Methods and Tools for GDPR Compliance through Privacy and Data Protection Engineering

Yod-Samuel Martín
Universidad Politécnica de Madrid
Madrid, Spain
samuelm@dit.upm.es

Antonio Kung
Trialog
Paris, France
antonio.kung@trialog.com

Abstract— **In this position paper we posit that, for Privacy by Design to be viable, engineers must be effectively involved and endowed with methodological and technological tools closer to their mindset, and which integrate within software and systems engineering methods and tools, realizing in fact the definition of Privacy Engineering. This position will be applied in the soon-to-start PDP4E project, where privacy will be introduced into existent general-purpose software engineering tools and methods, dealing with (risk management, requirements engineering, model-driven design, and software/systems assurance).**

Keywords—GDPR; Privacy by Design; Privacy engineering; Risk management; Requirements engineering; Model-driven engineering; Software and systems assurance; Privacy Impact Assessment; PDP4E

I. BACKGROUND: CURRENT PRIVACY PERSPECTIVES

Despite the relevance that privacy and data protection have gained in the recent years in the regulatory, organizational, and technological and management fields, their perspectives are still disconnected from one another and, more relevantly, from the engineering practice, as we will show in this section. In the rest of paper we defend a position to vouch for the need of methods and tools integrated within mainstream engineering practice, particularize their application to specific engineering disciplines, and introduce a project that will implement it.

A. Regulating data protection

Data protection regulatory frameworks have long existed, and evolved along with the society and technology. The General Data Protection Regulation (GDPR) [1], in force since 2016 and mandatory in May 2018, sets a novel array of binding data protection principles, data subjects' rights, and legal obligations so as to ensure the protection of personal data of EU citizens. Such regulatory innovations do have an impact on the technological products that must abide to them, and on the engineering process followed for their creation. For instance, products must implement any functionality needed to support data subject requests to enforce their rights. Other regulatory innovations affect directly the processes, e.g. accountability and data protection impact assessment. Thus, this legal approach need come along with technical measures to protect privacy and personal data in practice: as it is often said, "[software] code is law", in that the support of technological features regulates what we can do as much as the legal framework. However, such 'privacy-by-policy' [2] approach leaves the responsibility to comply with regulation in the hands of legal staff; while *the engineers are not prepared to deal with the related concepts and lack tools to translate those regulations into the products they create*.

B. Promoting Privacy by Design

The principles of Privacy by Design (PbD) plead for the proactive consideration of privacy since the onset of a project, throughout all the activities involved during the design and development of products, services and systems, rather than as an afterthought. This approach has been openly embraced by Data Protection Authorities [3], [4], legally required by GDPR, and supported by the European Commission in order to foster the data economy [5], [6]. For PbD to be viable, engineers must be effectively involved in the loop, as they are ultimately responsible for creating their products. Otherwise, PbD risks becoming in practice a mere slogan, a bare principle without any real impact; or even worse, being voided of its content as a fashionable term subject to false claims by pretenders [7]. However, PbD has not yet gained widespread, active adoption in the engineering practice, due to a mismatch between the legal and the technological mindsets [8]. Indeed, *from the engineers' mindset* [9], *privacy is usually considered just from the perspective of data security, if any; and they tend to disregard privacy and data protection on the technical designs* and architecture, relying instead on privacy policies for compliance. Instead, a hands-on approach is needed [3] that provides specific guidance to developers [10] and engineers overall, as key parts to achieve effective data protection [11].

C. Crafting Privacy-Enhancing Technologies

From the purely technological arena, a plethora of solutions have long been researched and elaborated to create Privacy-Enhancing Technologies (PETs), with varying degrees of maturity, which foster data protection and respond to privacy concerns. The systematization of such knowledge has been tackled by several reviews, handbooks and surveys [12]–[14]. However, *Privacy-Enhancing Technologies remain unknown for most engineers*, due to the uncoupling between the PETs and the practice of systematic engineering and development; which makes engineers unaware or unknowledgeable of the proper applicability of such solutions. In practice, when engineers need to face privacy issues, they resort to crafting tailored solutions (if any), rather than choosing the systematic and economic application of existent solutions drawn from the state of the technique.

