

Ontology-Based Privacy Compliance Checking for Clinical Workflows

Saliha Irem Besik and Johann-Christoph Freytag

Humboldt-Universität zu Berlin, Department of Computer Science,
Unter den Linden 6, 10099 Berlin, Germany
{besiksal,freytag}@informatik.hu-berlin.de

Abstract. Data privacy is an essential human right to determine what, when, and how personal data is communicated to various recipients. In the healthcare domain, it is an important and challenging issue how to safeguard data privacy of patients. Healthcare providers have to process sensitive medical data compliantly with binding privacy regulations such as the European Union General Data Protection Regulation. Clinical workflows play an important role in healthcare domain by outlining the tasks must be done for the delivery of clinical services. However, in general, they do not support privacy constraints in an adequate way. In this paper, we propose an ontology-based privacy compliance check approach to detect the possible privacy violations in clinical workflows. In order to analyze the potential applicability of our methodology, we describe a Newborn Screening scenario where we show how to apply semantic reasoning to support building privacy-awareness.

Keywords: Data Privacy · General Data Protection Regulation (GDPR) · Privacy Policies · Privacy Preferences · Ontology · Ontology Reasoning · Business Process Compliance

1 Introduction

The European Union General Data Protection Regulation (GDPR) has come into force very recently to protect data privacy of all individuals within the European Union [1]. “Privacy by Design” (PbD) is a core principle according to the GDPR and it obliges the organizations to proactively embed privacy into their technology design [GDPR, Article 23].

In order to support PbD, the organizations can greatly benefit from the business process models via checking privacy compliance of their business process models during the design time. Business process modeling and management are mostly based on a *control-flow-centric* perspective which emphasizes on the sequencing of activities with ignoring the data aspects. However, we believe that business process models can also be considered as a means to capture how data

is transmitted by whom and for what purpose at the conceptual level. Business process models are usually captured using the Business Process Model and Notation (BPMN), hence, we use BPMN to model clinical workflows.

We represent privacy-awareness for the clinical workflows not only in terms of the regulatory compliance with the GDPR and the privacy policies of healthcare providers, but also in terms of the compliance with the privacy preferences of patients when sharing and/or processing their personal data. We have used an ontology-based reasoner to verify privacy compliance. Ontology development contributes to our research significantly in terms of sharing a common understanding among different domains, namely business process modeling, privacy, and clinical domains.

In this paper, we present our Privacy-aware Clinical Workflow (PaCW) ontology and we give a running example which is related to the newborn screening procedure applied in Germany to show how our reasoning approach works. We selected the clinical workflows as a case study because privacy in the clinical domain is particularly significant due to the high sensitivity of “data concerning health”. However, our proposed approach is domain-independent and can thus be applied to any other domain.

The rest of this paper is organized as follows: Section 2 explains our Privacy-aware Clinical Workflow (PaCW) Ontology in detail. Section 3 presents our reasoning approach to check privacy compliance and discusses the implementation details. Section 4 gives a running example in the clinical domain in order to show the usability of our reasoning approach. Section 5 reviews related works. Finally, Section 6 concludes this paper and discusses our future work and perspectives.

2 Privacy-aware Clinical Workflow (PaCW) Ontology

Our research brings business process modeling, data privacy, and clinical domains together. We used ontologies to bridge the gap and share a common understanding of these domains. We developed the Privacy-aware Clinical Workflow (PaCW) Ontology which consists of three ontologies which are Privacy Ontology, BPMN Ontology, and Clinical Domain Ontology. In this section, we propose Privacy Ontology and BPMN Ontology, as well as the mappings between them. We explain Clinical Domain Ontology while presenting our running example. We have developed our ontologies through the Unified Modeling Language (UML).

2.1 Privacy Ontology

Privacy compliance is mostly defined as a stakeholder’s accordance with established privacy policies and/or privacy regulations. We argue that privacy compliance has a broader definition considering also the data owners’ (data subjects’) personal privacy preferences. In this regard, we define “*privacy-awareness*” for clinical workflows as the compliance both with the privacy principles based on the GDPR and the privacy policies provided by healthcare providers, as well

as the compliance with the privacy preferences of patients on sharing or processing their personal medical data. In this section, we briefly introduce these three sources which contribute to our privacy-awareness definition and then we present our Privacy Ontology.

