PANEL 3.1

Security by Design/Privacy and Data Protection by Design

Privacy and Data Protection by Design: A Critical Perspective

Jörg Pohle

Alexander von Humboldt Institute for Internet and Society, Berlin, Germany
Email: joerg.pohle@hiig.de

Neither the political debate nor the legal debate nor the engineering debate properly reflects upon the essential contestedness of the underlying concepts of privacy and data protection. Talking about privacy and data protection by design without clarifying what is meant by "privacy" or "data protection" misses the point as much as talking about "democracy by design" without specifying which concept of democracy—direct, representative or semi-direct, parliamentary or presidential, just to name a few—is meant. Without such clarification, one cannot reasonably expect that privacy or data protection built into technical systems meets one's own expectations, those of the lawmaker of a particular law, the general public, or any other stakeholder, and will be accepted as being compliant with the applicable laws.

I. INTRODUCTION

In a talk on the first conference of our series, Ira Rubinstein sharply criticized the EU General Data Protection Regulation (GDPR) for its lack of clarity about both the very meaning of the data protection by design provision (Article 25) and "what steps a controller or processor should take to ensure compliance" (Rubinstein 2018, 106). On the other side, his description of the "five models of privacy by design" he identifies may create the impression that the engineering community is much clearer in what they mean and what is to be done than the legal community.

I want to take up at this very point and show that a similar critique can and should be formulated regarding the engineering side of the privacy and data protection by design field. If one looks into both the debate among engineers and between engineers and—amongst others—lawyers, the central question is what really is identified as being at stake if different parts of the engineering community talk about "privacy" or "data protection." While among themselves they are often pretty clear about what they mean and at times even very formally define the terms and concepts they use, this does not translate to their external communication, e.g., vis-à-vis lawyers or the general public, where they usually don't clarify that they have redefined terms like "privacy" or "anonymity, making them essentially incompatible to both the legal and the general public's understanding and use of these terms.

In the end, I will draw some conclusions on how to move forward and find a way to a common understanding of the problems at hand, leading to a mapping of existing engineering approaches and technical solutions to legal regimes and requirements, and to identifying gaps in the coverage and areas of further research and development.

II. PRIVACY AND DATA PROTECTION AS AN ESSENTIALLY CONTESTED CONCEPT

Although this has only been proven for privacy (Mulligan, Koopman, and Doty 2016), data protection is also what Walter Gallie termed an "essentially contested concept" (Gallie 1956). There is no agreement in the scholarly or political debate on the countless aspects that are essential for understanding the problem and for developing solutions (Pohle 2018). Even at the level of determining the range of phenomena to be assessed, there are massive discrepancies between the descriptions, classifications, and explanations provided by different parties, ranging from interpersonal relationships to the structural conditions of modern, functionally differentiated society.

It is therefore not surprising that there is also no agreement on what is to be protected—almost everything has been brought forward in the debate: individual needs, interests or rights, group or societal values, or structural characteristics of societies. The same applies to the possible reasons, triggers, or amplifiers of the hazard to the protected goods concerned and the harms associated with them. Virtually the only aspect that has seen a large majority gathering behind the same thing is the object of protection: "personal data" or "personally identifiable information"; even if there is both a dispute about its suitability and a dispute about its very meaning.

The regulatory architecture, on the other hand, is again fundamentally controversial: property-oriented approaches compete with contract-oriented ones, collection- or processing- with flow-oriented approaches, procedural with substantive ones, law- with market- or with technology-based approaches, and ones imposing duties on controllers with ones empowering data subjects. And last but not least, countless different identifiers are in use: (computer/information/data) privacy, data protection, (informational) personality rights, even (informational/digital) intimacy, or—although apparently not fitting well into this series—surveillance. This hodgepodge of names, areas of phenomena, and theories of explanation, unbelievable as it may seem at first sight, has produced myriad laws, each of which asserts a claim to validity and compliance. And even if there appears to be a consensus on a particular legal text, e.g. the GDPR, the very contestedness of the underlying values, theories and understandings resurface in its interpretation and application to specific contexts, situations, and information processing practices. More and more of these laws contain provisions demanding to implement privacy or data protection into technology—"privacy by design" or "data protection by design," "privacy engineering," "privacy-enhancing technologies"—and those approaches and technologies then inherit their originating concepts' contestedness.

