identity management systems are grouped into:

- Type 1: IDMS for account management, implementing authentication, au-thorization, and accounting;
- Type 2: IDMS for profiling users, e.g. detailed log file analysis or data warehouses which support personalized services or the analysis of cus-tomer behaviour;
- Type 3: IDMS for user-controlled context-dependent role and pseudonym management.

The following sections will introduce the three types of IDMS. The presentation is based on [Fritsch/Fuglerud/Solheim IDIS journal 2010].

2.2.1 Type-1 IDMS

A taxonomy of such systems is shown in Figure 2. In order to be able to use a large number of public and private services the user must be authenticated. A very basic requirement for e-inclusion is that the authentication methods can be used by as broad a range of users as possible. Common authentication methods include passwords and PINs, tokens, smart cards, and use of 3rd-party channels such as one-time codes from tokens or code generators. These methods can be difficult or impossible to use by different user groups.

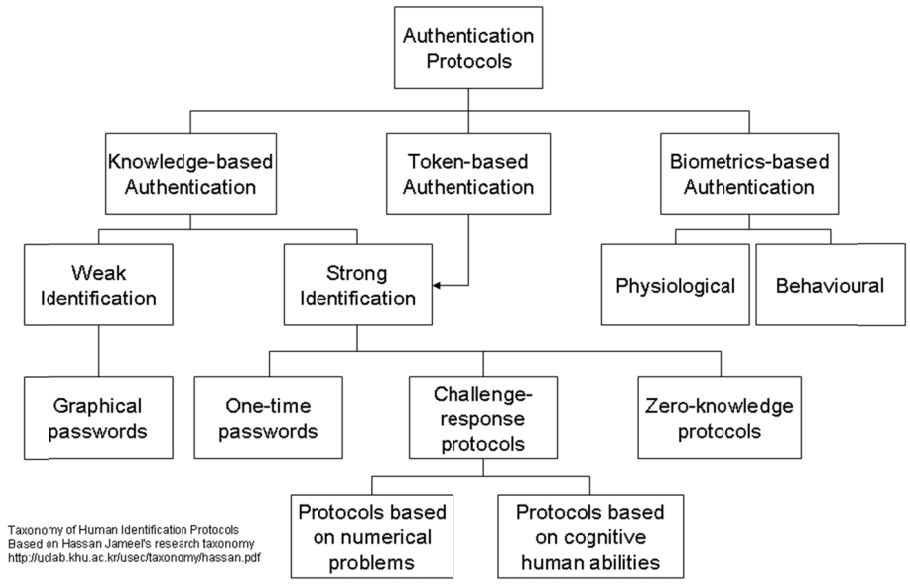


Figure 3: Jameel's taxonomy of identification methods [(ameel 2007).

According to the taxonomy in (Jameel et al. 2007), authentication methods can be divided in three categories:

- Knowledge-based authentication: Systems based on the knowledge of a secret, e.g. passwords or PIN/TAN.

- Token- or possession-based authentication: Systems based on the possession of a token (a physical or electronic unique authentication resource). This could for example be a cryptographic key or certificate, a smart card, a number sequence generator.
- Biometric authentication: The use of unique personal, physical traits as input for authentication.

2.2.2 Type-2 IDMS

This section relates to IDMS type 2, which are identity management systems for profiling of user data by an organization, e.g. detailed log files or data warehouses, which are used for personalized services, customer management, or the analysis of customers.

Profiling has been defined as "The process of constructing profiles (correlated data), that identify and represent either a person or a group/category/cluster" (Fidis 2005b). There are in principle two types of profiling: group profiling (e.g. use of data mining techniques to establish general, abstract profiles of a group), and personalized profiling which is focused on in this article.

On the general level, it is likely that that profiling technologies will have a pro-found impact on access to and participation in the Information Society, as profiles 'could possibly be used against individuals without their knowledge, thus shaping their access to facilities, goods and services, also potentially restricting their movement and invading personal space. In fact, this would regulate their access to, and participation in, the European Information Society' (Levi and Wall 2004).

The use of profiling techniques however poses a challenge to existing anonymization techniques, which mostly aim at avoiding profiles, and might render profiling difficult (Fritsch 2008).

2.2.3 Type-3 IDMS

This section relates to IDMS type 3, which are user-controlled context-dependent role and pseudonym management systems. In (Fidis 2005a) such IDMS are characterized as follows:

"The data managed are mainly personal data. Privacy protection therefore is a driving force for the development of IMS[4] of this type and a relevant unique selling proposition (USP). To implement certain functions, such as use of trusted pseudonyms or authentication (e.g. via credentials), in some cases the implementation of centralized third party services is necessary. In addition, the communication partner of the user, who is contacted via the managed identity, in many cases is an organization. "

In other words, type 3 IDMS enable the user to choose how identifiable he or she wants to be for a service or for other users. Such identity management has some important implications:

- users should be enabled to participate anonymously or pseudonymously
- users decide which of their personal attributes shall be revealed in which context
- users might like to keep track about what has been revealed

[4] IMS in FIDIS is equivalent to IDMS in this text.

- to engage in e-commerce, forms of payment that support IDM with type 3 IDMS can be necessary, e.g. anonymous payment mechanisms.

Such IDMS of type 3 are often called "user centric identity management".

2.3 What are e-ID's used for?

There are three generic uses for e-ID's. These uses are fundamentally different, and may be decisive for the security and privacy properties for the applications relying on the e-ID's.

2.3.1 Identification

e-ID's are used to identify a person or a computer. Here, the unambiguous identification is the purpose of an ID transaction. "Who is this?" is the question answered by identification.

What exactly is made known to a computer through an identification transaction is dependent on the application context and on the used e-ID system. Identification can use a person's real identity, or establish a pseudonym.