D. Automating privacy management

Privacy Management software tools aim at streamlining the process of compliance with data protection regulation by non-experts, driven by novel regulations such as GDPR [15]. Yet

Work funded by Spanish Government's Ministerio de Economía, Industria y Competitividad - Agencia Estatal de Investigación, ref. EUIN2017-87180.

these tools are often targeting user profiles other than engineering, and they are disconnected from Software Engineering practice: Privacy Program Management (PPM) tools (e.g. assessment managers) address the legal privacy office; while Privacy Enterprise Management (PEM) tools (e.g. data mapping) address IT departments, but just as a supporting role to legal departments again. Thus, these tools cannot be integrated within the engineering lifecycle. In consequence, research has shown [16]–[19] that *developers and engineers (who usually are not privacy-savvy at all), find privacy and data protection alien to their work and, most importantly, seldom use privacy management tools*, as they find these are more oriented to the legal arena rather than to the engineering activities. Same research has encountered that they will be more akin to take decisions that protect privacy and data protection when the process is embedded within their usual development workflow and tools.

II. Position: Methods and Tools for Privacy Engineering

As we have shown, regulation and policy ask engineers to stick to PbD principles and apply data protection solutions throughout their projects, and technological developments provide such solutions e.g. in the form of PETs. However, engineers are used to thinking in terms of systems and software: their usual skills include working with e.g. dataflow models, database structures, or deployment architectures of the system under development. Thus, they often feel lost in translating regulatory issues into operational work items and activities for the projects they manage: they are unsure at how the GDPR translates into backlog items, what specific threats their users face, what technical measures choose to meet user's rights, whether the rights to access or portability implies slit opening databases and revealing their witty algorithms, etc. If engineers have long applied systematic practice to deal with other categories of requirements (e.g. through secure software engineering), why the same approach cannot be applied to deal with privacy and data protection? Wouldn't it be possible to answer those doubts while keeping to the mainstream methods and tools that engineers already use in their daily work? In order to answer such challenges, *we argue that engineers must be endowed with methodological and technological tools* that allow the systematic application of data protection principles to attain the compliance with the regulatory framework, while keeping to an economy of resources, living with other requirements and achieving overall business goals.

This vision is realized in the field of Privacy (and Data Protection) Engineering, "*a nascent field of research and practice which pursues systematic approaches for the inception and application of privacy-oriented solutions throughout systems and software development processes*" [20]. It has been widely recognized as the practical complement to PbD by the EDPS (European Data Protection Supervisor, which has created the Internet Privacy Engineering Network - IPEN), ENISA (European Network and Information Security Agency) [21], and the proponents of PbD themselves [22]. The keystones of this field are precisely the methods employed by engineers to capture and address privacy issues during the different stages of the development of sociotechnical systems, and the software tools that support engineers to perform the engineering tasks or activities prescribed by such methods [23].

We posit that not all engineers need be privacy experts, yet they will face privacy issues in their regular work anyway, even if supported by specialist teams. Thus, adoption and acceptance of Privacy Engineering methods and tools entails their integration within the large heritage of software and systems engineering methods and tools, both with mature communities of practice that have generated well-established, widely accepted bodies of knowledge [24], [25] applied by engineers in their daily work. Engineers have a solid training on requirements elicitation, software analysis, architectural and detailed design, data modelling, software validation, quality assurance, etc. and they are used to applying software tools that help them systematically and economically perform such tasks. Privacy engineering can take advantage of all that savvy in the sake of privacy and data protection.

