Specification of Personal Data Protection Requirements
Analysis of Legal Requirements from the GDPR Regulation

Mário Fernandes[1], Alberto Rodrigues da Silva[1] and António Gonçalves[1,2]
[1]*INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, Portugal*
[2]*Escola Superior de Tecnologia de Setúbal, Instituto Politécnico de Setúbal, Portugal*

Keywords: Personal Data Protection, Requirements Specification, Rslingo, Regulation (EU) 2016/679, GDPR.

Abstract: The European Union establishes in the Regulation 2016/679, or GDPR (General Data Protection Regulation), a set of legal dispositions to achieve the protection of natural persons in what personal data processing and the free movement of such data is concerned. When those dispositions are considered in the development of information systems, the later become attainable for legal approval within that scope. This paper presents the methodology we are following to elaborate a reusable catalogue of personal data protection requirements aligned with the GDPR. Following a separation-of-concerns approach, the catalogue shall serve the purpose of constructing information systems able to communicate with those that process individuals' personal data, to materialize the regulatory data protection capabilities disposed in the GDPR. In that context, the elicitation of system requirements demands for the interpretation of a legal document by business analysts, which consists of a scientifically relevant challenge. This research is contextualized by the RSLingo initiative, a model-driven requirements engineering approach for the rigorous specification of system requirements. In particular this paper discusses the GDPR's requirements defined as a catalogue of both business goals and system goals.

1 INTRODUCTION

The need to protect valuable information, particularly of individuals, is a societal challenge imposed by many governments to organizations. Within the European scope, the General Data Protection Regulation (GDPR or Regulation 2016/679) was defined by the European Union, consisting of a set of legal dispositions to achieve the protection of natural persons in what personal data processing and the free movement of such data is concerned (EU, 2016). The guarantee of compliance with the GDPR, at the level of information systems, demands for an effort from business analysts when interpreting such legal document.

The goal of this paper is to propose a rigorous specification of requirements included in the GDPR, and consequently to promote a better and systematic interpretation of those legal requirements.

From the analysis of the GDPR document, the aim is to elaborate a *reusable catalogue of personal data protection requirements* to be used in the design and implementation of information systems able to communicate with those that process individuals' personal data, to 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 caracterization 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

Fernandes, M., Rodrigues Silva, A. and Gonçalves, A.
Specification of Personal Data Protection Requirements.
DOI: 10.5220/0006810603980405
In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)*, pages 398-405
ISBN: 978-989-758-298-1

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 RSLingo 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 some challenges addressed by our research, together with its main contribution, and critically elaborates on related work. Finally, Section 6 provides for some concluding remarks and future direction for our research.

2 BACKGROUND

RSLingo is a long-term research initiative in the Requirements Engineering area (Ferreira & Silva, 2012). It is a linguistic approach to improve the quality of requirements specifications. Although being the most common and preferred form of representing requirements, natural language is prone to producing ambiguous and inconsistent documents, hard to automatically validate or transform into other kinds of artefacts (Silva, 2017).