Privacy Principles based on the GDPR We have introduced the founding privacy principles for clinical workflows on the ground of the GDPR. Even though we have focused on the GDPR, we believe these principles are also relevant and valuable for other regulations.

1- Purpose Specification: “Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes [...]” [Article 5, §1(b)]

2- Data Minimization: “Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.” [Article 5, §1(c)]

3- Consent Check: “Processing shall be lawful if the data subject has given consent to the processing of his or her personal data for one or more specific purposes.” [Article 6, §1(a)]

4- Limited Retention Period: “Personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed [...]” [Article 5, §1(e)]

In summary, for data operations to be GDPR compliant:

1. data must be processed for a specified purpose,
2. the data owner has consented -unless the specific conditions such as “vital interest” or “public interest” are met for lawful processing-[Article 6],
3. be necessary to achieve the specified purpose,
4. and data must be erased when it is no longer necessary for the given purpose.

Privacy Policies According to the GDPR, service providers have to inform the data subjects about how their personal data is being used by specifying privacy policies. A privacy policy document is a high-level natural language description of the privacy practices of a service provider. Privacy policies describe what data is used, for what purpose and by whom, as well as how long the data will be retained. They might also contain the modality of data processing, whether it requires explicit consent or not.

Privacy Preferences We have constructed our privacy preference formulation based on Alan Westin’s definition of privacy. In his well-known book *Privacy and Freedom*, he defines privacy in terms of self-determination: “Privacy is right on decision of every individual: when, how and how much information will be available for storing and exchange between systems.” [2]. Therefore, we express privacy preferences as the right of data subjects to determine who can access their personal data and for what purposes.

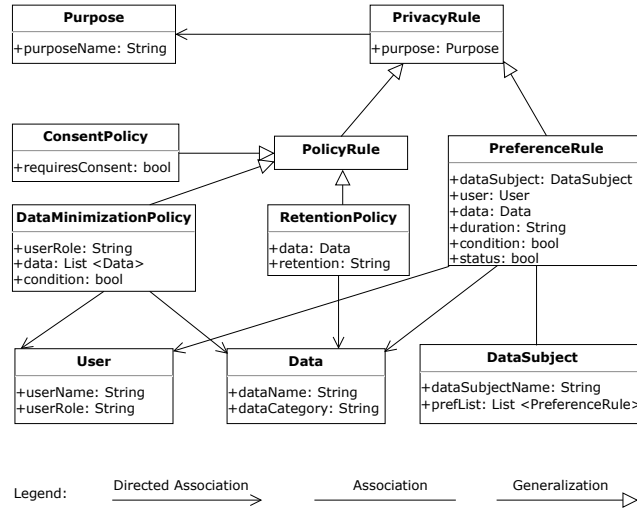


Fig. 1. The main components of privacy ontology.

We built privacy ontology according to these three sources. Figure 1 illustrates the main components of the privacy ontology: Both privacy policy statements (`PolicyRule`) and privacy preferences (`PreferenceRule`) are defined as `PrivacyRule`. Privacy rules are all purpose-focused, hence they are associated with the `Purpose` ontology class which specifies the reason for which data is collected, used, or disclosed. `PolicyRule` class has three sub-classes: `ConsentPolicy` ontology class defines which kind of `Purpose` instances require explicit consent. `RetentionPolicy` declares the retention period of data practices. `DataMinimizationPolicy` expresses the amount of personal data which should be revealed and processed by the users in different conditions. `User` is the set of individuals or organizations who accesses the personal data. `Data` is categorized to adopt the right privacy measures suitable for the type of data to be protected. We have used different data categories which are defined in the GDPR (*personal data*, *sensitive data*, *identification data*, *anonymous data*, *public data*, etc.) `DataSubject` refers to any individual person who can be identified. `prefList` represents a list of `PreferenceRule`.