III. THE INFLUENCE OF THEORIES AND CONCEPTS ON THE DESIGN OF TECHNOLOGY

In order to better assess the field of privacy and data protection by design, there is much to be learned from the security field because they operate under very similar circumstances and conditions. The central goal of security engineering is to design and develop secure information systems. What looks like a straightforward goal is—truth be told—a complex issue (Schneier 2004, Anderson 2008). It's well known in the IT security field what Bruce Schneier once said about claiming to have a secure product or a product producing security:

Inevitably, these claims are naïve and simplistic. […] The first questions to ask are: "Secure from whom?" and "Secure against what?" (Schneier 2004, 12)

Systems are not secure in an absolute sense but only against specific threats, attacks, and attackers. It's not that these systems are faulty, but that their design is based on conscious or unconscious assumptions and design decisions by the systems' designers. These decisions are about what threats are to be handled by the system and about what kinds of attacks the system is going to prevent or ignore They are dependent upon how much experience the designers have, but more importantly upon how much the designers know about the specific contexts where the system is

going to be deployed and operated. And especially relevant for issues like privacy or data protection is which theoretical understanding, ideological concept, or legal regime regarding privacy and data protection the designers take as a basis for their design decisions.

IV. MISSING THE TARGET–EXAMPLES ON THREE DIFFERENT LEVELS

The following section provides some examples that shed light on the consequences for the design and implementation of technical means and measures for protecting privacy, against surveillance or for data protection.

A. Different theories and understandings

There are myriad different, partially overlapping, partially contradictory privacy theories and frameworks that approach privacy in different ways, from different perspectives, and with different goals.

Some scholars see privacy as just another word for secrecy (e.g., Schneier 2004) while some equate privacy with confidentiality (e.g., Ware 1967). Alan Westin's definition of privacy refers to the individual's, group's, or organization's control of the dissemination of information about them (Westin 1967). Bernhard Hoffmann defined the goal as "preserving the original context" (Hoffmann 1991), while Ferdinand Schoeman conceptualized privacy's function in a similar vein as to "maintain the integrity of different spheres of life" (Schoeman 1992), which Helen Nissenbaum later reformulated as "contextual integrity" (Nissenbaum 2004). Ruth Gavison argued that privacy should be seen as accessibility to the individual (Gavison 1980), while others argue for privacy as a matter of fairness in the processing of personal information (e.g., US Department of Health, Education, and Welfare 1973; Rotenberg 2001), respect (e.g., Benn 1971; Parent 1983), interpersonal boundary regulation (e.g., Petronio 2002; Palen and Dourish 2003) or due process (e.g., de Vries 2013). Each of these understandings demands very different approaches, technical designs, and techniques when being implemented into IT systems–e.g., equating privacy with secrecy demands an all-or-nothing approach to the protection of privacy and a factual prevention of the processing of personal data, while other concepts formulate requirements on how and for which purposes personal data may be collected, processed, and used.

But there are not only privacy theories competing for the most convincing explanation of modern-day information practices' consequences for individuals, groups, organizations, and society. After the early German discussion about the privacy problem threw the term "privacy" and the public-private dichotomy into the dustbin of history (Pohle 2016), Adalbert Podlech defined data protection as "setting and enforcing the conditions under which the information practices of a given society may be acceptable for all parts of that society" (Podlech 1976, 313). Surveillance studies scholars also focus on analyzing real-world information practices, sometimes with and sometimes without taking any pre-scientific assumptions about privacy as the subject of protection (e.g., Lyon 2003; Marx 2015). Both of these concepts defy an easy mapping to design patterns, technical building blocks, or code snippets.

