Privacy Rule Definition Language -
A Multistakeholder Approach to ENDORSE Privacy

T. KURZ a, C. RUECKER a, T.J. LAMPOLTSHAMMER a and T. HEISTRACHER a

a Department of Information Technologies and Systems Management, Salzburg University of Applied Sciences, A-5412 Puch/Salzburg, Austria

Abstract. The right of privacy of personal data is fundamental to democratic societies and self-determined individuals. The legal fundamentals regarding personal data privacy, however, are not reflected in current technology platforms or data processing systems like databases in a systematic way that is entirely transparent to all the users of such systems. This work introduces PRDL (Privacy Rule Definition Language) as a basis for user-oriented notation of data processing rules that can be automatically processed by IT systems to certify data protection compliance in the long run.

Keywords. Data privacy, rule enforcement, legal compliance, data processing rules

Introduction

Data privacy is a topic which has become an important topic, not only concerning the data subjects (end-users) but also for the data controllers (organisations) and the legislative bodies. Furthermore, data protection agencies, consumer rights organisations and human rights advocates have a high interest in systematic processes towards protection of sensitive data. End-users are entitled to decide what is going to happen with their personal data and expect compliance with their decisions. Therefore, processing of personal data has to be done in a more transparent way [9]. These circumstances put pressure on affected organisations due to the fact that they have to design their data processes to be law-compliant and additionally to reduce the possibility of intentional and unintentional misuse by their employees. These statutory requirements lead to a financial and operational overhead, which can be a heavy burden for small and medium enterprises in terms of competitive global markets. This is where the ENDORSE project and the Privacy Rule Definition Language come into play.

1. Environment and Development

This part of the paper provides first insights into the challenges of the ENDORSE project, stakeholders as well as its requirements. In the description, the origin of the project is presented together with its internal structure and its main objectives. The section dedicated
to the stakeholders outlines the target audience and their corresponding goals. Following, the requirements of the project will be discussed in brief.

1.1. The ENDORSE Project

ENDORSE is funded by the European Commission within the Seventh Framework Programme and is concerned with providing a legal-technical framework for privacy preserving data management. The output of the project will be an open-source toolset, which provides guarantees to data controllers as well as data subjects (end users) that personal data is being handled in a legally compliant manner. The project will also produce a certification methodology to help to increase trustworthiness in ICT products with respect to privacy and data protection. ENDORSE has brought together a consortium of data protection legal experts, academic computer science partners, software implementers and interested industry players from Ireland, UK, The Netherlands, Spain, Austria and Italy.

To achieve these goals, the project is working on the introduction of a privacy rule definition language (PRDL), which will be used to express the appropriate European directives together with national legislative implementations. The language and these legislative instances along with the toolset, both released as open-source, will allow the user to create legally compliant privacy policies. A set of open-source technology adapters will take these policies as input and produce access control specifications for deployment within organisational infrastructures. Two industry players will perform trials using this toolset, a large multi-national insurance organisation and a start-up web based organisation providing voice and video communication services.

1.2. The Stakeholders

This section reflects the outcomes of the identification process regarding the main stakeholders and roles within and outside the project’s consortium including priorities and the resulting inputs for PRDL. For this task, each consortium partner was asked to write out a set of goals related to the type of user following this pattern:

As a <type of user>, I want <some goal> so that <some reason>.

As the ENDORSE consortium’s vision and the roles of stakeholders are considered essential for PRDL as well, the process was drafted here for defining the environment where PRDL will be used in the future. The consortium identified 20 roles, which are potential stakeholders of the ENDORSE project, namely:

Manager, Company, Controller, Employee, Regulator, IT Manager, Policy Creator, Company Executive, Interest Group, Consumer Group, Journalist, Police / Military, Doctors, Victims, Consultant, Person/ End-user, Researcher, Bad Guy, Lawyer, Developer.

Afterwards, stakeholders perceived most important were chosen, which resulted in three main groups: end-users, organisations and regulator/policy creator/law enforcement. The following describes the primary goals of these groups:

End-user: The end-user should be empowered through choice and control accompanied by transparent processing of user-related data to be aware of errors and possible countermeasures. The process of empowering can be achieved by the use of state-
of-the-art technology, European data protection laws as well as standardisation, its supervision and continuous improvement.

