Manifesto from Dagstuhl Perspectives Workshop 11061

Online Privacy: Towards Informational Self-Determination on the Internet

Edited by

Simone Fischer-Hübner[1], Chris Hoofnagle[2], Ioannis Krontiris[3], Kai Rannenberg[4], and Michael Waidner[5]

1 **Karlstad University, Sweden**
simone.fischer-huebner@kau.se
2 **University of California, Berkeley, U.S.A.**
choofnagle@law.berkeley.edu
3 **Goethe University Frankfurt, Germany**
ioannis.krontiris@m-chair.net
4 **Goethe University Frankfurt, Germany**
kai.rannenberg@m-chair.net
5 **TU Darmstadt, Germany**
michael.waidner@sit.fraunhofer.de

Abstract

While the collection and monetization of user data has become a main source for funding "free" services like search engines, online social networks, news sites and blogs, neither privacy-enhancing technologies nor its regulations have kept up with user needs and privacy preferences. The aim of this Manifesto is to raise awareness for the actual state of the art of online privacy, especially in the international research community and in ongoing efforts to improve the respective legal frameworks, and to provide concrete recommendations to industry, regulators, and research agencies for improving online privacy. In particular we examine how the basic principle of informational self-determination, as promoted by European legal doctrines, could be applied to infrastructures like the internet, Web 2.0 and mobile telecommunication networks.

Seminar 06.–11. February, 2011 – www.dagstuhl.de/11061
1998 ACM Subject Classification K.4.1 Privacy
Keywords and phrases Online Social Networks, Informational Self-Determination, Privacy Enhancing Technologies, Data Protection Directive
Digital Object Identifier 10.4230/DagMan.1.1.1

Online Privacy: Towards Inform. Self-Determ. on the Internet, *Dagstuhl Manifestos*, Vol. 1, Issue 1, pp. 1–20
Editors: S. Fischer-Hübner, C. Hoofnagle, I. Krontiris, K. Rannenberg, and M. Waidner

Dagstuhl Manifestos
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Executive Summary

Existing conceptions of privacy typically incorporate user control as a key component, or indeed describe privacy as a form of user control over information. However, the architecture and development of the Internet have driven individuals to lose control over the collection, use and transfer of their personal data online. Instead, the fundamental value exchange underlying the Internet economy is that services are provided free of charge in return for pervasive use of individuals' information. This business model remains opaque to many users, who willingly or unwillingly share massive amounts of personal data, with a myriad of parties online.

State of the Art. The subjective and contextual nature of privacy challenges any attempt to crystallize individual attitudes toward issues like information sharing on social networks or online behavioural advertising. Perceptions of privacy vary across cultures; there is often inconsistency between what people say about privacy and the options available to them to express their preferences; and privacy harms are difficult to measure. This complicates the understanding of social norms with respect to online privacy.

Privacy enhancing technologies (PETs), while existing for many years, have not been widely adopted by either industry or users. PETs include opacity tools, intended to "hide" personal data in accordance with the principle of data minimization, as well as transparency enhancing tools (TETs), providing users with information about privacy policies or granting them online access to their personal data.

Problems with PETs include low demand from both industry, which is fuelled by information, and users, whose awareness and interest are low; as well as lack of mechanisms for integration into large infrastructures, which were conceived and designed without privacy in mind, requiring properties like usability, scalability, efficiency, portability, robustness, preservation of system security, and more. Problems with TETs include the difficulty and lack of interest of individuals to comprehend complex data flows; and security problems arising from the provision of online access to personal data.

Engineering and Industry Options. Businesses have insufficient incentive to integrate privacy into the design and management of products and services, due to low demand from users; low awareness on the part of both users and industry; competitive business pressures to exploit information; and an absence of coherent, harmonized global regulation.

To improve the current environment, we must meet three challenges: First, transparency must be enhanced for users, through implementation of TETs in a privacy-friendly manner, open source development of TETs, and better user interfaces for transparency in complex environments. Second, PETs should be designed and delivered to end users by building blueprints and sample prototypes for key scenarios (e.g., delivery of service on mobile devices; or use of pseudonyms on communication networks), and deploying open source code to reduce market entry costs. Third, identity management should be promoted as a key technique to manage information while satisfying principles of data minimization and transparency; using minimum data to authenticate and authorize users; and giving particular emphasis to user centric identity management.

Improving Regulations. Unfortunately, neither the current European legal framework nor the United States approach of industry self-regulation has been effective in protecting privacy online. The main problems inflicting the current framework are the blurring distinction between personal and non-personal data; the erosion of consent as a sound basis for data processing; the (in)applicability of European law to websites and third parties based in

the United States; and the regulatory emphasis on ex post remedies in lieu of ex ante risk minimization.

Recognizing that privacy is regarded in Europe not only as an individual right but also as a societal good, which underlies values such as democracy, autonomy and pluralism, we must insist on the continued existence of a strong European legal instrument based on principles of data minimization and ex ante risk prevention. While the path of least resistance may be to make incremental changes to the Data Protection Directive, this may not succeed (and may result in slightly better but still ineffective regulation) absent rectification of fundamental conceptual shortfalls. First, recent examples of de-anonymization attacks have proven the futility of trying to distinguish between personal and non-personal data. Second, given the societal value of privacy as well as the inherently suspect nature of consent in many settings, the limits of consent must be clearly delineated preventing the use of watered down consent to legitimize intrusive processing activities. Third, policy makers should engage with industry not only through lobbyists and trade associations, which pursue a maximalist anti-regulatory agenda, but also with technical experts, system designers, computer scientists and engineers, whose approach towards privacy is more balanced. Fourth, given the global nature of the market for information and ubiquity of cross-border data flows, international enforcement must be coordinated by a central authority, advised by the Article 29 working party; and national privacy regulators should be staffed with not only lawyers but also computer scientists, economists, political scientists, and more, to veer away from their current bureaucratic culture and develop state of the art technological competence.

