IDIS (2010) 3:257–265
DOI 10.1007/s12394-010-0045-z

Built-in privacy—no panacea, but a necessary condition for effective privacy protection

Alexander Dix

Received: 1 November 2009 / Accepted: 8 February 2010 / Published online: 13 April 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Built-in privacy has for too long been neglected by regulators. They have concentrated on reacting to violations of rules. Even imposing severe fines will however not address the basic issue that preventative privacy protection is much more meaningful. The paper discusses this in the context of the International Working Group on Data Protection in Telecommunications ("Berlin Group") which has published numerous recommendations on privacy-compliant design of technical innovations. Social network services, road pricing schemes, and the distribution of digital media content have figured prominently in the group's latest working papers. More recently, a judgment of the European Court of Human Rights has thrown light on weaknesses in the protection of patients' data in hospitals that requires urgent action by designers of IT systems. Built-in privacy is no magic button, no panacea, but it has turned out to be a necessary condition for meaningful privacy protection.

Keywords Social networks · Road pricing · Locational privacy · Distribution of digital media content · Hospital IT systems · Web-based telemedicine · Electronic health records · International Working Group on Data Protection in Telecommunications · "Berlin Group"

"Locking the stable door after the horse has bolted"—this has been for too long a rather accurate description of what Privacy and Data Protection Commissioners and other regulators have been doing. They have tried to enforce rules and laws and to assist/support data subjects after their rights had been violated. Many regulators are lawyers, and this is what lawyers are trained to do and consequently used to doing, but it has turned out to be insufficient to protect privacy by relying solely on reactions and sanctions for not abiding by the rules of data/privacy protection. Very often data controllers claim to be unable to comply with the rules because there is no

A. Dix (✉)
Berlin Commissioner for Data Protection and Freedom of Information, Berlin, Germany
e-mail: dix@privacy.de

technology available to do so. Or they simply pay the fine and do not change their practices hoping that this will go unnoticed at least for some time.

What is necessary is a more preventative, systemic approach to the issue. The International Working Group on Data Protection in Telecommunications *(Berlin Group)* has on numerous occasions made proposals for building minimum privacy features into services, platforms and terminals.[1] This is highlighted here in the context of 1) social networks, 2) road pricing, 3) the distribution of digital media content, and 4) electronic health records, especially in hospitals.

Social networks

The working group was among the first to address privacy issues in social networks. In the Rome Memorandum of March 2008,[2] the Group described in detail the privacy risks in social communities such as Facebook and offered guidelines on how to tackle them. Social communities offer an "illusion of intimacy." It is therefore vital to look at the way in which users have to identify themselves when creating their profiles. Very often users are required or at least encouraged to use their real names. Under German law, there has to be at least an option to use such network services pseudonymously. The Working Group called on regulators to introduce such an option where it does not already exist. This is the classical "lock the stable door" approach.

However, the Group did not stop there. They also called on service providers to *encourage* the use of pseudonymous profiles. Neither Facebook nor their biggest German competitor StudiVZ has so far followed this advice. StudiVZ tolerates the use of pseudonyms but discourages it openly on their website. The platform providers argue that pseudonymous use lowers the quality of communications in the network and makes it difficult to identify cyber-stalkers. However, a network with built-in privacy in this respect would require real names at registration but not in the profiles visible to all "friends" where pseudonyms (nicknames) should be encouraged.

Second, one has to look at the default settings in a social community. In the big networks such as Facebook and StudiVZ, the default is that any newly created profile is visible to all members of the community until and unless the user restricts visibility of his profile to friends he has explicitly chosen. Here, the Berlin Group pointed to the well-known fact that only a minority of users signing up to a service will make any changes in the default settings—including privacy settings. As Ann Cavoukian's Privacy by Design Principle No. 2 rightly stresses: "If we can all be certain of one thing – the default rules !" The Berlin Group uses similar language: "Privacy-friendly default settings play a key role in protecting user privacy." Therefore, privacy has to be built into the system, by default. This means that the default settings should be privacy-friendly and restrictive, and this was the

[1] See: International Documents on Data Protection in Telecommunications and Media 1983–2006, available at http://www.datenschutz-berlin.de/attachments/334/IWGDPT_WP_brochure.pdf?1193754976

[2] http://www.datenschutz-berlin.de/attachments/461/WP_social_network_services.pdf?1208438491; see also on this issue Working Paper 163 by the Art. 29 Working Party, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2009/wp163_en.pdf

recommendation by the working group. It should be up to the user after registration to define and possibly enlarge the number of friends that he thinks should have access to his data. This has again met with opposition from the network providers since they take the view that it would run counter to their business model if users could not communicate with all members of the community from the start. So far, the big providers are rejecting the idea of privacy by default in this respect.