2.3.2 Authentication

In authentication, the user or computer does not only claim its identity, but must in addition prove the identity claim to the computer system he wishes to authenticate against. Many authentication transactions follow a protocol similar to this sequence:

1. User: "Hello, I'm here, I am User123"
2. Computer: "Hello User123, prove it by sending a credential"
3. User: (uses password, smartcard, one-time-code or secret key with cryptographic algorithm): "Here is my authentication code: 034044Xy".

Authentication is used to ensure that a person needs to know more than just a name before he or she gets access to a computer system. The extra knowledge is called an **authentication factor**. In most cases, users receive or set the authentication factor from the system when they register to use it (e.g. when they receive their password).

Some authentication systems use more than one factor. So-called "two-factor-authentication" requires two authentication factors, e.g. two different passwords, or a "secret" customer number and a password. Authentication systems might use an authentication factor that is a personal secret, and another factor that is a physical object that cannot be stolen through the Internet (a smart card, or a code generator). This is aimed at preventing ID theft, or the passing on of authentication secrets.

2.3.3 Authorization

Authorization is the delegation of permissions to a computer system (or even an organization owning the computer) to perform some action on behalf of the user. An authorization transaction collects explicit consent for the execution of certain actions from the user. A typical example is the transfer of money through the Internet-based electronic banking software. After logging in (identification & authorization), the banking software requires an additional authorization for each money transfer.

An authorization can be understood as an electronic equivalent of written signatures. They serve both involved parties' purposes. The user explicitly delegates privileges for a certain

action, while the receiving computer or organization can document explicit authorization to perform this action.

Authorization technology meets extra challenges when compared to authentication. Depending on the application context, authorizations must be identifiably related to a person. They might be part of archival laws, where they need to be auditable in an archive over periods of many years. In addition, based on the value of the authorized actions (order a cup of coffee or sell stock for millions); the authorization system should match strong information security requirements.

Various other qualities of authorization systems might be relevant. Multi-party authorization (where several persons "sign" an action before it can be executed), legally binding signatures (such as defined e.g. by the European Directive on Electronic Signatures), and authorization systems that actually generate user-archival receipts for its actions have been researched and built.

2.4 e-ID and information security

e-ID's are essentially digital data crated by identity management algorithms. To ensure that e-ID's have sufficient quality for the context they are used in, they should have certain security properties. Such properties make them robust

2.4.1 Identity theft

Identity theft is a term used for incidents where e-ID's are abused by other people. The term references stolen passports that are being used for other purposes. "ID theft" however is a misleading term - it is rarely a person's "identity" that is being stolen, but a person's "electronic identity card" - some data used as e-ID. Some researchers prefer "identifier theft", which is more appropriately expressing that some e-ID is stolen, duplicated, or used in unauthorized ways by someone else.

Identity theft is often caused by either weak information security in identity management systems, users breaching with security policies, or applications that use e-ID's inappropriately. The damage caused by identity theft normally occurs through the abuse of identifiers, for example to:

- Steal money or realize other profits
- Commit fraud on another person's behalf
- Steal secrets
- Vandalize or sabotage information systems for fun, revenge or profit

The majority of available statistics for damages are reports from banks, credit card businesses, and mail-order businesses that were exploited with stolen e-ID's and/or bankcards. Some reports mention vandalizing or mobbing by using other people's accounts on social media. National security agencies are increasingly worried about the potential of theft of secrets through foreign intelligence services using stolen or fabricated e-ID.

Generally, the risk for identity theft is influenced by a number of properties of the identity management technology used for e-ID's. Such properties are shown below in Table 2:

Risk contributing factors	*Parameters*
Secrecy of Authentication tools	Publicly known, inferrable, secret
Mobility of Authentication Tool	Copyable, remotely usable, concurrently usable, immobile
Claim type	single, multiple
Risks to IDM	loss, misuse, disclosure, disruption, theft, replacement value
Provisioning	creation, edit, deletion
Frequency and duration	Uses per year, total life time of identifier/transaction
Use/Purpose	Authentication, Authorization, Identification
Personal attributes	Forced, chosen, role, pseudonymity
Obligations & policies	Relationship to ID, Relationship to PI

Table 2: Factors contributing to security and privacy risks in e-ID's and their application (Paintsil and Fritsch 2010).

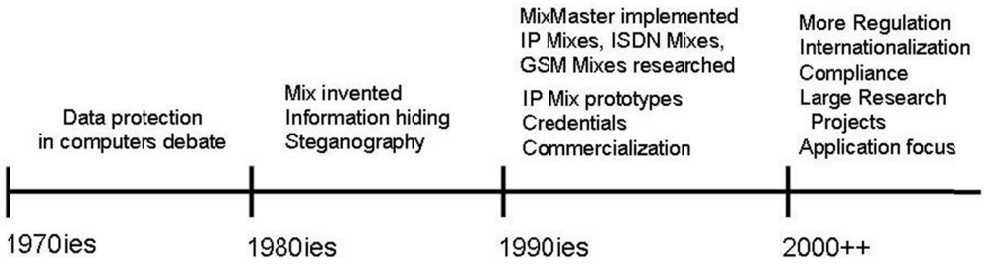
3 Privacy enhancing technology (PET)

PETs have been a topic in research since the 1980ies. This section is providing an overview of terms, concepts and research activities in the field of PET. It is based on a state-of-the-art report from the Petweb[5] project (Petweb I D2.1 2007) and an overview in (Fritsch, Abie 2008).

For the sake briefness, the reader is asked to refer to chapter 2 in (Petweb I D2.1 2007) for a survey on available, implemented PET tools available to users and developers.