Additional principles that must be better enforced are privacy by design, requiring comprehensive and iterative privacy impact assessments and implementation of PETs; transparency, providing users with online access to their personal data conveniently, securely, privately, and free of charge, including through the use of "privacy agents;" and accountability, meaning not only passive logging of activity but also the proactive policing and deterrence of abuse within organizations.

Recommendations for Research. First, we suggest research is undertaken to examine the deployment, integration and scaling of PETs in large open-ended networks with decentralized governance and control structures. Research should be multidisciplinary and grounded on empirical data documenting information flows in cloud based applications, ubiquitous computing, and online behavioural targeting. Second, research is needed to promote privacy friendly system engineering, including the transformation of privacy impact assessments from an art into a systematic and transparent process; as well as the integration of PETs through the entire protocol stack via a number of applications, engineering privacy into complete systems and examining methods of evaluation, criteria and metrics. Third, research should seek creative, innovative tools, such as "virtual care-takers," to empower users by enhancing transparency and informational self-determination. Finally, research should explore the "known unknowns", anticipating possible changes to the technological and social environment, such as the impact of quantum computing on cryptographic technologies and the availability of robust face recognition technologies and natural language processing.

Table of Contents

1 Introduction

The principle of informational self-determination is of special importance for online privacy due to the infrastructural and interactive nature of modern online communication and to the options that modern computers offer, even though it is much older than the notion of "Online Privacy". Well before the advent of Web 2.0, the term informational self-determination originated in the context of a German constitutional ruling, related to the 1983 census, making Germany the first country to establish the principle of informational self-determination for its citizens. The German Federal Constitutional Court ruled that[1]: "[...] in the context of modern data processing, the protection of the individual against unlimited collection, storage, use and disclosure of his/her personal data is encompassed by the general personal rights of the [German Constitution]. This basic right warrants in this respect the capacity of the individual to determine in principle the disclosure and use of his/her personal data. Limitations to this informational self-determination are allowed only in case of overriding public interest."

To put it simply, this provision gave individuals the right to determine what personal data is disclosed, to whom, and for what purposes it is used. Informational self-determination also reflects Westin's description of privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [13]. Despite this legal development, the path of the Internet did much to undermine these values and nowadays, individuals have effectively lost control over the collection, disclosure and use of their personal data.

With the evolution and commercialization of the Internet and the advent of Web 2.0, including its search engines and social networks, the environment, in which we need to support online privacy and informational self-determination, became more complex. Some new business models, like ad-financed "free services" for Internet users, rely on a wide-range collection of user data for various purposes, such as marketing of online shops or targeted advertising and include user profiling. It appears that many users of online services are unaware of the implications of this business model. In other contexts, data collected for commercial uses has been later employed for government purposes; this has been possible by the fact that the rules of "free services" place little restriction on reuse of data.

This manifesto is the main result of the Dagstuhl Perspectives Workshop 11061 that took place in February 2011. A primary challenge that it deals with is the correction of power imbalances arising from a loss of informational self-determination, as introduced above. The rest of the document is structured as follows: Section 2 discusses the current state of online privacy, existing technologies to protect privacy, as well as transparency in online information systems, in order for users to have leverage to protect their privacy. Section 3 analyzes what Engineering and Industry can do to improve online privacy. Section 4 gives recommendations on how regulators can improve online privacy, and Section 5 suggests more long-term research topics that are needed to improve online privacy. Finally, the Annex provides further details to these four main chapters.

[1] *BVerfGE 65,1 – Volkszählung*, available in English at http://en.wikipedia.org/wiki/Informational_self-determination.

2 State of the Art

The state of the art in online privacy includes extensive work in a broad spectrum of disciplines. The primary purpose of this section is to set the landscape for the manifesto and provide responses to clusters of specific questions that arise concerning the current state of work in the area of online privacy.

2.1 Understanding Online Privacy

Privacy is subjective, contextual and therefore hard to evaluate. In this regard, one of the main challenges that researchers are currently exploring is linked with the analysis of individual attitudes on privacy. For instance, research has shown that most users of websites with customizable privacy settings, such as Online Social Networks (OSNs), maintain the default permissive settings, which may lead to unwanted privacy outcomes [6]. The explanation to this behaviour is not necessarily that users do not care about their privacy[2]. Instead, existing studies demonstrate an ambivalence of the users' attitudes towards privacy [11, 3]. What makes it more difficult to interpret people's attitude against privacy is that the notion of privacy differs or changes, depending on the culture that individuals are coming from. So, there is still much need for experiments with individuals to allow a broader range of privacy related analysis to be tested and enable a better understanding of people's concerns and the actions they take to address these concerns.