2.2 BPMN Ontology

The Business Process Model and Notation (BPMN) is a widely used standard for business process modeling and also maintained by the Object Management Group (OMG). Therefore, we have selected BPMN 2.0 as the modeling notation for the clinical workflows.

Figure 2 shows the graphical representation of the core BPMN elements used in our clinical workflows. We have adapted the BPMN 2.0 Ontology presented in [3] to semantically represent these core BPMN elements.

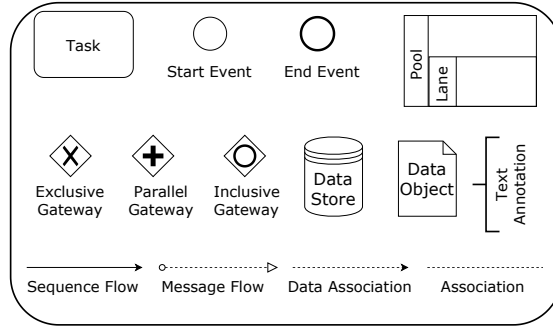


Fig. 2. Core BPMN elements.

Data privacy in clinical workflows is a subject of how data is handled in BPMN. Therefore, we have worked on different ways of data handling supported in BPMN (shown in Table 1) and we created new ontology classes to support data privacy in clinical workflows accordingly.

Figure 3 illustrates the main components of our BPMN ontology. It presents the important attributes and operations of the ontology classes and “*is-a*” hierarchy (generalization) among them. We omitted the association relations for the sake of illustration (except the one between BPMNModel and BPMNElement).

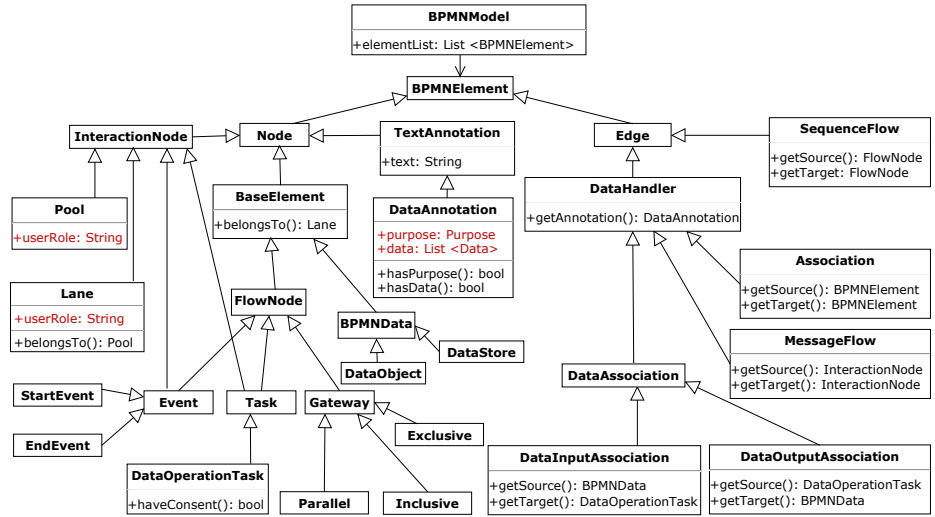


Fig. 3. The main components of BPMN ontology.

The ontology classes for the core BPMN elements are adapted from BPMN 2.0 Ontology [3]. The newly created ontology classes to support privacy-awareness

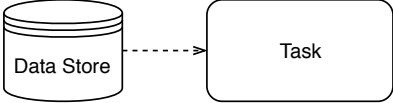
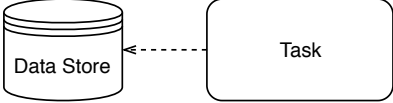
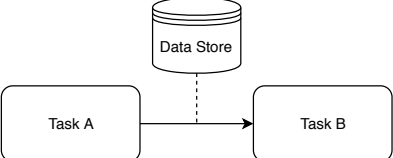
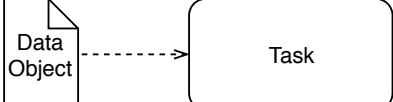