3.1 History of PET

PET as a research topic has been opened by David Chaum in 1981. In his MIX paper (Chaum 1981), he describes a method for anonymous and unobservable delivery of electronic messages called "Mix". Chaum uses security protocols and subsequent layers of encryption to provide privacy protection by "mixing" several people's e-mail traffic in encrypted form. The concept later was implemented in the MixMaster e-mail anonymization system (Möller, Cottrell, Palfrader and Sassaman 2004), which is the first practically available PET system.



The appearance of technological measures for privacy protection coincides with strengthening legal regulation of the use of personal data on information systems. Starting in the 1970ies, regulatory regimes were put on computers and networks. Starting with government data processing, along the lines of computerization of communication and workflows, explicit rules like the European Data Protection Directive (European Comission 2002) have been put in place.

With the adoption of Internet and mobile telephony in society in the past decade, the privacy challenges of information technology came to everyday life. Hence in the 1990ies, research efforts on PET increased, with Chaum's concept being adapted to internet data traffic (Pfitzmann and Waidner 1986), (Pfitzmann, Priftzmann and Waidner 1991), (Goldschlag, Reed and Syverson 1996) and call routing in ISDN (Jerichow, Müller, Pfitzmann and Waidner 1998) or mobile telephony (Federrath, Jerichow, Kesdogan, Pfitzmann and Spaniol 1997). Along with several publicly funded research projects (Lacoste, Pfitzmann, Steiner and Waidner 2000), (PRIME 2003), (FIDIS 2003), several companies turned privacy protection into a business model (Anonymizer.com, Zeroknowledgesystems.com, XeroBank, Anti-Spyware, Virus tools). Researchers investigated cryptography and information hiding technology to produce privacy-supporting protocols such as anonymous credentials (Camenisch and van Herreweghen 2002). A milestone in this development is the appearance of a "Handbook on Privacy-Enhancing Technologies" (Blarkom, Borking and Olk 2003) written by representatives of the regulatory authorities, not by Pet researchers or technicians.

With the globalization of the economy and the IT infrastructure supporting it, in the years staring the 3rd millennium privacy management has turned into a matter of corporate governance and compliance, with legislation targeting this issue (e.g. (European Comission 2002)). Standardization bodies and interest groups such as ISO, W3C and IETF (Müller 2004) initiate privacy technology standardization work. Global players such as IBM and HP target corporations with their privacy compliance services. In this context, recent efforts on using Trusted Computing (TCG 2007) to implement privacy-compliant data handling show the path to the future of information privacy as a matter of compliance.

3.2 Taxonomy of PET

3.2.1 Privacy

Privacy enhancing technology (PET) is about the protection of privacy in information systems. The term privacy is used in many contexts, and with many possible interpretations. In the context of PET, privacy is either viewed from a legal view – by the data protection community. Alternatively, it is viewed as a technical challenge to information security, which relates to the cryptography and computer security community. The specific challenges in information privacy are described in D. Solove's "A Taxonomy of Privacy" (Solove 2006), which has won the 2006 PET award. Here, the four basic challenges of information privacy are found to be:

- Information Collection: The collection of personal information by some party.
- Information Processing: The processing of personal information by some party.
- Information Dissemination: The distribution of personal information by some party.
- Invasion of privacy
- Intrusion of private spaces
- Influencing decisions

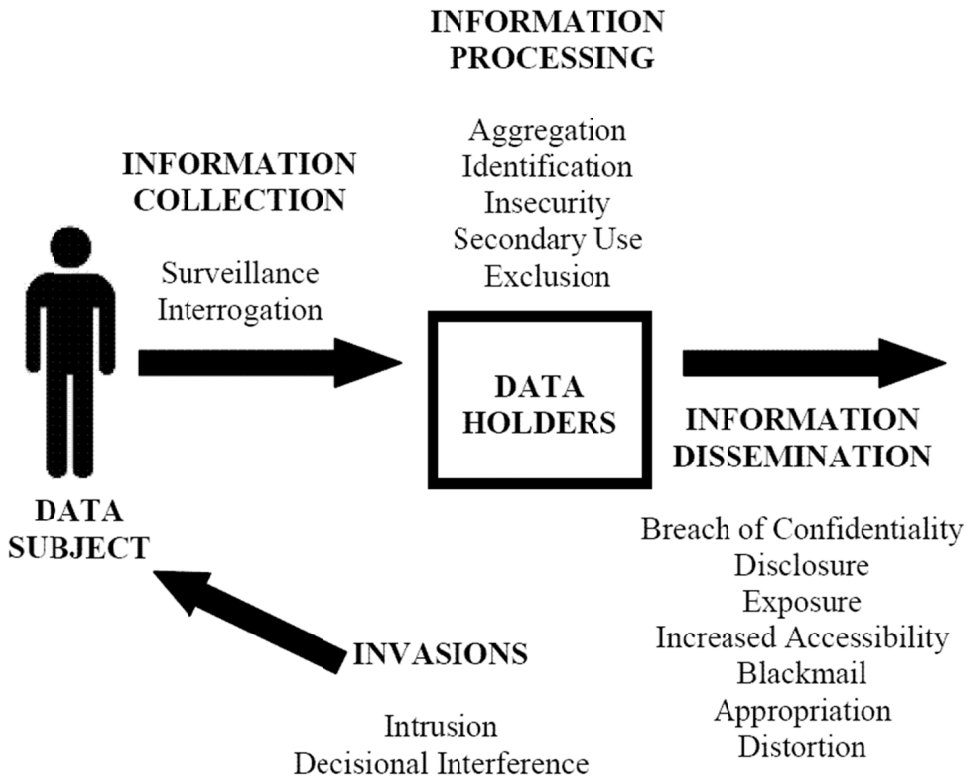


Figure 5: Taxonomy of privacy from (Solove 2006).