Organisation: The organisation has to be equipped with tools for handling sensitive data and being compliant with the European Data Protection Directive (DPD) [1]. These tools have to include a domain-specific language to express data access and data handling policies. There should be the possibility to express these policy statements in a human-readable form as well as tracking possibilities of data accuracy and data access.

Regulator / Policy Creator / Law enforcement: The regulators have to be able to trust the ENDORSE ‘seal’ assuring that there are no false positives/negatives. Furthermore, they should continuously work towards compliance in a transparent manner empowering end-users and companies to better understand the regulation process and making it easier to implement and affordable for them.

This first draft of goals and requirements acts as the basis for the project vision and subsequent work. To make the usage of planned components and the roles of stakeholders clearer, the following problem space is discussed in which new solutions are developed towards privacy policy handling and related automation. It should provide a general conspectus of the use cases and specific visions of the consortium. Furthermore, it acts as the basis for a first prioritisation process for identifying the starting points for the implementation of PRDL and the consequent ENDORSE components.

The description below provides an overview of the scenarios and the logical structure, based on the requirements so far. The example relates to the fact that users are often not aware of the risks of providing personal information. A user Stefania registers at a webpage and provides personal information in order to consume a service. This process starts to become an issue, if a third party Kevin has access to this data and especially, if Kevin uses this data to bother or harm Stefania. In order to avoid data misuse, national and international legislation and policy makers, e.g. Poliandro, introduced directives and national implementations for data privacy and therefore is acting in the interest of Stefania. Issues here are the differences in national implementation and its complexity, especially the fundamental challenges like ambiguity, traceability and accountability of law as such. The challenge, the project is focused on, is ambiguity, which can be distinguished into three different categories [8]:

Logical ambiguity: This type references to words e.g. in English that can be logically interpreted in different ways.

Attributive ambiguity: This type describes the possibility to assign phrases within a sentence to one or several parts within that sentence.

Referential ambiguity: This type occurs if it is not clear to what exactly a word or phrase is referencing and can therefore offer multiple interpretation.

Many initiatives in the last years tried to formulate laws in a set of computable rules [2,3,4]. Besides the complexity of law, the liability of translations from law to rule languages is critical. In other words: Who is liable for a translation into a rule set, if it does not match 100% of the legislation due to ambiguity.

After the end-users’ requirements, the requirements of companies/businesses were identified as second important ones. While the user has an interest in minimising the data necessary for a registration process, the organisation typically wants to learn as much
as possible about its users in order to provide customised services. Organisations are obliged to be compliant with the national and international data protection directives and laws. The representative of the stakeholders of the legal departments, Giulia, is responsible for the legal compliance and interpretation of the law. Massimo represents the technical implementation stakeholders at the organisations, those need to implement, e.g. data access at the back-end systems. The main requirements of the organisations herein are: i) Rules need to be derived from national and international legislation and ii) ENDORSE components should regulate how applications can access data objects.

To complete the picture of the company, DocCharly visualises third party contractors, who need to access data of a data subject. Within the ENDORSE consortium, the multi-national insurance company provides a health care service name Docticare to their customers. Within Docticare, a data subject might want a doctor (e.g. DocCharly) to access its data. When Stefania travels to Germany and needs some medical advice, a local hospital or doctor could need to access her sensitive personal data in Docticare in order to perform a proper treatment. Based on the results of empirical research conducted within the project, ENDORSE will provide Stefania more transparent and easy-to-understand information on the usage and privacy of her data, when registering at an online system. One possibility is colour-coded symbols, which could be automatically encoded based on PRDL statements. The end-user verification tool will be the user front-end of this information and helps Stefania to understand the data privacy statements of ENDORSE-enabled online systems. The requirements of Stefania are i) easy to understand privacy statements, ii) transparency what happens with the data and, iii) assurance that the actions are implemented according to law and privacy statements of the service provider.

