A Model-based Conceptualization of Requirements for Compliance Checking of Data Processing against GDPR

Orlando Amaral[a], Sallam Abualhaija[a], Mehrdad Sabetzadeh[b,a], and Lionel Briand[a,b]
[a]SnT, University of Luxembourg, Luxembourg
[b]School of EECS, University of Ottawa, Canada
{orlando.amaralcejas, sallam.abualhaija}@uni.lu, {m.sabetzadeh, lbriand}@uottawa.ca

***Abstract*—The General Data Protection Regulation (GDPR) has been recently introduced to harmonize the different data privacy laws across Europe. Whether inside the EU or outside, organizations have to comply with the GDPR as long as they handle personal data of EU residents. The organizations with whom personal data is shared are referred to as data controllers. When controllers subcontract certain services that involve processing personal data to service providers (also known as data processors), then a *data processing agreement (DPA)* has to be issued. This agreement regulates the relationship between the controllers and processors and also ensures the protection of individuals' personal data. Compliance with the GDPR is challenging for organizations since it is large and relies on complex legal concepts. In this paper, we draw on model-driven engineering to build a machine-analyzable conceptual model that characterizes DPA-related requirements in the GDPR. Further, we create a set of criteria for checking the compliance of a given DPA against the GDPR and discuss how our work in this paper can be adapted to develop an automated compliance checking solution.**

***Index Terms*—Conceptual Modeling, Qualitative Research, Regulatory Compliance, Data Processing Agreements, General Data Protection Regulation (GDPR).**

I. Introduction

The General Data Protection Regulation (GDPR) is introduced by the European Union (EU) to harmonize the different data protection laws across Europe [1]. The regulation has been in place since May 2018. Regardless of whether they are inside or outside the EU, organizations are obliged to comply with the GDPR as long as they collect or process personal data about EU residents. Violating the privacy standards in the GDPR can result in penalties up to tens of millions of euros [2]. At the time when more people, for various reasons, are entrusting organizations with their personal data, the GDPR provides far-reaching capabilities to protect individuals' rights and increase organizational accountability for data breaches.

Understanding how to comply with the GDPR is a daunting task for organizations [3]. Though the GDPR is written in natural language (NL), it is large, covers diverse aspects of data protection and privacy issues, and is prone to ambiguity inherent in NL.

Modelling the requirements of the GDPR can be performed in order to help organizations understand what is needed for compliance checking. In this paper, we address the above challenges for compliance checking against the GDPR with a focus on data processing agreements (DPAs). Specifically, we help requirements engineers and other business stakeholders understand where to focus their GDPR compliance efforts. A DPA is required when a data controller subcontracts to a data processor certain services that involve processing personal data. Fig. 1 illustrates the flow of personal data in the context of a DPA. To illustrate, consider the following example about a DPA for payroll services. Let *Amazon.com, Inc.* in Italy be the data controller. The data subjects are Amazon employees in that branch. Amazon can subcontract payroll services to an accounting firm (a data processor) in the UK, e.g., *Simpson Wreford LLP*. To deliver payroll services, the accounting firm develops automated systems that process personal data including but not limited to: name, contact details, employee ID, bank account details, marital status, leave records, and contract of employment. The processor can also subcontract other processors. To ensure the protection of individuals' personal data, the GDPR imposes obligations on organizations (both data controllers and processors).

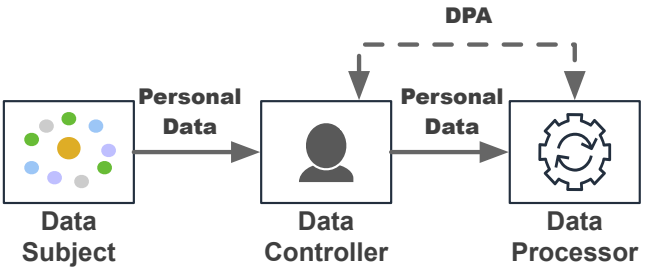


Fig. 1: The Flow of Personal Data in Context of DPA.

A DPA is a legally binding contract or legal act under Union or Member State law between a data controller and processor. The agreement sets out the obligations of the processor with regard to the controller and regulates the relationship between the two parties (the GDPR, Art.28[1]). Table I provides key definitions in the GDPR that are related to a DPA. The data controller is required by the GDPR to take measures to ensure the protection of personal data. Through a DPA, the controller and processor share legally-protected personal data as well as the consequences and costs of data breaches.

