PERSONA - A Personalized Data Protection Framework

Marina Egea, Federica Paci, Marinella Petrocchi, Nicola Zannone

Position Paper: PERSONA - A Personalized Data Protection Framework*

Marina Egea[1], Federica Paci[2], Marinella Petrocchi[3], and Nicola Zannone[4]

[1] Atos Research, Madrid, Spain
[2] University of Trento, Italy
[3] Istituto di Informatica e Telematica, CNR, Pisa, Italy
[4] Eindhoven University of Technology, Netherlands

Abstract. The European Directive on Data Protection recognizes the right of data subjects to control the usage of their information. However, to date there are no data protection solutions that involve data subjects in the definition and enforcement of data protection policies. In this paper we present the foundation of a novel approach to personalized data protection in which users play a central role in the authoring and enforcement of the policies governing the access and usage to their data. We discuss the challenges of designing a personalized data protection framework using *personalized medicine* as an illustrative scenario.

1 Introduction

The European Directive 95/46/EC, and its recent reform IP/12/46 of 25 January 2012, recognize the right of data subjects (i.e., the identifiable persons to whom the personal data refers) to have a central role in deciding (i) which personal data can be collected, (ii) for which purposes it can be collected and disclosed, and (iii) to whom it is disclosed. However, to date there are no data protection solutions that give data subjects an active role in the definition and enforcement of the policies which govern the access and usage of their data. After the data subjects have authorized the collection of their data, their influence in the specification of data protection policies is marginal; the task of specifying policies is usually left to security administrators.

Data protection solutions should not only allow data subjects to agree upon predefined data protection policies, but also be customizable with respect to their privacy requirements. In fact, privacy is an individual concept which may vary from person to person [33]. This calls for a new way to conceive data protection and privacy management that is based on customizing data protection specification and enforcement to data subjects' needs. Personalized data protection is the key to empower data subjects with the control over their data, and to assist data controllers (i.e., the natural or legal person responsible for data processing) in enforcing such policies as well as legal requirements.

Implementing a framework for personalized data protection requires addressing several challenges. First, data subjects are usually not privacy experts, and thus the authoring of data protection requirements can be difficult, if not impossible, for them. In addition, user requirements are specified at a too high level to be enforced directly, and thus

* Work partially funded by EU-FP7-IST-NoE-NESSOS, EU-EIT-ICTLabs, EU-FP7-SEC-SECONOMICS, PAT-TRISE projects and by Dutch program COMMIT (THeCS project).

to ensure compliance they need to be refined into enforceable policies [30]. Moreover, data may reveal information about more than one data subject (e.g., genetic data) and, thus, they can be regulated by (possibly) conflicting policies specified by the different data subjects or imposed by legal and business constraints of data controllers [17,20]. Data protection policy enforcement alone is not sufficient to prevent information leakages. Most privacy breaches are caused by human errors and organizational flaws rather than by policies. Therefore, in-depth privacy solutions are needed to mitigate privacy risks. Last but not least, enforcing personalized data protection policies does not eliminate privacy risks completely. For example, data re-purposing cannot be avoided using preventive means especially in dynamic data sharing systems [5].

In this paper, we present PERSONA, a user-centric framework inspired to the principles of privacy by design: it supports data subjects and data controllers in the realization of a data protection enforcement system which allows to embed privacy in the design and architecture of data management systems. In fact, PERSONA allows data subjects and data controllers to specify (high-level) privacy requirements in a controlled natural language which, together with regulatory requirements, are automatically transformed into (low-level) enforceable policies. PERSONA also provides a guided support to policy conflict detection and resolution to guarantee that data are managed in ways that are intended by all data subjects. In addition, the framework relies on the generation of customized GUI and application interfaces to ensure privacy compliance also at design and implementation level. Finally, the framework relies on a-posteriori verification to analyze the actual user behavior and determine the risks of policy infringements.