Solove describes the four areas in further detail, whereby he identifies particular actions that produce threats to privacy (see Figure 5). A classification of privacy risks and the cost induced by these risks has not been done in convincing ways. Privacy risks are not well defined in the literature. Too low quality of a particular protection technology might destroy particular applications, as Friedmann shows in (Friedmann and Resnik 1999). In (Gellman 2002), the business and consumer side of privacy risks and costs is examined. The author classifies risks and provides an example with monetary figures on how much cost is imposed on the average U.S. family through privacy breaches. The suggested risks are listed in Table 3. Noteworthy is the distinction in risks not only to the consumer, but also to businesses. Odlyzko agrees that a lack of privacy in consumer commerce settings leads to financial losses due to price discrimination (Odlyzko 2003).

Businesses	Consumers
· Sales Losses Due to Lack of Privacy	· Higher Prices
· One Retailer's Loss Is Another Retailer's Opportunity	· Junk Mail, Telemarketing
· Lost International Opportunities	· Identity Theft
· Increased Legal Costs, Investor Losses	· Internet Effects
	· The Dossier Society

Table 3: Privacy risks from (Gellman 2002).

3.2.2 Terms and Definitions

Terminology in the PET community is sometimes confusing. This section defines the most important terms and concepts that are used in this report. They are mostly taken from or inspired by Hansen & Pfitzmann's long-term terminology effort (Pfitzmann and Hansen 2003), which is also a good source for the translation of the terms into many other languages beyond English.

Term	Definition
Anonymity	Anonymity means that a subject is not identifiable within a set of subjects.
Identity	A person's identity is either the person's self-perception, or the person's external categorization using attributes that are observable. In the sense of PET, the identity is a set of externally observable attributes and properties that – when taken all together – allow for the identification of a subject among others. The term "partial identity" is used to point out the fact that a subject in a certain role might use – or be identified by – a subset of his personal, externally visible attributes.
Identity management	Identity management is the process of administration of various partial identities of a subject. Privacy-preserving identity management systems keep distinct partial identities of a subject separate from each other, and thus unlinkable.
Privacy	Privacy in the sense of PET is the autonomy of a subject over his personal information. Privacy in information systems hence is the control over personal information that is being released to other parties. Additionally, transparency about what happens with the information at the other party and ways to limit actions on the information is considered a part of information privacy.
Pseudonym	A pseudonym is an alias name or other form of identifier that removes a subject's real name, but serves as a means of relating to that subject. Pseudonymity is the state of using a pseudonym as an identifier. Pseudonyms can model roles, transactions, persons, relationships with different degrees of anonymity.
Unlinkability	Unlinkability of a pseudonym or a subject's actions refers to a situation where an n actions or appearance of a subject on a system cannot be identified to belong to any other action of this subject.
Unobservability	Unobservability means that · a data object / transfer is not observable to parties uninvolved in the transaction;

- the involvement of the subjects in the aforementioned data transfer is not observable to any other parties.

3.2.3 Classification of PET systems

In recent research in the FIDIS project (FIDIS 2003), a functional distinction of privacy and identity protection in transparency tools and opacity tools was introduced (FIDIS 2007).

Transparency tools are intended to create insight into data processing. Their effect is a better understanding of procedures, practices, and consequences of personal data processing at a data processor. Because they enhance understanding and visibility, they are called transparency tools. Opacity tools are intended to hide a user's identity or his connection to personal data that occurs at a data processor. As they hide identities, reduce visibility, or camouflage connections, they are called opacity tools.

	Transparency tool	Opacity tools
Definition	Tools that show clearly to a person what personal data are being processed, how it is processed, and by whom it is processed.	Tools that hide a person's identity or his relationship to data as it is processed by someone else.
Non-technical example	Legal rights to be informed about data processing; Privacy audits.	Pseudonymous access to on-line services; Election secrecy.
Technical example	Database audit interfaces; Audit Agents, **Log files.**	MixMaster anonymous e-mail; TOR anonymizing web surfing; Pseudonyms.

Table 4: Transparency and opacity tools.

This classification originally conceptualized tools as legal framework and technical practice. But its adaption to a technical classification of PET systems only is useful. The distinction is introduced in Table 4.

The distinction above can be further elaborated by the analysis of PET functionality. A study for the Danish Government (Meta Group 2005) divides privacy technologies in the two groups of "privacy protection" and "privacy management", where the description of the technologies grouped by the two concepts goes along the transparency-opacity distinction. In Table 5, "privacy protection" lists opacity tools, while "privacy management" aims at the transparency tools.

Category	Subcategory	Description
Privacy Protection	Pseudonymizer Tools	Enabling e-business transactions without requiring private information.
	Anonymizer Products and Services	Providing browsing and email capability without revealing the user's address and identity.
	Encryption Tools	Protecting email, documents and transactions from being read by other parties.
	Filters and Blockers	Preventing unwanted email and web content from reaching the user.
	Track and evidence erasers	Removing electronic traces of the user's activity.
Privacy Management	Informational tools	Creating and checking Privacy Policies.
	Administrative Tools	Managing user identity and permissions.

Table 5: Privacy protection classification from (Meta Group 2005).

However, the PET community will not agree with certain aspects in Table 5, as user-centric identity management aims at a user's informational self-determination, and thus clearly is an opacity tool (Pfitzmann and Hansen 2007). Nonetheless, the Danish study proceeds with the analysis of the core protection mechanisms provided by the classified PET techniques, with a distinction of the functions in unobservability, unlinkability and anonymity. In addition, the target of the mechanism is identified to be of informative, or curative nature. This once again reflects the transparency-opacity nature of PETs.

