A Process for Assisting Privacy-by-Design Software Engineering

Selena Lamari, Nadjia Benblidia, Chouki Tibermacine, Christelle Urtado, Sylvain Vauttier

Started date: April 2021 Expected defense date: June 2024

A Process for Assisting Privacy-by-Design Software Engineering

Selena Lamari[1], Nadjia Benblidia[1], Chouki Tibermacine[2], Christelle Urtado[3], and Sylvain Vauttier[3]

[1] LRDSI, Univ. Blida 1, Blida, Algeria
{selena.lamari,nadjia.benblidia}@univ-blida.dz
[2] LIRMM, Univ. Montpellier, CNRS, Montpellier, France
{chouki.tibermacine}@lirmm.fr
[3] EuroMov Digital Health in Motion, Univ. Montpellier & IMT Mines Ales, Ales, France
{Christelle.Urtado,Sylvain.Vauttier}@mines-ales.fr

Abstract. Today, the mine vast troves of personal data contained in applications raises the issue of user privacy. Indeed, privacy is increasingly threatened by the spread of unethical practices by device and service providers. Despite the existence of privacy preservation standards such as the European General Data Protection Regulation (GDPR), effective since 2018, there is no widely adopted architectural solution for modeling and assessing privacy by design (PbD) of personal data, as proposed in the various principles of the GDPR. This article presents PRIvacy Assessment Model (PRIAM), which is an approach composed of a GDPR metamodel tooled with a Domain Specification Language and supports a process to protect personal data. The metamodel can be instantiated by architects and integrated in the design of their system, with minimum additional efforts to ensure compliance.

Keywords: Privacy by design, Personal data, GDPR metamodel, Domain Specific Language, Architecture design

1 Introduction

As the use of intelligent environments increases, they process a huge amount of data which is often personal and confidential. These environments offer a better comfort to users. For example, an environment can offer users remote nutritional coaching, combining devices such as a connected scale, a connected watch and a connected bike. An illustration of this example is shown in Figure 1. The data collected from these objects is sent directly to the provider's server, which considers himself the owner of this data and gives himself the possibility to exploit them for derived uses. These uses include: user data extraction, data transmission to other data consumers, automatic user profiling or personalized prospecting proposal. These are considered illegal practices without the consent of the data subjects.

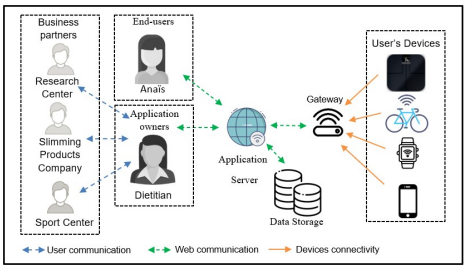

Fig. 1. Remote nutritional coaching application example.

Ethical actions must be taken to ensure the protection of users' privacy. We can refer to the European GDPR (General Data Protection Regulation [3]), effective since 2018. The principles of the GDPR are not always respected [2]. The data subject cannot enforce his/her rights mentioned in the GDPR, such as the right to rectification, the right to be forgotten, the right to object, the right to restrict processing.

Since regulations are legislative texts, it is very difficult to translate them into software practices. Cavoukian [5] emphasizes the notion of privacy by design and proposes a set of functional requirements in the form of seven principles. These principles should be taken into consideration for any conceptual approach aiming to protect personal data as an integral component of any application. For this, we propose PRIAM, a user-centric architecture that ensures the ethical use of personal data according to the GDPR and this from the design stage. The proposed PRIAM approach implements GDPR compliance while at the same time being generic, thus allowing to take any privacy policy into account.

The remainder of this paper is as follows. Section 2 gives an overview of the proposed method. Section 3 discusses the related works on privacy management. Section 4 summarizes the contributions of this paper and draws some perspectives.

2 PRIAM: An architecture privacy enforcement approach

In order to improve any system with a privacy protection mechanism according to the GDPR from the design phase, we propose a method consisting of three parts: a metamodel encompassing a description of GDPR concepts, a **Domain Specification Language** that implements the **metamodel** and a **process** ensuring privacy requirements, supported by the DSL.

2.1 The PRIAM metamodel for architecture privacy enforcement

We propsed at first PRIAM metamodel (*cf.* Figure 2). It represents the structural part of the privacy requirements of applications as prescribed by the GDPR.