1.3. The Requirements

The upcoming part discusses the heterogeneous stakeholders in this project and builds a common ground on the usage of PRDL. By defining deliverables on requirements, language specification and rules engine in a very short time period, a fast prototyping approach was taken in order to test the first implementations as soon as possible. As the interest and the requirements of the stakeholders of the project are very heterogeneous, we want to summarise the most important ones structured by stakeholder groups. In the following paragraphs, a short summary of the collected requirements is given:

- PRDL should be machine-readable
- PRDL should be based on and be compliant with common standards
- PRDL should be extensible
- PRDL should hide the technical details of the back-end, e.g. rules engine and therefore be human-readable
- PRDL or a transformed version of PRDL should support rules processing, e.g. execution, conflict handling and combination of rules

The results of the survey performed by the social science team provided interesting insights in the expectations and capacities of the end-users. Although PRDL is targeted at describing rules of privacy policies based on legal regulations and not on privacy statements, we could imagine a direct extraction of information from PRDL, which will be presented by text, symbols or icons to the end-users. The business requirements are collected mainly from the two companies in the project consortium. One of them also ranked the requirements and identified the two most desired ones:
a. PRDL rules should be derived from legislation
b. PRDL rules should regulate data access in day-to-day business.

Summarising the requirements for rules, those can be structured in general rules, cron job rules (periodic processing) and access rules. General rules are only for manual lookup; therefore they do not trigger actions, e.g. minimum length of passwords in the company. Cron job rules trigger actions without a triggering action, e.g. data needs to be deleted after 10 years. Access rules are standard data access rules which are triggered by a request and trigger themselves other actions or rules. Legal requirements provided mainly the legal consortium members, focus on the language elements needed for formalising legal texts into rules and additionally show which metadata and tagging information is necessary to identify and track the related legal texts and their hierarchy. Concluding the requirements process so far, the two most critical points seem to be the user interfaces and the difference of expressions in legislation on the one hand and computation and access control on the other hand. The ease of using and handling the ENDORSE toolset can be considered essential for the usage of the tools. Even if there are already XML-based languages available, the implementation of editors which are usable for non-technical experienced users lack behind. Additionally, the automatic transformation of PRDL expressions which define the data access, into something like icons or symbols which are easy to understand for end-users is challenging. One of the problems faced is while policies can formulate rules with modalities, an IT system is not able to interpret these. Additionally, the policy rules might be ambiguous or use terms which are not computable. This is also reflected by the fact that other projects in the data protection field are neither concentrating on the formulation of legislation, e.g. MetaLex [5] nor the formulation of access rules based on policies, e.g. XACML [6]. The automatic, semi-automatic or even manual (supported by ENDORSE) matching of policies and access rules therefore is the second main challenge of PRDL.

2. Languages and Evaluation

After the introduction of the project itself, its stakeholders and its requirements, this part of the paper focuses on the evaluation process of existing languages for the formulation of access rules based on policies. Therefore, a wide range of candidates was evaluated in order to find the most suitable one. These are:

- X-Author, FASTER, XrMi, ODRL, XACL, SAML, XACML and SecPAL

Due to restrictions of the project - for instance to use only open-source software - some candidates had to be eliminated. Furthermore, for some candidate languages it turned out that they are focussed on different application areas not compatible with the targets of the project.

Concerning the remaining XML-based framework candidates, the following part will summarise features of the languages regarding the key aspects system, approach and security. Combining the intended goals from the project and stakeholder description it can be stated that a suitable solutions should include several of the following properties:
After consideration of all listed aspects of the former described candidates, it can be stated that XACML and SPL are the best fitting proposals. In direct evaluation, XACML has some drawbacks which could be solved. One of the major aspects here is in the security area and distribution. This could be resolved by the use of XML databases for secure storage and distribution. Another aspect is the validation of semantics, which could be addressed by the editor with the help of dropdown fields, providing correct syntax and ensure correct semantic. In summary, the XACML solution prevails due to the demands to PRDL regarding the prioritisation of the stakeholders and their goals. There it is stated that standardisation has a very high priority, combined with interoperability. Due to the fact that XACML is an internationally accepted industrial standard, the demand could be matched best with XACML as PRDL candidate.

3. Privacy Rule Definition Language (PRDL)

To be able to represent privacy policies, PRDL has to be designed to have enough expressivity to cover the most important concepts of data privacy. It should avoid a large vocabulary which probably makes the usage too complicated and unhandy. With the usage of PRDL, the two main objectives of the ENDORSE project will be tackled. These are more transparency for the data subject (the one whose data are processed) and legal compliance with European directives and national legislative implementations for data controllers (responsible personnel in the organisations). PRDL has been designed to find the right balance between expressive power and efficiency - expressive power in terms of being able to formulate large parts of the relevant data privacy regulations and privacy policies, efficiency meaning the usability and applicability for small and medium-sized enterprises (SME) and consequently for all types of users.