The paper is organized as follows. In the next section, we discuss the challenges to be addressed for the design of a personalized data protection framework using a scenario on personalized medicine. Section 3 presents the PERSONA framework along with its components. Section 4 discusses related work and Section 5 concludes the paper.

2 A Motivating Example: Personalized Medicine

Personalized medicine exploits the advances in genetic studies to provide specific treatments and therapeutics best suited for an individual's genetic make-up. In particular, it opens the possibility to predict an individual's illnesses and design a personalized treatment plan to prevent or detect these diseases in early stages. Personalized medicine thus offers a new paradigm for the development of drugs and medical practices.

A typical example of personalized medicine involves biobanks. Biobanks represent key resources for clinico-genomic research and advances in personalized medicine. Biobanks collect, store, and distribute biological materials like organs, tissue, blood samples, cells and other body fluids containing traces of DNA or RNA that allow genetic analysis. Physicians and research groups can access data stored in the biobank for clinical trials, personalization of treatments or research purposes. Moreover, biobanks allow research communities to share biomaterial and related analytical data with other research groups which carry out similar research studies.

Since personalized medicine requires building data sharing networks to distribute genetic and medical data across different actors – healthcare providers, research and government institutions, and industry – a key issue is the lack of trust and control over data usage and disclosure. After patients have given their consent to collect genetic data samples, they have limited control to whom data are shared with and how data are used. Patients may not be willing to participate in personal medicine because they are concerned about genetic data being used against them, for example, to deny health insurance or as a basis for hiring decisions [3]. Knowing the benefits of genetic testing for diagnosis and treatment does not always relieve the concerns of patients on privacy risks. To lay the groundwork for privacy in personalized medicine, it is necessary to design a data protection framework which gives patients confidence that the usage and disclosure of their genetic data and related clinical information are "*under their control*".

The design of the framework does not come for "free". A first step is the elicitation of patients' requirements as well as legal and business requirements on the collection and processing of genetic data. However, the authoring of these requirements can be a tedious and time-consuming process for non-expert users such as patients and doctors. It is necessary to develop a user-friendly authoring tool for requirements specification.

A second concern is that patients' requirements are usually too high level to be enforced and thus they need to be translated into enforceable policies which are compliant with patients and healthcare providers' requirements. It may be the case that a patient wants to share the result of a genetic test for breast cancer only with her therapist, but instead the data protection policies enforced at the biobank allows to share the results also with researchers. To ensure compliance with requirements, the policies enforced by the system should be dynamically generated from the patient's requirements.

Moreover, genetic data does not only carry information about a patient but also about her family. Conflicts may arise among the policies generated from patients' and her family's requirements, as well as from constraints imposed by health care providers to ensure that data access is compliant with data protection regulations. For example, a patient may want to take a genetic test to verify her chances of having a cardiac arrest which has affected most of the members of her family and she wants to make available the results to her insurance company. However, other members of the family may be against that because they are afraid that the health insurance company will increase the prize if the test results are disclosed. The conflicts among the policies authored by patients, their families, and healthcare providers should be identified and solved before policies are enforced. Moreover, the system should be transparent and make end-users aware if their policy is not applied in favor of another.

In addition to these issues, errors in the design of the applications' interfaces through which users access the genetic and clinical data stored at biobanks may lead to undesirable leakage of information: the results of a patient genetic test may be accidentally shown to one of her relatives, due to design errors. This calls for approaches that embed data protection policies enforcement into GUIs' design.

The enforcement of data protection policies guarantees that only authorized users can access the data. However, enforcement mechanisms are usually preventive and they do not check how data is actually processed. This poses risks of data re-purposing, as users might process the data for purposes other than those for which the data was originally collected [26,28]. For instance, a doctor may legitimately access her patients' genetic data for providing treatment, and later share this data with her colleagues for