This analysis becomes particularly difficult, since frequently there is no immediate damage for individuals. Even though in some cases, an individual may directly experience an offence, if harassed, manipulated or embarrassed as a result of a prior privacy violation, more frequently the consequences may occur only later or not at all, as for example in third-party tracking of online behaviour for targeted advertisements. Even though tools that try to limit this tracking by third parties exist, with varying degree of effectiveness, these tools do not reveal the impact of processing and usage of personal data from third parties for their own purposes. Evidence also exists to show that third-parties are not only receiving browsing behaviour information about individuals, but are in a position to link this behaviour to personal information via sites such as OSNs [7].

While being subjective and contextual, privacy as a concept has a larger function in society. In this manifesto we discuss privacy under the light of collection and usage practices, but a more general discussion on the basic values that are challenged by the changes brought by a networked society, remains open. What are the effects on our democratic societies of massive-scale data collection, trend prediction and individual targeting? Are people forced into higher conformance? Is conformance pressure affecting the building of political opinions? A scientific approach to these questions cannot rely on the repetition of an old mantra saying that data collection is bad, but will undertake research into the new power relations as they form in the new networked landscape.

[2] On the contrary, several polls and surveys support the opinion that individuals care about their privacy. For example, such a collection for US consumers is presented in http://www.cdt.org/privacy/guide/surveyinfo.php.

2.2 Privacy Technology Landscape and Technology Transfer

In this section, we briefly sketch the current landscape of privacy-enhancing technologies (PETs) and then try to shed light on the reasons for lack of wide-spread adoption of PETs. There is a growing amount of research in the field of PETs, proposing technologies for solving various aspects of the privacy problem; yet PETs are not widely adopted in practice. One cannot expect a simple explanation to this, as online privacy is a complex and interdisciplinary issue. Therefore, we will revisit this issue in the next sections, from the perspective of different disciplines separately with the goal to suggest specific actions. But first, in this section, we set the landscape from a more general view. More specifically, we elaborate on the following reasons, with the understanding that this is an incomplete list of issues:

- current economic environment fosters personal data collection in some business models,
- user awareness of the privacy problems, as well as demand for transparency of data usage and information processing is low,
- today's PETs still lack usability, scalability and portability in many cases,
- regulatory and technical agendas lag behind new data collection practices and data flows,
- integration of many new PETs require costly changes in the existing infrastructure.

After more than 20 years of research in the area of privacy and PETs, there exists a wide variety of mechanisms [4]. Broadly speaking, we could distinguish between opacity tools and tools that enforce other legal privacy principles, such as transparency, security or purpose binding[3]. Opacity tools can be seen as the "classical" PETs, which "hide information", i.e. striving for data minimization and unlinkability. They cover a wide variety of technologies, ranging from cryptographic algorithms and protocols (e.g., [homomorphic] encryption, blind and group signatures, anonymous credentials, oblivious transfer, zero-knowledge proofs etc.) to complex systems like user-centric identity management. Opacity tools can be further characterized depending on whether they focus on data minimization at the network layer or at the application layer. Proposals for achieving sender or recipient anonymity at the network layer comprise protocols such as Chaumian Mixes, DC-Net, etc. At the application layer, a much greater variety of technology proposals exists, such as private information retrieval, privacy preserving data mining (random data perturbation, secure multiparty computation), biometric template protection, location privacy, digital pseudonyms, anonymous digital cash, privacy-preserving value exchange, privacy policies etc.

Transparency-enhancing tools (TETs) belong in the second category of PETs and focus on enforcing transparency, in cases where personal data need to be processed. By transparency we mean the informative representation to the user of the legal and technical aspects of the purpose of data collection, how the personal data flows, where and how long it is stored, what type of controls the user will have after submitting the personal data, who will be able to access the information, etc.

TETs frequently consist of end-user transparency tools and services-side components enabling transparency [10]. The end-user tools include, among other techniques, (1) tools that provide information about the intended collection, storage and/or data processing to the users when personal data are requested from their system (via personalized apps or cookies) and (2) technologies that grant end-users online access to their personal data and/or to information on how their data have been processed and whether this was in line with privacy laws and/or negotiated policies[4].

[3] Purpose binding means that personal data should be relevant to the purposes for which they are to be used and to the extend necessary for those purposes, and should not be usable in other contexts.

[4] A third type of TETs, which has so far only been discussed in the research community, include tools with "counter profiling" capabilities helping a user to "guess" how her data match relevant group profiles,

Examples are the Google Dashboard[5] or the Amazon's Recommendation Service, which grant users online access to their data and allow them to rectify and/or delete their data. However, these are server-side functions and not user-side tools and they usually grant users access only to parts of their data and not to all the data that the respective service processes. An example of user-side transparency enhancing tool is the Data Track developed in the EU project PrimeLife [12], which gives the user an overview of what data have been sent to different data controllers and also makes it possible for a data subject to access her personal data and see information on how her data have been processed and whether this was in line with privacy laws and/or negotiated policies.

In the current state, once the data has been submitted to an online information system, individuals get no knowledge about any further processing. But, even if we assume that the data processing of such complex systems like Facebook, Apple iTunes or Google Search could be transparent to the public, it would be hard or impossible for ordinary individuals to understand what happens with their data. Full transparency of data movements also increases security problems in such environments, if misused with malicious intent. Consequently, this limitation leads to the observation that it is more important for individuals to understand the outcome and implications of data flows in complex online information systems than understanding the full data movements. One technique, among others, that can achieve this kind of transparent outcome-based approach is the creation of ad-preferences by some third-party advertisers, where users are allowed to see the set of outcomes, based on which the data has been forwarded to the third-party (examples here would include Google Ad Categories[6] or the Deutsche Telekom Privacy Gateway for location-based services).

