available at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodinf.htm

Information Security Technical Report

Identity management throughout one's whole life☆

Marit Hansen[a,*], *Andreas Pfitzmann*[b,1], *Sandra Steinbrecher*[b,2]

[a]*Independent Centre for Privacy Protection Schleswig-Holstein (ULD), Holstenstraße 98, 24103 Kiel, Germany*
[b]*TU Dresden, Department of Computer Science, 01062 Dresden, Germany*

Keywords:
Identity
Privacy
Identity management
Privacy-enhancing technologies
PET
User-controlled identity management system
Privacy-enhancing identity management system

ABSTRACT

Identity management has to comprise all areas of life throughout one's whole lifetime to gain full advantages, e.g., ease-of-use for all kinds of digital services, authenticity and authorisation, reputation and user-controlled privacy.

To help laying the foundations for identity management applicable to people's whole life, we describe the formation of digital identities happening numerous times within one's physical life, i.e., their establishment, evolvement and termination, and derive building blocks for managing these digital identities from the needs of individuals and of society.

The identity attributes occurring and developing can be categorised according to their sensitiveness and the security requirements individuals have regarding them. We give an analysis of the sensitivity of identities and their attributes w.r.t. privacy and security both from a legal and individual's perspective. This leads to how systems for identity management throughout one's whole life should be designed using the building blocks derived.

1. Introduction

The identity of an individual begins to form after birth at the latest. As a baby, self-perception is limited. Usually at that time, the parents' perception of the baby's identity is the most comprehensive perception that exists. Self-perception as a person increases when the child grows up, making the child more and more independent from the parents' perception. With the growing perception of the own identity, the rights an individual has over his personal data, i.e., data relating or relatable to him, increases as well.

A converse process might apply to old-age people: in many cases self-perception might decrease, and in case an individual is no longer able to care for his own interests, rights and obligations concerning his personal data have to be transferred to others such as legal guardians. Such cases result in a situation similar to childhood when the parents are responsible.

Personal data (or if we include human perception: personal information) can be represented by so-called *digital identities* consisting of attributes, i.e., sets of personal data. Only subsets of these attributes – depending on the situation and the

☆ The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 216483 for the project PrimeLife. The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The PrimeLife consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law.

* *Corresponding author*. Tel.: +49 431 988 1214.
E-mail addresses: marit.hansen@acm.org (M. Hansen), pfitza@inf.tu-dresden.de (A. Pfitzmann), sandra.steinbrecher@tu-dresden.de (S. Steinbrecher).
[1] Tel.: +49 351 463 38277.
[2] Tel.: +49 351 463 38436.

1363-4127/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.istr.2008.06.003

context both in the physical and digital worlds – are needed to represent an individual, so-called (digital) *partial identities* (pIDs) (Pfitzmann and Hansen, 2008). An individual typically appears under different partial identities for work, others for leisure activities (e.g., doing sports or with the family), or dealing with companies (e.g., a bank, a store). Fig. 1 shows an example of partial identities. Partial identities contain information that interaction partners could know about a person in typical situations of daily life. An individual might want to decide which partial identity to use, depending on the situational context and the interaction partner, even though in most cases this decision is not done explicitly. Occasionally, it is suitable to remain entirely anonymous, e.g., when just surfing the web. In other cases, it is necessary to reveal identifying personal data, e.g., when paying for a product in a shop by credit card. Often, neither complete anonymity is acceptable to the other interaction partners nor identifiability of an individual is required, but only a share of personal data (typically reliable and may be certified) is needed. Appropriate partial identities and pseudonyms, which act as identifiers thereof, can fulfil the needs of the situational contexts and the parties involved. This may comprise the entire field between and including anonymity and identifiability (Pfitzmann and Hansen, 2008).

Much research and development has been done during the past couple of years to assist users in managing their partial identities in the digital world by several types of identity management (Bauer et al., 2005). But current concepts for identity management systems implicitly focus on the present (including the near future and recent past) only. The sensitivity of many identity attributes and the resulting need to protect them to enable privacy-aware identity management throughout the whole life is currently not dealt with.

After this short introduction, in Section 2 we describe which needs arise for personal data to evolve throughout one's whole life and which implications this evolvement has for the management of these identity attributes. In particular we sketch various building blocks for user-controlled identity management. In Section 3, we discuss the sensitiveness of various identity attributes occurring and developing and elaborate relevant categories with respect to privacy and security. Based on the results of the previous parts, Section 4 applies the findings to identity management throughout life. This comprises different areas and stages of life as well as the issue of the lifelong time period which has to be covered. Finally, we conclude and give an outlook on the open issues in Section 5.

2. Development and management of identity

Identity and especially digital identity of a human being begin to develop at latest at birth. Identity as well as digital identity can be generally modelled as a **set of attributes**. This set of attributes contains subsets representing **partial identities** in different areas of life the individual wants or must take part in.

In the following, we firstly describe phases of the formation of partial identities and show a few examples for the individual and the society perspective. Then we stepwise derive building blocks for user-controlled management of one's own partial identities.

2.1. *Formation of partial identities*

The lifecycle of partial identities consists of different phases:

''**Establishing a partial identity**'' means that the partial identity is created by or assigned to a person.
''**Evolving a partial identity**'' includes the usage of the partial identity both by the holder and by others. The usage of a partial identity by others can be done by observing or storing it and possibly by applying all kinds of data processing operations.

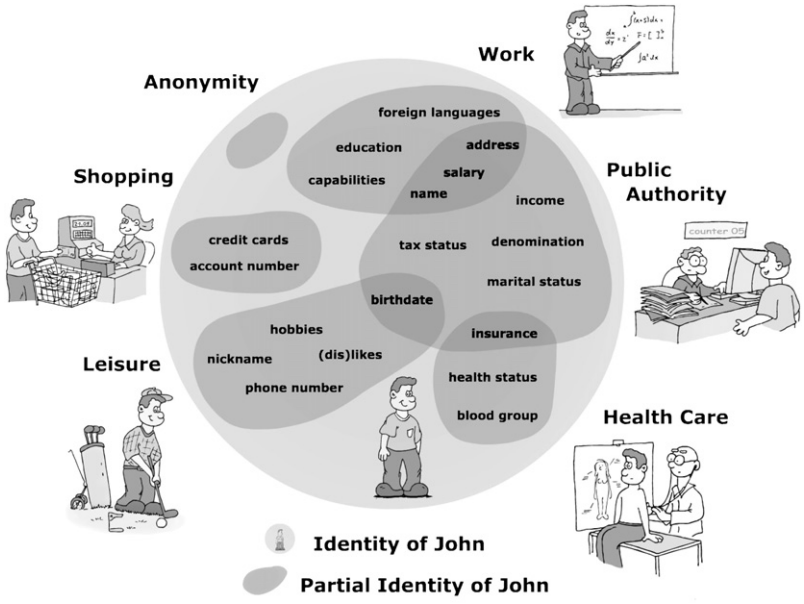

Fig. 1 – Partial identities of John (Borcea-Pfitzmann et al., 2006).