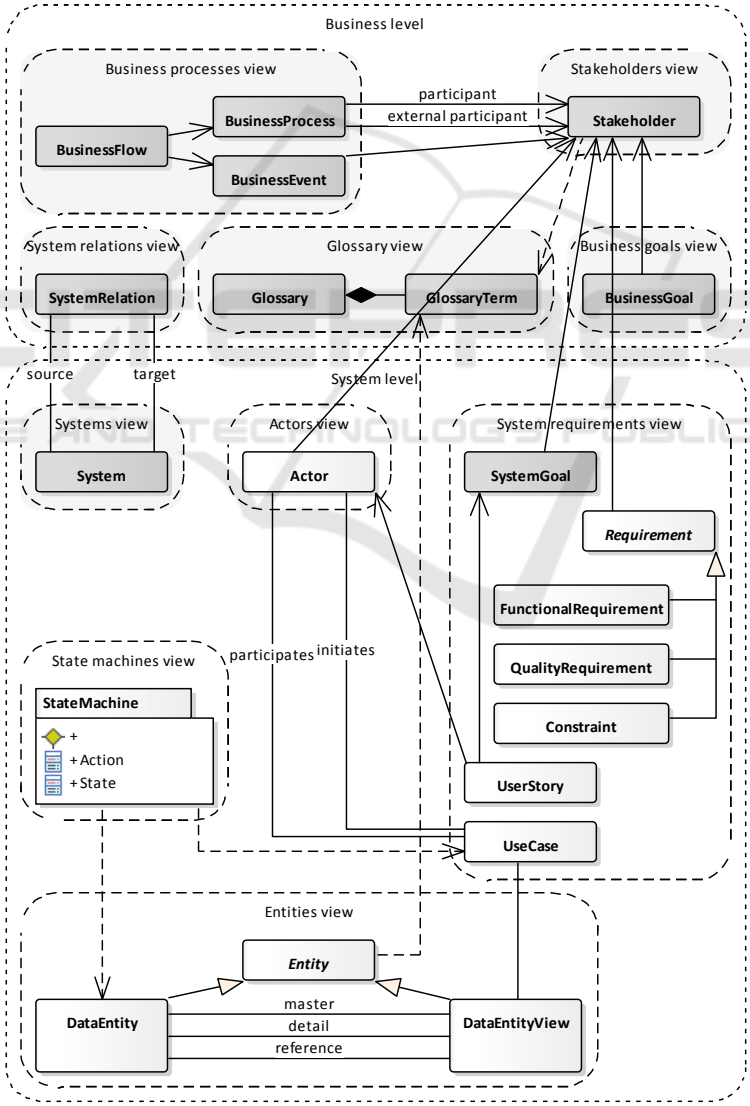

Figure 1: RSL metamodel (simplified). The constructs in dark grey background are applied in the scope of this paper.

RSLingo RSL. (Requirements Specification Language) (Silva, 2017, 2018) is a comprehensive domain-specific language designed to address general-purpose RE activities such as the rigorous specification, automatic validation, persistence and management of system requirements. RSL is based on other languages such RSL-IL (Ferreira & Silva, 2013), RSL-IL4Privacy (Caramujo & Silva, 2015; Silva, Caramujo, Monfared, Calado, & Breaux, 2016), *Pohl's* (Pohl, 2010), *XIS** (Ribeiro & Silva, 2014a, 2014b; Silva, Saraiva, Silva, & Martins, 2007) and *SilabREQ* (Savić et al., 2015). In a more detailed way, RSL is a controlled natural language to support the elaboration of SRSs (Systems Requirements Specifications) in a systematic, rigorous, consistent and less ambiguous way, namely by using the RSL Excel Template, which is based in RSL itself. Representing domain knowledge with such SRSs has in itself a way of providing business stakeholders with a better understanding of natural language statements that represent requirements. RSL is a language that includes a rich set of fundamental RE-specific constructs logically arranged into views, according to two dimensions: *abstraction levels* and *RE concerns*, which means the architecture of RSL is *bidimensional* and *multiview* (Silva, 2017).

Figure 1 depicts a simplified version of the RSL metamodel. The two abstraction levels are illustrated in the diagram: the business level and the system level. The *business level* groups the views closer to the business perspective, whereas the system level groups the views closer to the system perspective. Each view gathers one or more constructs. System constructs may depend on some business constructs (e.g. Entities may depend on Glossary terms and Actors may depend on Stakeholders). These constructs are defined as linguistic patterns and represented textually by mandatory or optional fragments (text snippets) (Silva, 2017). The RSL is structured according a large set of constructs, such as: glossary term, stakeholder, business goal, business process, business event, business flow, system, actor, data entity and data entity view, system goal, functional requirement, quality requirement, constraint, use case, user story, state machine (Silva, 2017).

The personal data protection requirements (extracted from the GDPR) are under specification with a set of RSL constructs selected according to the chosen linguistic style, appropriate for the early stages of the requirements specification. That set is composed of some business level views together with the system goals view. (However, we could have specified these requirements at the system level following other styles by using constructs such as functional requirements, quality requirements and constraints; use cases or user stories).