In general, most, if not all, of the proposed PET solutions lag behind the real world situations. They still need to overcome the shortcomings of current approaches, as real world solutions require properties like usability, scalability, efficiency, portability, robustness, preservation of system security, etc. Today, only a patchwork of mechanisms exists, far from a holistic approach to solve the privacy problems. The interaction between these mechanisms and their integration in large scale infrastructures, like the Internet, is not well understood.

Our infrastructures have not been designed with privacy in mind, and they evolve continuously and rapidly integrating new data collection practices and flows. Current privacy mechanisms, not only have difficulties in catching up with these developments, but they also collide with some security and business requirements. A redesign of the system in question can often resolve the collision of interests, but this sometimes requires costly investments.

At the same time, the demand of users for PETs is rather low today. One reason for this is the lack of user awareness with respect to privacy problems, which can be partly attributed to missing transparency of data acquisition and the related information processing, as emphasized above. A second reason lies in the complicated and laborious nature of control imposed on persons, as no legal standards or general consumer protection rules exists. Finally, PETs do not always take into consideration the evolution of privacy models caused by the rapid creation of new technologies and communication models.

Yet another important reason for the lack of adoption of existing PETs lies in some models in data commerce that are based on access to personal data. In the current eco-system, doing nothing about privacy or even aggressively collecting data sometimes pays off, as some companies seem to acquire new clients with new features based on creative data use

which may affect her future opportunities or risks [5].

[5] https://www.google.com/dashboard/

[6] http://www.google.com/ads/preferences

and serendipity. Furthermore, for some players, implementing complex data minimization schemes is costly and time consuming and makes information filtering catered to the end-user much harder, if not impossible. It is important to note here, however, that this approach is not adopted by all industry players. The next section takes a closer look at the problem of adoption of PETs from the industry and suggests addressing specific challenges to overcome this problem.

3 Engineering and Industry Options

Generally speaking, there is a lack of clear incentives for enterprises to manage personal data in a privacy-respecting manner, to design privacy-preserving products, or to make the use of personal data transparent to the data subject[7]. We identify the following root-causes for this current situation:

1. Lack of customer (individuals, business partners) and market demand for privacy respecting ICTs, systems, services and controls (beyond punishments for breaches and other excesses). Usage models for privacy-enhancing technologies cannot currently be targeted to customer demand;
2. Some industry segments' norms, practices and other competitive pressures that favour exploiting personal data in ways contrary to privacy and the spirit of informational self-determination (resulting in diffusion of transparency and accountability);
3. Poor awareness, desire, or authority within some industry segments on the operationalization of privacy (e.g., to integrate existing PETs, to design privacy-respecting technologies and systems, and to establish, measure and evaluate privacy requirements); and
4. Lack of clarity, consistency, and international harmonization in legal requirements governing data privacy within and across jurisdictions (avoided, for example, by migrating data somewhere up in the cloud).

To improve the current environment, we need to increase awareness across users, industry and technologists regarding

- the protection of privacy of users across different media,
- the transparency for processing of personal data,
- the acceptance and incorporation of improved privacy-enhancing technologies by technologists outside of the "privacy community".

To support this goal, we recommend that industry addresses three mid-term challenges, which we discuss in the rest of this section.

3.1 Challenge 1: Promoting Transparency

3.1.1 Transparency-enhancing tools

Transparency enhancing technologies (TETs), which have been developed in the recent years within research projects and by the industry, can help end-users to better understand privacy implications and thus help to increase the user awareness, as we demanded. On the other

[7] We make a disclaimer here that these deficiencies do not apply across the board to all enterprises.

hand, allowing users to control and correct their data processed at services sides will also lead to better data quality for the respective industries.

Challenges for practical TETs that still remain, include the following:

- Providing transparency in a privacy-friendly manner means that TETs should work for pseudonymous users. Industry should consider integration of existing research prototypes and concepts of such privacy-friendly TETs, like the PrimeLife Data Track [8], in real world processes and IT systems.
- The open source development of transparency-enhancing technologies and end-user tools needs to increase, in order to lower market entry costs.
- Better use interfaces for transparency tools in complex environments will need to be created. Also, user-friendly display of data handling practices by "hidden" data processors will play an important role.

3.1.2 Transparency within industrial organizations

Industry needs to foster in-house transparency and awareness for the risks of system-imminent privacy issues in order to effectively enhance privacy in the developed products and services. Principles, such as data minimization and purpose-binding, have to become design principles for processes, IT, service and product design. Industry needs to consistently consider privacy issues, risks, and privacy principles in internal guidelines. These guidelines need to be communicated to engineers, developers, etc. to create a "culture of privacy".

3.2 Challenge 2: Designing and Delivering Privacy Respecting Products to End-users

3.2.1 Demonstrating the power of PETs by blueprints and sample prototypes

When building applications, engineers often lack practical knowledge on incorporating PETs to achieve security and privacy protection. To support engineers in employing privacy-enhancing technologies, we propose to build blueprints and sample prototypes for key scenarios and for different industries. Examples for such prototypes include the following:

- A service that can be delivered to a user on a mobile device, such that the parties involved are able to deliver their parts and are paid for their service, while the user is ensured that every such party receives and stores only minimal data. The user is provided with transparency and control of his own data flows, while data dispersion is minimized, e.g. by attribute-based access-control[8].
- A communication platform that offers its users a convenient communication and collaboration environment with simple and secure user privacy controls to set the audience for certain private data dependent on different social roles and the support of user pseudonyms. The prototype must further demonstrate its economic viability by proper business models that do not conflict privacy requirements.