Main Category	Subclasses	Typical Features	I	1	2	3	S
Privacy Protection	Pseudonymizer Tools	CRM personalization			X		
		Application Data Management			X		
	Anonymizer Products and Services	Browsing pseudonyms				X	
		Virtual Email addresses				X	
		Trusted third Parties			X	X	
		Surrogate Keys			X		

Main Category	Subclasses	Typical Features	I	1	2	3	S
	Encryption Tools	Encrypting email		X			
		Encrypting transactions		X			
		Encrypting documents		X			
	Filters and Blockers	Filtering email spam					S
		Filtering web content					S
		Blocking pop-up windows					S
	Track and evidence Erasers	Spyware detection and removal		X	X	X	
		Browser cleaning tools		X	X		
		Activity traces eraser		X	X		
		Harddisk data eraser		X	X	X	
Privacy Management	Informational tools	Privacy Policy generators	I				
		Privacy Policy readers/validators	I				
		Privacy Compliance scanning	I				
	Administrative Tools	Identity management				X	
		Biometrics				X	
		Smart cards		X		X	
		Permission management		X		X	
		Monitoring and Audit tools		X			S
		Forensics tools					S

Table 6: PET mechanisms classified in (Meta Group 2005).
1. Unobservability – making private information invisible or unavailable to others
2. Unlinkability – preventing others from linking different pieces of observed information together
3. Anonymity – preventing others from connecting observed information with a specific person
I. Information tools
S. Secondary protection targets (countermeasures)

A closer look at the intention of, and functions provided by existing PET reveals an almost even distribution of unobservability, unlinkability and anonymity support (which suggests that none of these properties can be reached alone). Some of the tools surveyed target specific risks posed

by on-line systems, such as spyware or cookies. Few of the tools are classified as "information tools" – or transparency tools. Table 6 lists the privacy-enhancing properties of the surveyed systems from (Meta Group 2005).

Roger Clarke has suggested categories for PET systems in (Clarke 2007):

- Pseudo-PETs: Privacy seals, P3P
- Counter-Technology: Counters one specific privacy threat, e.g. SSL encryption or spyware removal.
- Savage PETs: Will provide untraceable anonymity
- Gentle PETs: Balanced pseudonymity tools with accountability, identity management

However, no sharp definition of the classes and no classification of real systems are given.

3.3 PET in information ecosystems

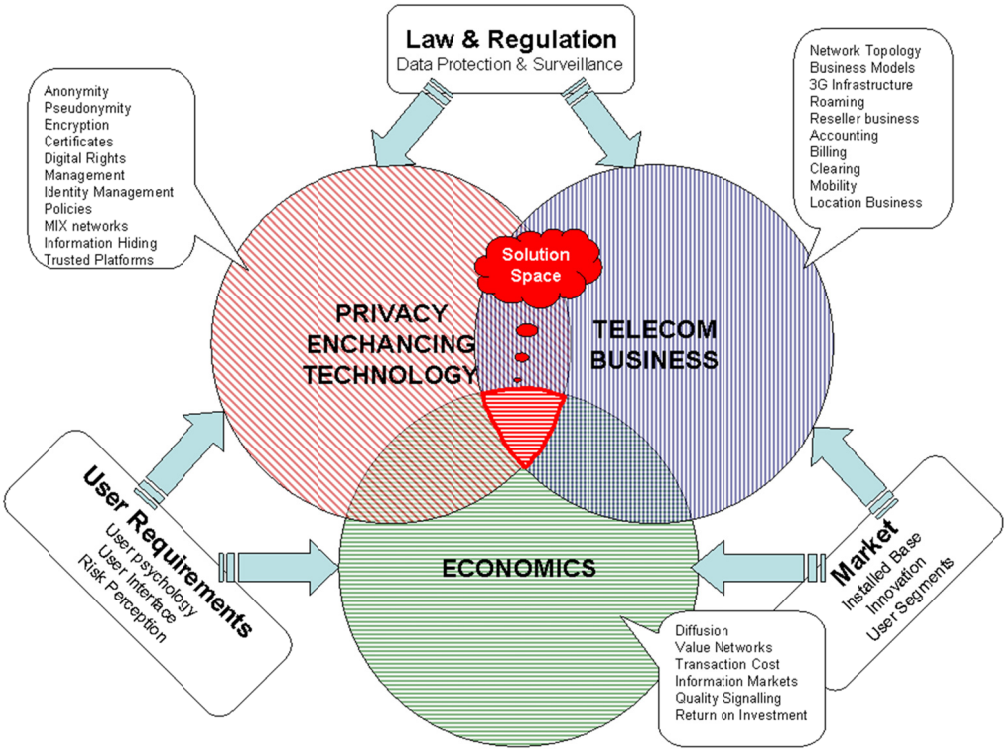


Figure 6: PETs in their information ecosystem (based on (Fritsch 2007)).

Privacy in information systems is not restricted to technological matters. Information systems have a large context that is defined by all stakeholder designing, using, regulating or being influenced by the information system. A deployment of PETs and their meaning to a certain group of stakeholders, a broad analysis of the system's environment and purpose is helpful. This environment is called an "information ecosystem" in this study. At first, we will present the environment PETs are deployed into. Next, we examine the systematic approaches on how technological measures that are in favor of privacy are being handled in terms of technical standardization. Finally, certification schemes and audits are examined.