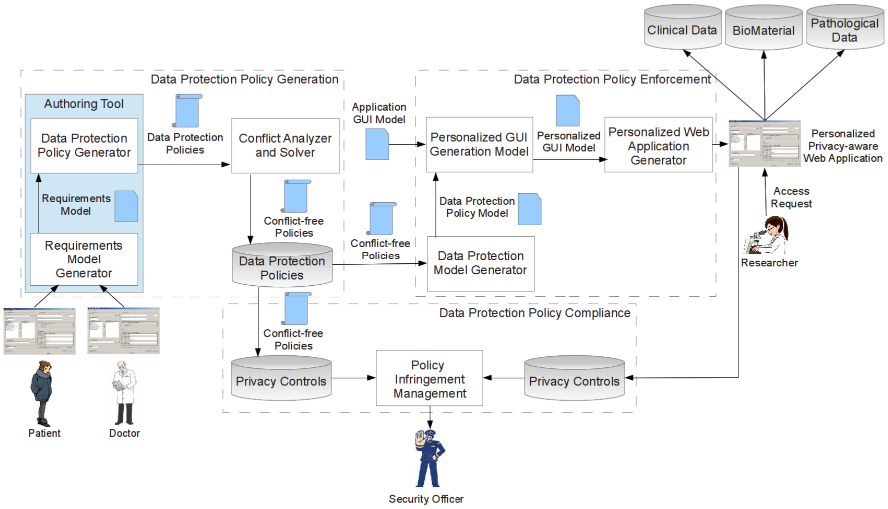
Clinical Data
BioMaterial
Pathological Data
Data Protection Policy Generation
Data Protection Policy Enforcement
Authoring Tool
Data Protection Policy Generator
Data Protection Policies
Conflict Analyzer and Solver
Application GUI Model
Personalized GUI Generation Model
Personalized GUI Model
Personalized Web Application Generator
Personalized Privacy-aware Web Application
Requirements Model
Requirements Model Generator
Conflict-free Policies
Data Protection Policies
Conflict-free Policies
Data Protection Policy Model
Data Protection Model Generator
Access Request
Researcher
Data Protection Policy Compliance
Conflict-free Policies
Privacy Controls
Policy Infringement Management
Privacy Controls
Patient
Doctor
Security Officer

Fig. 1: PERSONA Architecture

research purposes without patient consent. To detect policy infringements, we need methods for analyzing user behavior and identifying deviations from policies.

3 Towards a personalized data protection architecture

Figure 1 shows the functional architecture of the PERSONA framework. The architecture consists of three main components: *Data Protection Policy Generation*, *Data Protection Policy Enforcement*, *Data Protection Policy Compliance*.

Data Protection Policy Generation. One key innovative aspect of the PERSONA framework is the empowerment of data subjects with control over their data through personalization of data protection policy specification, verification, and enforcement. To this end, the framework provides a user-friendly requirements *Authoring Tool*, which allows non-expert users to specify requirements using a constrained natural language founded on goal-oriented requirements languages [18]. We have selected goal-oriented requirements languages because they are effective in capturing stakeholders' needs. The *Requirement Model Generator* component translates the elicited requirements into a formal model amenable for verification. Then, the *Data Protection Policy Generator* generates the data protection policies from the requirements model by applying a set of axioms to derive the elements of the policies from the model. The tool has one main advantage over existing approaches: the derivation is automated based on a set of rules, while other available techniques need the requirement engineer to extract, from a requirements specification in natural language, the policy elements based on a set of heuristics. The dynamic generation of data protection policies from different data subjects' requirements calls for conflicts identification and resolution strategies. The *Conflict Analyzer and Solver* is a transparent engine, which assists data subjects and controllers in the identification and resolution of policy conflicts. It is based on multi-criteria decision making techniques [27]. The engine assures that either data is accessed by end-users according to original requirements, or it notifies why some policies have been assigned priority over others.