There are huge differences with respect to the means, including the technical means, for protecting privacy or the values it serves: from access control mechanisms (e.g., Hung 2005) to information flow control measures (e.g., Ortmann, Langendörfer, and Maaser 2007), from allocating control options to data subjects (e.g., Lazaro and Le Métayer 2015) to imposing duties like purpose-binding onto data holders (e.g., Massacci and Zannone 2004), from data minimization (e.g., Antignac, Sands, and Schneider 2017) to simply using computers instead of people to process data (e.g., Posner 2005), from educational programs and privacy icons or labels (e.g., Hansen 2009; Kelley et al. 2009) to protection goal-based operationalization (e.g., Hansen, Jensen, and Rost 2015). Many scholars treat anonymity as a means for achieving privacy (e.g., van Rossum et

al. 1995), some as a broken means (e.g., Ohm 2010), while others treat them as disparate concepts (e.g., Shmatikov 2011). A majority of scholars still treat sensitivity as a property of information (e.g., Bing 1972, Ohm 2015), while this has long been refuted (e.g., Miller 1969, 1188) and called a fiction (Simitis 1990). Depending on the design goals as defined by the different theories and understandings, each means might be necessary and sufficient, or just helpful, or even counterproductive to and undermining the issue at stake.

Maybe the shortcomings of the engineering debate on privacy and data protection by design can best be seen in the very widespread equation of "privacy of data" and "privacy of people" in publications by engineers and software developers.

B. Misidentifying the applicable law

While compliance to the applicable law is mandatory for public and private parties, compliance to non-applicable law—or even non-legal sets of norms or standards—is irrelevant and cannot be used for justifying non-compliance with the applicable law. Data holders, or controllers under the GDPR, therefore need to know which legal regime or particular law was used as a reference for eliciting legal, organizational, and technical requirements in order to assess whether or how far a particular IT system would make or help them to be compliant.

Identifying the applicable law seems to be hard for engineers. For more than one decade, European requirements for engineering projects referred to the EC Data Protection Directive of 1995 for eliciting legal requirements in order to translate them into technical requirements before Kiyavitskaya, Krausová, and Zannone (2008) explained to them that EC directives are not applicable laws—they address Member States, demanding from them that they implement the directives into national laws that only then would be applicable. In the US, the Health Insurance Portability and Accountability Act's (HIPAA) Privacy Rule played a quite comparable role in the discussion on designing information systems with legal compliance in mind—over time, it became more or less the single point of reference for legal compliance, often used interchangeably with "legal requirements" (e.g., Massey et al. 2010).

In the engineering community's international debate, the problem is even bigger: many privacy engineering and privacy by design approaches take either the Fair Information Practice Principles (FIPPS) (cf. U. Department of Health, Education, and Welfare 1973; e.g., Dennedy, Fox, and Finneran 2014) or—less often—the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (e.g., Čas 2005) as their basis, while selling the systems then built as compliant to all privacy and data protection laws. The particular shortcomings of these approaches with regard to applicable laws are seldom addressed (the few exceptions include Cavoukian 2000 regarding the applicability of the OECD Guidelines and Grimm and Roßnagel 2000 criticizing the Platform for Privacy Preferences Project's (P3P) very poor compatibility with provisions in applicable laws).

Last but not least, there are a huge number of privacy and data protection by design approaches, design patterns, technologies, components, and other building blocks based on undisclosed requirements or ones with no, unclear, or questionable provenance, or which seem to be made up out of thin air. It is therefore virtually undecidable for any of these approaches or technologies whether they would be of help for data holders or controllers to comply with and to demonstrate compliance with applicable laws.

C. Conceptual differences regarding particular aspects

In addition to the more general mismatches between different theories, understandings, and laws, there are also conceptual differences in various details and particular aspects.

For example, the engineering debate on anonymization techniques in the statistical disclosure

community focuses on third-party attackers only, contrary to e.g., the data protection law's focus on the data holder itself as an attacker. In addition, data protection law does not differentiate between confidential and non-confidential, or vulnerable and non-vulnerable data, as the engineering debate does, regarding the scope of the law's application (Hölzel 2018). A similar mismatch can be seen regarding differential privacy that protects only against true identification, but not against false identification, while the GDPR makes no such difference, "otherwise the right to rectification in art 16 GDPR would be pointless" (Hölzel, 2019). The consequence of a mismatch between the legal and the engineering debates cannot be to simply declare the engineers' understanding of the issue preferable, as Nissim and Wood (2018) do with regard to the anonymization debates.