[8] For example see the ABC4Trust project (https://abc4trust.eu/).

3.2.2 Open or shared-source developments

Sharing source code which can be reused and adopted easily, allows market entrants to lower development costs. One example is the WebKit library[9]. An open-source suite of privacy-enhancing tools can lower market entry costs for companies, which want to offer privacy products and support the emergence of non-commercial software that integrates privacy-protecting functions.

3.3 Challenge 3: Identity Management as a Key Technique

It has been pointed out that identity management is instrumental to the implementation of online privacy management [10]. We also believe that identity management can be used to manage handling of data relevant to satisfy privacy requirements, such as data minimization and transparency.

The scope of identity management is quite broad, comprising authoritative information about legal persons, customer or user relationships, self-issued claims, pseudonyms and anonymous credentials. A minimum of personal data must be conveyed to the service in order to authenticate and authorize the accessing subject.

The service-side storage of personal information without transparent and traceable relation to identities creates fundamental asymmetries in the relationship between the users and the industry and erodes transparency, confidence and trust. Therefore, we propose user-centric identity management systems, which can restore this balance and confidence.

User-centric identity management in this context implies that personal data – even in cases that is created by a service – is always handed back to the user upon completion of the service. If the user desires consistency across service invocation, it is her decision to hand over the data again to the same or another service. This way, individuals can supervise and limit personal data disclosure and exercise rights of access to their data held by third parties.

User-centric identity management allows users to detect any linkages to third parties created from the primary relationship. Enterprise policies and procedures should support user-centric identity management as well, to prevent unwanted linkages and inadvertent disclosures of personal data.

4 Recommendations for Improving Regulations

4.1 Current Regulatory Framework Insufficient

Neither the current European legal framework, nor the US approach toward private sector self-regulation, has been effective for the protection of privacy online, particularly with regard to new business models, such as behavioural targeting, user profiling, social networking and location-based services. Key weaknesses in the EU framework include that: 1) services based predominantly in the US are effectively outside European jurisdiction 2) European users have little choice but to "consent" to companies' terms of use and privacy policies in the absence of alternatives of comparable functionality, 3) the concept of "personal data" is currently the necessary trigger for the applicability of the Data Protection Directive (DPD) and 4) there

[9] http://www.webkit.org/

seems to be too much reliance on ex post securing of data rather than on ex ante elimination of privacy risks through data minimization (for example the recent Art.29 WP Opinion on smart metering[10] omitted entirely any consideration of radical data minimization through cryptographic methods[11]).

4.2 Distinctive European Privacy Values

The European culture of privacy incorporates values of democracy, autonomy and pluralism. The European views on privacy as a societal good and as a factor of public interest lead to a more prominent role of the State in this domain. This conception is not widely shared outside Europe, where the notion of privacy is strongly linked to "the right to be let alone". Consequently, countries such as the US do not necessarily establish the same balance between economical needs and privacy protection.

European approaches protect privacy through consumer protection interventions, instead of reliance upon contract. For instance, it is conceivable that a European national government might prohibit certain extremely privacy-invasive practices, like long-term storage of online search requests for commercial purposes. Unlike contract approaches, such prohibitions can never be waived by acquiring the consent of the users[12]. It has to be recognized that this European view is not shared by legislators in other parts of the world.

4.3 Surveillance Society and Blanket Retention of Data

An important issue of principle for the future Internet of things is the legitimacy of the blanket retention of traffic data (or metadata). In so far as such data relates to individuals, it constitutes a "map of private life" [2]. Case law of the European Court of Human Rights (ECtHR) establishes that "merely" storing such data engages the right to privacy. The troubling exception to this rule is the Data Retention Directive (DRD), requiring storage of certain telecommunications and Internet traffic data. However, the legitimacy of the DRD remains controversial and the concept of indiscriminate continuous retention of data about the entire population has been ruled unconstitutional in its entirety by the Romanian Supreme Court[13], because it "makes the essence of the right disappear".

4.4 A Strong European Legal Instrument Remains Useful

Notwithstanding the fact that a global harmonization in this area is not yet possible, a strong and effective European legal instrument has the potential of having an impact on the global online context. The essential question in conceiving a unique, strong and effective European legal instrument is the goal we want to achieve. The first fundamental objective should be the prevention of privacy-endangering information-processing practices at all levels.

[10] http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp183_en.pdf
[11] http://research.microsoft.com/en-us/projects/privacy_in_metering/
[12] Art. 8.2, a) of the Directive provides that in certain cases the prohibition to process sensitive personal data may not be lifted by the data subject's giving his consent
[13] http://www.legi-internet.ro/english/jurisprudenta-it-romania/decizii-it/romanian-constitutional-court-decision-regarding-data-retention.html

This instrument must guarantee real protection against actual and potential risks taking into account technological developments and not merely offer formalistic legal assurances. Therefore, it is crucial to take maximum advantage of the opportunity offered by the current proposed review of the European directive to maximize its impact.

The path of least resistance is to make incremental changes in the existing directive. However this may not succeed in rectifying some serious conceptual defects. There is a risk that the review of the directive will merely result in a slightly better but still largely ineffective regulatory solution.