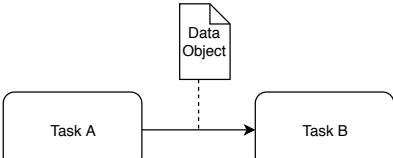
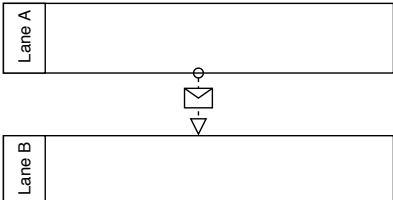
Data Flow	Graphical Representation	Meaning
Data Association		The task retrieves information from the data store
Data Association		The task writes information to the data store
Association		Task A writes information to the data store and Task B retrieves information from the data store
Data Association		The task receives data object
Data Association		The task sends data object
Association		Task A sends data object and Task B receives data object
Message Flow		Lane A sends message to Lane B in other words Lane B receives message from Lane A

Table 1. Data Handling in BPMN

in clinical workflows are `DataOperationTask`, `DataHandler`, and `DataAnnotation`. `DataOperationTask` is a type of `Task` which has a data association. As it is illustrated in Table 1, data is generally handled either via `DataAssociation` or via `MessageFlow`. Alternatively, data may be directly associated with a sequence flow via `Association`. `DataAnnotation` is a type of `TextAnnotation`. We assume that for each data operation in a BPMN workflow, it is known which data is used for which purpose explicitly. For this we use `DataAnnotation`, where `purpose` referring to the purpose of accessing data and `data` referring to a set of data which is accessed. `DataOperationTask` has `haveConsent()` function which aims to find out whether a data operation task has consented. The attributes written in red indicate that they are associated with the ontology classes from Privacy Ontology. The foundational components of Privacy Ontology, namely `User`, `Data`, and `Purpose` are represented in clinical workflows via a set of BPMN constructs. `User` is mapped onto the BPMN `Pool` and `Lane` elements. `Data` and `Purpose` are represented via the BPMN `Data Annotation` element.

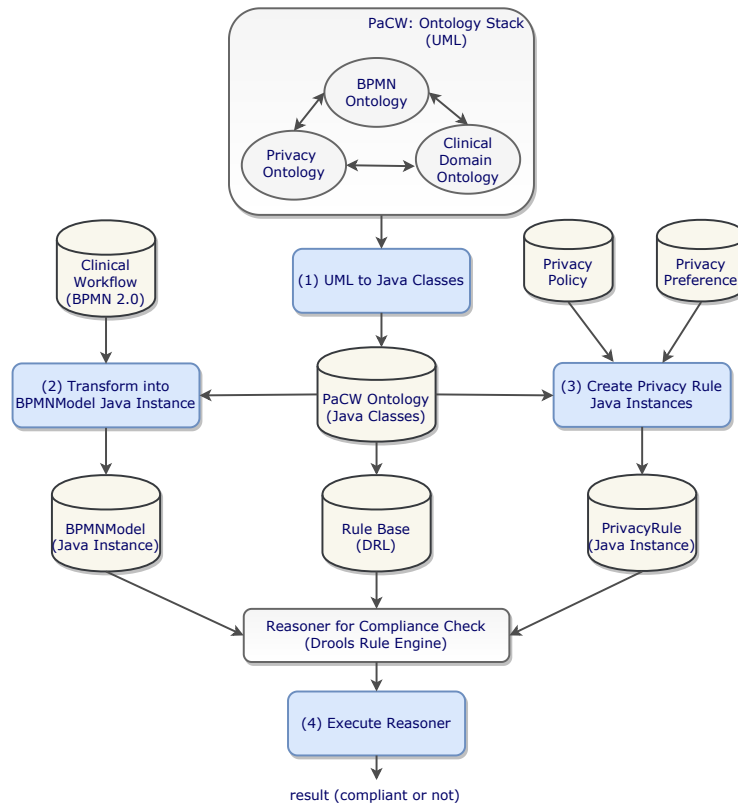


Fig. 4. Reasoning Approach.