Another example is some scholars' assumption that one would not fall under the law if the processing of personal data is done on the user's device and the personal data does not leave this device (e.g., Hartzog 2009; Holtz 2010), while the law clearly defines the controller as the one who "determines the purposes and means of the processing of personal data" (Article 4.7 GDPR). This belief in user-centric computation as a solution is quite surprising if one considers that already in the DRM debate, it became clear that the processing's location and the control over the purposes and means of processing may easily differ substantially.

V. HOW TO MOVE FORWARD

In order to move forward in building common approaches for cybersecurity, privacy, and data protection in a globalized world, I would like to propose the following research and work program.

First of all, we need to accept the essential contestedness of privacy, surveillance, and data protection and therefore refrain from hiding behind the veil of ambiguous terms and instead clarify what is meant specifically if we use particular terms.

Secondly, we should start in earnest to translate legal requirements—or requirements from non-legal theories and understandings—into a language that engineers understand. This translation effort would first of all be an obligation on the part of the legal and social science community in order to ensure that the source language is well understood.

Third, we need to identify and clarify—and preferably prove, e.g., using certification mechanisms—which legal and non-legal requirements are to which extent addressed by particular technical design approaches or fulfilled by particular design patterns, technologies, components, and other building blocks. To provide for this would be an obligation of the engineering community.

Fourth, in order to move from theory to practice, we need to map the field of legal and non-legal requirements and technical implementations in order to provide data holders and controllers with trustworthy information on viable solutions and best practices, but also limitations that need additional measures to be taken in order to be compliant with applicable laws.

Finally, we would then be able to use the map to identify gaps in the existing technology's coverage of legal and non-legal requirements and push research and development in this field in order to close the identified gaps in the near future.

REFERENCES

Anderson, Ross. 2008. *Security Engineering: A Guide to Building Dependable Distributed Systems*. Vol. 2. New York: John Wiley & Sons.

Antignac, Thibaud, Sands, David, and Schneider, Gerardo. 2017. "Data Minimisation: A Language-Based Approach." In *ICT Systems Security and Privacy Protection*, edited by Sabrina De Capitani di Vimercati and Fabio Martinelli, 442-456. Cham: Springer.

Benn, Stanley I. 1971. "Privacy, Freedom, and Respect for Persons." In *Privacy*, edited by J. Roland Pennock and John W. Chapman, XIII:1-26. NOMOS. Yearbook of the American Society for Political and Legal Philosophy. New York: Atherton Press.

Bing, Jon. 1972. "Classification of Personal Information with Respect to the Sensitivity Aspect." In *Data Banks and Society*, 98-141. Oslo: Universitetsforlaget.

Čas, Johann. 2005. "Privacy in Pervasive Computing Environments: A Contradiction in Terms?" *IEEE Technology and Society Magazine* 24 (1): 24-33.

Cavoukian, Ann. 2000. *Should the OECD Guidelines Apply to Personal Data Online? A Report to the 22nd International Conference of Data Protection Commissioners (Venice, Italy)*. Ontario: Information and Privacy Commissioner/Ontario.

de Vries, Katja. 2013. "Privacy, due process and the computational turn: A parable and a first analysis." In *Privacy, Due Process and the Computational Turn*, edited by Mireille Hildebrandt and Katja de Vries, 9-38. Abingdon: Routledge.

Dennedy, Michelle, Fox, Jonathan, and Finneran, Tom. 2014. *The Privacy Engineer's Manifesto: Getting from Policy to Code to QA to Value*. New York: Apress.

Gallie, Walter Bryce. 1956. "Essentially Contested Concepts." *Proceedings of the Aristotelian Society*, New Series, 56: 167-198.

Gavison, Ruth. 1980. "Privacy and the Limits of Law." *The Yale Law Journal* 89, no. 3 (January): 421-471.

