Catalog of Security Requirements Aligned with The General Data Protection Regulation (GDPR) (EU) 2016/679

Mário António Fernandes
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa
Lisbon, Portugal

ABSTRACT

The European Union establishes in Regulation 2016/679, or GDPR (General Data Protection Regulation), a set of legal provisions to achieve the protection of natural persons in relation to the processing of personal data and the free movement of such data. When such provisions are considered in the development of information systems, the latter become attainable for legal approval within that scope. This dissertation presents the methodology used to elaborate a reusable catalog of personal data protection requirements aligned with the GDPR. Following a separation of interests approach, the catalog should serve the purpose of constructing information systems capable of communicating with those processing personal data of individuals to materialize the regulatory data protection capabilities set out in the GDPR. In this context, the elicitation of system requirements requires the interpretation of a legal document by business analysts, which is a scientifically relevant challenge. This research is contextualized by the ITLingo initiative, a model-based requirements engineering approach to the strict specification of system requirements, which supported the extraction of GDPR requirements and the transformation of these requirements into a catalog of data protection requirements detailed in business objectives and systems, which can be adapted by organizations and ensure alignment with the new regulation.

Keywords: Personal data protection; data processing; requirements specification; ITLingo; Regulation (EU) 2016/679; GDPR

1. INTRODUCTION

The interconnection of information systems via the Internet shows the increasing scale of the public electronic market in the world, which makes accessible to untrusted users. Attacks can be locked anonymously and from a safe distance, if the internet is to provide the platform for business transactions, it is vital that confidential information (such as credit card numbers or cryptographic keys) is stored and transmitted securely (Jürjens, 2004). Therefore, laws and regulations have been established in terms of the elaboration of legal texts that have an impact on the requirements engineering itself in the development of information systems, with the purpose of guaranteeing the protection of the data processed by these systems, through the application of measures techniques. In the view of (Breaux, 2009) in a society based on the rule of law, individual and organizational actions are governed by laws that serve to achieve social goals. Laws governing information systems represent significant compliance challenges for relevant stakeholders, including corporate executives, lawyers, and software engineers.

The European Union (EU) establishes in Regulation 2016/679, the General Data Protection Regulation (GDPR), a set of legal provisions to guarantee the protection of natural persons with regard to the processing of personal data and the free movement of data (European, 2014). This legislative act is a legal piece immediately applicable and enforceable by all Member States, with entry into force scheduled for May 2018. Therefore, all information systems for data processing must be in line with the provisions established in this regulation and the organizations must adapt or develop their systems in accordance with these obligations characterized by legal requirements, otherwise they will be subject to penalties for non-compliance with the law.

In this context, obtaining software requirements requires the interpretation of a legal document by business analysts, which is a scientifically relevant challenge. This Dissertation proposes a Security Requirement Library aligned with the General Data Protection Regulation (GDPR), which should serve as a catalog for the construction of concrete information systems that process personal data. This tool should be useful to requirements engineers and systems developers in ways that adapt the requirements contained in their systems to ensure data protection in accordance with the regulations and turn them attainable for legal approval within the scope of personal data protection, specifically within the provisions of the GDPR.

This research is contextualized within the RSLingo initiative, a model-driven requirements engineering approach for the rigorous specification
of system requirements (Ferreira & Silva, 2012; Silva, 2015). The analysis of the GDPR involved several tasks, namely reading, manual knowledge extraction, and characterization of many concepts and sentences expressed in that legal document. The analysis was supported by the RSLingo RSL language (Silva, 2017, 2018), in which requirements are defined at different abstraction levels, yet they always represent an expression of stakeholders’ needs. Stakeholder, glossary of terms, business process and business goal constructs allow to express a high-abstraction-level overview of stakeholders’ needs and concerns. From those, system goal, functional requirement, quality requirement, constraint, use case or user story constructs can be used to specify their concerns at system level. From the later, test cases can be derived and defined in order to drive a verification process (Silva, Paiva, & Silva, 2018). At a particular point in time, especially if the software development project is in its earlier stages, it may be just enough to specify these requirements using only business and/or system goal constructs.