PRDL in general has a very broad field of application which spans from the end-user over the organisation-user to the policy creators. Moreover, the language has to meet all the user skills concerning usability, understand ability and extensibility. As the focus of the language lies mainly on designing rules for data handling of privacy appreciable data, all the relevant constructs within data privacy have to be considered. The rules should be capable of expressing the usage of data within a business context. In addition, the relevant legal aspects of such a data handling should be expressible. The language has to support purpose [10] within the data handling context, as data collection has to be purpose-driven by law.

The design of PRDL is accomplished in two meta-models which are the Meta-access Model and the Meta-Rule Model. On the one hand, the meta-access model explains and
illustrates the creation of a request for data to the ENDORSE system and the relevant data roles in this requesting process. On the other hand, the meta rule model depicts the layout of the PRDL rule and all the language constructs needed to express a data privacy related rule.

3.1. Development Process

PRDL heavily benefits from the different and sometimes even opposite viewpoints of the stakeholders. At the same time, we want to emphasise that the communication and the requirement gathering process in a distributed multi-stakeholder project is extremely challenging. Beside the parallel research on already existing approaches for computable rule languages and tools [11], the company members of the consortium provided the first trial. This was followed by two on-site meetings with the trial partners where the requirements, documented in the following sections, were identified, gathered, discussed and revised. Additionally, a set of scenarios was developed. Research on already existing approaches, regular telephone conferences and chats, Emails, consultancy of experts in the data privacy field as well as collaboration on the project’s wiki supported the requirement gathering process, which led to a set of requirements. One of the main difficulties was the fact that many dependent deliverables which act as a source for the requirement gathering process were due at the same time or even two months after. Thanks to the other partners, we could get the necessary information already in advance. The reason for the directly following preliminary specification is the need for a first set of language elements to start the developing of the Rule Editor Prototype as well as of other dependent tools and components, e.g. the language repository and rule engine implementation. After the first presentation to the project reviewers and the project officer in autumn 2011, the priorities for PRDL will be revised by the consortium and a cyclic prototyping process can start. We suggest having several iterations of a) extending the language specification and prototyping the respective tools, and b) review and field-testing the tools against real world policies and regulations, based on the relevant meta-model.

3.2. Meta-model

In a nutshell, the meta-access model, shown in Figure 1, describes how a request to the system has to be composed in order to enable the rules, developed with PRDL to evaluate the request. The model should fit into the architecture developed and should provide help with the understanding of the rule enforcement. The model will be used for showing the interactions between the PRDL rules and a request within the planned demonstrator. To give an overview the model is described shortly in the following as an explanation to the given Figure 1.

The root element of the meta-access model is the Request. A Request contains the three elements which are the RequestPurpose, the requested Data and the reference to the DataController, who executes the Request. The rule engine will have to handle this Request and either grant or deny the access by evaluating the Request against the stored PRDL rules. The model includes two types of Purposes, the RequestPurpose and the Purpose. These two types are needed to distinguish from the Purpose for that Consent has been given by a DataSubject. A RequestPurpose can also be legitimated by legal obligations and other reasons. A DataController always has to justify a Request for data
by adding a RequestPurpose. A Request can include more than one RequestPurposes, though has to have at least one.

A Request naturally has to include a Data element. This element represents the Data elements which the DataController wants to use for a specific Purpose. It is composed out of a Consent, Constraint, Purpose, ProcessingGround and a list of DataObjects. Data is additionally a container for a subset of DataObjects that describe the Data in greater detail. A Data element must have at least one DataObject and Consent and can have one or more Constraints.

The DataObject object is the atomic representation of a part of personal data and it represents the smallest part of information within a Data element. (E.g. telephone number, first name, last name, etc.). In case of special circumstances, that prevent certain access or actions on the Data element, a Constraint element is defined. The Consent element in every Data element grants that the DataSubject has given his consent. As Purpose is not the only element that can legitimate the processing of data the ProcessingGround element is introduced. It specifies further reasons for processing the Data such as legal obligations etc. Additionally, the element Purpose specifies the reason why Data is collected and processed. It represents the action to be performed on Data that the DataSubject or another legitimation has granted. There are three actor elements in meta-access model. These are the DataSubject, DataProcessor and DataController. They are all related to Data and therefore the DataRole super element was created to interlink these elements. The DataSubject has one or many DataElements assigned to it. As a central actor, the DataController creates a request for Data. Additionally every DataController has a DataProcessor assigned which is responsible for the executed actions on the Data. The meta-access model is under active development and might be change to tackle future challenges within the project.
3.3. PRDL XML Base Structure and Rule Templates