Maybe a recent incident in Germany makes them think again: The company offering the StudiVZ community promises their user on the Website: "Your data are safe in this community." Two users of StudiVZ—having registered on the platform—used a "crawler" to automatically collect as many profiles as they could (names, pictures, data on interests, hobbies, etc., but no e-mail addresses or phone numbers) and exported them to the open Internet. These were all profiles accessible to all members of the community by default. Technical barriers in place ("captchas") were easily circumvented. All in all, more than 1 million datasets where thus harvested. One of the users had benevolent motives—he simply wanted to highlight this weakness—while the other tried to blackmail and was arrested. This would not have had such a devastating effect on the trustworthiness of the company had they followed our advice that the default settings should be privacy-friendly and restrictive, leaving it to the users to open their profiles for designated friends. Also, the use of pseudonyms would have mitigated considerably the effect of this data harvesting. Since the company has adopted privacy-friendly default settings competitors such as Facebook still have to learn this lesson.

The Rome Memorandum also called for the creation of means allowing for user control over third party use of profile data, which is vital especially to addressing risks of identity theft. Although there are at present only limited means to control information once it has been published online, the Berlin Group called on service providers to strengthen research activities in this domain.

Existing and promising approaches include research on the "semantic" or "policy-aware web," encrypting user profiles, decentralized storage of user profiles (e.g., with users themselves), the use of watermarking technologies for photos, the use of graphics instead of text for displaying information, and—particularly important—the introduction of an expiration date to be set by users for their own profile data. Providers should also take effective measures to prevent spidering, bulk downloads (or bulk harvesting) of profile data. Specifically, user data should only be "crawled" by (external) search engines if a user has given his explicit, prior, and informed consent. The Article 29 Working Party of European Data Protection Authorities went into greater detail in their recommendations regarding user-mediated third party access.[3] When offering "Application Programming Interfaces," social network services should let the user choose a level of access for third parties that is only just sufficient to perform a certain task (e.g., install a game in the profile, read and post messages to the network from a mobile phone, etc.). Some of these questions are yet unsolved, but the preventative protection of privacy greatly depends on how soon positive answers to them can be found.

[3] See note 2 above: Working Paper 163, p. 9; see also the Canadian Privacy Commissioner's findings in CIPPIC v. Facebook Inc. http://www.priv.gc.ca/cf-dc/2009/2009_008_0716_e.pdf

This is all the more important since social networks are the platforms where people increasingly define their digital identity. This identity is most vulnerable in an environment which claims to provide for high minimum standard (default) privacy protection without in fact doing so.

Road pricing

The second example to illustrate the importance of built-in privacy is road pricing. People using vehicles normally travel anonymously, i.e., without being registered by any authority or private entity as long as they follow the rules (a possible exception being modern fleet management systems where the employer tracks the movements of trucks and drivers). Indeed, the right to move freely (without being registered routinely) in public spaces (the *right to locational privacy*) applies to drivers as to pedestrians alike. With the advent of electronic toll systems for road pricing, however, this could change fundamentally, but here, again, much depends on the design of the system.

The Berlin Group adopted the Sofia Memorandum[4] in March 2009, which contains a number of findings and recommendations addressing this issue:

Privacy issues only arise where personal data are processed by electronic toll systems. This excludes vignettes, anonymous tags and beacons, and non free-flow toll booth-based systems where you have to stop your car and pay cash. The desired outcome of an electronic road pricing scheme is the ability to charge for actual use (pay as you go, i.e., the more you drive, the more you pay) depending on the time of journey (e.g., less during off-peak periods) and with a varying tariff according to the chosen road. Traffic flow may even be enhanced because drivers do not have to stop any longer at toll booths in such schemes, which is desirable from an ecological point of view.