The paper is structured as follows. Section 2 introduces the ITLingo and RSL language, which supports this research. Section 3 explains our methodology of systematically analysing the GDPR based on the background research. Section 4 presents some results obtained from applying this methodology, namely through the presentation of some produced work products. Section 5 discusses a comparative study between ISO27001 and GDPR. Finally section 6 provides for some concluding remarks and future direction for our research.

2. BACKGROUND

ITLingo is the main package of RSLingo, which is a long-term initiative to research, develop and apply rigorous IT specification languages. And the RSLingo (sub-package) is a long-term research initiative in the area of Requirements Engineering that recognizes that natural language, although it is the most common and preferred form of representation used in requirements documents, is prone to produce ambiguous and inconsistent documents that are difficult to validate or transform automatically (A. Silva, 2017). Initially, RSLingo proposed an approach to natural language processing (NLP) techniques as well as human techniques to capture relevant information from natural language requirements specifications and then apply light analysis techniques to extract coded domain knowledge in them (Ferreira & Silva, 2012). RSLingo is based on two languages and the mapping between them: RSL-PL (Pattern Language), designed for coding of specific linguistic standards of Requirements Engineering, and RSL-IL (Intermediate Language), a language specific to the domain to be addressed the Concerns of Requirements Engineering (Ferreira & Silva, 2013). Mapping these two languages allows the initial knowledge written in natural language to be extracted, analyzed and converted to a more structural format, which reduces its original ambiguity and creates a more stringent system requirement specification document (A. R. Da Silva, 2015). This approach therefore envisages the entire process of extracting knowledge and converting to a more rigorous representation as a way of helping stakeholders to better understand the set of declared natural language statements that represent their real needs (A. Silva, 2017).

RSLingo RSL (Requirement Specification Language) is a natural control language to help the production of system requirements specification in a more systematic, rigorous, and consistent way, and includes a rich set of logically distributed constructs in views (A. Silva, 2017). These constructs represent fundamental concepts according to requirements engineering concerns that may exist at levels of business abstraction and systems. RSL is an independent process and tool that can be used and adapted by multiple users and organizations with different processes, methodologies and supported by various types of software tools (A. Silva, 2017). RSL was implemented with the Xtext Framework (Bettini, 2016) and therefore its specifications are strict and can be validated and automatically transformed into other representations and formats. On the other hand, the constructs are defined by a linguistic standard and represented textually according to a particular linguistic style. These constructs are based on two aspects, namely: (1) Levels of Abstraction - composed in a manner organized by the Business Level and System Level; (2) The specific concerns of requirements engineering - composed and organized by concepts, context, active structure, behavior, passive structure and requirements (A. Silva, 2017).

The Table 1 illustrates in detail the levels of RSL abstraction, with the Business Level comprising the following points: (1) Stakeholders (active structure), defines the most important source of requirements, provides relevant information for the system of interest to be designed; (2) Glossary (passive structure), includes all the terms of the system domain, its main objective is to reduce negative effects of imprecision of the natural language, through ambiguity resolution techniques; (3) Business objective (requirements), this view allows to establish a bridge between the system of interest and the context of the business (A. Silva, 2017). The needs presented by the interested parties are considered as requirements to satisfy this view; (4) Business process (behavior), represents set of processes that define the behavior of the system and the flow of information between them. As for the System Level, it comprises the following points of
view: (1) Actors (active structure), describing the actors that interact with the system; (2) State machine (behavior), allows to describe the behavior of the system; (3) Structural view (passive structure), allows to describe the structure of the system based on the data entity; (4) Cases of use (passive structure), allows to describe the functionalities of the system, through case-diagram of use cases; and (5) other type requirements such as system objectives, quality requirements, functional requirements, constraints and usertory, which should be defined in the context of the system (A. Silva, 2017). The specification of requirements with RSL is based on the RSL Excel Template model, tool structured according to the views of RSL constructs represented in Table 1. The intersection between the level of abstraction and a requirement engineering concern corresponds to a view of the RSL.

