Enhancing Business Process Modelling with Data Protection Compliance: An Ontology-based Proposal

Cesare Bartolini1, Antonello Calabró2 and Eda Marchetti2

1University of Luxembourg, Interdisciplinary Centre for Security, Reliability and Trust (SnT), Luxembourg
2Istituto di Scienza e Tecnologie dell’Informazione “A. Faedo”, Consiglio Nazionale delle Ricerche (CNR)

Keywords: Business Processes, BPMN, General Data Protection Regulation (GDPR), Privacy by Design, Legal Ontologies, LegalRuleML.

Abstract: The research and industrial environments are struggling to identify practical approaches to highlight the (new) duties of controllers of personal data and foster the transition of IT-based systems, services, and tools to comply with the GDPR. In this paper, we present a solution for enhancing the modelling of business processes with facilities to help evaluate the compliance with the GDPR. The proposal is based on a model describing the constituents of the data protection domain: a structured form of the legal text, an ontology of data protection concepts, and a machine-readable translation of the GDPR provisions. An example of application is also provided.

1 INTRODUCTION

The technical evolution of the last decades has significantly changed the environment in which data protection rules operate, in particular by blurring the distinction between the controller and the data subject, thus making it difficult to understand how to apply those rules (Van Alsenoy et al., 2009). The General Data Protection Regulation (GDPR) aims to bring such rules on par with technology (Reding, 2010) and harmonize them across the European Union. Further objectives are to enhance individuals’ rights, give them more control over their personal data, simplify the regulatory environment for businesses, and set the foundation for the Digital Single Market (European Commission, 2015).

The GDPR raises concerns for the development of software systems that can provide services compliant with it, and in line with the novel principle of data protection by design1. Unfortunately, designing systems compliant with data protection rules is no easy task.

Within computer science, data protection is often referred to as privacy and considered a subset of the security domain (Pfleeger et al., 2015; Massacci et al., 2005). However, privacy and data protection mean different concepts in legal doctrine. There is indeed an overlapping between the two terms, as privacy measures can contribute to the protection of personal data, but privacy alone does not fulfill the GDPR requirements.

Problems may become more stringent due to the layering of the different services. A significant amount of them could be located in different parts of the world, especially in the United States, and thus not subject to the GDPR (Jaeger et al., 2009).

To facilitate the introduction of GDPR requirements in software design processes, automating the enforcement of its requirements is very important. A crucial step is therefore to try and rewrite the GDPR requirements into a machine-readable representation.

Indeed the automated processing of the GDPR can support compliance, in the sense of assisting in the design, development, maintenance, and verification of a system in order to comply with the GDPR requirements, including the detection of possible violations, with the objective to minimize the risk of sanctions being issued by the supervisory authority. The integration of compliance-supporting approaches in development tools and methodologies is an application of the data protection by design principle.

In view of developing and adopting supporting approaches in the development process, industrial or-
Organizations are nowadays moving towards the use of Business Process Model and Notation (BPMN) (Object Management Group, 2011). As highlighted in (Fleacă et al., 2016), the main benefits of BPMN commonly rely on the possibility of having a clear and standard notation for creating a description of processes (in terms of participants and activities) and develop executable frameworks for the overall management of the process itself. Directly integrating the GDPR requirements into the business process execution represents a key aspect both for privacy management and validation.

Existing works (Bertoli et al., 2013; Calabró et al., 2015) focus on defining key performance indicators (KPIs), including time-based and cost-based parameters, to assess the BPMN execution. Conversely, this paper presents an approach that uses the legal model of the GDPR to enrich a business process with annotations that express data protection requirements. In other words, the business process will be extended with appropriate connectors to introduce the legal requirements, as they are expressed in the data protection model. The immediate benefits can be summarized as follows:

- controllers of personal data would have a clearer view of their duties with respect to data protection in the context of their business;
- the auditors would have a first-look model to verify the GDPR requirements implementation;
- supervisory authorities would have a structured approach to detect potential violations.

Concerning the structure of the paper, Section 5 provides a survey of existing literature concerning legal models for data protection and privacy, and existing techniques to express legal requirements in software engineering. Section 2 presents the legal representation of data protection legislation, explaining how to describe data protection rules by coordinating three separate models to express the concepts, the legal text, and the rules, respectively. Section 3 introduces the extension of business processes to accommodate data protection rules. An example in Section 4 shows how the proposed approach can be used. Finally, Section 6 gives a set of conclusions and the envisioned future work.

2 BACKGROUND

The present work aims at integrating the legal model for the GDPR into a business process. It is true that there are many possible ways in which the two might be integrated, depending on the specific purpose. Whatever the purpose and means of the integration, it is founded on two building blocks:

- the legal model of the GDPR;
- the model of the business process.