3 Reasoning Approach and Implementation Details

Figure 4 gives an overview of our reasoning setting. The implementation steps can be summarized as follows:

(1) Transform UML ontology classes to Java classes: We worked with Eclipse Modeling Framework to be able to manipulate the instances of BPMN 2.0 and UML ontology. We used UML as an ontology developing language because it is widely adopted by the software engineering community and Eclipse provides an automatic transformation library from UML to Java programming language (Eclipse UML Generators¹) We created Java classes regarding for each ontology classes in our PaCW ontology. For instance; below you can see the Java class for the `ConsentPolicy` ontology class after the transformation by the UML Generator.

```
public class ConsentPolicy extends PolicyRule {
    private boolean requiresConsent;
    /* getter and setter functions*/
    // Constructor
    public ConsentPolicy(Purpose p, boolean requiresConsent){
        super (p);
        this.requiresConsent = requiresConsent;
    }
}
```

(2) Parse Clinical Workflow and Create a BPMNModel Instance accordingly: Our input clinical workflows are in BPMN 2.0 format. In order to parse these files, we used Camunda BPMN Model API². We accessed the BPMN elements via this Java library and we created `BPMNModel` Java Class instance by using the constructor of `BPMNModel` ontology class.

(3) Create Privacy Rule Instances: Our privacy rules are coming from either privacy policies or privacy preferences. We created `PrivacyPolicy` and `PrivacyPreference` Java Class instances by using the constructors of these ontology classes. For instance; below you can see a Java class instance for the `ConsentPolicy` ontology class.

```
public class PrivacyRuleInstances{
    ConsentPolicy p1 = new ConsentPolicy (new
        Purpose("hearing-screening"), true);
}
```

(4) Execute Compliance Check Rules via Rule Engine: We implemented our rule engine by using Drools Business Rule Engine.³ Eclipse offers a plugin for Drools which offers a development environment for rules manipulation. Drools

¹<https://eclipse.org/umlgen/>.

²<https://docs.camunda.org/manual/7.7/user-guide/model-api/bpmn-model-api/>.

³<https://www.drools.org/>.

is a Java-based open source business rule management system with a forward-chaining and backward-chaining inference based rules engine. Drools has its own rule language called Drools Rule Language (DRL). In DRL syntax, every rule has a “when” section which defines the conditions to be fulfilled to trigger the rule and a “then” section which defines the actions to be executed when the rule is triggered. DRL uses instances of Java classes, thus all concepts in the UML ontology are represented as plain Java classes. We created compliance rules regarding each privacy principles which we have proposed. Below, you can find two examples of DRL rules as “Purpose-Specification-Check” and “Consent-Check” DRL rules. “Purpose-Specification-Check” checks for each `DataHandler` whether there is a data annotation. It also finds out whether the data annotations have a purpose. If there is no data annotation or if there is no purpose for the data annotation, the rule gives a failure message.

```
rule "Purpose-Specification-Check"
  when
    BPMNModel.DataHandler(!hasDataAnnotation() ||
      !dataAnnotation.hasPurpose())
  then
    System.out.println("Compliance Check fail: Purpose-Specification");
  end
```

“Consent-Check” DRL rule checks whether the data associations requiring consent according to privacy policy rules have consented in the BPMN model. For this purpose, we use `haveConsent` function. It searches whether the consent check pattern (illustrated in Figure 5) precedes the data associations. If `haveConsent` function does not encounter the consent check pattern preceding a data association, DRL rule returns a failure message.