We vouch for the seamless inclusion of privacy and data protection functions into general-purpose software and system engineering tools of customary use by engineers (a demand which has been captured by ENISA [21]), rather than forcing them to learn to use unfamiliar tools. Likewise, *we defend the integration of privacy and data protection activities into the different stages of the SDLC (System Development Lifecycle)*, and into the methods and workflows followed by engineers, rather than considering them unconnected activities. Also, for everyday application, *we plead for a readily available body of knowledge with the wisdom amassed by the privacy community* of practice and research, and in terms compatible with the engineering mind-set. Thus, privacy and data protection knowhow should be brought into mainstream practice of software and systems engineering, by providing engineers with methods and tools that are closer to their expertise.

III. Privacy within Software and Systems Engineering

We also vouch for the introduction of privacy and data protection into specific software and system engineering disciplines, due to their relevance to the regulatory framework and the maturity of applicable related privacy methods. Such disciplines are, namely: risk management, requirements engineering, model-driven design, software/systems assurance.

A. Risk management

Risks are the negative effects that deviate from the expected objectives, and which arise as the consequence of an uncertain event. Risk management supports the execution of Privacy Impact Assessments (PIA) from the engineering perspective, aligned to the legal requirements (e.g. GDPR art. 35 and WP29 guidance [26]), to deal with the impact of risks on the data subjects (rather than on e.g. business profits), which may arise from processing activities that deal with their personal data [27]. Risk management starts out from the categories of personal data handled and the processing activities to which they are subject. Then, the potential threats are identified, and risks are estimated based on different factors that affect their likelihood and impact [28]. Next, risks are evaluated, prioritized, and different actions are taken, including specific countermeasures (privacy controls) or solutions (PETs) to mitigate them. Finally, risks are documented and monitored. One privacy risk assessment method is LINDDUN [29], which uses Data Flow Diagrams (DFDs) to model the flow of personal data across processing activities, and the

realms of the processors and controllers. Besides, following the approach of multilateral security, risk management can integrate security impact assessments, include the impact of privacy risks on business areas [30], [31], and be leveraged for Vendor Relationship Management (VRM).

B. Requirements engineering

Software/system requirements define the needs of the stakeholders, including the functions to be covered and any other conditions that must be met (to satisfy some contract, standard, regulation, etc.) Privacy and data protection are an example of the latter case ("non-functional requirements"). The sources for privacy and data protection requirements specification are twofold: first, regulations and standards (e.g. GDPR, ISO29100 [32]) prescribe procedures, guidelines, and principles (e.g. lawfulness, fairness, transparency, consent, etc.); and second, privacy protection goals describe privacy properties to be held (e.g. unlinkability, transparency, and intervenability) [33] —both approaches being indeed complementary [34], [35]. Requirements elicitation shall take into account the context of the system under development, for which methods such as PROPAN [36] propose the use of 'problem frames'. With respect to management of privacy requirements, a relevant approach is the operationalization of high-level privacy goals and data protection regulations into privacy controls [34], [37], [38, Ch. App. 5](i.e. more concrete and fine-grained requirements that can be mapped to the original goals). This goal-driven process is complementary to risk management [39] and supports the appropriate choice of solutions (PETs and privacy patterns) linked to such controls.

C. Model-driven design

A development process moves through different models, from an overall, abstract understanding of a system, towards fine-grained details and designs from different perspectives (architectural, functional, data, etc.) Model-driven design supports engineers to analyze and design the systems under development. From a privacy perspective, first and foremost, a proper system model involves a data mapping and inventory (i.e. identify and categorize) of the personal data elements that the system will process. In general, structural system models can be enriched with information about the categories of personal information represented and relevant properties for data protection (purpose, visibility, granularity, and retention), behavioral models with information about the processing activities, and architectural models with information about the allocation of processing activities to components. Besides, the architectural models allow the systematic analysis and reasoning about the space of design solutions and the choices that best fit privacy and data protection [21] Strategies have been described that can be followed to shape a privacy-friendly architecture [40]; going into deeper design, minimization, separation or aggregation strategies can be implemented through the application of model transformation techniques. Last, automated Model Based Testing (MBT) can be particularly relevant for the verification of the correct application of mechanisms for access control to personal data.