The privacy-related risks caused by road pricing schemes are based on the fact that they bring together location data, identification data, and charging data. This would entail knowing who was where at what time and charging them for it. In order to enable the "pay as you go" principle in free-flow traffic and also to achieve interoperability between the systems, it is clear that road pricing schemes have the potential to become infrastructures of massive surveillance with regard to the movements of individuals (vehicle owners, drivers, employees, etc.). It is easy to imagine the considerable value of a centralized database of drivers' movement data and various function-creep scenarios where data might be used for purposes other than those they were originally collected for (i.e., road pricing). A number of Information Commissioners and Data Protection Commissioners have already issued detailed opinions and guidance on privacy protection in electronic road pricing schemes (e.g., Ontario/Canada, the Netherlands, Victoria/Australia, Norway, and Slovenia[5]).

Indeed, discussions in a number of countries such as Germany have shown that it is difficult - if not impossible - to prevent the secondary use of such data for purposes such as crime prevention and law enforcement once the data have been

[4] http://www.datenschutz-berlin.de/attachments/596/Roadpricing_english.pdf?1245751410
[5] See Sofia Memorandum, p. 2, for detailed references.

collected in a personalized way and stored centrally. There may be regulatory provisions to prevent such secondary use, but these provisions are subject to political debate and change by the regulators. Therefore, it is vital to install from the start a technical infrastructure which strictly limits the processing of personal data and especially excludes the storage of detailed movement data.

Two mainstream technologies are currently being used or envisaged for road pricing: short range communications (tag-beacon systems) and global navigation satellite systems (GPS, GLONASS, Galileo). The former is still more widespread but seems to be losing ground to the latter for practical reasons: Large-scale implementations are not possible or too costly on a tag-beacon basis because of the infrastructure needed. Topographic and geographic conditions also limit the possibilities of terrestrial systems. Satellite-based systems are likely to be rolled out in the future. They do not necessarily lead to satellite-based global databases since all known systems (GPS, GLONASS, and the future Galileo system) are based on passive receivers, which only calculate the location of the vehicle using satellite data. There is no "up-link" communication back from the car to the satellite. However, an all encompassing database of location and identification data could be built "on the ground" in the control centers. This is the main concern of the International Working Group.

Often road pricing schemes are being compared with mobile communications networks. This is incorrect since whereas the phone user is always in a position to switch off his phone, a driver cannot opt-out of a road pricing scheme. Otherwise, it would be easy to evade costs on payable roads. It is all more important to deal with the privacy implications of road pricing schemes because they—unlike mobile telephony—will not and cannot allow for a simple switch-off. Where road pricing is in place, it cannot be avoided (unless you stop driving a vehicle on a payable road).

There are two principal models for road pricing schemes: the thin client approach and the smart client approach. Hybrid systems also exist but are not dealt with in detail here.[6] Basically, the thin client approach using on-board units (OBUs) is least favorable in terms of privacy protection. OBUs transmit all data concerning the location of the vehicle, the distance traveled, and the roads used to a control center where the toll charge is then calculated. Thin clients do not allow for user control of the data.

On the other hand, smart clients can preserve the anonymity of the driver because they will do all the data collection and calculation of charges under the control of the driver. The total sum of the charge due is then transmitted to the control center. Only in the case of irregularities will the driver have to identify himself. The control center does not receive any data on the location of the vehicle; it only checks if the smart client device operates correctly. Obviously, the system requires technical measures to protect the smart clients from fraudulent manipulation. Neither the thin nor the smart client should allow for being switched off while the driver is on a payable road.

Enforcement has to take place in a privacy-compliant manner as well. It is not acceptable that all participants in a road pricing scheme are identified and tracked in

[6] See the Sofia Memorandum for a detailed discussion of these models.

order to identify drivers who are breaking the rules (and not paying the charges). Only in the event that there is evidence or at least a concrete suspicion that such a case has happened can the identity of the driver be ascertained by the system. The principle of proportionality would require that, as a first step, it is established whether the toll system device is present and functioning faultlessly in the vehicle. If that is the case, then the control unit should take no further steps to identify either the vehicle or the driver. Only if the absence or malfunctioning of the device is detected should the authorized body proceed with the identification of the driver, e.g., via plate number recognition. Access to the personal data stored under the driver's control should only be allowed in exceptional circumstances (e.g., if the device has been tampered with), and a complete audit trail should be made to prevent unauthorized access to the data in the device.