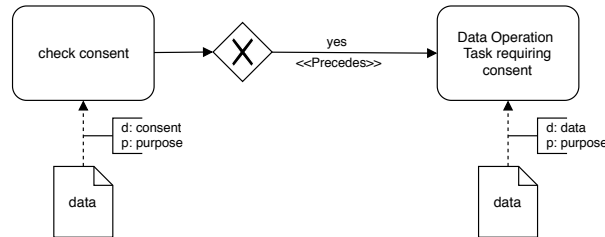


Fig. 5. Consent Check Pattern

```
rule "Consent-Check"
  when
    ConsentPolicy(requiresConsent = true, $p: purpose)
    BPMNModel.DataInputAssociation ($d: dataAnnotation, $d.getPurpose()
      = $p, !source.haveConsent()) ||
```

```

BPMNModel.DataOutputAssociation ($d: dataAnnotation, $d.getPurpose
    = $p, !target.haveConsent())
then
System.out.println("Compliance Check fail: Consent-Check");
end

```

4 A Running Example

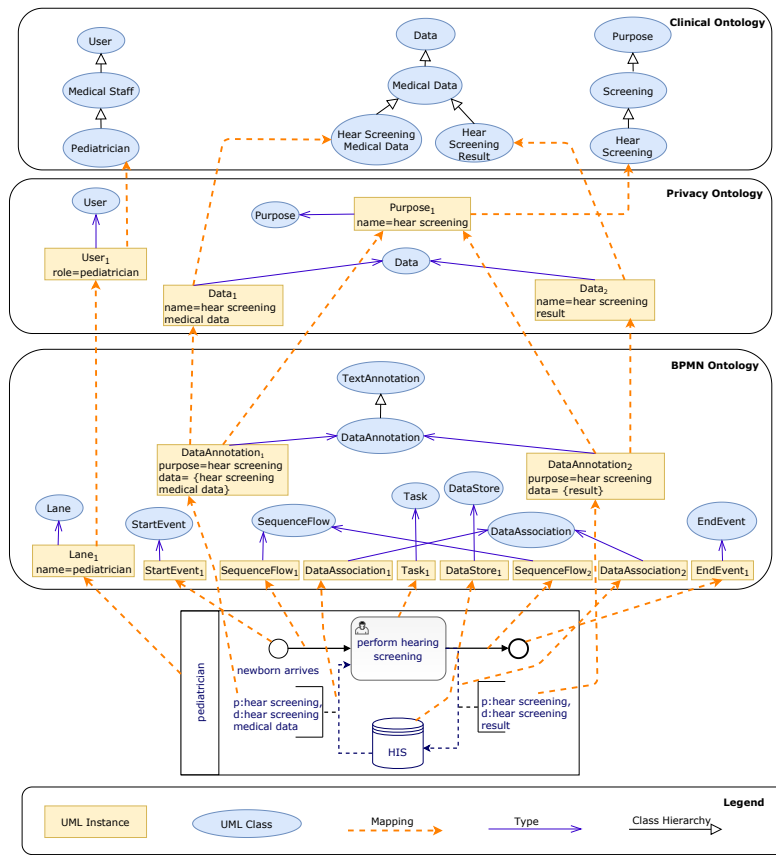


Fig. 6. The mapping between a sample BPMN Model and PaCW Ontology.

Figure 6 illustrates a simple BPMN model for a newborn hearing screening procedure. It also shows the mappings when we parse this model and create a `BPMNModel` instance accordingly. The mapping include some concepts from Clinical Ontology. For instance, according to the ontology *pediatrician* is a *medical*

staff. When a privacy rule is regarding *medical staff* as *user*, it is also about its subclasses including *pediatrician*.

In order to check privacy compliance, the compliance rules given in DRL are executed. In this section, we look at the compliance with the “Purpose-Specification-Check” and “Consent-Check” DRL rules. In the running example, there are two data associations which are also `DataHandler`. “Purpose-Specification-Check” rule checks the failure cases. It checks both data associations whether they have `DataAnnotation` and whether their data annotations have a purpose. Since both of the conditions return false, “then” part is not executed. Hence, compliance check for the purpose principle succeeds. Both data associations require consent according to the consent policy which we have shown as privacy rule instance. “Consent-Check” DRL rule checks both data associations whether they have consented. `haveConsent` function searches whether the consent check pattern precedes data associations. Since the condition is satisfied, compliance rule for the consent check returns true and it gives a failure message.