3.3.1 Context of Privacy-enhancing technology

PETs are connected to many disciplines. PETs are deployed into a larger context of information systems, which in turn are governed by societies' requirements and business requirements. Few complete frameworks for PET-related contexts or approaches have been published, namely KPMG's model (KPMG Canada 2003), a security framework (Zuccato 2005), and a design process (Fritsch 2007). Work on risk modeling (Hong, Ng, Lederer and Landay 2004) also provides insight on requirements engineering. In particular, the interdisciplinary nature calls for a model that provides a frame for knowledge in important disciplines as well as a way of integration of application-specific knowledge. In most on-line scenarios, the application specific communities can be identified as telecommunications, PET and Economics (see **Error! Reference source not found.**). These communities are influenced by law and regulation, by the situation on the market of needs and related products, as well as by the user requirements from various disciplines respectively. They all influence the need for, and the deployment of PETs, which in Figure 6 is illustrated by the "solution space" - the union of all communities in the diagram's center. Any PET development and deployment must be made in awareness of such a context.

3.3.2 Technical standards

Very few technical standards for privacy protection exist. Those that have been specified usually lack relevance in practice. Many industrial associations have published their own hands-on standards that are intended to comply with new regulation, e.g. with the treatment of location data in mobile phone networks (e.g. the OMA/LIF privacy guidelines (Oinonen 2002)). On the level of IETF, some preparatory work has been done to standardize a large geo-spatial privacy framework called "Geopriv" (Müller 2004). The World Wide Web consortium keeps publishing specifications for privacy preferences selection and other privacy-related description languages. Their focus is web-centric, their relevance in practical application uncertain.

On the international level, there are some ISO activities, but so far the application of ISO 15408 'Common Criteria' (ISO 1999) for privacy evaluation is only under research in PRIME (Kohlweiss, Fritsch, Radmacher, Hansen and Krasemann 2004) and in a special study period at ISO/IEC/JTC1/SC27/WG3 (Brand 2005)). Current developments there are described in (Bramhall, Hansen, Rannenberg and Roessler 2007), however it will take some time until the ISO will actually describe a technical standard. What might come from that direction however could be an extension for the application of the Common Criteria. Protection profiles for privacy-related security properties could be expressed as illustrated for the case of MIX remailers in (Rannenberg and Iachello 2000).

3.3.3 Audit and Guidelines

Many countries have proposed frameworks for privacy audits. Complementing commercial privacy seals aim at confirming privacy properties of e-commerce web site. The major difference in these schemes is their goal. The governmental schemes target at the implementation of the legal privacy principles (consent, purpose of data processing, transparency). The commercial seals are used for marketing purposes, and usually intend trust building with the businesses' customers.

Many of the schemes provide checklists and guidance for audits that follows closely the legal frameworks. Often, the methodologies used are intended to detect the state of a system, but not to suggest improvements of the system using PET.

A number of audit & seals schemes can be found in Table 7.

Name	Issuer	Description	Reference
Privacy Audit Manual	The Australian Privacy Commissioner	This manual outlines the policies adopted by the Privacy Commissioner for the performance of Privacy Audits, describes the Privacy Audit process and the concepts underlying it, and provides guidance as to the audit procedures that should be applied.	http://www.privacy.gov.au/publications/ippam1a.pdf
Privacy Audit Framework under the new Dutch Data Protection Act (WBP)	Co-operation Group Audit Strategy	The Privacy Audit Framework was set up to carry out Privacy Audits in organizations where personal data are processed. Privacy Audits must be carried out in careful consideration: not every organization is initially ready to undergo a Privacy Audit. A thorough analysis to assess whether a Privacy Audit has added value for an organization must take place in advance. This is to prevent disappointing the client with regard to the Privacy Audit's results. If the aforementioned analysis shows that a Privacy Audit has insufficient added value for the organization at that time, then the organization must take proper measures first. The WBP Self-assessment can be used for this purpose if so desired. The auditor can help an organization by giving advice during the improvement process.	http://www.dutchdpa.nl/downloads_audit/PrivacyAuditFramework.pdf
Datenschutz-Gütesiegel (Privacy Seal)	Independent Centre for Privacy Protection (ICPP; Unabhängiges Landeszentrum für Datenschutz)	The aim of the project is to persuasively strengthen the confidence of consumers, particularly in the Internet. This Privacy Seal certifies that the compatibility of the product with the regulations of privacy and of security was assessed in a formal process. This process is enacted in the State Data Protection Act of Schleswig-Holstein.	https://www.datenschutzzentrum.de/guetesiegel/eria/information-sheet_icpp_privacy_seal.pdf
TrustE and BBBOnline commercial seals	TrustE, BetterBusinessBureauOnline	Both companies offer privacy seals for e-commerce web sites. Truste has the highest market share among the seals, listing 1,374 Web sites to BBBOnLine's 701. Truste has	http://www.truste.org/ http://www.bbbonline.org/

	nearly a 2-to-1 edge over BBBOnLine on the top 50 Web sites, and a 3-to-1 edge among Safe Harbor members.

Table 7: Privacy Audit and Privacy Seals.

Concerning the commercial privacy seals, some scientific results in favor of the acceptance of privacy seals exist. In (Cranor, Reagle and Ackermann 1999), the authors state that a combination of a privacy seal and a privacy policy on a web page has a similar trust building effect as a privacy audit.

3.4 Current research in PET

Current research in the area of PET focuses on several topics:

- The integration of PET into application frameworks;
- The interplay of PET and identity management systems in large, meshed-up application worlds;
- The improvement of security in the handling of personal data;
- The increasing transparency of use of personal information.

The integration of PETs into applications is researched in the PRIME project (PRIME 2003). Here, an interdisciplinary framework for the application of PET components to IT systems is developed and explored in prototypical implementations. PRIME has produced trial prototypes in three application areas. Upcoming projects are intended to research privacy and PET usage on collaboration platforms and within Web 2.0 communities. Some research focuses on the application of newer cryptographic protocols for the purpose of privacy protection, for example for hiding location information in geo-spatial, mobile applications (Kohlweiss, Gedrojc, Fritsch and Preneel 2007).