''**Termination of a partial identity**'' means deletion or suspension of the partial identity. Note that in some specific cases it can be possible to re-establish suspended partial identities.

All phases are relevant for formation of partial identities. These phases occur often during a person's lifetime. In the following typical areas of life, we sketch a few partial identities with their lifecycles and outline consequences on society as a whole.

2.1.1. *Individuals*

Soon after a child is born, several attributes and unique identifiers for major parts of life are **created automatically** in most European countries.

- **Government**: A birth certificate is issued by the registry office that usually contains the names inherited resp. given, gender, date and place of birth and information on the (biological) parents. In Europe typically neither the newborn nor his parents can prevent this automatic generation. With his birth the child becomes a resident of the state he is born in or his parents are residents of. Thus, the respective national law applies to the individual with all the rights and duties. Many of the respective laws require the disclosure of specific identity attributes (like gender, age, place of residence). For example, when a child grows up to a certain age, he has to go to school. Another example is the family benefits which parents get as long as the child is below a certain age. Thereby new partial identities with other governmental or also private institutions may be created.
- **Health care system**: The first data of the medical record created usually contain the information of the birth certificate as well as medical data like height/weight, specifics of the mother's pregnancy, birth of the child and his first physical examinations. Possibly a health insurance number has to be added. When the child grows older, data about additional examinations, possible vaccinations and diseases are added.

In addition to these two areas of life, parents start to make somewhat deliberate choices in which surroundings, called application contexts in IT jargon, their child will be known – this also means to create partial identities. This may be done, e.g., in the following typical parts of life:

- **Education**: Parents might apply for courses or a day-care centre. After having concluded the contract with the day-care centre, the child is authorised to participate in certain courses. Usually personal data on the child and the parents are disclosed to the day-care centre, and the way how to prove the authorisation for attending the courses is communicated – e.g., by simply stating one's name or by showing an assigned chipcard. As soon as these partial identities are created and the child himself begins to use them like in the case of visiting the day-care centre, he begins to further develop those partial identities – and thereby also to manage identity – himself.
- **Companies and associations**: Parents might register (often via Internet) their child with a commercial organisation like a nappy company to get some benefits from it like gifts. Typically parents also consult photographers who take (typically also digital) photos of the child. Also they may contract several insurances for their child.

Given that it is very hard – if not impossible – to erase widely used digital data, a digital identity usually is **only growing**, never shrinking. This means data parents allow others to collect about their child will still exist when the child – being more mature – takes over the care of his right of informational self-determination. From the legal perspective, the European Article 29 Working Party has made clear that children require legal representation to exercise most of their privacy rights (Art. 29 WP, 2008). This group demands a decision by the child when attaining majority: If the processing of a child's data began with the consent by the representatives, the matured child concerned may revoke the consent or give explicit consent. But for data disclosed on the Internet, a revocation of consent probably has little effect.

A grown-up individual whose partial identities have been partly or mainly developed by others so far, can take over more control to form them. Still in many cases, the individual concerned has no full autonomy w.r.t. use and further development of partial identities. In particular in governmental processes, the degrees of freedom for the individual are limited as data processing mostly is regulated by law.

Partial identities often do not terminate. For example, when active usage of a partial identity will stop (e.g., when a child leaves a school), often some of the data associated with that partial identity will have to be certified and transferred to other partial identities (e.g., when the child just changes to another school). Even if an individual dies, some data of partial identities will remain and may be transferred to other people. For example, the social insurance number will be used for paying pension to the surviving dependant.

2.1.2. *Society*

Society itself only exists because individuals, who live in it, are connected to each other.

Society needs unique identifiers to make a partial identity unique and addressable in parts of life. In a day-care centre, last name and first name usually will be sufficient to identify and address uniquely an individual while insurance companies usually assign a unique number to each client.

In every-day life unique identifiers are not suitable for speaking to individuals. For human beings, names usually are easier to remember than long numbers. Although several children with the same first name might exist in a day-care centre, the teacher will usually only use the first name when speaking to a child.

First names are difficult to change in most European countries while last names might be changed: An individual usually can willingly assume the family name in case of adoption or marriage, and in case of divorce, the name before marriage can be assumed again. At least governmental institutions keep track on the series of last names an individual and also his parents had. This is deemed necessary to prevent marriages between relatives.

Society subsists because it is not only alive in a single moment, but has ties to its past and plans – or at least options – for its future. Society itself is relatively stable

because it is rooted in its past, having history to learn from as well as traditions to keep, to further develop and to pass on. But society has a continuous interest in building life-stories of individuals that have relationships to each other through which the values of society can be transferred as well as the rights of individuals (like children or elderly people) can be asserted by others. Society cannot exist without links between individuals. But an individual needs self-determination, specifically informational self-determination, as stated in the 1983 ruling of the German Federal Constitutional Court, demanding that each person can at any time ascertain who knows what about him. This concept is particularly relevant in the data protection area that bases on transparency (in the meaning of clarity) of what others can know about oneself and the possibility for the individual to control that at least in a given range of options, which of course also have to be transparent to the individual. Partial identities and their processing by others are vital for understanding risks to one's private sphere and for enabling individuals to exercise their right to informational self-determination.

2.2. *Building blocks for managing partial identities*

Starting from *persons*, their attributes, identities and relationships, which persons commonly exhibit for centuries in the *physical world*, we derive **building blocks** for privacy-enhancing user-controlled identity management in a systematic way.

Each person has many *attributes*. Usually, attributes have a unique value which may be the value ''undefined'', e.g., the name of one's husband. Some *attribute values* might change over time, e.g., the colour of one's hair. Usually, it is possible to define attributes in a way such that they don't change, e.g., the colour of one's hair on May 1, 2008 at noon.