D. Software and systems assurance

The concept of assurance comes from the field of safety in critical systems, where it represents the actions that must be planned to ensure confidence in the safety of a product, system, etc., and it has been later extended to other fields e.g. cybersecurity. Such concept is quite close to the principles of accountability and transparency in the field of data protection, and to the goal of privacy intervenability, which makes easy to translate assurance methods and tools into privacy engineering. An assurance process departs from (reusable) models of the data processing and data protection activities of an organization, and of the regulatory framework (GDPR and its interpretation through WP29 guidance, codes of conduct, privacy standards [27], [32], [41], corporate policies, etc.), including relevant roles and processes, compulsory activities and formal requirements. The assurance process supports the demonstration of compliance with regulation and the observance of the principle of accountability through systematic capture of evidences, their association to requirements and artefacts, traceability to the regulation, and argumentation of compliance derived from such evidences. Besides, data protection certification schemes can also take advantage of the evidences collected.

IV. Upcoming work: the PDP4E project

The authors are endeavoring to realize this position through the coordination of PDP4E, a European innovation project due to start in May 2018 that will provide engineers with methods and tools to systematically apply data protection principles in the projects they carry out, so that their comply with GDPR and bring the principles of Privacy by Design to practice.

Existing, general-purpose software engineering tools have successfully demonstrated the applicability to safety or security of risk management (MUSA DST[1]), requirements engineering (Papyrus Requirements), model-driven analysis, design and verification (Papyrus[2], Diversity[3] and Sophia[4]), and software and system assurance (Opencert[5]). Thus, rather than creating a set of tools from scratch, PDP4E will integrate privacy and data protection engineering functionalities into those other tools, already in use by engineers, leveraging the project efforts and ensuring a seamless adoption of its results. Likewise, PDP4E will also innovate in integrating data protection methods (LINDDUN [29], PRIPARE [37], PROPAN [36]) into mainstream, existent software and systems engineering methodologies and process models, specializing them to operationalize GDPR compliance, and integrating the current work on standards and methods (e.g. OASIS PMRM [42], ISO 29134 [27], or ISO 27550 –under development). PDP4E will deliver a set of evolvable knowledge bases (operational data protection requirements; data protection risks, threats and solutions; privacy patterns; assurance reference frameworks) which distil the existing knowledge in the field, providing engineers with guidance at hand they can use during the engineering activities. In order to facilitate the adoption of the project results, PDP4E will release most of its outcomes through open licenses, pivoting around the open-source Eclipse ecosystem. Likewise, in order to maximize their interoperability, reusability and adaptation, the toolset will

[1] http://www.musa-project.eu/tools
[2] https://www.eclipse.org/papyrus/
[3] https://projects.eclipse.org/projects/modeling.efm
[4] https://www.polarsys.org/esf/
[5] https://www.polarsys.org/projects/polarsys.opencert

stick to mainstream, standard modelling frameworks, languages, and interchange formats.

PDP4E will involve different stakeholders so as to ensure that the results respond to the widest range of engineers. First, the community of developers hosted by Eclipse will be targeted to address their needs. Second, the outcomes will be validated and demonstrated on engineering projects in two innovative application domains, viz. smart grid and fintech, both of which make intensive, novel uses of personal data which pose specific problems, and must abide by sectoral-regulation.

In the long term, PDP4E will set up an ecosystem of research and practice to boost the adoption of data protection practices in software and systems engineering, providing the open-source PDP4E methods and tools, accompanying training material, a body of knowledge for this emerging field, and a meeting point to serve as reference for the whole community. All in all, the application of PDP4E methods and tools will ease the engineering of GDPR-compliant products, which shall lead to a widespread creation of products, systems and services that better protect the privacy and personal data of EU citizens.