[1]Throughout the paper, we will use the abbreviation Art. for Article.

TABLE I: DPA-related Key Definitions in GDPR.

Concept	Definition	Reference
Data Subject	A natural person who can be identified, directly or indirectly, by reference to an identifier like a name or location data.	Art.4(1)
Personal Data	Any information relating to a data subject.	Art.4(1)
Processing	Any operation or set of operations which is performed on personal data, such as collection, recording, storage, disclosure by transmission, restriction, erasure or destruction.	Art.4(2)
Data 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.	Art.4(7)
Data Processor	A natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.	Art.4(8)

To comply with the GDPR, both the controller and processor are obliged to implement DPA-related requirements. Compliance also includes implementing these requirements in the systems that process personal data. For example, to ensure an appropriate level of data protection, the personal data can be pseudonymized or encrypted (the GDPR, Art.32(1)).

Through qualitative methods, we present an approach for checking compliance of DPAs against the GDPR. In particular, the paper makes three primary contributions.

(1) We draw on model-driven engineering (MDE) [4] to build a machine-analyzable conceptual model that describes the requirements of data processing as stipulated in the GDPR.

(2) We create a set of criteria to check whether a DPA is GDPR compliant.

(3) We conduct a preliminary evaluation of our compliance criteria on 24 DPAs. This work provides a first step towards the development of future automated methods for assessing the compliance of DPAs according to the GDPR.

The paper is organized as follows: Sec. II provides an overview of the related work. Sec. III discusses our approach for building the conceptual model and criteria for compliance checking. Sec. IV illustrates a usage scenario of our proposed approach for checking the compliance of 24 DPAs. Sec. V concludes the paper and presents future directions.

II. Related Work

In the context of regulatory compliance, modelling requirements from normative text in general and the GDPR in particular has been widely studied in the requirements engineering (RE) literature [5]–[8]. Torre et al. [9] propose the use of model-driven engineering as a platform for GDPR-compliance automation. Caramujo et al. [10] target privacy policies from the web and mobile applications, and propose a domain-specific language along with model transformations for specifying privacy-policy models. Bhatia et al. [11] propose an incompleteness detection method based on semantic frames in privacy requirements. Torre et al. [12] present an artificial intelligence-enabled automation for compliance checking of privacy policies based on a conceptual model that characterizes the privacy requirements from the GDPR. Vanezi et al. [13] propose a graphical modeling language for GDPR privacy policies and a methodology for transforming them into formal definitions. Pullonen et al. [14] present a multi-level model to be used as an extension of the Business Process Model and Notation to enable the visualization, analysis, and communication of the privacy-policy characteristics of business processes.

As far as the analysis of documents and specifications pertinent to IT systems is concerned, RE has focused primarily on privacy policies. Our work, in contrast, concentrates on DPAs, which are yet another highly relevant class of legal documents with a bearing on privacy requirements in IT systems. In this paper, we take preliminary steps in the direction of automating the compliance checking of DPAs according to the GDPR.

III. A Qualitative Study of DPA-related requirements in the GDPR

In this section, we present our qualitative study to derive: (1) a conceptual model describing the data processing requirements in the GDPR; and (2) a set of criteria for checking the compliance of DPAs according to the GDPR.

A. Conceptual Model of Data Processing Requirements

A DPA is a text document which, to be GDPR compliant, has to include the requirements that regulate the processing of personal data. The conceptual model, depicted in Fig. 2, captures and categorizes a total of 45 information types that can be stated in a DPA. The model is hierarchical and structured in three levels. Level 1 contains 12 information types (shaded grey), level 2 contains 16 information types (shaded blue) and level 3 contains 17 information types (shaded white). Some of these information types are critical, i.e., being absent leads to non-compliance (in font color black). Following best practices, the model also includes additional information types that are optional but commonly mentioned in a DPA (font color blue).