The first steps within the PRDL developed resulted in an XML schema leading to a base structure for the rules. Additionally, rule templates in natural language were created to emphasis the usability of the PRDL approach. Before starting with the PRDL rule templates and according rule examples in XML, the base structure is given in the following. In the subsequent examples the XML header will be dropped to assure a better readability.

In the following, the above-mentioned templates are presented. They are the results from a requirements analysis within the sector of privacy data management and provided inputs by the project consortium. An example for the first template will be given in the next section. First, elements within [] are rule elements implemented in the language; second, elements within {} are single choice elements also implemented within PRDL; and third, elements covered within () are fill elements used to provide understandable natural language. As PRDL is developed in an on-going process, the templates may be changed and extended in the future.

Template for normal privacy data access rule

\[
\text{[DataController]} \ \text{MUST, MAY VIEW, ADD, DELETE, MODIFY, STORE, ACTION} \\
\text{[Data Object]} \ (\text{FOR}) \ [\text{Purpose}]
\]

Template for a normal access rule with a constraint

\[
\text{[DataController]} \ \text{MUST, MAY VIEW ENTER, DELETE, MODIFY, ACTION} \ [\text{Data Object}] \ (\text{FOR}) \ [\text{Purpose}] \ (\text{IF}) \ [\text{Constraints}]
\]

Template for data access rules with time constraints

\[
\text{[DataController]} \ \text{MUST, MAY VIEW, ADD, DELETE, MODIFY, STORE} \ [\text{Data Object}] \ (\text{FOR}) \ [\text{Purpose}] \ (\text{WHEN}) \ [\text{TimeConstraint}]
\]

4. Conclusion and Outlook

The paper at hand presents the on-going work on PRDL and shows how the ideas and inputs of the preliminary requirements are formed into a novel language. Essential in this process was the bidirectional communication with the project consortium in terms of trial scenario development, further language inputs and reality checks concerning existing languages. Finally the paper presents the initial steps towards a framework which ensures the fair and lawful usage of personal data in organisations. Bridging the gap from the requirements, the language section deals with candidate languages and their impact in the computer law society as well as requirements and goals these languages focused on.

One aspect that all languages had in common was the usage of XML-based schemata as a base for the language implementation. Although a lot of work was conducted in the area of legal normalisation and computerisation, none of the candidate languages complied with all the requirements that PRDL had. Therefore the decision to create an own XML-based language as a first step was taken. As the development of a new language always involves the creation of an editing environment an XML-based language and XML editing environments can never be seen as user-friendly. Initially PRDL was envisaged as a language that is useable even for non-technical end-users outside of the information technology business. To meet these requirements additional syntax templates which are parse-able into PRDL XML were created. These can be included into the first editor
prototype to enable the real world testing of the language. The template development is not finished and there will certainly be some elements or even new templates added to enable a complete coverage of occurring use cases. To emphasize the valid definition and interpretation of the language elements that occur in the templates a Glossary was created. All the relevant elements were added and the ENDORSE consortium added the definitions that are considered as the right ones in the context of the project.

To cover the main goal of PRDL which is to describe rules which can be executed with a rule engine, a further step was unavoidable. In respect to the available resources, an existing rule engine has to be used rather than developing an own, proprietary one. Therefore it was decided to transfer the XML represented rules into the Drools [7] rule language as a first step towards a working prove-of-concept. As Drools is fully generic in terms of used meta-model it was ideal for a first step towards a fully functional rule language.

The next steps will be to show that all the relevant constructs needed for implementing privacy policies in PRDL can be covered by Drools or maybe another rule engine. These analysis and try-outs will be the foci of the future of the ENDORSE project. Finally, this paper presents the way that PRDL has evolved from a language defined by the elements and constructs needed to represent privacy policies in a computerised way into a XML based language which can be written using syntax templates and which is parse-able into an existing rule language. The future work will include the development of a full rule live cycle starting with the creation of rules through a legal person over to the execution within the rule engine and all consequences triggered from the rules.

References