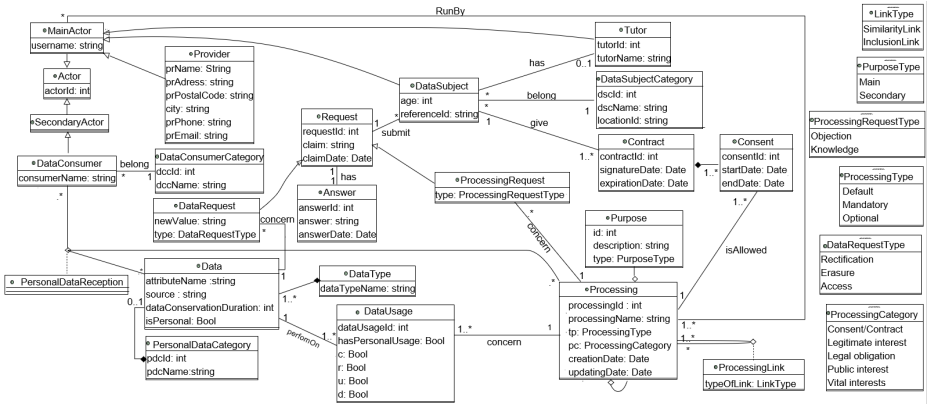

Fig. 2. The PRIAM privacy management metamodel.

The proposed metamodel concepts will be instantiated to **model the privacy enforcement components**, **generate the digital contract** between the provider and the user and **record processing activities**.
In our Metamodel we distinguish several concepts, like Actor, Processing, Data and Consent.

The actors of a system are represented by the *Actor* class. A distinction is made between *MainActor* (internal actor to the system) and *SecondaryActor* (external actor to the system). The *Provider*, the legal entity embodied by its legal representative, and the *DataSubject*, the end user whose personal data may be collected, are the main actors. In case the data subject is under 16, the person with parental responsibility must be designated, this actor is represented by the *Tutor* class. A *DataConsumer* is a secondary actor. He can be a natural or legal person entitled to obtain communication of the collected, recorded and/or produced data. He/she can be a partner, a processor, a legal authority, a supervisory authority or a third party.

All application data are identified and classify it as personal or non-personal. We hold the description of each of the data collected, stored, processed and/or disseminated by the application. We also keep all the links between the processings and their used data, personal or non-personal, and listing the authorized operations on this data.

The functionalities of the application are represented by the Processing class. Processing is an operation or a set of operations performed on (personal) data. According to the GDPR, a processing belongs to a *ProcessingCategory* (Art 6). The purposes of each processing operation must be defined in a clear and precise way. We add the notion of type and *ProcessingLink*. Three types of processing are distinguished. *Default processings* use only non-personal data or use personal data but do not require the consent of the data subject. *Mandatory Processings* use personal data and are necessary to the application. *Optional Processings* use personal data and the data subject can accept or refuse to give his/her consent,

withdraw consent or oppose to legal uses of its personal data as long as it does not contravene public interest (GDPR, Art 21.6). The *processingLink*, which can be a link of similarity or inclusion, allows better management of consents. For both links, a single consent is sufficient.

The data subject's *consent* must be given to allow the provider to run a processing using personal data for a defined period of time. The data subject must be able to withdraw consent any time. He/She may also request to enforce his/her other rights: rectification, erasure or portability of data and restriction or objection to processing. A *request* is followed by a response from the provider as a Notification (GDPR, art. 16-21).

2.2 PRIAM DSL

To address the challenges of effectively enforcing GDPR requirements throughout applications lifecycle, we envision leveraging commodity modeling tools used specifically to a domain. We selected JetBrains MPS [1] as one of the core technologies which is one of today's most popular metaprogramming environments. The language definition starts by defining its abstract syntax. Each concept can have a definition in one or more aspects of the language such as structure, editor or generator. The PRIAM DSL is used to:

1. Define GDPR concepts (corresponding to classes in the metamodel), their properties and relations. Figure 3 shows the definition of the Processing concept.

```
concept Processing extends    BaseConcept
                   implements INamedConcept

  instance can be root: true
  alias: processing
  short description: <no short description>

  properties:
  id           : integer
  tp           : ProcessingType
  pc           : ProcessingCategory
  createDate   : string
  updatingDate : string

  children:
  purposes : Purpose[1..n]
  dataUsed : DataRef[1..n]

  references:
  << ... >>
```

Fig. 3. Definition of Processing concept

2. Generate SQL statements to create an instance of PRIAM metamodel (part of the USER's personal data annotation presented in Figure 4). Annotate all needed information about the application (actors, processing, personal data, etc).