B. DPA Compliance Checking Criteria based on the GDPR

We created a set of 14 criteria for checking compliance of a given DPA against the GDPR, presented in Table II. More details are given in the next sub-section. Since optional information types do not affect the final decision as to whether a DPA complies with the GDPR, the criteria cover only the information types that are required. A criterion is considered satisfied only if the information types associated with that criterion are mentioned in the DPA, otherwise the criterion is violated. Accordingly, the DPA is considered compliant if all of the criteria are satisfied.

Some of these criteria are concerned only with the content of a DPA, while some others concern the systems used for processing personal data. In particular, the processor's obligation includes removing or returning the personal data to the controller after the processing service is terminated. For example, in the case of cloud-service providers acting as processors, it might not be possible to delete the data that is stored in archives immediately but in the next deletion

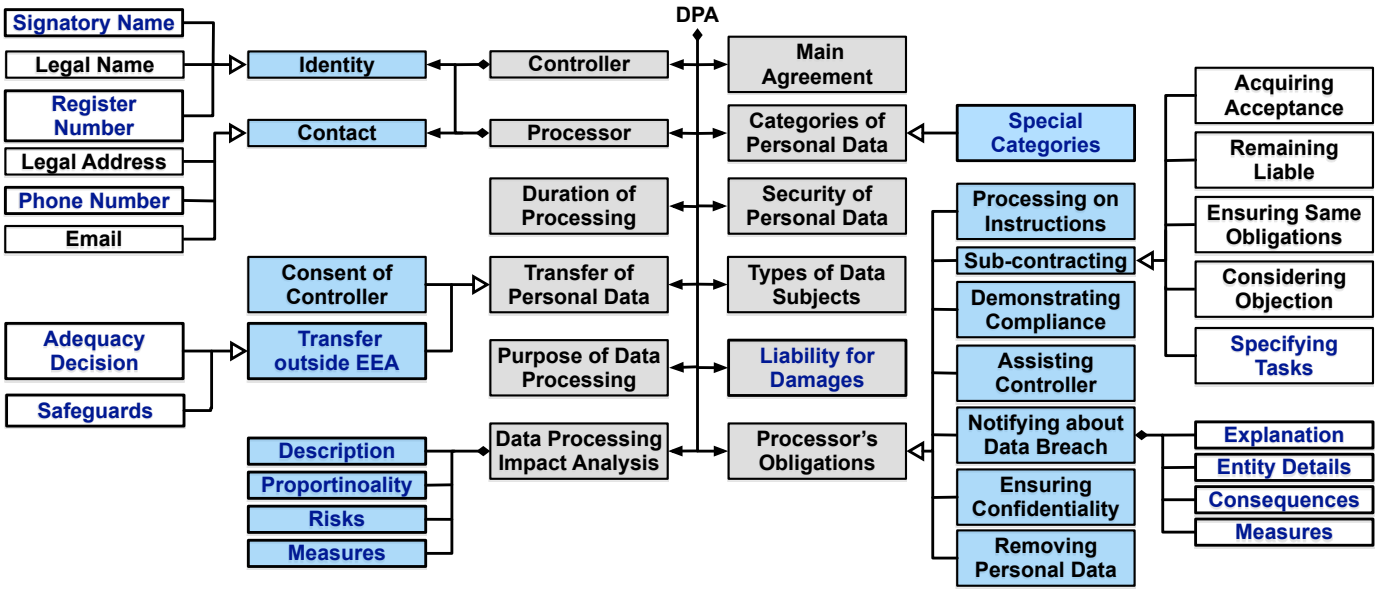


Fig. 2: Conceptual Model of Data Processing Requirements in the GDPR

round. Further, the processor must ensure the protection of personal data by applying security measures like encryption. For transferring personal data, e.g., for storage or processing purposes by a third-party, the processor has to acquire a written authorization from the data controller. If the above constraints are not satisfied by the system used by the processor, then this is a violation that can lead to penalties.