Data Protection Policy Enforcement. To prevent information leakages, the GUIs through which end-users access data and the database where the data are stored need to be configured with respect to the regulating data access and usage. PERSONA follows a model-driven security approach to enforce data protection policies at the GUI and database level. In particular, PERSONA ensures that data protection policies are correctly enforced by having the GUI displaying data according to the data protection policies derived from data subjects and data controllers' requirements. To this end, the *Data Protection Model Generator* derives a data protection policy model from the conflict-free data protection policies. This model, together with the GUI model, is used by the *Personalized GUI generator Model* to derive a personalized GUI model which defines which objects in the generic GUI model can be visualized. The *Personalized Web Application Generator* applies model transformation techniques to the personalized GUI model to automatically generate a privacy-aware data management web application where data protection policies are enforced in-depth.

Data Protection Policy Compliance. The personalization of data protection policy specification and enforcement does not eliminate privacy risks like data re-purposing [26,28]. To this end, the PERSONA framework provides an *infringement management mechanism* for assessing the risks of data misuses after data has been disclosed. To identify privacy risks, the framework relies on conformance checking based on alignments [2]. These techniques aim to detect deviations of the observed behavior (recorded in an event log) from the specified behavior (represented by process models), and to locate and explain the root causes of deviations. Intuitively, data protection policies are mapped into *privacy controls* that define the boundaries of compliant behaviors. Compliance of user behavior is determined by aligning the event data recorded by the system against privacy controls. However, not all deviations may correspond to an infringement. For instance, doctors may deviate from the specified behavior to react to an emergency. An auditing technique that reports every deviation as an infringement would lead to a large number of false alerts (i.e., situations reported as infringements although they are admissible). The PERSONA framework employs metrics to measure to what extent user behavior deviates from the specification. This way we can determine whether a deviation is indeed an infringement.

4 Related Work

The literature relevant to our work is related to the following topics.

Requirements-Driven Policy Generation. Several approaches have been proposed to derive access control policies from requirement specifications, but none of them focuses on the automatic generation of data protection policies. He and Anton [13] propose Requirements-based Access Control Analysis and Policy Specification (ReCAPS). The method provides guidelines for identifying access rules elements from the requirements specifications and for detecting and resolving conflicts among the rules based on a set of heuristics. Fontaine [11] employs KAOS, a goal-based requirements acquisition and elaboration method, to map requirements specifications into access control policies expressed in Ponder [10]. Liu et al. [16] apply the i* framework [34], a goal-based requirements analysis method, to support access control analysis by modeling the dependencies among actors, tasks and resources. This approach assumes that roles and privileges have been previously defined and derived. Massacci and Zannone [19] propose a formal framework built upon the SI* requirements engineering framework to assist policy makers in the specification and analysis of access control policies. As the PERSONA framework, this approach supports the automatic generation of access control rules from a SI* requirements model. However, the policies generated do not consider elements such as conditions, purposes, and obligations that are relevant for data protection enforcement.

Conflicts Detection and Resolution. Data protection policy analysis is essential to detect inconsistencies and conflicts before the actual enforcement. Matteucci et al. [21] propose a controlled natural language for formally specifying Data Sharing Agreements (DSA). Subsequent work [22,20] deals with DSA usability by presenting an authoring tool for a user-friendly and cooperative editing of DSA and a conflict detection tool. Work in [17] proposes a refined conflict detection technique. In [8], it is shown that the Event-B language (www.event-b.org) can be used to model obliged events. The Rodin platform provides an animation and a model checking toolset for analyzing specifications written in Event-B, thus leading to capability of obligations analysis [4]. Policy conflict detection is generally followed by conflict resolution. Current approaches, like the one adopted by XACML [25] are based on standard, predefined rule-combining algorithms. Jin et al. [14] propose a resolution strategy based on high level features of the policy as a whole (such as the recency of a policy). The approach in [24] proposes a strategy based on a finer definition of the policy specificity. PERSONA aims to overtake existing approaches based on the analysis of fixed patterns of data protection policies and on the application of static kind of resolution strategies. The goal is to increase user awareness on policy enforcement when conflicts are unavoidable.