Grimm, Rüdiger, and Roßnagel, Alexander. 2000. "Can P3P Help to Protect Privacy Worldwide?" In *Proceedings of the 2000 ACM Workshops on Multimedia*, 157-160.

Hansen, Marit. 2009. "Putting Privacy Pictograms into Practice: A European Perspective." In *INFORMATIK 2009: Im Focus das Leben*, edited by Stefan Fischer, Erik Maehle, and Rüdiger Reischuk, 154: 1703-1716. Lecture Notes in Informatics (LNI). Bonn: Gesellschaft für Informatik.

Hansen, Marit, Jensen, Meiko, and Rost, Martin. 2015. "Protection Goals for Privacy Engineering." In *Proceedings of the International Workshop on Privacy Engineering (IWPE)*.

Hartzog, Woodrow. 2009. "The privacy box: A software proposal." *First Monday* 14 (11). http://www.firstmonday.dk/ojs/index.php/fm/article/view/2682.

Hoffmann, Bernhard. 1991. *Zweckbindung als Kernpunkt eines prozeduralen Datenschutzansatzes*. Baden-Baden: Nomos Verlagsgesellschaft.

Holtz, Leif-Erik. 2010. "Datenschutzkonformes Social Networking: Clique und Scramble!" *Datenschutz und Datensicherheit* 34 (7): 439-443.

Hölzel, Julian. 2018. "Anonymisierungstechniken und das Datenschutzrecht." *Datenschutz und Datensicherheit* 42 (8): 502-509.

(CONT.)

Hölzel, Julian. 2019. "Differential Privacy and the GDPR." *European Data Protection Law Review* 5 (2): 184-196.

Hung, Patrick C.K. 2005. "Towards a Privacy Access Control Model for e-Healthcare Services." In *Third Annual Conference on Privacy, Security and Trust, October 12-14, 2005 Proceedings*.

Kelley, Patrick Gage, Bresee, Joanna, Cranor, Lorrie Faith, and Reeder, Robert W. 2009. "A 'Nutrition Label' for Privacy." In *Proceedings of the 5th Symposium on Usable Privacy and Security*, Article 4.

Kiyavitskaya, Nadzeya, Krausová, Alzbeta, and Zannone, Nicola. 2008. "Why Eliciting and Managing Legal Requirements Is Hard." In *Requirements Engineering and Law (RELAW'08)*, 26-30.

Lazaro, Christophe, and Le Métayer, Daniel. 2015. *The control over personal data: True remedy or fairy tale?* Technical report. Inria, Research Centre Grenoble – Rhône-Alpes.

Lyon, David, ed. 2003. *Surveillance as Social Sorting: Privacy, risk, and digital discrimination*. London: Routledge.

Marx, Gary T. 2015. "Surveillance Studies." In *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed., edited by James D. Wright, 733-741. Amsterdam: Elsevier.

Massacci, Fabio, and Zannone, Nicola. 2004. "Privacy Is Linking Permission to Purpose." In *Security Protocols*, edited by Bruce Christianson, Bruno Crispo, James A. Malcolm, and Michael Roe, 179-191. Berlin: Springer.

Massey, Aaron K., Otto, Paul N., Hayward, Lauren J. and Antón, Annie I.. 2010. "Evaluating existing security and privacy requirements for legal compliance." *Requirements engineering* 15 (1): 119137.

Miller, Arthur Raphael. 1969. "Personal Privacy in the Computer Age: The Challenge of a New Technology in an Information-Oriented Society." *Michigan Law Review* 67 (6): 1089-1246.

Mulligan, Deirdre K., Koopman, Colin, and Doty, Nick. 2016. "Privacy is an essentially contested concept: a multi-dimensional analytic for mapping privacy." *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 374 (2083).

Nissenbaum, Helen. 2004. "Privacy as contextual integrity." *Washington Law Review* 79:101-139.

Nissim, Kobbi, and Wood, Alexandra. 2018. "Is privacy privacy?" *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 376 (2128).

Ohm, Paul. 2010. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization." *UCLA Law Review* 57: 1701-1777.

Ohm, Paul. 2015. "Sensitive Information." *South California Law Review* 88: 1125-1196.