C. Qualitative Method

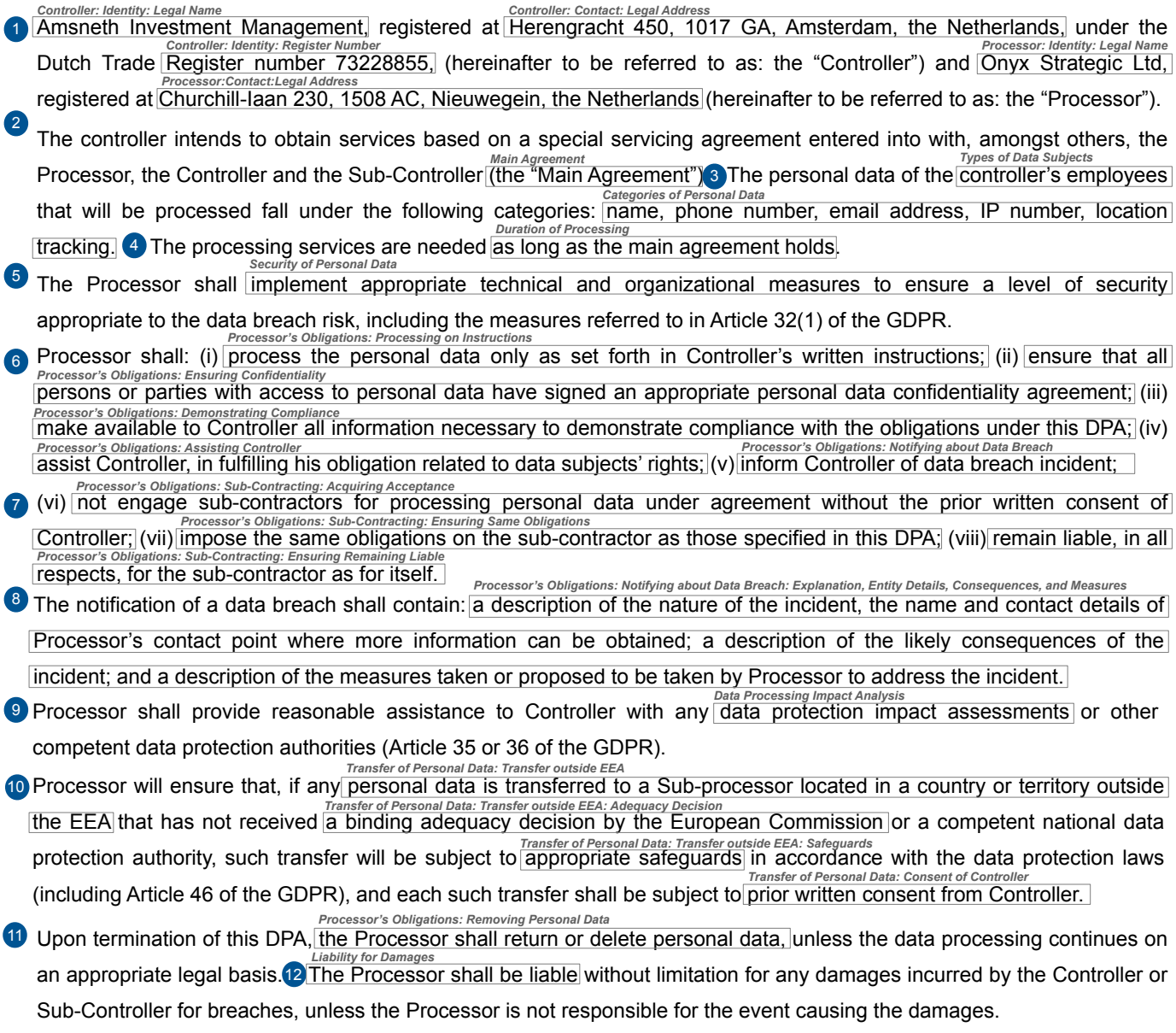
We created the above artifacts in close interaction with legal experts from *Linklaters LLP*, a global law firm headquartered in London and based in Luxembourg. To do so, our qualitative study followed an iterative and incremental process divided into two steps, as we explain next.

The first step involved reading DPA-relevant articles in the GDPR as pointed out by Linklaters. We started with Art.28 which is the main article in the GDPR about DPA. Further, we read through all referenced articles in Art.28, namely Articles 32–36, 40, 42, 43, 63, 82–84, and 93(2). During this step, we applied an in-vivo coding [15] to identify an initial set of codes representing the concepts in the GDPR. These codes are the information types that we present in Fig. 2. The initial set of codes were carefully discussed with Linklaters and refined according to their feedback. To enrich our conceptual model, we also applied sub-coding [15] to specialize the codes we had found in the GDPR articles with sub-codes (second and third levels in Fig. 2). For instance, Art.28(3)(a) states that *"the processor shall process the personal data only on documented instructions from the controller, including with regard to transfers of personal data to a third country or an international organisation [...]"*. From this text, the information type *Transfer of Personal Data* was first encoded and further specialized with *Controller Consent*. According to the feedback from Linklaters, the documented instructions entail a written authorization for transferring personal data. In this step, we also discussed with the legal experts the compliance checking criteria listed in Table II.