3 GDPR ANALYSIS

The GDPR is targeted at the protection of natural persons regarding the processing of their personal data by automated or manual means, only in the case that those data are contained or intended to be contained in filing systems. According to the GDPR, a filing system is set of structured personal data accessible via specific criteria, whether centralized or not. However, the GDPR mentions not only filing systems, but also (secure) systems to which controllers may grant remote and direct access of data subjects to their own personal data.

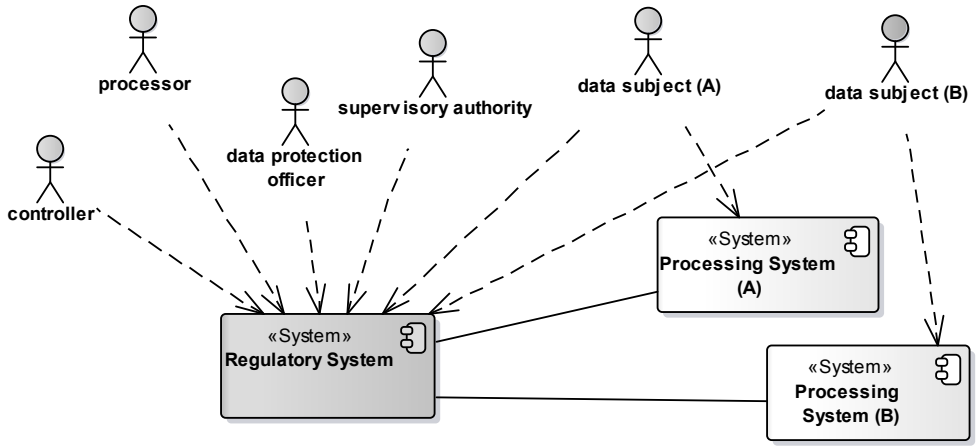

Figure 2: Systems and stakeholders defined in the scope of GDPR.

Nevertheless and based on an interpretation of the GDPR, other systems may also be considered, namely systems where controllers or processors could register violations of data, where data subjects could request for the rectification of data, where controllers or processors could record processing activities and where data protection officers could monitor that registration.

For the purpose of this paper, the GDPR is an elicitation source of not only business but also system requirements to consider when developing information systems able to communicate with those that process individuals' personal data, in order to materialize the regulatory data protection capabilities disposed in the GDPR. This may imply the existence of two types of systems: *regulatory systems* and *processing systems*, whose operation includes, yet it is not restricted to processing individuals' personal data. Figure 2 shows the two types of systems mentioned based on the GDPR's Introductory items 15 and 16, and Article 2(1). This paper presents not only business and system requirements of regulatory systems, but also a RSL-based methodology to extract these requirements from legal documents. In the context of personal data protection, system requirements state the way in which the business requirements shall be met in the application domain i.e. how the data shall be protected regardless of the technology used by those systems to implement that protection.