Moreover recent results in the field of de-anonymization [1, 8] suggest that some data may be impossible to anonymize (e.g. social networks) and it is difficult to predict the vulnerabilities for and consequences of re-identification when contemplating the release of pseudonymous data [9]. It seems that a better approach would be to make the application of the legal framework dependent on an evaluation of the actual and potential privacy risks related to any data processing.

4.5 Consent Must not Overrule Everything

From a European perspective, explicitly given user consent should not be accepted as a waiver for privacy-intrusive online practices. A European legal instrument should clearly emphasize that individual privacy is not purely a matter of the individual concerned, but of the society as a whole. Moreover in many situations the voluntariness of the user's consent can be put into question because of the lack of reasonable alternatives for commonplace services, which meaningfully adhere to European Data Protection. US consumer protection law recognizes many situations where consumer consent cannot waive risks. This policy has not yet mobilized into the law of privacy. Consent should expire, according to the scope and extent of processing. When asking for consent, data controllers should make explicit what is revocable and what is irrevocable and how it is possible to revoke that consent. Legislation may prohibit processing that would have irrevocable consequences. Otherwise a European-wide warning system for specific risks or breaches, similar to governmental travel warnings for dangerous regions, could be advisable.

4.6 Effective Implementation and Enforcement Is Crucial

Privacy lobbying has been concentrated amongst a few law firms and trade associations. These interests pursue a maximally anti-regulatory agenda, even in situations where their clients would admit that they could comply with privacy rules proposed. Retention of information in network advertising companies is a key example – while on the lobbying level, it is often argued that this data must be retained for very long periods of time, the engineers at network advertising companies will admit that data becomes less valuable after a very short period of time, and is often unused for ad targeting purposes within a few months. However, since policy makers rarely engage beyond trade associations, they get a jaundiced view of the actual requirements that businesses have for data. Too often, regulators pose technical and implementation questions to attorneys rather than the technical experts who design systems. We recommend that to the extent practicable, regulators invite relevant actors, rather than their representatives, to public fora, consultations, and other consensus-building events around privacy.

The US Federal Trade Commission recently employed two technical experts on a short term basis to assist in the evaluation of technologies, and the agency has, also on a short term basis, employed a senior computer scientist to assist with policy analysis. We believe that technical expertise is increasingly necessary for policy makers and regulators, and we recommend they look more to in-house technical expertise to assist in their rule-making and investigations.

So far the DPD agenda for reform has not sufficiently considered basic limitations on the effectiveness of enforcement when 27 national authorities must reach consensus. Competence for international enforcement actions should be given to a central authority, advised by the Art. 29 WP, leaving national DPAs better able to focus on national-level issues. Current technological knowledge must be an indispensable part of the professional competence of DPA administrations. But at most a few percent of these officials have any relevant postgraduate scientific competence, and overwhelmingly DPAs have an irredeemably bureaucratic culture. A complete renewal of these institutions is necessary. A minimum of one third of staff should be experts in the computer science of privacy, as well as first rate talent in law, economics, political sciences, sociology and philosophy. Access to justice through privacy litigation is out of reach for most people today. Data Protection Authorities should evolve into Information Privacy Ombudsman (IPO), explicitly acting to uphold privacy rights. IPOs must expect to show intellectual excellence in every relevant field, and earn their authority through merit, or be dissolved.

4.7 Privacy by Design

A privacy by design approach can be mandated (or otherwise encouraged) by legal or regulatory provisions, if scientific discoveries demonstrate that a service can be offered practicably in a more privacy protecting way. This could involve, for example, requiring that comprehensive and iterative privacy risk and impact assessments be carried out and that state-of-the-art privacy technologies be adopted.

4.8 Transparency for Data Subjects

In order to give meaningful effect to the right to informational self-determination, it is clearly necessary for users to have the possibility of "information self-awareness". Its importance has been emphasized in all previous sections already, together with the limitation in the corresponding transparency enhancing tools. Because invoking existing "subject access" rights is cumbersome, slow, and often incomplete, these rights should be strengthened to provide a right to comprehensive online access to data about an individual's use of an online service, conveniently, securely, privately, and free of charge. This should henceforth be regarded as an indispensable aspect of the human right to privacy in the 21st century. To provide such data genuinely in "intelligible form", more disclosure of algorithms will be necessary (also for automated processing or anonymization), whether these act on personal data or can affect the individual through the application of statistical data models ("red-lining").

Consumers' ability to designate "privacy agents" as proxies for exercise of their rights should be recognized by firms and governments. Consumer privacy agents are now a viable business, but they are frustrated by organizations that question the authority of the agent to act for the consumer, and by systems that attempt to obfuscate the invocation of rights

to opt out or gain access to personal information. Collective negotiation through "privacy unions" potentially is also an important democratic mode of political expression and must be protected from harassing lawsuits.

4.9 Transparency by Design for Auditors

A further important aspect of transparency is the need to design mechanisms, which allow the flows of data in a system to be documented and verified by internal and external auditors, including algorithms used to perform profiling and social sorting.

4.10 Accountability

A core reason for defining the notion of a data controller was to assign clear legal responsibility for compliance. However the complex mesh of legal relationships, which have since arisen, often do not allow a controller to guarantee any effective operational performance of such obligations. Mere logging of system activity is insufficient to counter insider threats – active policing of such logs is required. "The Principle of Accountability" should be understood to mean not merely a passive ability to provide an account, but the creation of an effective deterrent against abuse. Moreover the creation of detailed logs about data subjects itself is prejudicial to privacy and therefore all logging activity must be assessed from the point of view of the interests of privacy protection as well as justifiable security goals.