Model-driven security. Mitigating the risks of data leakage requires adopting in-depth security solutions. Model-driven security [6] has been proposed based on the idea that security needs to be built redundantly into systems. For example, in a web-application, access control may be enforced at all tiers: at the web server, in the back-end databases, and in the GUI. Such defenses in depth are necessary to prevent data access in unanticipated ways, for instance, directly from the database by circumventing the web application server. The work in [7] presents a methodology for designing security-aware GUI which supports the consistent propagation of an access control policy from data models to GUI models and, via code generation, to GUI implementations. A few tools like WebRatio [31], Olivanova [9], and Lightswitch [23] support rapid development of secure data management applications. They apply UI generation patterns to the data model. These patterns enable data retrieval, data editing, data creation, and database search through the UIs. Moreover, these tools support the definition and generation of RBAC policies at different granularity levels. However, existing model-driven methodologies only focus on access control. The PERSONA framework extends them by including concepts that are necessary to capture data protection policies.

Regulatory Compliance and Auditing. Adherence to regulations is becoming a major issue for organizations. This has spurred the design of several frameworks and methods for regulatory compliance. For instance, Accorsi et al. [1] and El Kharbili et al. [15] address the issue of automated certification for compliant business processes. Compliance management is performed at design-time by verifying the process specification against compliance rules. Several compliance checking techniques [12,26,29,32] have been developed to check recorded event data with predefined rules/process models. For instance, van der Aalst et al. [29] propose an LTL-based approach to check whether a process execution satisfies a set of properties. Weidlich et al. [32] propose an approach based on behavioral profiles. Given a set of process executions and a process model, this approach encodes the behavior of both the model and traces in metric representation for comparison. Although these techniques can detect deviations, they do not explicitly identify their root causes, making it difficult to quantify their severity. Moreover, they only locate deviations based on the tasks executed and their control flow. The PERSONA framework uses conformance checking based on alignments to determine the root causes of deviations and the privacy factors characterizing data protection policies to assess their severity.

5 Conclusions

In this paper, we highlight the need of new approaches to data protection and privacy enforcement, where data subjects are in control of who can access their data and for which purposes. We investigate the particular challenges that need to be addressed to give data subjects an active role in the management of their privacy using personalized medicine as an illustrative scenario. We propose the PERSONA architecture which empowers data stakeholders with tools that support them in the specification and enforcement of data protection policies, and the a-posteriori verification of policy infringements.

References

1. Accorsi, R., Lowis, L., Sato, Y.: Automated certification for compliant cloud-based business processes. Business & Information Systems Engineering **3**(3) (2011) 145–154
2. Adriansyah, A., Sidorova, N., van Dongen, B.F.: Cost-Based Fitness in Conformance Checking. In: Proc. of ACSD, IEEE (2011) 57–66
3. Anderson, H.: Personalized Medicine and Privacy - Pairing Genetic Information, EHRs Raises Concerns (2010)
4. Arenas, A., et al.: An Event-B Approach to Data Sharing Agreements. In: Integrated Formal Methods, Springer (2010) 28–42
5. Banescu, S., Petkovic, M., Zannone, N.: Measuring privacy compliance using fitness metrics. In: Business Process Management. LNCS 7481, Springer (2012) 114–119
6. Basin, D., Clavel, M., Egea, M.: A decade of model driven security. In: Proceedings of the 16th Symposium on Access Control Models and Technologies, ACM (2011) 1–10
7. Basin, D.A., et al.: Model-driven development of security-aware GUIs for data-centric applications. In: FOSAD VI. LNCS 6858 (2011) 101–124