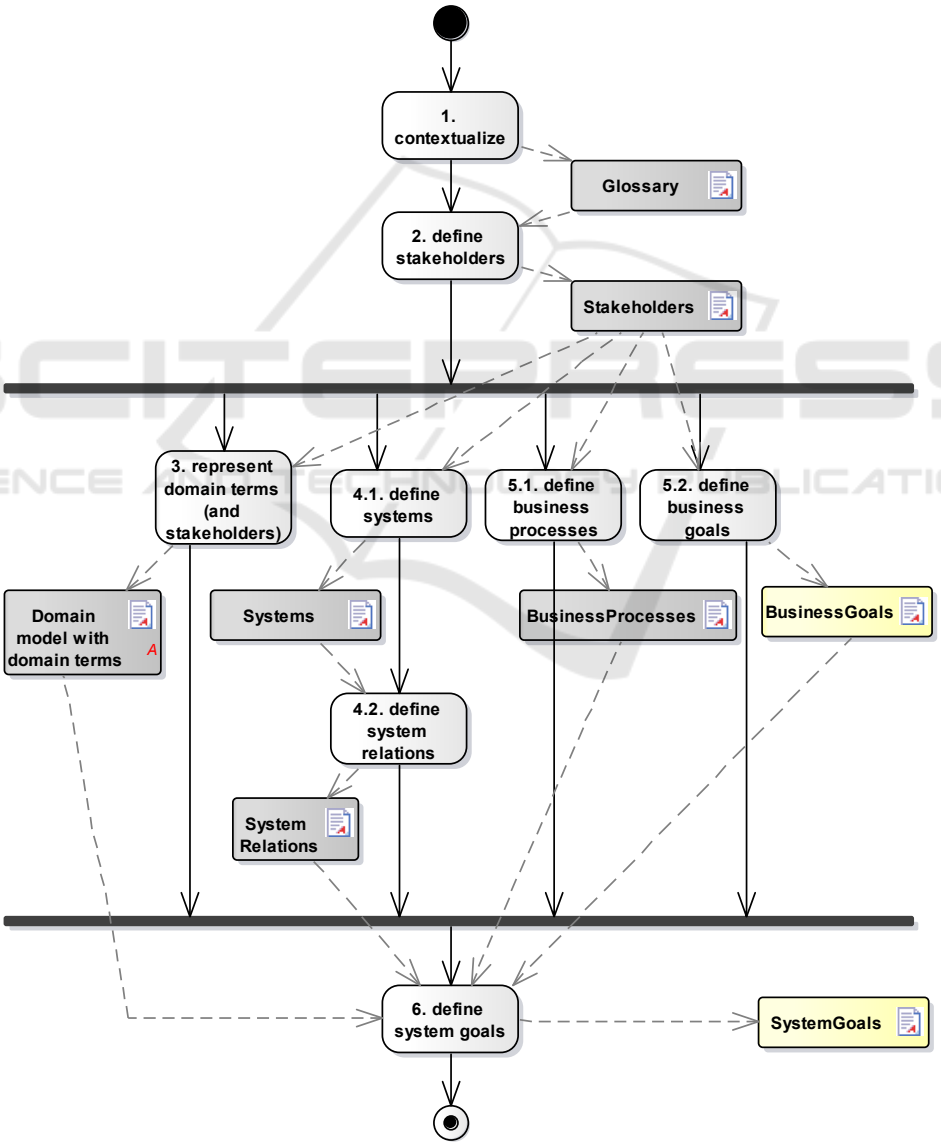

Figure 3: Methodology for analysing the GDPR.

The analysis of the GDPR involved a set of tasks as depicted in Figure 3. The objects shown in this figure represent work products as partially include in a RSL specification.

Task 1: Contextualization. This task consisted of the familiarization of the business analyst with the domain of personal data protection. The work product of this task is a list containing the definition of the business domain terms to better support the understanding of regulatory systems, from a business perspective.

Task 2: Definition of Stakeholders. This task consisted of identifying and characterizing the parties interested in the development and operation of regulatory systems.

Task 3: Representation of Structure Composed of Domain Terms (and stakeholders). This task consisted of modelling the relation between the main domain terms identified in the GDPR, along with their attributes. A UML class diagram can be elaborated to complement that understanding.

Task 4: Definition of Systems and System Relations. This task consisted of **(subtask 4.1)** identifying each system involved in the business domain of personal data protection, regardless of being an in-scope or an out-of-scope system, and characterizing its type e.g. whether system or subsystem; and **(subtask 4.2)** characterizing the interactions between source and target systems involved in the business domain of personal data protection, whether those systems being internal systems (in-scope) or external systems (out-of-scope).

Task 5: Definition of Business Processes and Business Goals. This task consisted of **(subtask 5.1)** identifying and characterizing the business processes, based on an ordering of tasks concerned with regulating the processing of personal data; and **(subtask 5.2)** identifying and characterizing the businessgoals to be achieved with the application of the provisions in the GDPR to regulatory systems. Although not represented in the diagram, this task also included the identification and characterization of business flows to express control flows between business processes or between business processes and business events – business flows can be of type sequential, conditional (equivalent to the BPMN exclusive gateway) or parallel (equivalent to the BPMN parallel gateway).