5 Related Work

Our research provides a holistic approach by combining privacy, business processes, and ontologies. We grouped the related works into three research areas:

1. **Privacy in Business Processes:** [4][5][6] propose BPMN extensions in order to capture the security and privacy needs. These works are focused on modeling privacy into business processes; however, they do not represent any reasoning to enforce privacy constraints. [7] introduces a framework for enforcing data privacy in the context of Data Analysis Workflows. It provides an ontology to incorporate the Data Analysis workflows and traditional privacy-preserving algorithms. Their privacy definition significantly differs from our definition because they focus only on privacy-preserving algorithms, not the regulatory compliance.
2. **Semantic Approaches in Business Process Compliance:** [8] proposes an ontology-based approach to detect possible semantic errors in business processes designed by using Coloured Petri Net (CPN). The work represents the business ontology in Ontology Web Language (OWL) and the business rules in Semantic Web Rule Language (SWRL). It checks the compliance in between through a reasoner. [9] checks the semantic correctness of CPN processes with the SPARQL query language. However, these works do not include privacy.
3. **Semantic Approach for Privacy Compliance:** Rahmouni et al. built an ontology of privacy requirements for sharing medical data between different healthcare organizations in European countries [10]. Belaazi et al. provide an ontology-based approach for access control [11]. These works do not include business processes.

6 Conclusion and Future Work

In this paper, we introduced our Privacy-aware Clinical Workflow (PaCW) ontology and we provided our semantic reasoning approach for privacy compliance through a running example from the clinical domain. Our focus was detecting privacy violation via ontology-based reasoning. As a future work, we are also interested in providing corrective actions in terms of how to fix the causes of violation and transform the non-privacy-aware workflow into a privacy-aware workflow accordingly.

References

1. EU General Data Protection Regulation (GDPR). Regulation (EU) 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, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119.
2. Alan F Westin and Oscar M Ruebhausen. *Privacy and freedom*, volume 1. Atheneum New York, 1967.
3. Christine Natschläger. Towards a BPMN 2.0 ontology. In *International Workshop on Business Process Modeling Notation*, pages 1–15. Springer, 2011.
4. Wadha Labda, Nikolay Mehandjiev, and Pedro Sampaio. Modeling of privacy-aware business processes in BPMN to protect personal data. In *Proc. of the 29th Annual ACM Symposium on Applied Computing*, pages 1399–1405. ACM, 2014.
5. Cesare Bartolini, Robert Muthuri, and Cristiana Santos. Using ontologies to model data protection requirements in workflows. In *JSAI Int. Symposium on Artificial Intelligence*, pages 233–248. Springer, 2015.
6. Alfonso Rodríguez, Eduardo Fernández-Medina, and Mario Piattini. A BPMN extension for the modeling of security requirements in business processes. *IEICE transactions on information and systems*, 90(4):745–752, 2007.
7. Yolanda Gil, William K Cheung, Varun Ratnakar, and Kai-kin Chan. Privacy enforcement in data analysis workflows. In *Proceedings of the 2007 International Conference on Privacy Enforcement and Accountability with Semantics-Volume 320*, pages 41–48. Citeseer, 2007.
8. Tuan Pham and Nhan Le Thanh. A Ontology-based Approach for Business Process Compliance Checking. In *IMCOM'16-the 10th Int. Conference on Ubiquitous Information Management and Communication*, pages 1–6. ACM SIGAPP, 2016.
9. Thi Hoa Hue Nguyen and Nhan Le Thanh. Ensuring the semantic correctness of workflow processes: an ontological approach. In *Proceedings of 10th Workshop on Knowledge Engineering and Software Engineering (KESE10) co-located with 21st European Conference on Artificial Intelligence (ECAI 2014)*, volume 1289, 2014.
10. Hanene Boussi Rahmouni, Tony Solomonides, Marco Casassa Mont, and Simon Shiu. Privacy compliance and enforcement on European healthgrids: an approach through ontology. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926):4057–4072, 2010.
11. Maherzia Belaazi, Hanen Boussi Rahmouni, and Adel Bouhoula. An Ontology Regulating Privacy Oriented Access Controls. In *Int. Conf. on Risks and Security of Internet and Systems*, pages 17–35. Springer, 2015.