On the identity management frontier, research came up with anonymous credentials and the IDEMIX system (Camenisch and van Herreweghen 2002) for secure, pseudonymous attestation. This approach enables unlinkability of identity and other credentials.

Concerning transparency, a recent development called “sticky policies” aims at establishing trustworthy computing environments with respect to privacy. By using a Trusted Computing platform in combination with a policy-based data processor, this research seeks to build computers that cannot process personal data in any other way than expressed in a policy attached to it – hence the name “sticky policy” (Casassa Mont, Pearson and Bramhall 2003).

Some research on transparency focused on early notification of people upon their private information leaking out to the internet. With a specialized “privacy search engine”, an approach in (Deng, Fritsch and Kursawe 2006) shows how to keep track of potentially compromising digital photos somebody else has made.

The EU FP7 project PICOS[6] aims at the development of technology for privacy-friendly mobile social networks, while the FP7 PrimeLife[7] project conceptualizes "life-long privacy and identity management", while trying out their test community ""

4 References

Blarkom, G. W.; Borking, John and Olk, J.G. (2003) Handbook of Privacy and Privacy-Enhancing Technologies, College bescherming persoonsgegevens, The Hague,.

Bramhall, Pete; Hansen, Marit; Rannenberg, Kai and Roessler, Thomas (2007) User-centric identity management, : New trends in standardization and regulation." IEEE Security & Privacy (5:4), pp. 64 - 67.

Brand, Sheila (2005) ISO/IEC/JTC1/SC27/WG3 COMMITTEE MEETING, http://www.incits.org/tc_home/CS1/2005docs/cs1050163.htm, accessed 16.Nov. 2006.

Camenisch, Jan and van Herreweghen, Els (2002) Design and Implementation of the Idemix Anonymous Credential System: Research Report RZ 3419, IBM Research Division,IBM Zürich Research Lab, Zürich.

Casassa Mont, Marco; Pearson, Siani and Bramhall, Pete. (2003) Towards Accountable Management of Identity and Privacy: Sticky Policies and Enforceable Tracing Services,Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03), IEEE Computer Society, pp. 377.

Chaum, David (1981) Untraceable electronic mail, return addresses, and digital pseudonyms, Communications of the ACM (4:2), pp. 84-88.

Clarke, Roger. (2007) Business Cases for Privacy-Enhancing Technologies, in: R. Subramanian (Eds.): To appear in: Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions, 12-Jun-2007, Hershey, USA, IDEA Group Publishing,.

Cranor, Lorrie Faith; Reagle, Joseph and Ackermann, Marc S. (1999) Beyond Concern: Understanding Net User's Attitudes About On-line Privacy: AT&T Labs-Research Technical Report TR 99.4.3,.

Deng, Mina; Fritsch, Lothar and Kursawe, Klaus. (2006) Personal Rights Management,: Taming camera-phones for individual privacy management, in: G. Danezis and P. Golle (Eds.): Privacy Enhancing Technologies - Proceedings of the 6th workshop on privacy-enhancing technologies PET2006, 29.Jun.2006, Berlin, Springer,.

European Comission (2002) Directive 2002/58/EC of the European Parliament and of the council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

[6] See http://www.picos-project.eu

[7] See http://www.primelife-project.eu

FIDIS (2003) Future of Identity in the Information Society,: The IST FIDIS Network of Excellence, www.fidis.net, accessed 6.11.2006.

Fidis. (2005a). FIDIS Deliverable D3.1: Structured Overview on Prototypes and Concepts of Identity Management Systems. 15. September 2005. European Union IST FIDIS Project.

Fidis. (2005b). FIDIS Deliverable D7.2: Descriptive analysis and inventory of profiling practices. European Union IST FIDIS Project, 2005.

FIDIS(2007) FIDIS Deliverable D7.5: Profiling the European Citizen,:Cross-Disciplinary Perspectives, European Union IST FIDIS Project,.

Federrath, Hannes; Jerichow, Anja; Kesdogan, Dogan; Pfitzmann, Andreas and Spaniol, Otto. (1997) Mobilkommunikation ohne Bewegungsprofile, in: A. P. G. Müller (Eds.): Mehrseitige Sicherheit in der Kommunikationstechnik, Addison-Wesley-Longman, pp. 169-180.

Friedmann, Eric J. and Resnik, Paul (1999) The social cost of cheap pseudonyms, Journal of Economics and Management Strategy (10:2), pp. 173-199.

Fritsch, Lothar (2007) Privacy-Respecting Location-Based Service Infrastructures: A Socio-Technical Approach to Requirements Engineering, Journal of Theoretical and Applied E-Commerce research (2:3), pp. 1-17.

Fritsch, Lothar. (2008) Profiling and Location-Based Services, in: M. Hildebrandt und S. Gutwirth (Eds.): Profiling the European Citizen - Cross-Disciplinary Perspectives, April 2008, Dordrecht, Springer Netherlands, pp. 147-160.

Fritsch, Lothar und Abie, Habtamu. (2008) A Road Map to the Management of Privacy Risks in Information Systems, in: Gesellschaft f. Informatik (GI) (Eds.): Konferenzband Sicherheit 2008, Lecture Notes in Informatics LNI 128, 2-Apr-2008, Bonn, Gesellschaft für Informatik, pp. 1-15.

Fritsch, Lothar; Scherner, Tobias and Rannenberg, Kai (2006) Von Anforderungen zur verteilten, Privatsphären-respektierenden Infrastruktur, Praxis in der Informationsverarbeitung and Kommunikation (PIK) (29:1), pp. 37-42.