5 Recommendations for Research

The up-scaling of privacy-enhancing technology to larger systems and its integration with existing systems fails, mainly because systems aspects and the related interdisciplinary issues are not taken into account. In this section we address this by recommending research into:

- scalability and integration on a large scale,
- technologies to support privacy-enhanced systems engineering and
- research to enable systems for individual-level privacy protection.

Finally, we recommend research into the "known unknowns" of the technological and societal environment that privacy technology exists in.

5.1 Web-Scale Integration, Deployment and Infrastructures

As discussed in Section 2, over the last 20 years, the privacy community has developed a large pool of tools and primitives. Yet, we do not see deployment in large scale infrastructures such as the Internet and the World Wide Web, and the interaction between individual technological tools is ill-understood.

To address this limitation, we recommend research and experimentation focusing on integration and deployment: How do privacy-enhancing technologies scale, in terms of deployment on large and open-ended networks with decentralized control and governance structures, large populations, and qualitatively different scales of data collection and processing? Research instruments to address this question may include: mathematical modelling of interdependencies

and integration effects, test beds and demonstrators that enable research and demonstration, as well as private-public partnerships focusing on adoption and deployment of experimental technologies. This approach can also foster infrastructure and product development through pre-commercial procurement. Also other incentives for deployment should be analysed. Specific fields, in which these approaches should be tried, include (but are not limited to):

- Privacy-enhanced identity management infrastructures;
- Techniques for minimal data disclosure;
- Data governance and policy language approaches;
- Accountability in data disclosure and processing, including transparency and auditability, as well as real-time detection and investigation of privacy breaches and data abuse;
- Technological approaches that help to reconcile privacy interests and business models;
- Privacy-protected communications.

Research approaches towards these questions need to be multidisciplinary. Relevant disciplines include economics, psychology, sociology, business administration, law and political studies, as well as various fields within the discipline of computer science (ref. Annex for examples).

Within this research agenda of understanding large-scale, system-level interactions of technological and social phenomena, the empirical data about the evolution of data collection practices and data flows on the Internet and the Web become a critical asset. Relevant data flows include data treated by cloud-based applications, sensors that interact with the physical environment and users' behaviour as they interact with online services. Regulatory and technical agendas need to be informed by empirical understanding of these data flows. We recommend creating an observatory for these flows and interdependencies, taking existing research work to a systematic new level, and creating the basis for more rigorous analysis of the technical *status quo*.

5.2 Towards Privacy-friendly System Engineering

The privacy enhancing building blocks available today need to be integrated into an overall privacy engineering environment, so as to enable adequate evolution of privacy concepts and the required privacy friendly technology and systems in the future. Requirements for privacy need to be analysed, especially when new systems are coming up that can have a negative impact on privacy. Consequently, a Privacy Impact Assessment (PIA) is needed. PIA requires research about methods to develop it from an art into a systematic and transparent process, which also allows the comparison of different development alternatives and their privacy impact. When privacy enhancing technologies are deployed on servers, network infrastructures and devices, multiple independently developed technologies are brought together. The way in which the privacy properties of these modules interact with each other and with the surrounding system through the entire protocol stack and via a number of applications, is often ill-understood. For example, integration of different systems can lead to surprising effects (e.g., unwanted data flows) resulting in the violation of privacy policies or assumptions implicit in privacy technologies.

Further research in the composability (e.g. considering the current research in "differential privacy") of these tools and systems is needed to develop suitable best practices. First, this research will contribute to the development of methodologies and guidelines that contribute to the ability to engineer practical privacy in complete systems with the help of a multi-stakeholder community. In order to evaluate the privacy assurances given by these approaches,

further research into evaluation methods, criteria and metrics is necessary. Second, the research direction proposed here will also facilitate re-engineering processes of deployed systems to take privacy aspects into account.

5.3 Individual Protection

In the area of individual protection, we can frame many privacy concerns in terms of power imbalances between data subjects and data processors, and we can frame privacy enhancing technologies as tools to assure or restore an adequate power balance. Research should continue towards tools that assist individuals' informational self-determination and permit users to learn, e.g. when they share data and may not know about the consequences. Those tools should leverage progress in machine learning. We could imagine relevant tools ("care-takers"), as for example:

- advisers, helping users before they engage in privacy-relevant activities online,
- bodyguards, assisting users as they act online,
- litigators that might be able to help users reconcile breaches of their expectations afterwards.

A crucial element of individual protection and autonomy is individuals' ability to understand and act on their context, assisted by appropriate and intuitive tools. Related to the observations on scaling in the previous sections, research should address how the implications of massive-scale data collection and processing can be made comprehensible and practically manageable for individuals. Research topics here range from usability of technology to the development of philosophical and psychological models for the consequences of data processing. Additionally, empirical experiments should be designed to better understand what users' privacy interests and assumptions are, and to what extent they are (or are not) able to take action using the tools available today.

5.4 Known Unknowns: Possible Changes to the Technological and Societal Environment

Research agendas in privacy need to address the evolution of underlying technologies and the surrounding societal and business landscape, in particular in cases where that evolution might create qualitative changes to the privacy landscape. Cryptographic technologies build the foundation for controlling access to data and are also used as a primitive in many privacy enhancing tools. When there are risks that cannot be articulated, such as whether Quantum Computing will lead to negative consequences for cryptographic primitives as available today, the question is raised whether we are prepared for the consequences of changing the underlying assumptions that today's technology is built on.