TABLE II: DPA Compliance Checking Criteria.

ID	Criteria
C1	At least one *Controller*:*Identity*:*Legal Name* has to be mentioned in the DPA.
C2	At least one *Controller*:*Contact*:*Legal Address* or *Email* has to be mentioned in the DPA.
C3	At least one *Processor*:*Identity*:*Legal Name* has to be mentioned in the DPA.
C4	At least one *Processor*:*Contact*:*Legal Address* or *Email* has to be mentioned in the DPA.
C5	The applicable *Main Agreement* has to be referred to in the DPA.
C6	The *Purpose of Processing* has to be mentioned in the DPA.
C7	The *Duration of Processing* has to be mentioned in the DPA.
C8	The *Security of Personal Data* has to be mentioned in the DPA.
C9	The *Types of Data Subjects* have to be mentioned in the DPA.
C10	The *Categories of Personal Data* have to be mentioned in the DPA.
C11	The *Transfer of Personal Data*: based on *Consent of Controller* has to be mentioned in the DPA.
C12	The *Processor Obligations*:*Processing on Instructions*, *Sub-Contracting*, *Demonstrating Compliance*, *Assisting Controller*, *Notifying Data Breach*, *Ensuring Confidentiality*, and *Removing Personal Data* have to be mentioned in the DPA.
C13	The *Processor Obligations*:*Sub-Contracting*:*Acquiring Acceptance*, *Remaining Liable*, *Ensuring Same Obligations*, and *Considering Objection* have to be mentioned in the DPA.
C14	The *Data Processing Impact Analysis* has to be mentioned in the DPA.

In the second step, we applied the refined set of information types on a sample of DPAs, what is referred to as



1 Controller: Identity: Legal Name: Amsneth Investment Management, registered at Controller: Contact: Legal Address: Herengracht 450, 1017 GA, Amsterdam, the Netherlands, under the Dutch Trade Controller: Identity: Register Number: Register number 73228855, (hereinafter to be referred to as: the "Controller") and Processor: Identity: Legal Name: Onyx Strategic Ltd, registered at Processor:Contact:Legal Address: Churchill-laan 230, 1508 AC, Nieuwegein, the Netherlands (hereinafter to be referred to as: the "Processor").

2 The controller intends to obtain services based on a special servicing agreement entered into with, amongst others, the Processor, the Controller and the Sub-Controller Main Agreement: (the "Main Agreement").

3 The personal data of the Types of Data Subjects: controller's employees that will be processed fall under the following categories: Categories of Personal Data: name, phone number, email address, IP number, location tracking.

4 The processing services are needed Duration of Processing: as long as the main agreement holds.

5 The Processor shall Security of Personal Data: implement appropriate technical and organizational measures to ensure a level of security appropriate to the data breach risk, including the measures referred to in Article 32(1) of the GDPR.

6 Processor shall: (i) Processor's Obligations: Processing on Instructions: process the personal data only as set forth in Controller's written instructions; (ii) ensure that all Processor's Obligations: Ensuring Confidentiality: persons or parties with access to personal data have signed an appropriate personal data confidentiality agreement; (iii) Processor's Obligations: Demonstrating Compliance: make available to Controller all information necessary to demonstrate compliance with the obligations under this DPA; (iv) Processor's Obligations: Assisting Controller: assist Controller, in fulfilling his obligation related to data subjects' rights; (v) Processor's Obligations: Notifying about Data Breach: inform Controller of data breach incident;

7 (vi) Processor's Obligations: Sub-Contracting: Acquiring Acceptance: not engage sub-contractors for processing personal data under agreement without the prior written consent of Processor's Obligations: Sub-Contracting: Ensuring Same Obligations: Controller; (vii) impose the same obligations on the sub-contractor as those specified in this DPA; (viii) remain liable, in all Processor's Obligations: Sub-Contracting: Ensuring Remaining Liable: respects, for the sub-contractor as for itself.