Building blocks of identity management have to do with representing persons, their attributes and attribute values, their identities and relationships in computer systems, i.e., the *digital world*. To derive them in a systematic way, we consider first the functionality in the physical world and then answer the question how this kind of functionality (or even more of it) can be supported in the digital world.

Let's start from elementary to more complex, i.e., from properties of single persons to properties of two persons, then to properties of three or more persons.

2.2.1. *Basic building blocks*

Each person has his *attributes and attribute values* not only just as one large unstructured set, but attributes and attribute values are *structured in subsets*. These subsets are not necessarily disjoint and comprise those attributes and their values which are relevant for some parts of life, e.g., being mother, wife, employee, and many more. With some of these attributes and their values as well as with some subsetting, i.e., structuring the attributes in not necessarily disjoint sets, we deal consciously. With others we deal unconsciously, e.g., several people unconsciously dress differently for different kinds of activities, managing the visibility of some of their attributes that way. Subsets of attributes and attribute values, which are considered useful in particular situations, are called *partial identities*.

Therefore, an identity management system should support the **definition and representation of attributes and attribute values** as well as of the **explicit and implicit subsetting** within the large set of all our attributes and attribute values to help persons (or their proxies) to **establish** and **evolve** those **partial identities** of them which are useful for acting in the digital world.

Partial identities are much easier to handle for us if we have an intuitive way to *name* them.

Therefore, an identity management system should support **naming of partial identities**.

The naming used by human beings, e.g., in personal communication, is usually done by **easy-to-remember names**. In contrast to those names, within computer systems partial identities may be identified by **unique identifiers**, e.g., index numbers in databases.

In addition, communication and interaction partners might wish to *address* and interact with (partial) identities at their initiative and have their communication and/or interaction be interpreted and treated in the context of the (partial) identities addressed.

Therefore, an identity management system should support **addressing of (partial) identities**. This may comprise different layers of an ICT (information and communication technology) system, e.g., network addresses at the communication layers and naming of entities within different applications. In addition, addressing should trigger the corresponding context popping up, so helping the addressed person to communicate and interact accordingly.

While communicating or interacting, we want to be in control which attributes and attribute values are revealed to whom. Communicating or interacting using one partial identity at a time usually gives help and guidance on this.

Therefore, an identity management system should support **deciding which attributes and attribute values are revealed to whom**.

Often, we do not like if others can relate our partial identities to each other and that way jointly evaluate larger sets of our attributes and attribute values. We want to enjoy some *privacy* by determining by ourselves the subsetting of our identities into partial identities. This is why we usually try to avoid globally unique, easy to store and easy to communicate attributes and attribute values, e.g., nobody enjoys an engraved SSN (social security number) within his face. And our appearance in different settings is easy to recognise for human beings, but the ability to recognise usually is not at all easy to transfer to others.

Therefore, a privacy-enhancing identity management system should enable the individual to control the properties of **pseudonymity** during a transaction (Hansen et al., 2004).

Pseudonymity means the use of (not necessarily digital) pseudonyms as unique identifiers for persons. An important

property is the degree of identifiability of a person when using the pseudonym: If the pseudonym has been assigned by a service provider or a third party to a person who identified himself to them, they can always link the pseudonym to the holder. If the pseudonym is being used and reused over a longer timespan, an observer can deduct information on the holder by aggregating data disclosed by each usage. If the pseudonym is not randomly generated, but reveals information itself, e.g., when choosing the name of a superhero, this also yields a piece of knowledge which makes it easier to spot the holder. With **transaction pseudonyms**, the use of the pseudonym is limited to one transaction only, and data disclosed in different transactions are not linkable by the pseudonyms which may enable anonymous usage. A privacy-enhancing identity management system should offer anonymous usage as default when there is no need for personal information, e.g., when only browsing the Internet without any legal relevance. It is clear that usage of one particular pseudonym as substitute for the civil identity of the holder in all facets of life does not leave much room for anonymity against observers. Identity management systems should provide for **contextual pseudonymity** which means that each pseudonym is only used within a certain context, e.g., depending on the current role of its holder or the relationship to the current interaction partner.

For communication and interaction partners, (partial) identities are more useful if actions done under the authority of a (partial) identity are *authenticated* w.r.t. that (partial) identity.

> Therefore, an identity management system should support authentication of actions w.r.t. (partial) identities. This can be done e.g., by **digitally signing** messages constituting or causing actions w.r.t. public keys of digital signature systems.

Both the need for *authentication and authorisation as well as* the need for *privacy* should be fulfilled.

> Therefore, a privacy-enhancing identity management system should support completely distinct authentication for each (partial) identity. That is, it has to support **digital pseudonyms**, e.g., unrelated public keys of digital signature systems.
> Based on authentication of partial identities others can grant these partial identities **authorisation** to perform certain actions.

Some (partial) identities evolve and are used over *very long periods of time*. During such periods, the technological infrastructure might change substantially.

> Therefore, an identity management system should support **migration to other** technologies, i.e., migration to other user devices and other communication infrastructure as well as use for new applications. In addition, the identity management system should **maintain usability** and **help to evolve (partial) identities**, which is not only needed to avoid errors, but also to perceive one's own digital life as something of continuity. Therefore, an identity management system should support **long-term evolvement and maintenance** of (partial) identities including their attributes, attribute values, subsetting, naming, authentication and addressing.

To evolve our partial identities in a way which reflects both our wishes w.r.t. the development of our identity as well as is consistent with our expectations w.r.t. privacy, we need to be aware *which attributes and attribute values we communicated to whom in which context*, i.e., in relation to which (partial) identities.

> Therefore, a privacy-enhancing identity management system should support its user by offering to store and to make easily accessible the **history** which attributes and attribute values have been communicated to whom in which context.

So far, we have considered properties required by single persons or by direct interaction of single persons. Next, we have to consider more complex properties.

2.2.2. *Complex building blocks*

Sometimes, people will try to cheat with respect to their attribute values, but others want to be quite certain that the communicated attribute values are true. W.r.t. attribute values which the communication partner cannot check immediately (others are, e.g., capability to communicate in a certain language, politeness, responsiveness to requests, and the like), this brings into play *certification of attribute values by third parties*. We all know this in various forms: some of our attribute values are contained in our passports or driver's licenses and that way certified by the public administration of our country. Other attribute values are certified by our school reports or employers' letters of reference.