Data subjects' rights should be respected also in cases where there is a dispute about charges. Access to the driver's personal data should only be allowed (and technically possible) at his own explicit request. Road pricing systems can and should be designed in such a way that the detailed trip data are fully and permanently deleted from the system after the charges have been settled and any period for disputing the charges has elapsed (as is the case, for example, in the London Congestion Charge System).

In the Sofia Memorandum, the Working Group takes the view that centralized processing of personal data (especially detailed traveling routes) for the purpose of road pricing is not necessary and therefore unjustified. Strong privacy protection can and should be built into road pricing systems from the start so that the information transmitted to the control center would only relate to the bulk charges and would not include detailed data on time and place of travel. By adopting such a design principle, road pricing schemes would offer better privacy protection than existing credit card and mobile phone systems, where the provider knows the individual purchases made as well as the phone numbers called and the rather precise locations. This view has been supported by the US National Surface Transportation Infrastructure Financing Commission in a recent report.[7] In order to prevent function-creep effects, the detailed trip data stored under the individual's control should be deleted from the system immediately after the charges have been settled. The driver's anonymity should be preserved throughout the system unless the driver has violated the rules of the road pricing scheme. Processing of personal (including location) data for other purposes (e.g., pay-as-you-drive insurance) should only be possible with the individual's clear and unambiguous consent.

Road pricing schemes certainly require the processing of personal data, but they do not require the centralized processing and storing of such data (as long as no offense has been committed). Neither do they require disproportionate processing of and access to personal data nor ubiquitous surveillance. The fundamental principles of personal data protection and locational privacy strive for maintaining the anonymity of the driver. Technology can and should be used to support these principles.

[7] Paying Our Way, a New Framework for Transportation Finance, February 24, 2009, <http://www.itif.org/index.php?id=227>

Distribution of digital media content

A third example of privacy by design relates to the area of media distribution and consumption. The Berlin Group has addressed this area in an earlier working paper (adopted in September 2007) which dealt with "Privacy Issues in the Distribution of Digital Media Content and Digital Television."[8]

The increasing delivery of television and other audio and video services as digital signals over broadband data networks significantly changes the patterns of media production, distribution, and consumption. It involves a convergence of networks as well as the appearance of an increasing number of static or portable media devices—the divergence of devices. The introduction of new navigation paradigms (e.g., video search engines, peer-to-peer distribution) allows for access to an explosively growing amount of available video media (divergence of content). At the same time, this development allows for the collection and processing of personal data gathered from different sources, for example in multiple-play (e.g. triple-, quadruple-play) services.

This new personalized approach to television—providing anybody with anything, anytime, anywhere, and on any device—allows for new services such as T-Commerce (television-based commerce), video-on-demand, home-banking, and distance learning. At the same time, it introduces new threats to the protection of the privacy of viewers. The new digital interactive television systems are very often based on a sealed "black-box" controlled by companies giving the user little or no control. It is difficult, if not impossible, even for advanced users to identify what the system is doing.

When the television service is offered by an Internet Service Provider within a triple- or quadruple-play service, the TV program is either viewed on a TV set or a PC. In both cases, the channel may be retrieved on demand (when the user selects the channel), and the provider can therefore identify precisely which user is watching a specific program at any given moment. The same applies in the case of WebTV where the content is provided via a website, and the video stream is downloaded on demand. Personal data can in this case be collected by the website operator and also by the Internet Service Provider. Finally, some systems even allow individual users of a bidirectional service to upload their own content on to a video-on-demand platform (where it can be accessed by other users), or users may also broadcast their own live video streams on a dedicated video-on-demand TV channel.

The Berlin Working Group recalls in this respect that the possibility of anonymous use of television content must be maintained in the digital age. Anonymous payment methods (e.g. using prepaid cards) should be offered at least as an option at no additional cost.

Incidentally, this is a requirement which applies not only to digital television but to any paid content distributed online. There have recently been several announcements by media corporations that they will stop distributing content on the Internet for free and introduce some form of charges instead for economic reasons.

Information systems set up to deliver digital television (and other digital content) have to be designed, built, and configured to promote and assure anonymity or at

[8] http://www.datenschutz-berlin.de/attachments/349/digit_en.pdf?1201702193

least minimization of the use of personal data. To this end, a privacy impact assessment should be performed in advance.

If personal data are collected, it may only be for legitimate reasons and to a legitimate extent. The amount of data has to be relevant and not excessive in respect of the purpose to be achieved. Allowing individuals to choose content should not inevitably require them to be identified.