Table 1 - Classification of RSL views: level of abstraction versus requirements engineering specific concerns

<table>
<thead>
<tr>
<th>Concerns Levels</th>
<th>Package</th>
<th>Context</th>
<th>Active Structure</th>
<th>Behavior</th>
<th>Passive Structure</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
<td>package-system</td>
<td>System (Objects)</td>
<td>Actor</td>
<td>StateMachine (State, Transition, Action)</td>
<td>DataEntityView</td>
<td>SystemGoal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>QR Constraint</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FR UserCase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>UserStory</td>
</tr>
</tbody>
</table>

3. ANALYSIS OF THE GENERAL DATA PROTECTION REGULATION 2016/679 OF THE EUROPEAN UNION

The analysis of the European Regulation 2016/679 was made in phases based on the constructs and views of the RSL, and from its levels of abstraction, we selected the constructs that best fit this analysis. At the Business Level abstraction level, we select the views (glossary, stakeholder or stakeholders, and business objectives), from the System Level we select the views (actors, data entities, state machines) and requirements view systems objective, and use cases).

Phase 1: Contextualization - This phase consisted of the framing of the theme "protection of personal data" addressed in the regulation. The result of the work at this stage was a list containing the definition of commercial domain terms (hence a glossary) to support the development of data processing systems, from a business perspective.

Phase 2: Definition of Stakeholders or Stakeholders - This phase consisted in identifying and characterizing stakeholders in the process of developing and implementing data processing systems, according to their categorical distribution.

Phase 3: Definition of Business Objectives - This phase consisted in identifying and characterizing the business objectives to be achieved with the application of the provisions of the Regulation, both for treatment systems and for regulatory systems.

Phase 4: Structural Representation of the System Domain - This phase consisted in modeling the relationship between the main data entities and the role that each entity will play on the system, along with its attributes and operations. A UML class diagram has been developed for this purpose.

Phase 5: Definition of system objectives - This phase consisted of identifying and characterizing the objectives of the system to be achieved by applying the provisions of the Regulation to data processing systems. These objectives compose a list of requirements that have been elicited through a use case diagram and state machines, which allowed to understand in more detail at operations level the functionalities of these systems.

Phase 6: Definition of systems and relationship between systems - This phase consisted of identifying each system involved in the protection of personal data and characterizing the interaction, by type, between source and target systems involved in the protection domain regardless of whether these systems are internal systems (in scope) or external systems (out of scope).

3.1. SPECIFICATION OF DATA PROTECTION REQUIREMENTS WITH THE RSL

Figure 1 describes a simplified version of the RSL model, applied in this analysis in which the two levels of abstraction according to which the RSL is stratified are illustrated in the diagram: The level of business and the level of the system. The business level groups the views closer to the business sphere, while the system level groups the views closest to the sphere of the software system. Each view gathers...
one or more RSL constructs. Instances of system constructs depend on some instances of business constructs. For example; entities depend on the terms of the Glossary, and instances of constructs of the same level of abstraction depend on each other. (1) actors are dependent on stakeholders or stakeholders. Other dependencies were not represented in the diagram for reasons of readability, namely: (2) System objectives depend on Business Objectives; (3) Requirements are dependent on Business Objectives; (4) Actors, Entities, State Machines and Use Cases are system dependent. The relationships between (business events, business processes and business objectives) and relationships between (business objectives, system objectives, functional requirements, quality requirements, restrictions, use cases and user stories) as well as any concerns of ER, were also not represented here in the diagram for the sake of simplicity. Signs with a dark gray background in the diagram, like the constructs, marked with dark gray background are those that were applied in the analysis of the regulation, and which are in the scope of this document, the views and constructs, marked with white background correspond to future work.