> Therefore, an identity management system should support **third-party** certification **of attribute values** of partial identities. This can be done, e.g., by digitally signing some attribute values as part of a partial identity w.r.t. the third party's public key, possibly with some qualifications like expiration date.

Sometimes, the trustworthiness of the third parties may be put into doubt, particularly if we consider interactions where also the third parties would like to enjoy some privacy. Then first the reputation of the certifying party has to be determined and then the reputation granted by certifying the attribute values follows.

> Therefore, a privacy-enhancing identity management system should support a **privacy-respecting reputation system** (Steinbrecher, 2006).

Ideally, each communication or interaction partner only gets attribute values, i.e., personal information, he may act on whatever he deems fit. In practical life, we tell others personal information assuming they will use it only in particular ways agreed implicitly or even explicitly with us. In other words, we hope they will stick to a *policy* agreed. Sometimes, we are able to check whether they behave as we expect

them to. Often, we are not able to check or will only get to know if it is much too late to do anything about it. Therefore each of us learns during one's own life whom to trust w.r.t. which kind of policy – taking into account that even the legal framework might change which might imply changes to policies.

> Therefore, a privacy-enhancing identity management system should support policy **definition, policy negotiation**, and, as far as possible, **policy enforcement** w.r.t. how to use attribute values received. If agreed policies can no longer be enforced, e.g., a change of law requires a change of policy or a security breach occurred, all parties relying on the enforcement of the policy should be **informed of that change**. The latter gets the more important the longer the attribute values are stored or the more legal domains are involved.

For decades, when interacting with public administrations or larger companies, our attribute values get *input to larger workflows*. It is only now that we get able to negotiate workflows with public administrations or larger companies using computers to be at equal par with them. In the future, we might even define workflows when interacting with other people privately. Then, the very workflow definition may become some attribute value of ours if it is used only by us. Therefore, sharing our workflows with others or sanitising personal attributes from workflows may be needed.

> Therefore, a privacy-enhancing identity management system should support workflow **definition, modification, negotiation**, and – most importantly – **sharing and sanitising**.

For millennia, persons know that if they cannot do something by themselves or if they want not to, they can try to *delegate duties and the corresponding authorities* to someone else. This comprises delegating authority to authenticate actions to another person's digital pseudonym starting and ending under certain conditions (e.g., certain points in time or by revocation of authority). Delegation of course includes solving the problem to find out who is an appropriately qualified and trustworthy person to delegate to, comprising aspects of certified attributes of that person or at least his reputation, characterising his ability and willingness to stick to agreed policies and take part in agreed workflows.

> Therefore, a privacy-enhancing identity management system should support delegation **of duties and authorities from pseudonyms to pseudonyms**.

There are three possible situations that might occur regarding delegation from the legal perspective:

Firstly, delegation of rights might be made by law automatically for a certain time frame (e.g., for children to their parents). Secondly, delegation might be made willingly by an individual to others for a certain time frame (e.g., delivering mail to others during holidays). Thirdly, delegation of rights of an individual might be initiated by other individuals to achieve delegation of his rights to them or others (e.g., in the case of incapacitating a person), which presumably requires thorough juridical investigation before divesting the person of a right.

Today there is no ICT system which may provide the sketched building blocks to a sufficient extent. Various of the described functions are still subject to research, and the necessary interplay lacks standardisation for interoperability.

3. Sensitivity of identity

When focusing on specific aspects of identity management throughout life, clearly the degree of sensitiveness of identity attributes – single or combined – plays an important role. To begin with, we outline the legal perspective and contrast it to the individual perspective. Then we elaborate the most important categories for sensitivity with respect to privacy (Section 3.2) and to security (Section 3.3).

3.1. *Perception of sensitivity*

3.1.1. *Legal perspective*

For Europe, Article 8 of the Data Protection Directive 1995/46/EC defines special categories of data for which Member States shall prohibit the processing unless specific conditions are met: personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life. This legal provision reflects the societal consensus of sensitive data, in particular because of the long-term discrimination potential of this information: There is no guarantee that future data usage is bound to the purpose, the data were collected for, e.g., after change of government (Seltzer and Anderson, 2008).

This is also reflected in Human Rights Charters such as the Universal Declaration of Human Rights, adopted and proclaimed by General Assembly of the United Nations in December 1948, and the Charter of Fundamental Rights of the European Union from 2000. In addition to general statements on the need for protecting human dignity and the integrity of the person, which are clearly relevant for some aspects of identity, they define certain areas of life which specifically should be safeguarded: private and family life, home and communications.

Not only privacy, but also security issues are regulated in data protection law, in particular the need for confidentiality, integrity and also availability. For example, Article 6 of the Data Protection Directive 1995/46/EC focuses on quality of personal data which has to be accurate and, where necessary, kept up to date. Article 16 demands confidentiality of processing, and Article 17, dealing with ''security of processing'' in general, stipulates the implementation of ''appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access''.

In 2008, the German Federal Constitutional Court has derived from the fundamental legal premises of the German constitution a new basic right to ''the guarantee of confidentiality and integrity of IT systems''. This ruling is notable because it explicitly extends the room worth or even necessary to be protected by IT systems under control of an individual.

All these legal references give clues on what is typically estimated as sensitive from a society perspective. Still the individual perspective can differ.

3.1.2. *Individual perspective*

What a person estimates as sensitive or non-sensitive data, can be quite subjective: for example, the shoe size may be quite sensitive for some people while for others it is not. Of course this also may depend on the context: being served in a shoe shop may require giving the shoe size, which a person is not willing to disclose in other settings. Personal diaries usually are considered sensitive information in personal life – it would be highly inappropriate to have a look in a diary book unless being explicitly invited by the author. With publicly accessible blogs written as personal diary this is different because readers assume that the public access rights are assigned on purpose.

The estimation of sensitivity and the willingness to disclose certain information to specific people (or even the public) may vary over time.

Sensitivity of certain attributes has not only to be considered with respect to privacy, but also with respect to security of these data. The assignment to a person or pseudonym might have to be done and if it is done it has to be done authentic. For example, if one is only able to lend a book from the university library with a student ID, the certification of the student status is seen as sensitive information while it may not when just telling others about the student status.

In the following we try to subsume the sensitivity of common attributes, their possible attribute values and their possible need for certification.