Task 6: Definition of System Goals. This task consisted of identifying and characterizing the systemgoals to be achieved with the application of the provisions in the GDPR to regulatory and processing in-scope systems.

The execution of these tasks is guided by a control flow that suggests the adopted GDPR analysis methodology. Figure 3 is complemented by the illustration of dependencies between the tasks and the RSL constructs that are progressively specified throughout that flow.

4 SPECIFICATION OF DATA PROTECTION REQUIREMENTS WITH THE RSL

This section demonstrates part of the work products originated from each task of the GDPR analysis methodology described in Section 3. Some columns from the tables in the RSL Excel Template were omitted due to space restrictions (e.g. the Description or Priority BusinessGoals' attributes were omitted in the corresponding table).

4.1 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 1: Stakeholders identified by GDPR.

Name	Description	Type
Data Subject	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.	Person
Controller	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	Team
Processor	A natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller	Team
Supervisory Authority	An independent public authority which is established by a EU Member State.	Organi-zation
Data Protection Officer (DPO)	An enterprise security leadership role responsible for overseeing data protection strategy and implementation to ensure compliance with GDPR requirements.	Person

Stakeholders are original sources of requirements, despite that, in the case of this paper, requirements were elicited from the GDPR itself. Table 1 contains some stakeholders interested in the development of regulatory systems. Some stakeholders will be users of those systems, hence will be constituted *Actors* in the *Actors view*. A *Stakeholder* is characterized, at least, by its name and type, as stated in the table.

4.2 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 2 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(entifiers) and express no particular order in what the items or articles of the GDPR are concerned.)

Table 2: Some business goals extracted from the GDPR.

Name	Part of
bg_1: Facilitate the exercise of the data subject's rights	
bg_1.1: Right to obtain from the controller confirmation as to whether or not personal data concerning him/her are being processed	bg_1
bg_1.2: Right to transmit personal data to another controller without hindrance from the controller to whom they have been provided	bg_1
bg_6: Lawfulness, fairness and transparency of personal data processing (art. 6)	
bg_7: Conditions for consent (art. 7)	
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.	bg_7
bg_7.2: The data subject shall have the right to withdraw his or her consent at any time.	bg_7

4.3 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 if system goals are met, business goals are equally met, hence the provisions in the GDPR are respected by the regulatory system. Table 3 depicts the definition of some system goals extracted from the GDPR for regulatory systems. Additionally, to allow a comparative insight over both types of system, Table 4 presents some system goals also extracted from the GDPR for processing systems. The semantics associated with each column in both tables is the same as the one associated with the column of same name in Table 2.

Table 3: System goals extracted from the GDPR for regulatory systems.

Name	Part of
sg_1: Record personal data processing activities	
sg_1.1: Maintain a record of the personal data processing activities under the responsibility of each controller	sg_1
sg_1.2: Maintain a record of all categories of personal data processing activities carried out by each processor on behalf of a controller	sg_1
sg_2: Personal data breaches shall be notified to the competent supervisory authority and, eventually, the data subject	

Table 4: System goals extracted from the GDPR for processing systems.

Name	Part of
sg_1: Ensure a level of security appropriate for the data subject's rights and freedoms	
sg_1.1: Pseudonymize and encrypt personal data	sg_1
sg_1.2: Keep the confidentiality, integrity, availability and resilience of processing systems and services	sg_1
sg_1.3: Restore the availability and access to personal data, in a timely manner, in the event of a physical or technical incident	sg_1
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	sg_1

5 DISCUSSION

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.

6 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 paper illustrated the approach with a sample of some work products originated from executing the analysis methodology.

Future work includes further analysis of the GDPR from the system perspective, namely involving 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 will require 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. In future research efforts, the methodology to analyse legal documents presented in this paper may be applied to other regulatory documents that impose requirements to information systems specification and further development e.g. ISO 27000.