ACKNOWLEDGMENT

The authors thank the partners of PDP4E (Trialog, Universidad Politécnica de Madrid, CA Technologies, Commissariat à l'Energie Atomique et aux Energies Alternatives, Fundación Tecnalia, KU Leuven, Eclipse Foundation Europe, and University of Duisburg-Essen) for their work in the definition of the project and the position herein reflected.

REFERENCES

[1] The European Parliament and The European Council, "General Data Protection Regulation," Off. J. Eur. Union, vol. 2014, no. October 1995, pp. 20–30, 2016.
[2] S. Spiekermann and L. F. Cranor, "Engineering privacy," IEEE Trans. Softw. Eng., vol. 35, no. 1, pp. 67–82, 2009.
[3] Art 29 Working Party and Working Party on Police and Justice, "The Future of Privacy," 2009.
[4] "Resolution on Privacy by Design," Jerusalem, 2010.
[5] C. From et al., "Towards a thriving data-driven economy," Eur. Comm., vol. COM(2014), no. 442, 2014.
[6] Secretary-General of the European Commission, "Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace," 2013.
[7] S. Davies, "Why Privacy by Design is the next crucial step for privacy protection," 2010.
[8] M. Birnhack, E. Toch, and I. Hadar, "Privacy Mindset, Technological Mindset," SSRN Electron. J., Jul. 2014.
[9] I. Hadar et al., "Privacy by designers: software developers' privacy mindset," Empir. Softw. Eng., pp. 1–31, Apr. 2017.
[10] Article 29 Data Protection Working Party, "Opinion 02/2013 on apps on smart devices," October, no. February, pp. 1–11, 2003.
[11] "Warsaw declaration on the "appification" of society."
[12] J. Heurix, P. Zimmermann, T. Neubauer, and S. Fenz, "A taxonomy for privacy enhancing technologies," Comput. Secur., vol. 53, pp. 1–17, Sep. 2015.
[13] J. Argyrakis, S. Gritzalis, and C. Kioulafas, "Privacy Enhancing Technologies: A Review," Springer, Berlin, Heidelberg, 2003, pp. 282–287.
[14] V. Seničar, B. Jerman-Blažič, and T. Klobučar, "Privacy-Enhancing Technologies—approaches and development," Comput. Stand. Interfaces, vol. 25, no. 2, pp. 147–158, May 2003.
[15] International Association of Privacy Professionals (IAPP), "2017 Privacy Tech Vendor Report," 2017.
[16] R. Balebako, A. Marsh, J. Lin, J. Hong, and L. Faith Cranor, "The Privacy and Security Behaviors of Smartphone App Developers," in Proceedings 2014 Workshop on Usable Security, 2014.
[17] R. Balebako and L. Cranor, "Improving App Privacy: Nudging App Developers to Protect User Privacy," IEEE Secur. Priv., vol. 12, no. 4, pp. 55–58, Jul. 2014.
[18] S. Jain and J. Lindqvist, "Should I Protect You? Understanding Developers' Behavior to Privacy-Preserving APIs," in Proceedings 2014 Workshop on Usable Security, 2014, no. February.
[19] Y. S. Van Der Sype and W. Maalej, "On lawful disclosure of personal user data: What should app developers do?," in 2014 IEEE 7th International Workshop on Requirements Engineering and Law, RELAW 2014 - Proceedings, 2014, pp. 25–34.
[20] Y. S. Martín and J. M. Del Álamo, "A metamodel for privacy engineering methods," in CEUR Workshop Proceedings, 2017, vol. 1873, pp. 41–48.
[21] G. Danezis et al., "Privacy and data protection by design - from policy to engineering," 2015.
[22] A. Cavoukian, S. Shapiro, and R. J. Cronk, "Privacy Engineering: Proactively Embedding Privacy, by Design," Toronto, 2014.