3.2. *Sensitivity with respect to privacy*

Some attributes and attribute values usually need *more privacy protection* than others, e.g., those which are not easy to change, do not vary over time, are given attributes, or contain side information. We distinguish the following properties of identity attributes which alone or in combination pose specific risks to privacy when being disclosed:

1. **Static or changes quite accurately predictable**: Attributes which are static over time can be sensitive if being disclosed over and over again in different situations, because thereby they enable linkage of related data. Examples for static attributes are the birth date or – in most cases – the sex. It is not always possible to coarsen the detailedness of given information, e.g., only stating the year or the age range, or to lie. Disclosure of correct static attributes enables observers to link those situations and gather related data which possibly can be used to identify the individual whose attributes are being monitored. Similar to static attributes are those which are quite accurately predictable or guessable because they follow some rules, e.g., transitory modifications which tend to return to a natural state, such as hair colour, or persistent modifications which tend to remain in place, such as a tattoo (Dixon, 2005) or data following mathematical rules like the number of children that will only remain or increase. If static identity information is being used for purposes such as authentication, this bears a risk because these data cannot easily be revoked and substituted: For example, the use of finger prints with biometric access systems.
2. **(Initial) determination by others**: There are attributes values which the individual concerned cannot determine himself (e.g., the first name). Here especially the initial setting of identity attribute values is relevant as it may persist or it may take time or great effort to change it. A special case is the inheritance of attribute values from others, e.g., the DNA being inherited by the natural parents or the last name from one's family. Further all partial identities for a child created by the parents are initially determined by them, because the child is not capable of caring for the own informational self-determination yet.
3. **Change by oneself impossible or hard to achieve**: Wilful changes of attribute values can put the individual in control of his identity, but this is not always possible, e.g., if attributes are static (see above) or if a given value of the attribute is not under the individual's control, e.g., when processed in an organisation. Thus, the autonomy of the individual may be limited concerning the values of those attributes.
4. **Inclusion of non-detachable information**: Some attributes cannot be disclosed without simultaneously also disclosing some side information tied to the attribute. Examples are simple sequence numbers for identity cards which often reveal sex, birth data and at least a rough time frame of when the identity card was issued, or biometrics such as the face image which reveals information on some diseases, possible drug usage or the stress level of the individual concerned (Hansen and Meissner, 2007). In many cases it is very hard if not impossible to detach that side information, and usually the individuals concerned are not aware of revealing extra information if they are asked to disclose some piece of data.
5. **Singularising**: If an attribute or attribute value provides the possibility to singularise a person within a group, this can affect an individual. The person may still be anonymous, but may be recognised in different contexts, and then his privacy may be invaded by tracking or locating. If a direct contact can be established, e.g., by calling the person, sending e-mail or posing personalised advertisements, this can be conceived as intrusion into one's privacy, even if the intruder has no clue about one's name.
6. **Prone to discrimination or social sorting**: Even without the necessity to be singularised or establishing the direct link to the person, negative effects on the individual are possible. There is no attribute which is definitely resistant against a possible discrimination for ever and ever. Often only specific attribute values are prone to discrimination, e.g., when revealing (potential) poorness, diseases or negative personality traits. Usually one can assume that negative consequences from discrimination are less likely for behaviour which is compliant to current laws or a societal norm (at least discrimination from the state or society itself; cf. Section 3.1.1). When discussing discrimination, usually only direct negative effects on the individual concerned are considered. But they may also have an effect on other individuals: for example, if some people disclose information and thereby positively stand out from an unknown mass, this may negatively discriminate all others who do not disclose their identity attribute values.

We distinguish between the following types of attributes and add the numbers of the above-mentioned properties:

- **Attributes of possession** are all kinds of certificates given by others starting from the birth certificate over school reports to death certificates. Certificates are naturally determined by others (2) and hopefully static and difficult to change for integrity reasons (3). Although they might not be issued to persons, but to pseudonyms they might allow for singularising (5) and discrimination (6). Also they might contain non-detachable information (4) like the information who issued the certificate. For example, in the case of a school report or passport, this might reveal where a person had lived or had been. For hiding stamps from hostile countries, additional passports can be issued so that context-specific usage is supported.
- **Attributes of knowledge** cover both special experiences and knowledge. This ranges from long-learned experience in a specific scientific field to the simple by-heart-learned knowledge of a passphrase or bank account number. Usually knowledge is not static or predictable. But it might be difficult to forget knowledge. Knowledge might contain non-detachable information (4). Only someone who has personal experience with prisons might have been there. This might allow for singularising (5) and discrimination (6). To a certain extent, knowledge is transferable.
- **Attributes of interests** cover all kinds of things a person does, e.g., his shopping behaviour gives information about his interests. All behaviour might contain non-detachable information (4) and therefore allows for singularising (5) and discrimination (6).
- **Attributes of characteristics** cover all kinds of biometric attributes ranging from DNA to age. These attributes are static or predictable (1), initially determined by others (2) and very difficult or impossible to change (3). Therefore they allow for singularising (5) and discrimination (6). Also they might contain non-detachable information like diseases (4).

3.3. *Sensitivity with respect to security*

The handling of identities and the communication of identity attributes to others needs to adhere to – inter alia – integrity and availability requirements. Identity management should assist adhering to these requirements. Within the basic building blocks of identity management during communication, the following steps may be performed:

1. Demanding the declaration of certain attribute vales from a communication partner.
2. Showing a self-certified attribute to a communication partner.
3. Showing an attribute to a third party for certification and binding to a pseudonym.
4. Showing a third-party-certified attribute to a communication partner.

These four steps as well as the corresponding storage of identity attributes and certificates have integrity and availability requirements.

Certified attributes should only be linked to the users the attributes really belong to. Whenever false assumptions by others of the link between an identity attribute and the related individual are made or can be made, security of the respective attribute in the sense of integrity and authenticity is damaged. We distinguish the following possible threats to security during the actions above:

- **Wrong authentication possible**: If someone else misuses attributes in the sense that he shows these attributes to a third person as his attributes, one speaks of (partial) identity theft.
- **Wrong assignment of an attribute to a person/pseudonym**: If an attribute is assigned to a person who does not have this attribute, bad mouthing and stoning as well as wrong praising and ballot stuffing of someone (possibly also oneself) are possible.
- **Denial/repudiation of certification**: If no third party is willing to sign an attribute as belonging to a person/pseudonym the last two actions above are not possible.
- **Missing or unclear trust in certifying party**: There is only an advantage of third-party-certified attributes over self-certified attributes if a person trusts in the certifying party that it checked the correctness of the link between certified attribute and the respective pseudonym.
- **Deniability of attributes possible:** If someone is able to deny that a certain attribute he does not like, but that might be important for someone else to know belongs to him, one speaks of whitewashing.
- **Non-revocability of wrong assignment**: In this case a person might not be able to change attributes to correct values or a third party, who certified attributes, may not be able to revoke the certificate.