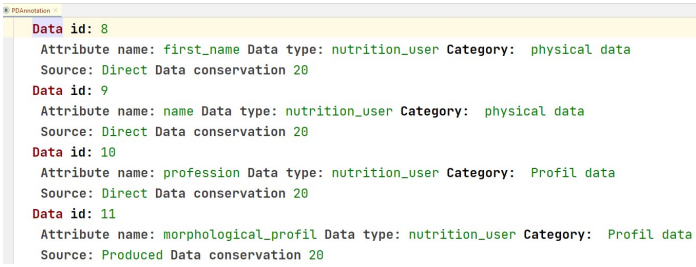

Fig. 4. Personal data annotation (extract)

3. Generate documentation containing new requirements as generic user stories corresponding to a data subject's rights according to the GDPR User stories (Figure 4 shows an example of a generated user story) .

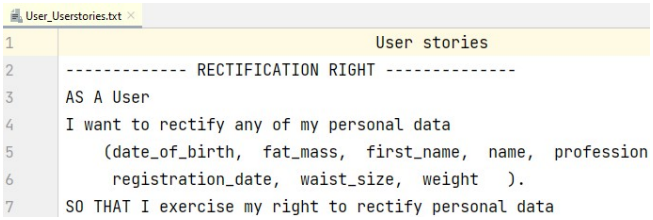

Fig. 5. A User story generated for assisting privacy enforcement in software development

Then, we evaluated and tested our language created with a sandbox tool, which contains the final user code. In order to ensure the protection of users' privacy of the application, we tested our DSL on a nutrition coaching application.

2.3 The PRIAM privacy enforcement development process

PRIAM process is presented in Figure 3. The process has two main parts: Design time (Privacy-enhanced Architecture design) and Runtime (End-user configuration and exploitation).
The design time part of the process is divided into three steps:

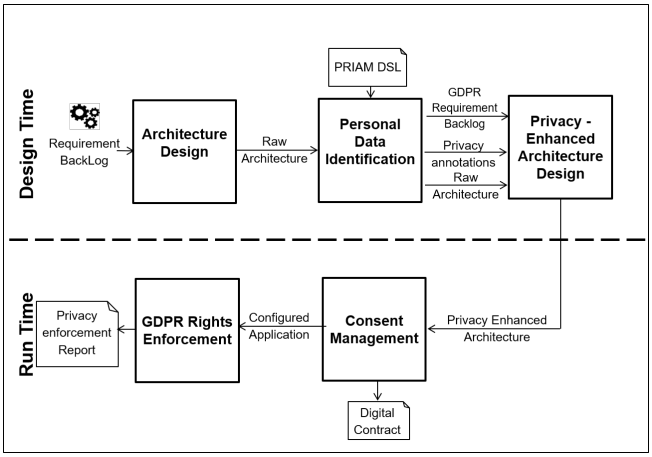

Fig. 6. Overview of the PRIAM privacy enforcement development process.

1. **Architecture design.** Architecture design is a traditional development step that enables to elaborate solutions to a set of functional and extra-functional requirements. The result of the architecture design step is a software architecture ready for privacy enhancement. The following steps will refine and extend the architecture to enforce the privacy of the users' personal data.
2. **Personal data identification.** Based on the PRIAM DSL, the architect is assisted to identify and annotate all information needed to the GDPR compliance (data subjects, business data classes whether they are personal or non-personal, processings, etc.).
   At the end of this step, the part of the PRIAM Metamodel related to annotations is instantiated. The other part of the metamodel, related to users' preferences regarding the protection of their privacy, will be instantiated at runtime during the Consent management step
3. **Privacy-enhanced architecture design.** The *GDPR requirements backlog* is composed of generic requirements related to the GDPR principles and the management of user rights. These are designed and integrated into the architecture as modules that interact with the business processes to manage privacy enforcement.
   Module design may be assisted in different ways, for instance thanks to prescribed architectural styles and patterns associated with generic GDPR backlog elements or the reuse of configurable modules in a software product-line approach.

The runtime part of the process has two steps:

1. **Consent management.** The data subject is informed about all the processing performed on his/her personal data by the application. The data subject must explicitly give his/her consent for each processing. Consent allows a processing to access specific personal data. Otherwise, the processing

is blocked. The user preference part of the meta-model is instantiated to define the consents and configure how privacy is managed by the application. Once the data subject's privacy preferences are defined, a digital contract between the provider and the data subject is generated. It can evolve at runtime each time the data subject changes his/her privacy options (*e.g.*, withdrawal of consent or erasure).