Ortmann, Steffen, Langendörfer, Peter, and Maaser, Michael. 2007. "Enhancing Privacy by Applying Information Flow Modelling in Pervasive Systems." In *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*, edited by Robert Meersman, Zahir Tari, and Pilar Herrero, 4806: 794-803. Lecture Notes in Computer Science. Berlin: Springer.

Palen, Leysia, and Dourish, Paul. 2003. "Unpacking 'Privacy' for a Networked World." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 129-136.

Parent, W.A. 1983. "Privacy, Morality, and the Law." *Philosophy & Public Affairs* 12 (4): 269-288.

Petronio, Sandra Sporbert. 2002. *Boundaries of Privacy*. Albany, New York: State University of New York Press.

Podlech, Adalbert. 1976. "Gesellschaftstheoretische Grundlage des Datenschutzes." In *Datenschutz und Datensicherung*, edited by Rüdiger Dierstein, Herbert Fiedler, and Arno Schulz, 311-326. Köln: J. P. Bachem Verlag.

Pohle, Jörg. 2016. "Die kategoriale Trennung zwischen »öffentlich« und »privat« ist durch die Digitalisierung aller Lebensbereiche überholt – Über einen bislang ignorierten Paradigmenwechsel in der Datenschutzdebatte." In »Worüber reden wir eigentlich?« Festgabe für Rosemarie Will, edited by Michael Plöse, Thomas Fritsche, Michael Kuhn, and Sven Lüders, 612-625. Berlin: Humanistische Union.

Pohle, Jörg. 2018. "Datenschutz und Technikgestaltung: Geschichte und Theorie des Datenschutzes aus informatischer Sicht und Folgerungen für die Technikgestaltung." PhD diss., Mathematisch-Naturwissenschaftliche Fakultät, Humboldt-Universität zu Berlin.

Posner, Richard A. 2005. "Our Domestic Intelligence Crisis." 21 December 2005, *The Washington Post*.

Rotenberg, Marc. 2001. "Fair Information Practices and the Architecture of Privacy (What Larry Doesn't Get)." *Stanford Technology Law Review* 1. http://stlr.stanford.edu/pdf/rotenberg-fair-info-practices.pdf

Rubinstein, Ira S. 2018. "Procedural, Institutional, Technical and Management Devices: A U.S. Perspective." In *Privacy and Cyber Security on the Books and on the Ground*, edited by Ingolf Pernice and Jörg Pohle, 103-106. Berlin: HIIG.

Schneier, Bruce. 2004. *Secrets and lies: digital security in a networked world*. Paperback Edition. Indianapolis: Wiley Publishing.

Schoeman, Ferdinand. 1992. *Privacy and social freedom*. Cambridge: Cambridge University Press.

Shmatikov, Vitaly. 2011. "Anonymity Is Not Privacy." *Communications of the ACM* 54 (12): 132.

Simitis, Spiros. 1990. "'Sensitive Daten' – Zur Geschichte und Wirkung einer Fiktion." In *Festschrift zum 65. Geburtstag von Mario M. Pedrazzini*, edited by Ernst Brem, Jean N. Druey, Ernst A. Kramer, and Ivo Schwander, 469-493. Bern: Stämpfli & Cie.

US Department of Health, Education, and Welfare. 1973. *Records, Computers, and the Rights of Citizens*. The Massachusetts Institute of Technology.

van Rossum, Henk, Gardeniers, Huib, Borking, John, Cavoukian, Ann, Brans, John, Muttupulle, Noel and Magistrale, Nick. 1995. *Privacy-Enhancing Technologies: The Path to Anonymity*. Den Haag: Information/Privacy Commissioner/Ontario, Canada & Registratiekamer, The Netherlands.

Ware, Willis H. 1967. "Security and privacy: similarities and differences." In *Proceedings of the April 18-20, 1967, spring joint computer conference, 287-290. AFIPS '67 (Spring)*. Atlantic City, New Jersey: ACM.

Westin, Alan F. 1967. *Privacy and Freedom*. New York: Atheneum.