Figure 1 - RSL metamodel (simplified). The constructs in dark grey background are applied in the scope of this paper.
3.1.1. STAKEHOLDERS VIEW

Stakeholders are people and organizations that influence the development of an information system or will be affected by its operation, in this case, the operation of regulatory systems.

Table 2 – Stakeholders identified by GDPR

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Subject</td>
<td>An identifiable natural person who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity.</td>
<td>Person</td>
</tr>
<tr>
<td>Controller</td>
<td>A natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.</td>
<td>Team</td>
</tr>
<tr>
<td>Processor</td>
<td>A natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.</td>
<td>Team</td>
</tr>
<tr>
<td>Supervisor Authority</td>
<td>An independent public authority which is established by a EU Member State.</td>
<td>Organization</td>
</tr>
<tr>
<td>Data Protection Officer (DPO)</td>
<td>An enterprise security leadership role responsible for overseeing data protection strategy and implementation to ensure compliance with GDPR requirements.</td>
<td>Person</td>
</tr>
</tbody>
</table>

3.1.2. BUSINESS GOALS VIEW

Business goals can be faced as the value that the system represents for the business, in this case that regulatory system represent to the protection of personal data and the free movement of such data. Table 3 lists the definition of some business goals that, once achieved with the development and operation of regulatory systems, allow warranting the regulatory data protection capabilities provisioned by the GDPR. The identification and characterization of business goals constitutes a starting point for the identification and characterization of system goals. The values in the column Part of indicate the aggregation relations that may exist between BusinessGoals, which means that they can be decomposed into sub-BusinessGoals. (The tags preceding the BusinessGoals are their id(identifiers) and express no particular order in what the items or articles of the GDPR are concerned.)

Table 3 - Some business goals extracted from the GDPR

<table>
<thead>
<tr>
<th>Name</th>
<th>Part of</th>
</tr>
</thead>
<tbody>
<tr>
<td>bg_1: Facilitate the exercise of the data subject’s rights</td>
<td>bg_1</td>
</tr>
<tr>
<td>bg_1.1: Right to obtain from the controller confirmation as to whether or not personal data concerning him/her are being processed</td>
<td>bg_1</td>
</tr>
<tr>
<td>bg_1.2: Right to transmit personal data to another controller without hindrance from the controller to whom they have been provided</td>
<td>bg_1</td>
</tr>
<tr>
<td>bg_6: Lawfulness, fairness and transparency of personal data processing (art. 6)</td>
<td>bg_7</td>
</tr>
<tr>
<td>bg_7: Conditions for consent (art. 7)</td>
<td>bg_7</td>
</tr>
<tr>
<td>bg_7.1: Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.</td>
<td>bg_7</td>
</tr>
<tr>
<td>bg_7.2: The data subject shall have the right to withdraw his or her consent at any time.</td>
<td>bg_7</td>
</tr>
</tbody>
</table>

3.1.3. SYSTEM GOALS VIEW

System goals can be derived from business goals, since the first operationalize the later in a particular system context (in the case of this paper, the context of the regulatory system). This means that the system goals are met, business goals are equally met, hence the provisions in the GDPR are respected by the regulatory system. Table 4 depicts the definition of some system goals extracted from the GDPR for general system objectives.