Other examples can be found in the rapid advance in the availability and quality of face recognition technology and natural language processing. Some of these progresses are further aided by the increasing availability of large data sets. The implications of these effects are likely compounded by the availability of more powerful mobile devices. Finally, we should also include unpredicted events and disasters to the factors that may change societal attitudes toward privacy in the future. More generally, blue-sky research should be undertaken to identify and prepare for changes in underlying technologies and broader science and societal context that we might not foresee today.

Annex A: Examples for research approaches on new privacy technologies

As mentioned in Section 5, research approaches on new privacy technologies should be multidisciplinary. Examples for the principles and use cases to check for multidisciplinary questions include:

- Cryptographic feasibility
- Scalability
- Usability/acceptability
- Regulatory
- Business models. Is the new technology approach compatible with the future?

For example, the above principles should be checked on the following upcoming technologies in privacy preserving or privacy-friendly distributed systems and activities:

- Privacy preserving distributed data mining and processing. In this category falls for example the application of Peer-to-Peer architectures on online social networks, in order to avoid control over user data and behaviour by a single entity, such as the service provider. Another example could be technologies targeting the protection against Spam or DDoS, in existing anonymous transport networks.
- Privacy preserving distributed data collection. In this category falls for example the sensing and collection of environmental data that are connected with the context of specific people (e.g., their location). This is the case, when sensors embedded in mobile devices are used for such a collection. While this is an upcoming technology, the privacy implications have been hardly studied.

Annex B: Participants and Observers

Participants

- Andreas Albers
Goethe University
Frankfurt, DE
- Caspar Bowden
Microsoft WW Technology
Office, GB
- Sonja Buchegger
KTH Stockholm, SE
- Johannes A. Buchmann
TU Darmstadt, DE
- Jacques Bus
Digitrust.EU – Brussels, BE
- Jan Camenisch
IBM Research – Zürich, CH
- Fred Carter
IPC – Toronto, CA
- Ingo Dahm
Deutsche Telekom AG, DE
- Claudia Diaz
K.U. Leuven, BE
- Jos Dumortier
K.U. Leuven, BE
- Simone Fischer-Hübner
Karlstad University, SE
- Dieter Gollmann
TU Hamburg-Harburg, DE
- Marit Hansen
ULD SH – Kiel, DE
- Jörg Heuer
Deutsche Telekom AG
Laboratories, DE
- Stefan Köpsell
TU Dresden, DE
- Ioannis Krontiris
Goethe University Frankfurt, DE
- Michael Marhöfer
Nokia Siemens Networks –
München, DE
- Andreas Poller
Fraunhofer SIT – Darmstadt, DE
- Kai Rannenberg
Goethe University Frankfurt, DE
- Thomas L. Roessler
W3C, FR
- Kazue Sako
NEC, JP
- Omer Tene
Israeli College of Management
School of Law, IL
- Hannes Tschofenig
Nokia Siemens Networks –
Espoo, FI
- Claire Vishik
Intel – London, GB
- Michael Waidner
TU Darmstadt, DE
- Rigo Wenning
W3C / ERCIM, FR
- Alma Whitten
Google London, GB
- Craig E. Wills
Worcester Polytechnic Inst., US
- Sven Wohlgemuth
National Institute of Informatics –
Tokyo, JP

Observer

- Jesus Villasante
European Commission, BE

References

1 Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*, pages 181–190, Banff, Alberta, Canada, 2007.

2 C. Bowden. Closed circuit television for inside your head: Blanket traffic data retention and the emergency anti-terrorism legislation. *Computer and Telecommunications Law Review*, March 2002.

3 L. Brandimarte, A. Acquisti, and G. Loewenstein. Privacy concerns and information disclosure: An illusion of control hypothesis. In *Proceeding of the 9th Workshop on the Economics of Information Security (WEIS 2010)*, June 2010.

4 G. Danezis and S. Gürses. A critical review of 10 years of privacy technology. In *Proceedings of Surveillance Cultures: A Global Surveillance Society?*, London, UK, April 2010.

5 Mireille Hildebrandt. Behavioural Biometric Profiling and Transparency Enhancing Tools. Technical Report FIDIS Deliverable D7.12, March 2009.

6 Balachander Krishnamurthy and Craig E. Wills. Characterizing privacy in online social networks. In *Proceedings of the 1st Workshop on Online Social Networks (WOSN '08)*, pages 37–42, 2008.

7 Balachander Krishnamurthy and Craig E. Wills. On the leakage of personally identifiable information via online social networks. *ACM SIGCOMM Computer Communication Review*, 40:112–117, January 2010.

8 Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *Proceeding of the IEEE Symposium on Security and Privacy*, pages 111–125, 2008.

9 P. Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57:1701, 2010.

10 K. Rannenberg, D. Royer, and A. Deuker, editors. *The Future of Identity in the Information Society – Challenges and Opportunities*, page 507. Springer, 2009.

11 J. Turow, C. J. Hoofnagle, D. K. Mulligan, N. Good, and J. Grossklags. The federal trade commission and consumer privacy in the coming decade. *I/S: A Journal of Law & Policy for the Information Society*, (723), 2007–08.

12 J. E. Wästlund and S. Fischer-Hübner. End User Transparency Tools: UI Prototypes. Technical Report PrimeLife Deliverable D4.2.2, June 2010.

13 A. Westin. *Privacy and freedom*. New York: Atheneum, 1970.