8 The notification of a data breach shall contain: Processor's Obligations: Notifying about Data Breach: Explanation, Entity Details, Consequences, and Measures: a description of the nature of the incident, the name and contact details of Processor's contact point where more information can be obtained; a description of the likely consequences of the incident; and a description of the measures taken or proposed to be taken by Processor to address the incident.

9 Processor shall provide reasonable assistance to Controller with any Data Processing Impact Analysis: data protection impact assessments or other competent data protection authorities (Article 35 or 36 of the GDPR).

10 Processor will ensure that, if any Transfer of Personal Data: Transfer outside EEA: personal data is transferred to a Sub-processor located in a country or territory outside the EEA that has not received Transfer of Personal Data: Transfer outside EEA: Adequacy Decision: a binding adequacy decision by the European Commission or a competent national data protection authority, such transfer will be subject to Transfer of Personal Data: Transfer outside EEA: Safeguards: appropriate safeguards in accordance with the data protection laws (including Article 46 of the GDPR), and each such transfer shall be subject to Transfer of Personal Data: Consent of Controller: prior written consent from Controller.

11 Upon termination of this DPA, Processor's Obligations: Removing Personal Data: the Processor shall return or delete personal data, unless the data processing continues on an appropriate legal basis.

12 Liability for Damages: The Processor shall be liable without limitation for any damages incurred by the Controller or Sub-Controller for breaches, unless the Processor is not responsible for the event causing the damages.

Fig. 3: Illustrative Example of an Annotated DPA.

hypothesis coding [15]. The intuition of this step is to check the applicability of our conceptual model on a real example and also to ensure that the conceptual model covers all the necessary information types required for compliance checking. While performing hypothesis coding, we kept track of encoded additional information types that were mentioned in more than five DPAs but were not in our conceptual model. According to feedback from Linklaters, these additional information types are considered best practice and should be included in DPAs. Therefore, we added such information types (those in blue font) to our conceptual model.

IV. Compliance Checking of DPAs

In this section, we illustrate how our proposed artifacts (presented in Sec. III) can be utilized for checking the compliance of DPAs against the GDPR. The first step to check whether a DPA is GDPR compliant is to extract and categorize the information type according to our conceptual model (in Fig 2). To do so, we analyzed a set of 24 DPAs manually. Note that this manual analysis has been a part of our qualitative method as explained in Sec. III. For a given DPA, we annotated each sentence with one or more information types. We skip the sentences that do not mention any such types. Illustrative examples of annotated sentences in a DPA is provided in Fig. 3. The second step is to apply the criteria in Table II to check whether the DPA is compliant. The DPA in Fig. 3 is not compliant because it violates C6, since it does not state the purpose of processing. Table III shows the results of our manual compliance checking of 24 DPAs according

to our proposed artifacts in Sec. III. Given the granularity level of C12 and C13, covering respectively seven and four information types, it is notable that 20 out of 24 DPAs violate C13 while only two DPAs violate C12. This high number of violations for C13 is due to this requirement covering level-3 specializations, which are less often detailed in the DPAs that we analyzed. Among the 20 DPAs violating C13, the information type *Ensuring Same Obligations* is missing in eight DPAs, *Considering Objection* in another eight, *Acquiring Acceptance* in three, and finally *Remaining Liable* is missing in only one DPA.

TABLE III: Results of Manual Compliance Checking.

ID	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
NV[1]	4	10	4	9	2	5	4	4	10	2	7	2	20	4

[1] NV: number of DPAs violating the corresponding criterion.

V. Conclusion and Future Directions

In this paper, we built two artifacts from the General Data Protection Regulation (GDPR) for compliance checking of Data Processing Agreements (DPAs). Drawing on model-driven engineering and using qualitative methods, we first developed a machine-analyzable conceptual model which characterizes DPA-related requirements as stipulated by the GDPR. The conceptual model consists of a total of 45 information types that should be addressed in a DPA. Building on this conceptual model, we then created a set of 14 criteria for compliance checking of DPAs against the GDPR. These two artifacts were developed in close interaction with legal experts.