All attributes stored have to be certified by someone having no control over the storage to guarantee that the one who stores them cannot damage the attributes' integrity. Usually only communication between communication partners is at the most integrity-protected during communication by SSL, but that does not protect against the wilful changing of attributes of the communication partners, but only against others. The storage typically is only protected by simple access control and logging, but that does not prevent anyone having access to the place of storage to change the stored attributes.

4. How to design systems for identity management throughout life

In the previous sections, we have discussed the relevance of identity management, and in particular of privacy-enhancing user-controlled identity management, for individuals in the information society. As our information society is based on usage of IT systems, they should support individuals' needs for identity management. Based on existing proposals for privacy-enhancing identity management as summarised in Section 4.1, the following subsections deal with the necessity to cover all areas of life (Section 4.2), all stages of life from childhood to becoming old age with more or less capability to maintain control over one's participation in the information society and over one's private sphere (Section 4.3) and

the ability of the identity management system to work for the full lifespan of an individual (Section 4.4).

4.1. *Mechanisms for user-controlled privacy*

Since the 1980s, basic mechanisms for privacy-enhancing identity management under control of the user have been proposed (cf., Chaum, 1985; Pfitzmann et al., 2000; Leenes et al., 2007). Control by the user requires – as already explained in Section 2.2 – that he firstly knows about actual and potential processing of his personal data and secondly that he in principle can decide case-by-case on data disclosure to specific parties, possibly in the limits given by law and society. The most effective, yet not always realistic way to protect one's privacy is data minimisation, i.e., to disclose as little personal data as possible. Relevant mechanisms are listed in the following subsections.

4.1.1. *Handling of partial identities*

All partial identities should be protected against misuse. Unfortunately this is not a trivial requirement: current computers with their operating systems and application software do not provide reliable confidentiality, integrity and availability of stored data. The personal computer in the hands of the user may be vulnerable when not being administered professionally including patching the system against new attacks. On the other hand, releasing data to organisations which may act as infomediaries for managing the user's identity on his behalf can also be critical because the collected information yields a detailed image of the individual's personality. Confidentiality breaches by employees of that organisation may go unnoticed, and infomediaries with a huge database are interesting targets for attackers. Partially, remedy may be achieved if personal data is encrypted by the user in a way that unauthorised persons cannot access the clear text. Many non-time-critical data functions can also be applied to encrypted data with the help of multi-party computation or secure function evaluation.

4.1.2. *Data minimisation*

Data minimisation does not only mean to reduce the amount of personal data to the least possible, but also to limit potential observation, linkage and identification when disclosing data. As explained before, context-specific pseudonyms can support the separation of multiple contexts, preventing overarching profiles of individuals. As re-use of partial identities can bear the risk of unwanted linkage, it should be supported to limit their usage, e.g., by creating partial identities which can only be used once or which expire after a predefined time period. Expiring e-mail addresses help users to control their reachability, e.g., can help against spam when disclosing communication addresses.

Undesired linkage when proving one's authorisation is prevented by private credentials which are certificates proving identity claims (e.g., ''being of age'') without revealing information that may identify the individual (Chaum, 1985; Camenisch and Lysyanskaya, 2000). Various private credentials can be derived from a single master certificate that are neither linkable to each other nor to the issuance interaction of the master certificate. Only in the case of misuse the identity of the user may be revealed.

Other technologies can be used to protect users against undesired observation: on the network layer, the communication may be encrypted and routed via several independent proxy servers (so-called Mixes, cf. Chaum, 1981) which can guarantee anonymity against other parties unless users disclose further information. Private information retrieval provides the possibility to hide which items from a database a user is interested in (Chor et al., 1998).

4.1.3. *Enforceable rules for data processing*

If personal data leave the area controlled by the user, he should know what will happen with them. Statements on planned data processing can be given by privacy policies which – if interpretable by software – can even be enforced automatically. Research is done in the area of sticky policies which are cryptographically tied to personal data and thereby can travel together with them after being disclosed (Karjoth et al., 2002; Casassa Mont et al., 2003).

Legal provisions foresee the possibility to withdraw consent on processing of personal data. However, often the already performed data processing and its consequences cannot easily be undone. In particular, there is no guarantee to delete personal data which has been transferred to other parties, possibly even locating in remote jurisdictions. A technical solution may be achieved when no copies of data are transferred, but only access to the user's data repository is granted so that the access may be denied after revoking consent. But that does not prevent anyone who had once access to the database to copy the data at that time and then re-using it later on. Some researchers plead for the implementation of forgetting in the digital world (Mayer-Schönberger, 2007).

4.1.4. *Transparency functionality*

Information of the user on planned and actual data processing is needed for informational self-determination. Thereby transparency – in the meaning of giving clear and understandable information – is a prerequisite to all kinds of control by the user. The history function as demanded in Section 2.2.1 should be able to cover all online communication and store when which personal data have been disclosed to whom and under which conditions. Also information on communication partners, their reputation, estimated trustworthiness or on the jurisdiction which apply to them, may be made visible to the user because this may be relevant before establishing a communicational relation.

The actual data processing may differ from the planned one, e.g., in the case of privacy or security incidents. In several states, Security Breach Notification Acts oblige organisations to inform people concerned on security breaches which affect their personal data. Other feedback processes can help improving the data quality and prevent conclusions of organisations based on inaccurate data or data containing non-detachable information. An identity management system should be able to collect and interpret all these kinds of data sources which are relevant to protection of the private sphere and user control. Also access to own personal data, as granted by European data protection law, should be supported (Hansen, 2008).

4.2. *Mechanisms for covering all areas of life*

Identity management which should cover all areas of life should act as the communicational gateway to the outside world. This means that, at best, all digital partial identities are handled, all digital communication can be logged in the history and there is comprehensive user support. This requires hardware and software interfaces to legacy and emerging systems, e.g., governmental eIDs, healthcards or SIMs of mobile phones. For supporting the user, the identity management system should be equipped with basic knowledge on typical or allowed processes for the covered areas of life. For example, the identity management system should not only provide storage space for school reports or diploma certificates, but it also should inform users on who is allowed to request or demand access to those documents and how sensitive those data are.