[23] S. Gurses and J. M. del Alamo, "Privacy Engineering: Shaping an Emerging Field of Research and Practice," IEEE Secur. Priv., vol. 14, no. 2, pp. 40–46, Mar. 2016.
[24] P. Bourque and R. E. Fairley, SWEBOK v3.0 - Guide to the Software Engineering Body of Knowledge. 2014.
[25] SEBoK Editorial Board, "SEBoK," The Guide to the Systems Engineering Body of Knowledge (SEBoK), v. 1.6, 2016. [Online]. Available:http://sebokwiki.org/wiki/Guide_to_the_Systems_Engineering_Body_of_Knowledge_(SEBoK).
[26] Article 29 Working Party, "Guidelines on Data Protection Impact Assessment (DPIA) and determining," 2017.
[27] "ISO/IEC 29134:2017 - Information technology -- Security techniques -- Guidelines for privacy impact assessment." ISO/IEC JTC 1/SC 27 IT Security techniques, p. 43, 2017.
[28] "Methodology for Privacy Risk Management: How to implement the Data Protection Act," Paris, 2012.
[29] M. Deng, K. Wuyts, R. Scandariato, B. Preneel, and W. Joosen, "A privacy threat analysis framework: Supporting the elicitation and fulfillment of privacy requirements," Requir. Eng., vol. 16, no. 1, pp. 3–32, 2011.
[30] Owasp, "OWASP Risk Rating Methodology," Owasp, pp. 1 – 5, 2013.
[31] R. a R. a. C. Caralli, J. F. Stevens, L. R. Young, and W. R. Wilson, "Introducing OCTAVE Allegro : Improving the Information Security Risk Assessment Process," Young, no. May, pp. 1–113, 2007.
[32] ISO/IEC JTC 1/SC 27, "Information technology -- Security techniques -- Privacy framework ISO/IEC 29100:2011," Geneva (CH), 2011.
[33] M. Hansen, M. Jensen, and M. Rost, "Protection goals for privacy engineering," in Proceedings - 2015 IEEE Security and Privacy Workshops, SPW 2015, 2015, pp. 159–166.
[34] R. Meis and M. Heisel, "Systematic identification of information flows from requirements to support privacy impact assessments," in 2015 10th International Joint Conference on Software Technologies (ICSOFT), 2015, vol. 2, pp. 1–10.
[35] A. Jøsang, W. Quattrociocchi, S. Fischer-Hübner, C. Lambrinoudakis, and G. Pernul, "Trust, Privacy and Security in Digital Business," Trust Priv. Secur. Digit. Bus., vol. 3592, no. July 2015, pp. 548 – 558, 2005.
[36] K. Beckers, S. Faßbender, M. Heisel, and R. Meis, "A Problem-Based Approach for Computer-Aided Privacy Threat Identification," Springer, Berlin, Heidelberg, 2014, pp. 1–16.
[37] N. Notario et al., "PRIPARE: Integrating privacy best practices into a privacy engineering methodology," in Proceedings - 2015 IEEE Security and Privacy Workshops, SPW 2015, 2015, pp. 151–158.
[38] R. M. Blank and P. D. Gallagher, "NIST Special Publication 800-53, Security and Privacy Controls for Federal Information Systems and Organizations, Revision 4.," Washington, DC, 2013.
[39] ISO, "ISO/Guide 73:2009(en), Risk management — Vocabulary," 2009.
[40] J.-H. Hoepman, "Privacy Design Strategies," in ICT Systems Security and Privacy Protection SE - 38, vol. 428, N. Cuppens-Boulahia, F. Cuppens, S. Jajodia, A. Abou El Kalam, and T. Sans, Eds. Springer Berlin Heidelberg, 2014, pp. 446–459.
[41] ISO/IEC JTC 1/SC 27, "ISO/IEC 29101:2013 - Information technology -- Security techniques -- Privacy architecture framework," 2013.
[42] J. Sabo, M. Drgon, and G. Magnuson, "Privacy Management Reference Model and Methodology (PMRM) Version 1.0." OASIS Open, 2016.