2.1 Modelling the GDPR

Modelling legal texts is a major problem that faces many difficulties due to the necessary interpretation of legal provisions. Additionally, the development of a machine-readable model of a law is hindered by the fact that the law can be viewed from several perspectives. Among the currently available proposals for modelling the GDPR, in this paper we refer the DAPRECO model, which is based on the interaction of three main components: i) a structured representation of the legal text; ii) a conceptual model describing the legal terms used in data protection legislation; iii) a machine-readable translation of the normative provisions.

Each of these components is built using a specific format, and the full GDPR model derives from the integration of all three. For the sake of completeness, a brief description of the components is shown below.

The structured representation of the legal text consists of the annotation of the legal text using specific tags to identify the various parts of the Regulation (e.g., chapter, article, paragraph, and so on). In the DAPRECO project, this is done by using the Akoma Ntoso\(^2\) language, an XML format designed for structuring legal texts (Palmirani and Vitali, 2011). Each part of the legal text is uniquely identified by an index, so that, using appropriate Uniform Resource Identifiers (URIs) containing anchors to the correct identifiers, it is possible to reference specific provisions, or combinations thereof, within the legal text. This referencing can be done from external files as well.

The second component is the conceptual model based on a legal ontology (Benjamins et al., 2005), which is a specific type of ontology specifically designed to represent concepts in the legal domain.

Ontologies can be formalized using several possible models. In the DAPRECO project, the conceptual model of the GDPR has been modeled using the OWL language (World Wide Web Consortium, 2012). OWL is a knowledge representation language that has several expression formats (called syntaxes), some of which are XML tags.

The legal ontology for the GDPR is called PrOnto (Palmirani et al., 2018b; Palmirani et al., 2019).
which stands for Privacy Ontology, since its purpose is to support not only the GDPR concepts, but also other data protection legislation and privacy.

The third and final component of the DAPRECO GDPR model is the machine-readable translation of the legal provisions. As the legal text is written in natural language, the obligations, permissions and prohibitions contained therein must be formalized in a machine-readable format. In DAPRECO, to express the legal rules of the GDPR, a set of logic formulae was used. The founding logic is called Reified Input/Output (RIO) logic (Robaldo and Sun, 2017), which is an extension of Input/Output logic (Makinson and van der Torre, 2000) with the addition of reification (Davidson, 1967). The DAPRECO logic formulae are the result of an interpretation of the legal provisions. The interpretation was performed with the assistance of legal experts, but still it is subject to be overcome by more authoritative sources (such as court decisions or opinions of supervisory authorities). The logic formulae in DAPRECO have been formalized using the LegalRuleML language (Palmitrani et al., 2011), which has the very purpose of expressing legal provisions. The formulae follow an if-then structure, i.e., when the preconditions are met (if part), a legal consequence follows (then part). The legal consequence could be an obligation, permission, or prohibition.

### 2.2 Business Processes at a Glance

A Business Process (BP) usually refers to any structured collection of related activities or tasks that are carried out to accomplish the intended objectives of an organization. Tasks within a BP may represent steps that can be performed either manually by a human, or automatically by an IT system (Gerth, 2013). A business model can be represented using one of the available Business Process Modeling Languages (BPMLs) (van der Aalst et al., 2003). In this paper BPMN is the formalism chosen to represent business models. BPMN is the de facto standard for process modeling. It is indeed a rich and expressive language (but also a complex one) used for the tasks associated with process modeling (Recker, 2010). BPMN has four categories of graphical elements that can be used to build the diagrams: Flow Objects are associated with the actions that can be performed in a BP and make up the behavior of the BP. They consist of Events, Activities, and Gateways. Connecting Objects can be used to connect elements to each other in three different ways: Sequence Flows, Message Flows, and Associations. Swimlanes provide the capability of grouping the primary modelling elements.
First off, classes in the ontology are referenced by the LegalRuleML formulæ, which in turn model GDPR provisions. The formulæ therefore provide a connection between the BPMN activities and the provisions that affect them, both in text (via the Akoma Ntoso document) and in the deontic rule. In this way, the extended model can provide assistance both to the legal expert and to the software designer. Additionally, the connection is not textual but semantic (as it is based on an ontology), so it is possible to further enrich it by leveraging on the relations between the ontological entities, by means of metadata and connections to other sources.

The integration of BPMN and the GDPR model (see box Enriched BP Model in Figure 1) also allows to clearly assign tasks to the various stakeholders, for example by highlighting the duties of the processor with respect to the controller, or the activities of the data protection officer. From a visual perspective, it can assist in identifying the GDPR prerequisites (with reference to the example above, the duties to request consent and to provide information to the data subject). Additionally, a BPMN is not only a static visual diagram, but also a formal model that can be executed by appropriate engines. This feature can be exploited for GDPR compliance. Referring again to the above example, the execution engine could keep track of the fact that the data subject has consented to the processing, and when personal data are to be collected, the engine might prevent from executing further if it cannot verify that prerequisite.