Table 4 – Some System goals extracted from the GDPR

<table>
<thead>
<tr>
<th>Name</th>
<th>Part of</th>
</tr>
</thead>
<tbody>
<tr>
<td>sg_1: Ensure a level of security appropriate for the data subject’s rights and freedoms</td>
<td>sg_1</td>
</tr>
<tr>
<td>sg_1.1: Pseudonymize and encrypt personal data</td>
<td>sg_1</td>
</tr>
<tr>
<td>sg_1.2: Keep the confidentiality, integrity, availability and resilience of processing systems and services</td>
<td>sg_1</td>
</tr>
<tr>
<td>sg_1.3: Restore the availability and access to personal data, in a timely manner, in the event of a physical or technical incident</td>
<td>sg_1</td>
</tr>
<tr>
<td>sg_1.4: Execute a process to regularly test, assess and evaluate the effectiveness of technical measures concerned with the security of personal data processing</td>
<td>sg_1</td>
</tr>
</tbody>
</table>
3.1.4. STRUCTURAL VIEW

This view represents the data entities that should be part of the system. This class diagram characterizes the System Domain Model, which allows us to perceive in a general way the role of each entity, attributes, purposes and operations associated with data processing. The entities represented in the diagram are only part of the system of treatments given, are not part of this, the entities of the regulatory system because the purpose of this dissertation is to model and elicit requirements for the data processing system.

![Figure 2 - Domain model of the data processing system](image)

3.1.5. USE CASE VIEW

In this view we can observe that the Stakeholder plays the role of system actor and the type actions that determine the iteration with the system. These uses cases in Figures 3 and 4 were adapted according to the articles, (12º points 1 and 5a), (13º, lines b), 16º and 17º with reference to the Data Subject and the controller, as stated in the GDPR. The data subject makes requests such as data query, rectification, data transfer or deletion of his data and a fee payment is added to these requests. The controller performs the operations illustrated in figure 4 to respond these requests and creates the conditions for the data subject to check the status of their requests and the purposes of the operations. In case of a violation that jeopardizes his data, he is notified together with the supervisory authority and a record is made of this violation by the processor. The use cases view represents functional requirements and help us define system behavior. With the application of this view of the RSL it was possible to define an overview of the type of behavior that should be applied to the data processing systems according to the GDPR requirements to ensure compliance.
3.1.6. **STATEMACHINE VIEW**

Figure 5 illustrates the state machine diagram, and the transitions that a request can do until the final state.

![State Machine Diagram]

**Figure 5 – Order Status**

4. **TRANSFORMATION AND VALIDATION OF RGPD REQUIREMENTS WITH ITLINGO-STUDIO**

ITLingo is the tool that allows transform and validate the requirements elicitation projects. The RSL Excel Template with the requirements specification extracted from GDPR, was imported into ITLingo-Studio. And it allowed us to automatically generate documents and diagrams properly structured in various formats and extensions, which makes the definition of requirements more systematic and rigorous.

After importing the file the next step was the validation of the project itself, which was not possible in the first place due to the fact that there were some errors that was needed to corrected. After the correction, it was possible to validate the project. The validation of the project only concerns the technical aspects of the tool, which means without the validation of the project it is not possible to carry out other operations at the tool level without it being validated.

After the project validation, the next step was the transformation and file exportation into more structured formats such as Word and Excel format. As well as the generation of diagrams of varied types for example: diagram of uses cases, state machines and class diagram. ITLingo-Studio also supports the production of diagrams through text, different from the direct manipulation of geometric forms to produce diagrams as it is done in other tools. The Figure 6 illustrates the RSL Excel Template structure, also generated by ITLingo, but does not differ from the document previously imported in terms of structure but has some changes in the content due to the normalization made directly in the code of ITLingo-Studio, figure 7.
5. MAPPING GDPR REQUIREMENTS VERSUS ISO/IEC27001 BEST PRACTICES

The certification of processes, services and products has been a constant practice of organizations. Therefore, with the introduction of the GDPR, organizations must adapt their systems, services or products in accordance with the GDPR. In this context we present the results of a comparative study that was based on the mapping between the requirements identified in the GDPR in general, and the good practice requirements of ISO/IEC27001, which aiming at, implementing, maintaining and continuously improving an organization’s information security management system (BSI, 2005). With this comparative study we try to answer the following question: Do the organizations that have ISO/IEC27001 certification fully comply with what is required by the GDPR?