The work presented in this paper provides a first step towards an automated solution for checking the compliance of DPAs against the GDPR. As future research directions, we plan to develop automated support for this task by combining natural language processing (NLP) and machine learning (ML). NLP and ML enable the extraction and classification of the different information types from DPAs according to our proposed conceptual model. To this end, we plan to, first, create an annotated dataset that serves as training examples based on which we can develop and evaluate an automated solution. Further, the compliance criteria have to be formalized into verifiable constraints to facilitate compliance checking.

Acknowledgments

This paper was supported by Linklaters, Luxembourg's National Research Fund (FNR) under grant BRIDGES/19/IS/13759068/ARTAGO, and NSERC of Canada under the Discovery, Discovery Accelerator and CRC programs.

References

[1] European Union, "General data protection regulation," *Official Journal of the European Union*, 2018, accessed on June 30th, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679

[2] EU-GDPR. EU GDPR portal. Accessed on June 30th, 2021. [Online]. Available: https://gdpr.eu/

[3] C. Tankard, "What the GDPR means for businesses," *Network Security*, vol. 6, pp. 5–8, 2016.

[4] M. Brambilla, J. Cabot, and M. Wimmer, *Model-Driven Software Engineering in Practice*, 2nd ed. Morgan & Claypool Publishers, 2016.

[5] J. C. Maxwell and A. I. Antón, "The production rule framework: developing a canonical set of software requirements for compliance with law," in *proceedings of the 1st ACM International Health Informatics Symposium*, 2010, pp. 629–636.

[6] J. García-Galán, L. Pasquale, G. Grispos, and B. Nuseibeh, "Towards adaptive compliance," in *2016 IEEE/ACM 11th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*. IEEE, 2016, pp. 108–114.

[7] P. Engiel, J. C. S. D. P. Leite, and J. Mylopoulos, "A tool-supported compliance process for software systems," in *2017 11th International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 2017, pp. 66–76.

[8] N. Zeni, E. Seid, P. Engiel, and J. Mylopoulos, "NómosT: Building large models of law with a tool-supported process," *Data and Knowledge Engineering*, vol. 117, pp. 407–418, 2018.

[9] D. Torre, M. Alférez, G. Soltana, M. Sabetzadeh, and L. C. Briand, "Model driven engineering for data protection and privacy: Application and experience with GDPR," *CoRR*, vol. abs/2007.12046, 2020. [Online]. Available: https://arxiv.org/abs/2007.12046

[10] J. Caramujo, A. Rodrigues da Silva, S. Monfared, A. Ribeiro, P. Calado, and T. Breaux, "RSL-IL4Privacy: A domain-specific language for the rigorous specification of privacy policies," *Requirements Engineering*, vol. 24, no. 1, pp. 1–26, 2019.

[11] J. Bhatia and T. D. Breaux, "Semantic incompleteness in privacy policy goals," in *26th IEEE International Requirements Engineering Conference, RE 2018, Banff, AB, Canada, August 20-24, 2018*, 2018, pp. 159–169.

[12] D. Torre, S. Abualhaija, M. Sabetzadeh, L. C. Briand, K. Baetens, P. Goes, and S. Forastier, "An ai-assisted approach for checking the completeness of privacy policies against GDPR," in *28th IEEE International Requirements Engineering Conference, RE 2020, Zurich, Switzerland, August 31 - September 4, 2020*, 2020, pp. 136–146.

[13] E. Vanezi, G. M. Kapitsaki, D. Kouzapas, A. Philippou, and G. A. Papadopoulos, "Diálogop - A language and a graphical tool for formally defining GDPR purposes," in *Research Challenges in Information Science - 14th International Conference, RCIS 2020, Limassol, Cyprus, September 23-25, 2020, Proceedings*, ser. Lecture Notes in Business Information Processing, F. Dalpiaz, J. Zdravkovic, and P. Loucopoulos, Eds., vol. 385. Springer, 2020, pp. 569–575.

[14] P. Pullonen, J. Tom, R. Matulevicius, and A. Toots, "Privacy-enhanced BPMN: Enabling data privacy analysis in business processes models," *Software & Systems Modeling*, pp. 1–30, 2019.

[15] J. Saldana, *The Coding Manual for Qualitative Researchers*. SAGE Publishing, 2016.