3.1 Sample Implementation

In this section, we present a proposal for the infrastructure enriching the Business Process with annotations able to express data protection requirements. Among the several available tools for developing BP models, the proposal of this paper relies on Eclipse Plug-In technologies. In particular, the implementation is a plugin of the Eclipse BPMN2 Modeler running on top of recent versions of Eclipse (tested on 4.7 and subsequent versions). Note that the proposed plugin would be only an example of a possible integration between BPMN and the GDPR model.

Developers of a BP can exploit the plugin facilities to enrich their design with tasks specifically conceived to model processing activities that involve personal data. The generic Task has been extended so as to support annotations extracted from the ontology. This new task, called Data Protection Task and recognized by a distinctive graphical appearance (marked with a red icon), has been instantiated with the operations and related properties extracted from the ontology specification (see Figure 2). Moreover, a new tab called Data Protection has been included in the palette area of the BPMN2 Modeler interface (see Figure 3).

Figure 2: Specific tasks for Data Protection.

Figure 2 shows the properties of the interface of the data protection tasks. In particular, the first drop-down menu allows user to select the specific activity retrieved from the ontology.

For example, a data protection task which represents a data transmission has a Processing type Transmit, and refers to the Transmit class (a subclass of the Action class) in the PrOnto ontology. Likewise, from the perspective of the ontology, a BPMN task of the Transmit type represents an individual of the Transmit class. The PrOnto ontology can also be used to retrieve the relevant information that can be used as parameters of the task.

More specifically, in the properties window called Data Protection, all the obligations involving the Transmit action are listed, while the select drop-down menu labelled Destination country can accommodate the information relative to the country towards which the data should be transferred.

-section 4 will show an example of usage of the new Data Protection Task Transmit, delving into the technical details of its application.

---

4 https://www.eclipse.org/.
5 https://www.eclipse.org/bpmn2-modeler/.
6 An ongoing version of the presented plugin can be found at https://github.com/guerret/tu.uni.eclipse.bpmn2/.
4 APPLICATION EXAMPLE

In this section, we provide an application example of the proposed infrastructure. Figure 4 shows a realistic, albeit invented, BPMN process for the creation of a user account for a service that relies on cloud storage. The collaboration diagram shows three participants, each having a respective role in the GDPR: the customer is the data subject, the service provider is the controller, and the cloud storage provider is a processor.

Several tasks in the lanes of the controller and processor are the extended data protection tasks, as they entail some activity that is related to data processing regulations. In particular, Display terms of service is such an activity, despite not involving an actual processing of personal data, because the terms of service must comprise the information pursuant to GDPR Articles 13 and 14. The use of a data protection task thus emphasizes the need to include such information in the terms of service. In the absence of such information, any processing would be unlawful. All other data protection tasks in the diagram involve some processing activity (such as collecting, reading, storing, transmitting, or deleting personal data) as per the definition in GDPR Article 4(2).

Considering the described scenario, we focus, for example, on the Transfer to cloud activity (executed by System). According to Article 4(2) of the GDPR, the transmission of personal data is a processing activity, which requires a set of safeguards to ensure that the data subject’s rights are properly protected. Such safeguards include, for example, the need to inform the data subject about the categories of recipients to which the data might be sent. Additionally, in case the data are to be transmitted to a country outside the European Union, the transmission is lawful only if it is supported by adequate measures (an adequacy decision by the Commission, binding corporate rules, or an express authorization).

Figure 3 displays one possible use of integrating the BP with the underlying GDPR model. In this example, the tool is used to extract all the logic formulae (in RIO logic format) that are pertinent for that type of action.

Though the one in Figure 4 is a basic example, it shows the connection between the various components of the GDPR model. In particular, the various data protection tasks are mapped to classes in PrOnto that are subclasses of an Action class. The Transfer to cloud action is mapped to the Transfer class in PrOnto, with IRI https://w3id.org/ontology/pronto#Transmit, or, using prefixes, prOnto:Transmit. This IRI is also used in the RIO formulae: The LegalRuleML file contains XML Rel elements with an attribute iri pointing to that PrOnto class. Consequently, in the specific example, the XML tag will be <ruleml:Rel iri="prOnto:Transmit"/>

The sample implementation then retrieves from the LegalRuleML file all the relevant formulae, i.e., those that contain a reference to that class in the if part of the formula (as shown in Figure 3). These are all the applicable rules in case a Transmit action must be put into place. In particular, the formula containing an obligation show what requirements must be met before, during or after (depending on the specific provision) the transmission for it to be lawful.