In order to answer this question, we have elaborated a list of requirements extracted from the GDPR and together with the controls of ISO/IEC27001, both were analyzed with the objective of identifying some relation between them. With this, it was possible to observe that there is a 48% coverage for ISO/IEC27001 certification on GDPR, which implies that only ISO/IEC27001 certification is not enough for organizations to comply with the GDPR.
6. RELATED WORK

Analysing a legal document like the GDPR encompasses a challenge for business analysts working in the development of information systems capable of processing personal data according to the legal dispositions in that document, as well as for business analysts working in the development of information systems responsible for providing the electronic means to regulate the processing of personal data.

The analysis effort required by the parties that must comply with the dispositions in the GDPR is reduced with the reuse of the work products from the analysis methodology reported in this paper. On one hand, reporting to the business level work products (e.g., Glossary, Stakeholders and BusinessGoals), they support a better understanding of the GDPR, from which regulatory systems can be further designed and implemented. They also augment the comprehensibility of the personal data protection business domain. On the other hand, reporting to the system level work products (e.g., Systems and SystemGoals), they provide for a rigorous interpretation of the GDPR, from which, specifically in the scope of our research, regulatory systems can be further designed and implemented.

Prior efforts of other authors have been undertaken to systematize past research concerned with the handling of legal texts for software systems development (Otto & Antón, 2007). The same authors who surveyed those efforts, together with Massey (Massey, Otto, & Antón, 2009), later reviewed specifications of legally compatible systems and produced requirements to foster legal compatibility. However, those authors focused on goals to specify requirements for the development of legally compatible systems. Our approach to the specification of such systems is broader in terms of the views over legal texts it considers (the RSL views).

Hoffmann, et al. (Hoffmann et al., 2012) presented in 2012 some research on the commonality within legal software requirements and proposed legal software requirements patterns (extracted from recurring legal requirements) to produce catalogues of that kind of requirements. The research reported in this paper involved dealing with linguistic patterns and linguistic styles, not for legal software requirements in the broad sense, but for the specification of personal data protection requirements, also extracted from legal documents to produce a personal data protection requirements catalogue.

7. CONCLUSION

The analysis of the GDPR is a starting point for the further analysis, design and implementation of information systems capable of ensuring the confidentiality, integrity, availability and resilience of the personal data they process. Due to technological evolution, several services currently share data and part of it relies on personal data related to banking, healthcare and other data domains. Those data require strict measures to protect them from diversion for improper purposes with irreversible consequences, yet maintaining the free movement of such data. The application of penalties for improper diversion purposes comes from the detection of infringements, so the extraction of business and system requirements from the GDPR is of extreme importance, in order to specify and further develop regulatory systems able to communicate with processing systems and operationalize the regulatory capabilities disposed in the GDPR. This paper presented a systematic approach to the analysis of the GDPR from both business and system perspectives, useful for the development and operation of regulatory systems.

The analysis of the GDPR from the system perspective, include the specification of Actors, DataEntities and StateMachines views, as well as developing automatic transformations of requirements into formats other than the RSL Excel Template, along with complementary diagrams. These model transformations required the use of RSL in its programmatic shape, therefore transforming the RSL-based SRS document in Excel format into an even more rigorous representation. The RSL-based specification of the GDPR will act as a catalogue of personal data protection requirements and the ultimate goal is for it to be reused in any project of personal data regulatory systems development. This research also includes a comparative study with results that help us understand the level of coverage of ISO certification, on GDPR namely: ISO27001, and this is very useful for organizations that use this certification because they will be able to know what are missing in their services to comply with the GDPR.

The requirements specification files of the GDPR analysis results under this dissertation as contributions are available at ResearchGate namely at [REF1] and [REF2].
REFERENCES


Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (2016).


Analysis of Privacy Policies: The RSLingo4Privacy Approach. 18th International Conference on Enterprise Information Systems (ICEIS), Rome, Italy.