ACKNOWLEDGEMENTS

This work was partially supported by national funds under FCT projects UID/CEC/50021/2013 and CMUP-EPB/TIC/0053/2013.

REFERENCES

Caramujo, J., & Silva, A. R. d. (2015). *Analyzing Privacy Policies Based on a Privacy-Aware Profile: The Facebook and LinkedIn Case Studies*. 17th IEEE Conference on Business Informatics (CBI), Lisboa, Portugal.

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).

Ferreira, D. d. A., & Silva, A. R. d. (2012). *RSLingo: An Information Extraction Approach toward Formal Requirements Specifications*. 2nd IEEE Model-Driven Requirements Engineering Workshop (MoDRE), Chicago, Illinois, USA.

Ferreira, D. d. A., & Silva, A. R. d. (2013). *RSL-IL: An Interlingua for Formally Documenting Requirements*. 3rd International Workshop on Model-Driven Requirements Engineering (MoDRE), Rio de Janeiro, Brazil.

Hoffmann, A., Schulz, T., Hoffmann, H., Jandt, S., Roßnagel, A., & Leimeister, J. M. (2012). *Towards the Use of Software Requirement Patterns for Legal Requirements*. 2nd International Requirements Engineering Efficiency Workshop (REEW), Essen, Germany.

Massey, A. K., Otto, P. N., & Antón, A. I. (2009). *Prioritizing Legal Requirements*. 2nd International Workshop on Requirements Engineering and Law (RELAW), Atlanta, Georgia, USA.

Otto, P. N., & Antón, A. I. (2007). *Addressing Legal Requirements in Requirements Engineering*. 15th IEEE International Requirements Engineering Conference (RE), New Delhi, India.

Pohl, K. (2010). *Requirements Engineering: Fundamentals, Principles, and Techniques*. Berlin Heidelberg: Springer-Verlag.

Ribeiro, A., & Silva, A. R. d. (2014a). Evaluation of XIS-Mobile, a Domain Specific Language for Mobile Application Development. *Journal of Software Engineering and Applications, 7*(11), 906-919.

Ribeiro, A., & Silva, A. R. d. (2014b). *XIS-Mobile: A DSL for Mobile Applications*. 29th Annual ACM Symposium on Applied Computing (SAC), Gyeongju, Korea.

Savić, D., Vlajić, S., Lazarević, S., Antović, I., Stanojević, V., Milić, M., & Silva, A. R. d. (2015). Use Case Specification Using the SilabReq Domain Specific Language. *Computing and Informatics, 34*(4), 877-910.

Silva, A. R. d. (2015). Model-driven engineering: A survey supported by the unified conceptual model. *Computer Languages, Systems & Structures, 43*(October 2015), 139-155.

Silva, A. R. d. (2017). *Linguistic Patterns and Linguistic Styles for Requirements Specification (I): An Application Case with the Rigorous RSL/Business-Level Language*. 22nd European Conference on Pattern Languages of Programs (EuroPLoP), Irsee, Germany.

Silva, A. R. d. (2018). Rigorous Requirements Specification: Specification of Use Cases with the RSLingo RSL Language. Lisboa: INESC-ID.

Silva, A. R. d., Caramujo, J., Monfared, S., Calado, P., & Breaux, T. (2016). *Improving the Specification and Analysis of Privacy Policies: The RSLingo4Privacy Approach*. 18th International Conference on Enterprise Information Systems (ICEIS), Rome, Italy.

Silva, A. R. d., Paiva, A., & Silva, V. (2018). *Towards a Test Specification Language for Information Systems: Focus on Data Entity and State Machine Tests*. 6th International Conference on Model-Driven Engineering and Software Development (MODELSWARD), Funchal, Madeira, Portugal.

Silva, A. R. d., Saraiva, J., Silva, R., & Martins, C. (2007). *XIS-UML Profile for eXtreme Modeling Interactive Systems*. 4th International Workshop on Model-Based Methodologies for Pervasive and Embedded Software (MOMPES), Braga, Portugal.