Without delving into excessive details, considering the Transmit to cloud action, the first recurrence of the prOnto:Transmit predicate connects both to the Internationalized Resource Identifier (IRI) GDPR:art_13__para_1__content__list_1__point_1, where the GDPR prefix represents the URI of the Akoma Ntoso serialization of the GDPR, and to a set of statements (statements44 in the LegalRuleML file) that contain the logic model of the rules in RIO logic.
By navigating to the identifier in the Akoma Ntoso file, as shown in Figure 5, it is possible to retrieve the exact text of the GDPR provision that concerns the transmission, i.e., Article 13.1(f).

The extended BP can also be used more extensively to filter the rules. In other words, the applicable formulae can depend on specific parameters, which might be instantiated by the designer or passed at runtime during the execution of the BP. In the specific example, a transmission of personal data is subject to a number of safeguards in case the personal data must be transmitted towards a third country (outside the European Union), whereas such warranties are not needed if the recipient is located in a Member State. The different requirements are exemplified in the implementation by adding a parameter to a Transmit operation, with the Destination country. If the country belongs to the European Union, then a reduced set of formulae will be displayed, as some provisions are not applicable and therefore not required for the purpose of GDPR compliance.

5 RELATED WORK

Due to the complexity and the importance of the GDPR application, in recent years a lot of attention has been devoted to the clarification of data protection principles, policies and regulations (IT Governance Privacy Team, 2017). At the same time, many supporting tools and applications have been developed to assist users in producing reports on GDPR compliance (Ferrara and Spoto, 2018).

Notwithstanding these important contributions, the integration of data protection rules into the commonly-used BPs is still an emerging challenge. Automatic assistance and guidance in the implementation of GDPR requirements is a key factor for its successful and effective application. To this end, recent proposals are mainly focusing on the improvement of privacy management, on the assessment of GDPR compliance, and on the integration of the GDPR into BPs. A complete overview of the active research fields about the GDPR application and implementation is out of the scope of this paper, as is an exhaustive literature survey of the most recent proposals. We will hereafter focus on the research activities closest to our proposal, and highlight the main improvements provided of our solution.

Concerning privacy management, there are several proposals in literature focusing on the inclusion of mandatory GDPR privacy aspects into the adopted business process (Notario et al., 2017; Buchmann and Anke, 2017; Sokolovska and Kocarev, 2018). These consider accountability and consent collection aspects in particular. The proposal of this paper is to extend these approaches including all of the GDPR requirements, not only those specifically related to the privacy ones.

Concerning the assessment of GDPR applications, several proposals focus on (model-based) approaches for privacy and security analyses in all stages of system development (Ahmadian et al., 2018; Bieker et al., 2018; Gonçalves et al., 2017). Other proposals (Basin et al., 2018; Duncan, 2018) target auditing with respect to the GDPR.

The proposal of this paper attempts to ease the assessment of GDPR requirements, by explicitly integrating them in the business process. Few proposals addressing the integration of the GDPR into BPs are currently available. An attempt to manually merge the GDPR formalisation into a specific business process has been provided by (Heuck et al., 2017). In (Ramadan et al., 2017), a model-based security engineering framework is presented for supporting a GDPR-compliant system design and implementation.

The proposal of this paper is to automate the GDPR requirements into BPMN specification and supervising activities that are qualified as the processing of personal data such as collection and transmission.

6 CONCLUSIONS

The GDPR represents a significant breakthrough in the digital economy, and brings a lot of changes to the way online services are delivered. This scenario calls for a new approach in the development of software systems, where legal requirements must be accounted for, just like the other requirements that a system must respond to. This work focuses on data protection requirements in particular, proposing a framework that
allows to express legal requirements from the GDPR into a Business Process. The legal requirements are modelled according to approaches used in legal informatics, such as legal ontologies. The integration of the two domains of legal informatics and business processes allows to not only express the legal requirements, but also associate them with specific activities that entail the processing of personal data. This connection can be used at all stages of the Software Development Life Cycle (SDLC): from the analysis of the requirements to the design, from development to testing, from deployment to monitoring.

The present work is a preliminary step to integrate legal requirements into the SDLC. Expressing the legal requirements is an added value by itself, as the BP can be used to coordinate activities in the organization and assign specific tasks with the related legal duties. However, this work needs to be thoroughly extended and validated with real case studies before it can unleash its full potential. For one, the logic formulæ expressing obligations, permissions and prohibitions still require supporting tools and methodologies.

ACKNOWLEDGMENTS

This work has been partially funded by the Luxembourg National Research Fund (FNR) CORE project C16/IS/11333956 “DAPRECO: DAta Protection REGulation Compliance”.

REFERENCES


Massacci, F., Prest, M., and Zannone, N. (2005). Using a security requirements engineering methodol-