Gellman, Robert (2002) Privacy, Consumers and Cost: How The Lack of Privacy Costs Consumers and Why Business Studies of Privacy Costs are Biased and Incomplete ,.

Goldschlag, David M.; Reed, Michael G. and Syverson, Paul F. (1996) Hiding Routing Information, in: R. Anderson (Eds.): Information Hiding, Berlin, Springer, pp. 137-150.

Hong, Jason; Ng, Jennifer; Lederer, Scott and Landay, James. (2004) Privacy risk models for designing privacy-sensitive ubiquitous computing systems, in: D. Benyon; P. Moody; D. Gruen and I. McAra-McWilliam (Eds.): Proceedings of the 2004 conference on Designing interactive systems: processes, practices, methods, and techniques, August 1, 2004, New York, ACM Press, pp. 91-100.

ISO (1999) ISO 15408 The Common Criteria for Information Security Evaluation.

Jameel (2007) Jameel, Hassan: research taxonomy, http://uclab.khu.ac.kr/usec/taxonomy/hassan.pdf

Jerichow, Anja; Müller, Jan; Pfitzmann, Andreas, Pfitzmann, Birgit and Waidner, Michael (1998) Real-Time Mixes: A Bandwidth-Efficient Anonymity Protocol, : Special Issue on "Copyright and privacy protection"." IEEE Journal on Selected Areas in Communications (16:4), pp. 495-509.

KPMG Canada(2003) A Retailer's guide to Privacy Risk Management,KMPG LLP, Canada.

Kohlweiss, Markulf; Fritsch, Lothar; Radmacher, Mike; Hansen, Marit and Krasemann, Henry (2004) Overview of existing assurance methods: PRIME Delivery D5.1.a,EU IST PRIME Project,.

Kohlweiss, Markulf; Gedrojc, Bartek; Fritsch, Lothar and Preneel, Bart. (2007) Efficient Oblivious Augmented Maps: Location-Based Services with a Payment Broker, in: N. Borisov and P. Golle (Eds.): Privacy Enhancing Technologies, 7th International Symposium, PET 2007 (LNCS 4776), Berlin, Springer, pp. 77-94.

Lacoste, Gérard; Pfitzmann, Birgit; Steiner, Michael and Waidner, Michael. (2000) SEMPER - Secure Electronic Marketplace for Europe, Springer, Berlin, ISBN 3540678255.

Levi, M. & Wall, D. S. (2004). Technologies, Security, and Privacy in the Post-9/11 European Information Society. Journal of Law and Society, 31: 194-220, 2004.

Meta Group (2005) Privacy Enhancing Technologies Ministry of Science, Technology and Innovation,Ministeriet for Videnskab, Teknologi og Udvikling, København, Denmark.

Möller, Ulf; Cottrell, Lance; Palfrader, Peter and Sassaman, Len (2004) Mixmaster Protocol Version 2, http://www.ietf.org/internet-drafts/draft-sassaman-mixmaster-03.txt, accessed 29-Dec-2004.

Müller, Markus (2004) Standards for Geographic Location and Privacy: IETF's Geopriv, Datenschutz and Datensicherheit (DuD) (28:5), pp. 297-303.

Odlyzko, Andrew (2003) Privacy, Economics, and Price Discrimination on the Internet: Extended Abstract, Digital Technology Center,University of Minnesota, Minneapolis.

Oinonen, Kari(2002) TR101 - LIF Privacy Guidelines.

Paintsil, Ebenezer and Fritsch, Lothar (2010): A taxonomy of Privacy and Security risk contributing factors in Identity Management. Privacy and Identity Management for Life - 6. Int. IFIP/PrimeLife Summer School , August 02, 2010.

Petweb I D2.1 (2007) : Fritsch, Lothar (2007) State of the Art of Privacy-Enhancing Technnology: Deliverable D2.1 of the PETweb project, Department of Applied Research in Information Technology (DART),Norsk Regnesentral, Oslo.

PICOS (2008): Schrammel, J. K., Christina; Weiss, Stefan; Kahl, Christian (2008). PICOS deliverable D2.2 Categorisation of Communities, PICOS FP7 EU project.

PRIME (2003) Privacy and Identity Management for Europe,: The IST PRIME Project, www.prime-project.eu, accessed 6.11.2006.

Pfitzmann, Andreas and Hansen, Marit(2003) Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology,:v0.21,.

Pfitzmann, Andreas and Hansen, Marit(2007) Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology,:v0.29, Dresden.

Pfitzmann, Andreas and Waidner, Michael. (1986) Networks Without User Observability – Design Options,Advances in Cryptology - EUROCRYPT '85: Proceedings of a Workshop on the Theory and Application of Cryptographic Techniques, Berlin, Springer, pp. 245.

Pfitzmann, Andreas; Priftzmann, Birgit and Waidner, Michael. (1991) ISDN-mixes: Untraceable communication with very small bandwidth overhead,In the Proceedings of the GI/ITG Conference on Communication in Distributed Systems, February 1991, pp. 451-463.

Rannenberg, Kai and Iachello, Giovanni. (2000) Protection Profiles for remailer Mixes -- Do the New Evaluation Criteria Help?Proceedings of the 16th ACSAC, IEEE Press, pp. 107-118.

Solove, Daniel (2006) A taxonomy of privacy, : GWU Law School Public Law Research Paper No.129." University of Pennsylvania Law Review (154:3), pp. 477.

TCG (2007) The Trusted Computing Group, http://www.trustedcomputinggroup.org, accessed 16-Jun-2007.

Zuccato, Albin. (2005) Holistic Information Security Management Framework, Karlstadt University Universitetstrycjeriet, Karlstadt, ISBN 91-85335-63-0.