2. **GDPR rights enforcement.** This step concerns the exploitation of all processings that have been consented by the data subject.
   At last, a report on all activities performed on personal data will be automatically generated to audit privacy enforcement.

3 Related work on privacy modeling

Several works have tried to put into practice solutions to ensure privacy by design. Others propose a formalization of the GDPR with a compliance checking mechanisms for existing applications.

[4, 6–8, 10] focus on **GDPR modeling**. They provide a formal representation of some rules of the GDPR to improve their understanding by architects and therefore ease GDPR compliance verification of their applications with the GDPR upstream. They support neither Privacy by Design, nor Privacy by Default, which makes them not compliant with Article 25 of the GDPR.

[9, 11, 12] specifically target **Privacy by Design**, basing their solutions on frameworks, privacy patterns or ontologies without specifically applying the rules and principles of the GDPR.
Our work is mainly motivated by the integration of legal requirements referring to compliance with GDPR and Privacy Enhancing. We propose PRIAM, a tool based on a metamodel enabling the integration of Privacy according to principles and users' rights required by GDPR to applications.

4 Conclusion

This paper proposes PRIAM, an architecture privacy enforcement approach for applications. This approach covers all GDPR concepts. They are formalized in our metamodel which supports the identification of privacy sensitive data and processings. This Metamodel is then tooled with a DSL to be used to assist the design and integration of GDPR related modules to enhance the architecture of applications with adapted privacy enforcement features. These features will enable users to express their consent to use their personal data in the context of each specific processing and application.

This paper is a preliminary description of our proposed PRIAM approach. Many perspectives remain open for future work. First, we aim to further develop the nutritional coaching example as an illustration application. We also plan to implement a proof of concept to demonstrate the feasibility of our method and refine it. In addition, we plan to identify architectural styles, patterns and modules that can be reused as generic design good practices in a software product-line approach for privacy enforcement.

References

1. MPS: The domain-specific language creator by jetbrains. `https://www.jetbrains.com/mps/`. (Accessed on 03/11/2022).
2. What is GDPR, the EU's new data protection law?- gdpr.eu. `https://gdpr.eu/what-is-gdpr/`. (Accessed on 12/07/2021).
3. General Data Protection Regulation (GDPR) compliance guidelines. `https://gdpr.eu/`, 2020. (Accessed on 12/07/2021).
4. Fabian Burmeister, Paul Drews, and Ingrid Schirmer. A privacy-driven enterprise architecture meta-model for supporting compliance with the general data protection regulation. *Proceedings of the Annual Hawaii International Conference on System Sciences*, pages 6052–6061, 2019.
5. Ann Cavoukian. Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada*, 5:12, 2009.
6. Vasiliki Diamantopoulou, Konstantinos Angelopoulos, Michalis Pavlidis, and Haralambos Mouratidis. A metamodel for GDPR-based privacy level agreements. In *ER Forum/Demos*, volume 1979, pages 285–291, 2017.
7. Sabrina Kirrane, Javier D Fernández, Piero Bonatti, Uros Milosevic, Axel Polleres, and Rigo Wenning. The special-k personal data processing transparency and compliance platform. *arXiv preprint arXiv:2001.09461*, 2020.
8. Raimundas Matulevičius, Jake Tom, Kaspar Kala, and Eduard Sing. A method for managing GDPR compliance in business processes. pages 100–112. Springer, 2020.
9. Sebastian Pape and Kai Rannenberg. Applying privacy patterns to the internet of things' (IoT) architecture. *Mobile Networks and Applications*, 24(3):925–933, 2019.
10. Jake Tom, Eduard Sing, and Raimundas Matulevičius. Conceptual representation of the GDPR: model and application directions. In *International Conference on Business Informatics Research*, pages 18–28. Springer, 2018.
11. Amri Toumia, Samuel Szoniecky, and Imad Saleh. ColPri: Towards a collaborative privacy knowledge management ontology for the Internet of Things. In *Fifth International Conference on Fog and Mobile Edge Computing*, pages 150–157. IEEE, 2020.
12. Fatima Zohra Benhamida, Joan Navarro Martín, Oihane Gómez Carmona, Diego Casado Mansilla, Diego López de Ipiña, Agustín Zaballos Diego, et al. PyFF: A fog-based flexible architecture for enabling privacy-by-design IoT-based communal smart environments. *Sensors*, 21(11), 2021.