Digital television providers should notify viewers beforehand about the exact purposes of the personal data collection and processing, the type of data collected as well as the place and the duration of storage.

The processing of viewers' profiles should require their informed prior consent ("opt in"). Specifically, the communication of viewers' data or profiles by digital television providers to a third party (e.g., for marketing purposes) may only be carried out with the free and informed consent of the data subject. Viewers should have the right to withdraw their consent at any time with effect for the future.

Viewers should have the right to access, inspect, and correct if necessary, preferably free of charge, all their personal data, including their profiles stored by digital television providers.

There is at least one recent example of German platform provider for high definition, satellite-based television that changed its technical distribution model from personalized registration to anonymous registration and use on a prepaid basis after having consulted with the German Data Protection Authorities.[9] This shows that privacy by design can support a realistic business model.

Electronic health records in hospitals

The final example concerns the processing of the most intimate and sensitive kind of personal data: health records. The Berlin Working Group addressed issues of web-based telemedicine as early as in 2002[10] and again in 2006,[11] recommending a number of measures that have to be taken before health records can be made available online.

However, a major issue with regard to the protection of patients' data appears to be largely unresolved even in offline settings. The European Court of Human Rights in July 2008 awarded pecuniary as well as nonpecuniary damages to a patient in a Finnish hospital.[12]

The patient had originally worked as a nurse on a fixed term contract in a hospital where she was treated in another part of this hospital for an infectious disease. On this occasion, she had been diagnosed as HIV-positive. Later, she began to suspect that her colleagues were aware of her illness. The entire hospital staff at that time had free access to the patient register which contained information on patients' diagnoses. After 3 years, the patient's contract of employment with the hospital was not renewed, and she changed her job. When she requested formally an examination

[9] HD+GmbH, Unterföhring (subsidiary of ASTRA, Luxembourg)
[10] http://www.datenschutz-berlin.de/attachments/184/wpmed_en.pdf?1177588172
[11] http://www.datenschutz-berlin.de/attachments/224/WP_HealthRecords_en.pdf?1200656797
[12] Case of I v. Finland, Judgment of the European Court of Human Rights of 17 July 2008

as to who had accessed her confidential patient record, it was revealed that although access had been restricted in the meantime, it was not possible to find out who, if anyone, had accessed her patient record. The data system revealed only the five most recent consultations (by working unit and not by person), and even this information was deleted once the file was returned to the archives.

In a remarkable judgment, the European Court ruled that Finland was in breach of Article 8 of the European Convention of Human Rights (human right to private life), in which the Finnish Government had not secured in the public hospital in question a practical and effective protection to exclude any possibility of unauthorized access to patients' records since it was impossible to retroactively trace the use of these records.

Obviously, this judgment will have far-reaching consequences, which have probably not yet been fully realized in Europe or elsewhere. At least two German supervisory authorities have found within their jurisdiction that this problem is certainly not limited to Finland. There are a number of large hospitals probably throughout Germany where the IT systems in place do not allow to check and to verify retroactively who has accessed which patient's record for what purpose. The reason is that the IT systems do not allow for these necessary audit trails.

This is a particularly alarming example of a lack of privacy by design. It also shows that even severe legal regulation (confidentiality of patients' data is protected by criminal law in most countries) does not guarantee compliance if the technology used is insufficient.

The German Data Protection Commissioners' Conference has therefore just set up a Working Party to look into hospital IT systems and work on recommendations and guidelines for manufacturers of such systems. The main task of this group will be to formulate minimum requirements of how patients' privacy can and should be designed into IT systems for hospitals.

Conclusion

All four examples above show that for too long, Privacy and Data Protection Commissioners have tried to tackle privacy issues in a reactive manner. The best rules and regulations to protect the identity of individuals lead to nowhere if the technology in place is not privacy-compliant. It will be difficult if not impossible for regulators simply to switch it off as being illegal (see the example of IT systems in hospitals). It is now important to focus on designers and manufacturers of technology in order to build privacy protection as far as possible into the systems from the start. However, such systemic data protection should not be mistaken for the magic "privacy button." There is and there will be no such thing. Privacy by design is no panacea because there is no simple technical fix for complex privacy challenges, but without privacy by design, it will be difficult if not impossible to achieve meaningful privacy protection in the twenty-first century. The door should be locked before the horse bolts.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.