As there is life outside the world of digital communication networks and IT systems, not all parts of life can be fully covered. Users may want to abandon their identity management system in intimate face-to-face meetings – it may rather be a sign of trust or of informality not to integrate a safeguarding identity management system in the communication with close friends. Another area which is difficult to comprise is the ubiquitous world of sensors and video surveillance as far as their existence is not communicated to the identity management system. Currently there is no reliable protection against hidden surveillance, thus the identity management system cannot prevent privacy intrusion.

What is typically difficult to prevent, but what the identity management should at least keep track of is the data others reveal that may be part of the own partial identity. Examples are the declaration of friendship or knowing each other like in social networks. Another example is consenting that others may analyse one's genetic data that can affect all biological relatives.

4.3. *Mechanisms for covering all stages of life*

As stated in Section 2.2, infants cannot decide on their own how they are involved in (the information) society and how their private sphere can be controlled. Adults, too, may have temporary or permanent needs that others support them or even act on their behalf concerning decisions on their partial identities. This can be implemented by different forms of delegation. Usually the delegate does not take over the identity of the individual concerned, but gets authorisations to act – often within defined ranges – on behalf. Cryptographic certificates can be employed for that purpose. In any case, technical processes have to be defined in accordance with legal procedures to grant and revoke delegations. Some mechanisms deserve particular attention when designing appropriate processes for acting on behalf of others: For example, group signatures allow each member of a group to anonymously sign messages on behalf of that group (Chaum and van Heyst, 1991). In secret sharing systems it is possible to define a threshold of people who have to cooperate before the delegation will be activated (Shamir, 1979). This may be used in emergency cases where the individual concerned cannot directly assign a delegate. Of course there should also be defined processes to terminate delegation, and to be informed about what the delegate has done on one's behalf, in particular in cases of temporary delegations. For achieving an atmosphere of trust, people should be able to choose their delegates and supporting parties, if they are in the mental state enabling such decisions.

Delegates have to pay attention that certification of attributes is done as soon as possible and is stored in a long-term fashion to prevent that the one they are delegates of has disadvantages from not having these certificates later on. This holds for example, for children's birth certificates or school reports. Also delegates may need to keep the same attention to the certificates they themselves have to prove after the end of delegation why they did or did not do something. There can be limits on what delegates can do because they have to pursue the good for the person they represent. Considering today's effects of potential lifelong storage of everything, which is posted on the Internet, it may be advised that parents object against all kinds of data disclosure on their children in the Internet. All data disclosure may negatively affect the children's future with respect to their private sphere.

4.4. *Mechanisms for covering the full lifespan*

A hard problem is that identity management systems should cover the full lifespan of a person and even some time afterwards. This means to keep personal data and their management for decades. From the perspective of today's IT system landscape, this is related to the archiving challenge to long-term availability and integrity of data. Remedy can be achieved by regularly migrating one's data to new hardware and software systems, but this may not be always possible, e.g., when data are bound to specific hardware for security reasons.

Similarly the degree of security against attacks on the data's confidentiality and integrity may not persist. On the one hand, there are – from today's point of view – quite academic discussions on the long-term robustness of cryptographic protection (cf. Buchmann et al., 2006). On the other hand, the actual security level of widespread IT systems is disastrous if not regularly (at least weekly) patched. And even immediate patching does not help against specifically tailored exploits or against adversaries taking advantage of information on security vulnerabilities before they are published. For specifically sensitive data, e.g., health data, a solution might be not to use IT systems at all, or at least not today's multi-purpose systems, but possibly specific, hardened hardware, running the identity management system in virtual machines with without online access or limited one. Today, this may be difficult for data being processed by user-controlled identity management: as single pieces of data, often these data are not that sensitive, but as a large compilation over several years, they are.

For achieving an appropriate level of security, there is the need of an ongoing security management process. Unfortunately this is not always considered – even with the highly sensitive, contactlessly readable ePass, the global standardisation from the International Civil Aviation Organization (ICAO), a comprehensive process is missing. It is advised not to plan for 10 years' validity of such machine readable travel documents because – as it was the case with the ePass – the

security flaws and privacy risks of the first generation will also persist for that long time instead of having the possibility to revoke these documents (cf. Meints and Hansen, 2006). By the way, because of the desired validity of 10 years, the technical solution had to be contactless as contact-based chipcards are said to have a shorter lifetime when being used intensively.

Finally, lifelong data protection yields the problem that possible privacy-relevant effects in the future are very hard to foresee. It is a principle that all consent according to the Data Protection Directive 1995/46/EC is defined as ''any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed''. But how well informed is the data subject, i.e., the individual? And how can other regulations – such as the limitation of processing to specific purposes or the obligation to erase data when they are not needed anymore –be enforced in a globalised world? Misuses of personal data, in particular breaches of confidentiality, can hardly be noticed by the individual concerned. Proving before court that misuse has happened and quantifying possible negative effects on an individual, are also difficult if not impossible. Here individuals need much better support from society, and organisations need real incentives to behave well regarding data protection.

5. Conclusions and outlook

To gain full advantages of identity management, it has to enable synergies between ease-of-use of all kinds of services, authentication and authorisation w.r.t. distinct partial identities, gaining reputation under diverse partial identities, and putting the user into control of managing his partial identities to enable trustworthy privacy. We described at a conceptual level how this can be done and what deliberations have to be taken into account to come up with appropriate compromises (e.g., between authorisation and reputation on the one hand and privacy on the other) and synthesis in building identity management systems.

To put identity management throughout one's whole life to practice, much has to be done: individuals have to get aware what identity management means if it has to comprise cyberspace and include both security and privacy and this again both for themselves as well as for others. Organisations have to learn that putting individuals in control of what kind of life they want to live is in the long term the most promising way of customer-relationship management. Last but not least, various kinds of interoperable identity management systems have to be implemented, tested (besides other aspects w.r.t. usability, security and privacy) and fielded at a large scale. And finally, all actors have to learn about the true demands of the others actors, since synergies can only fully arise then.

REFERENCES

Art. 29 Working Party. Working document 1/2008 on the protection of children's personal data (General guidelines and the special case of schools), 00483/08/EN, WP 147, adopted on 18 February 2008, http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp147_en.pdf; 2008 [current June 2008].