8. Bicarregui, J., et al.: Towards Modelling Obligations in Event-B. In: ABZ. (2008) 181–194
9. Care Technologies: Olivanova – the programming machine (2011) http://www.care-t.com.
10. Damianou, N., Dulay, N., Lupu, E., Sloman, M.: The Ponder policy specification language. In: Policies for Distributed Systems and Networks. POLICY '01, Springer (2001) 18–38
11. Fontaine, P.J.: Goal-Oriented Elaboration of Security Requirements. PhD thesis, Universite Catholique de Louvain (2001)
12. Goedertier, S., Martens, D., Vanthienen, J., Baesens, B.: Robust process discovery with artificial negative events. Journal of Machine Learning Research **10** (2009) 1305–1340
13. He, Q., Antón, A.I.: Requirements-based access control analysis and policy specification (recaps). Inf. Softw. Technol. **51**(6) (June 2009) 993–1009
14. Jin, J., Ahn, G.J., Hu, H., Covington, M.J., Zhang, X.: Patient-centric authorization framework for electronic healthcare services. Computers & Security **30**(2-3) (2011) 116–127
15. Kharbili, M.E., et al.: CoReL: Policy-Based and Model-Driven Regulatory Compliance Management. In: Enterprise Distributed Object Computing, IEEE (2011) 247–256
16. Liu, L., Yu, E., Mylopoulos, J.: Security and privacy requirements analysis within a social setting. In: Proceedings of 11th Int. Conf. on Req. Eng., IEEE (2003) 151–161
17. Martinelli, F., Matteucci, I., Petrocchi, M., Wiegand, L.: A formal support for collaborative data sharing. In: CD-ARES. (2012) 547–561
18. Massacci, F., Mylopoulos, J., Zannone, N.: Security Requirements Engineering: The SI* Modeling Language and the Secure Tropos Methodology. In: Advances in Intelligent Information Systems. Studies in Computational Intelligence 265. Springer (2010) 147–174
19. Massacci, F., Zannone, N.: A model-driven approach for the specification and analysis of access control policies. In: Proceedings of Confederated International Conferences On the Move to Meaningful Internet Systems. LNCS 5332, Springer (2008) 1087–1103
20. Matteucci, I., Mori, P., Petrocchi, M., Wiegand, L.: Controlled data sharing in E-health. In: STAST. (2011) 17–23
21. Matteucci, I., Petrocchi, M., Sbodio, M.L.: CNL4DSA: a controlled natural language for data sharing agreements. In: SAC. (2010) 616–620
22. Matteucci, I., Petrocchi, M., Sbodio, M.L., Wiegand, L.: A design phase for data sharing agreements. In: DPM/SETOP. (2011) 25–41
23. Microsoft: Visual studio lightswitch (2010) http://www.microsoft.com/visualstudio/en-us/lightswitch.
24. Mori, P., Matteucci, I., Petrocchi, M.: Prioritised execution of privacy policies. In: DPM, Springer (2012)
25. OASIS: eXtensible Access Control Markup Language (XACML) Version 3.0 (August 2010)
26. Petkovic, M., Prandi, D., Zannone, N.: Purpose control: Did you process the data for the intended purpose? In: Proc. Secure Data Management. LNCS 6933, Springer (2011)
27. Saaty, T.: How to make a decision: The analytic hierarchy process. European Journal of Operational Research **48**(1) (1990) 9–26
28. Spiekermann, S., Cranor, L.: Engineering privacy. TSE **35**(1) (2009) 67–82
29. van der Aalst, W.M.P., et al.: Process mining and verification of properties: an approach based on temporal logic. In: OTM. LNCS 3760, Springer (2005) 130–147
30. Vavilis, S., Petkovic, M., Zannone, N.: Impact of ICT on Home Healthcare. In: Proceedings of International Conference on Human Choice and Computers. IFIP Advances in Information and Communication Technology 386, Springer (2012) 111–122
31. Web Models Company: Web ratio – you think, you get (2010) http://www.webratio.com.
32. Weidlich, M., Polyvyanyy, A., Desai, N., Mendling, J., Weske, M.: Process compliance analysis based on behavioural profiles. Information Systems **36**(7) (2011) 1009–1025
33. Westin, A.: Harris-Equifax Consumer Privacy Survey. Report, Equifax Inc. (1991)
34. Yu, E.: Modeling organizations for information systems requirements engineering. In: Proceedings of IEEE Int. Symposium on Requirements Engineering. (1993) 34–41