D3.1: structured overview on prototypes and concepts of identity management systems. In: Bauer M, Meints M, Hansen M, editors. Deliverable 3.1 in the network of excellence FIDIS – future of identity in the information society. V1.1, Frankfurt/Main, Germany, Sep. 2005, http://fidis.net/fileadmin/fidis/deliverables/fidis-wp3-del3.1.overview_on_IMS.final.pdf; 2005 [current June 2008].

Borcea-Pfitzmann K, Hansen M, Liesebach K, Pfitzmann A, Steinbrecher S. What user-controlled identity management should learn from communities. Information security technical report, vol. 11. Elsevier; 2006. no. 3, p. 119–28.

Buchmann J, May A, Vollmer U. Perspectives for cryptographic long-term security. Communications of the ACM, vol. 49; Sep. 2006. no. 9, p. 50–55.

Camenisch J, Lysyanskaya A. Efficient non-transferable anonymous multi-show credential system with optional anonymity revocation, Research report RZ 3295 (# 93341). IBM Research; Nov. 2000.

Casassa Mont M, Pearson S, Bramhall P. Towards accountable management of identity and privacy: sticky policies and enforceable tracing services, HPL-2003-49. Trusted Systems Laboratory, HP Laboratories Bristol, http://www.hpl.hp.com/techreports/2003/HPL-2003-49.pdf; 2003 [current June 2008].

Chaum D. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, vol. 24; Feb. 1981. no. 2, p. 84–8.

Chaum D. Security without identification: transaction systems to make big brother obsolete. Communications of the ACM, vol. 28; Oct. 1985. no. 10, p. 1030–1044, http://chaum.com/articles/Security_Wthout_Identification.htm. [current June 2008].

Chaum D, van Heyst E. Group signatures. Advances in cryptology – EUROCRYPT '91, vol. 547. LNCS; 1991. p. 257–65.

Chor B, Kushilevitz E, Goldreich O, Sudan M. Private information retrieval. Journal of the ACM Nov. 1998;45(6):965–81.

Dixon M. Discovering identity: identity map – characteristics. Blog from 15 Nov. 2005, http://blogs.sun.com/identity/entry/identity_map_characteristics. [current June 2008].

Hansen M. Marrying transparency tools with user-controlled identity management. In: Proceedings of the IFIP & FIDIS summer school 2007. IFIP Publisher Springer Science and Business Media; 2008. p. 199–222.

Hansen M, Berlich P, Camenisch J, Clauß S, Pfitzmann A, Waidner M. Privacy enhancing identity management. Information security technical report, vol. 9. Elsevier; 2004. 1, p. 35–44.

Hansen M, Meissner S, editors. Verkettung digitaler Identitäten. Lulu Inc., https://www.datenschutzzentrum.de/projekte/verkettung/; 2007 [current June 2008].

Karjoth G, Schunter M, Waidner M. Platform for enterprise privacy practices: privacy-enabled management of customer data. In: Proceedings of 2nd workshop on privacy enhancing technologies (PET 2002), LNCS 2482. Springer; 2002. p. 69–84.

Leenes R, Schallaböck J, Hansen M, editors. PRIME white paper, https://www.prime-project.eu/prime_products/whitepaper/; 2007 [current June 2008].

Mayer-Schönberger V. Useful void: the art of forgetting in the age of ubiquitous computing, Faculty research working papers series no. RWP07-022. John F. Kennedy School of Government – Harvard University, http://ksgnotes1.harvard.edu/Research/wpaper.nsf/rwp/RWP07-022/$File/rwp_07_022_mayer-schoenberger.pdf; Apr. 2007 [current June 2008].

Study on ID documents. In: Meints M, Hansen M, editors. Deliverable 3.6 in the network of excellence FIDIS – future of identity in the information society, V1.1, http://www.fidis.net/

fileadmin/fidis/deliverables/fidis-wp3-del3.6.study_on_id_documents.pdf; Dec. 2006 [current June 2008].
Pfitzmann A, Hansen M. Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management – a consolidated proposal for terminology, Working paper v0.31, http://dud.inf.tu-dresden.de/Anon_Terminology.shtml; Feb. 2008 [current June 2008].
Pfitzmann B, Waidner M, Pfitzmann A. Secure and anonymous electronic commerce: providing legal certainty in open digital systems without compromising anonymity, IBM research report RZ 3232 (#93278) 05/22/00 computer science/mathematics. Zurich: IBM Research Division, http://www.semper.org/sirene/publ/PWP_00anoEcommerce.ps.gz; May 2000 [current June 2008].
Seltzer W, Anderson M. Using population data systems to target vulnerable population subgroups and individuals: issues and incidents. In: Asher J, Banks D, Scheuren FJ, editors. Statistical methods for human rights. Springer; 2008. p. 273–328.
Shamir A. How to share a secret. Communications of the ACM, vol. 22; Nov. 1979. no. 11, p. 612–3.
Steinbrecher S. Design options for privacy-respecting reputation systems within centralised internet communities. In: Fischer-Hübner S, editor. Security and privacy in dynamic environments, Proceedings of the IFIP TC-11 21st international information security conference (SEC 2006), IFIP, vol. 201. Springer; 2006. p. 123–34.

Marit Hansen is a computer scientist and Deputy Privacy Commissioner of Schleswig-Holstein, Germany, where she heads the ''Privacy-Enhancing Technologies (PET)'' Section. Since her diploma in 1995 she has been working on security and privacy aspects especially concerning the Internet, anonymity, pseudonymity, identity management, biometrics, multilateral security, and e-privacy from both the technical and the legal perspectives. In several projects she and her team actively participate in technology design in order to support PET and give feedback on legislation.

Andreas Pfitzmann is a professor of computer science at Dresden University of Technology. His research interests include privacy and multilateral security, mainly in communication networks, mobile computing, and distributed applications. He has authored or co-authored about 130 papers in these fields. He received diploma and doctoral degrees in computer science from the University of Karlsruhe. He is a member of ACM, IEEE, and GI, where he served as chairman of the Special Interest Group on Dependable IT Systems for ten years.

Sandra Steinbrecher is a scientific assistant of Computer Science at Dresden University of Technology. Since her diploma that she received from University of Saarland in 2000 she has been working in several projects and areas of privacy, computer security and cryptography. Her major research interests are the modelling and measurement of anonymity in distributed networks, privacy-enhancing identity management and the design of privacy-respecting reputation systems.