Sushant Agarwal and Simon Steyskal and Franjo Antunovic and Sabrina Kirrane

Legislative Compliance Assessment: Framework, Model and GDPR Instantiation

Conference or Workshop Item (Accepted for Publication)
(Refereed)

Legislative Compliance Assessment: Framework, Model and GDPR Instantiation

Sushant Agarwal, Simon Steyskal, Franjo Antunovic and Sabrina Kirrane

Vienna University of Economics and Business, Vienna, Austria
firstname.lastname@wu.ac.at

Abstract. Legislative compliance assessment tools are commonly used by companies to help them to understand their legal obligations. One of the primary limitations of existing tools is that they tend to consider each regulation in isolation. In this paper, we propose a flexible and modular compliance assessment framework that can support multiple legislations. Additionally, we describe our extension of the Open Digital Rights Language (ODRL) so that it can be used not only to represent digital rights but also legislative obligations, and discuss how the proposed model is used to develop a flexible compliance system, where changes to the obligations are automatically reflected in the compliance assessment tool. Finally, we demonstrate the effectiveness of the proposed approach through the development of a General Data Protection Regulatory model and compliance assessment tool.

Keywords: Compliance, GDPR, ODRL

1 Introduction

The interpretation of legal texts can be challenging, especially for people with non-legal backgrounds, as they often contain domain-specific definitions, cross-references and ambiguities [29]. Also, generally speaking legislations cannot be considered in isolation, for instance European Union (EU) regulations often contain opening clauses that permit Member States to introduce more restrictive local legislation. Additionally, depending on the legislative domain additional legislations may also need to be consulted. For example, when it comes to data protection in the EU, in addition to the General Data Protection Regulation (GDPR) [4], the upcoming e-privacy regulation (for e-communication sector) [5] or the Payment services (PSD 2) directive (for payments sector) [3] may also need to be consulted. As such, ensuring compliance with regulations can be a daunting task for many companies, who could potentially face hefty fines and reputation damage if not done properly. Consequently, companies often rely on legislative compliance assessment tools to provide guidance with respect to their legal obligations [8].

Over the years, several theoretical frameworks that support the modelling of legislation have been proposed [7, 10, 14, 22, 23, 25, 32], however only some of which were validated via the development of legal support systems [7,10,23,25,32]. One of the major drawbacks of such approaches is the fact that some do not consider concepts like soft-obligations (i.e. obligations that serve as recommendations rather than being mandatory) [22, 25] or exceptions (i.e. scenarios where the obligations are not applicable) [10, 29]. Additionally generally speaking the models are only loosely coupled with the actual legislation text, making it difficult to verify the effectiveness of such systems. More recently, a number of compliance assessment tools have been developed [18, 26, 28]. However, these systems are either composed of a handful of questions that are used to evaluate legal obligations [18] or do not filter out questions that are not applicable for the company completing the assessment [26, 28]. One of the primary drawbacks of existing compliance assessment tools is the fact that they do not currently consider related regulations.

In order to address this gap, we propose a generic legislative compliance assessment framework, that has been designed to support multiple legislations. Additionally, we extend the Open Digital Rights Language (ODRL) [34] (which is primarily used for rights expression) so that it can be used

to express legislative obligations. Both of which are necessary first steps towards a context dependent compliance system that can easily be adapted for different regulatory domains.

The contributions of the paper are as follows:(i) we devise a flexible and modular compliance assessment framework, which is designed to support multiple legislations; (ii) we propose a legislative ODRL profile that can be used to model obligations specified in different legislations; and (iii) we develop a dynamic compliance system that can easily be adapted to work with different legislations. The proposed framework is instantiated in the form of a GDPR compliance assessment tool, which is subsequently compared with alternative approaches.

The remainder of the paper is structured as follows: Section 2 presents different approaches that can be used to model data protection legislations, along with compliance assessment tools for the GDPR. Section 3 details our framework that decouples the legislative obligations from the compliance assessment tool. Section 4 introduces our legislative model and illustrates how it can be used to model the GDPR. Section 5 describes the compliance tool. In Section 6 we compare and contrast our proposal with alternative solutions. Finally, Section 7 concludes the paper and presents directions for future work.

2 Related work

Although the modelling of legal text has been a field of study for many years, in this section we discuss those that focus on the modelling of data protection related legislations, and present three different tools that have been developed to help companies to comply with the GDPR.

Barth et al. [7] present a theoretical model for the representation of privacy expectations that is based on a contextual integrity framework [27]. The approach is validated via the modelling of the Health Insurance Portability and Accountability Act (HIPAA)[1]. Broadly speaking, the modelling is based on two kinds of norms, positive (allowed) and negative (denied). Using their framework privacy provisions for the sharing of data with different actors can be represented. However, according to Otto et al. [29] actions and purposes are not well represented. For instance, it is possible to model if a company cannot share personal data with a third party, but it fails to include purposes such as statistical reasons whereby a company may be allowed to share data.

May et al. [25] also illustrate how their approach can be used to model the HIPAA. Conditions and obligations are represented as access control rules that allow/deny operations. Given that they use a formal modelling language called Promela [16], it is possible to leverage existing Promela tools, such as for query execution. However, their model can only represent specific access-control related obligations. Other obligations, which are not related to access-control such as providing information about the processing or ensuring appropriate security measures are difficult to model with their approach.

Apart from legislative texts, policies for privacy notice and data exchange have also modelled. The World Wide Web Consortium (W3C) has undertaken numerous standardisation initiatives which deal with the modelling of data related policies. The Privacy Preferences Project (P3P)[2] is one such initiative which deals with representing privacy preferences in a standard machine-readable format. Using P3P we can model different parts of a privacy notice such as what information is collected, how long is it stored and for what purposes it would be used [12]. Though use of P3P can improve transparency of data processing, it does not support representation of other data protection related obligations [15]. For instance, obligations such as for security, data portability and right to erasure are out of scope for the P3P. Open Digital Rights Language (ODRL) [34] is another W3C initiative which presents a standard language to represent permission and obligations for digital content. The ODRL has also been used for modelling data protection legislations, for example Korba et al. [22] have used it to model the older data protection directive of the EU [1]. They have, however, discussed a high level overview of the modelling process for the directive. As a result, it does not include specific

[1] https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/content-detail.html
[2] https://www.w3.org/P3P/

details to model components of the legislation such as soft-obligations (i.e., obligations that serve as recommendations rather than being mandatory) and exceptions to legal obligations.

In terms of the GDPR, the Information Commissioner's Office (ICO) in the UK have developed an online self-assessment tool [18]. It provides two separate checklists, one for controllers[3] and one for processors[4]. The applicable assessment questions are shown for a set of obligations. For every question the users have an option to see additional information. After the questions are answered, a report can be generated which summarises the compliance levels and suggests actions to ensure full compliance. The primary limitation of the tool is the fact that the questions do not assess the obligations in detail.

Microsoft has also developed a GDPR assessment tool [26]. Unlike the ICO tool, it is a spreadsheet based assessment i.e. users have to provide the input in the provided spreadsheet. The questions include references to the GDPR text for further reference. Questions are organised in a hierarchical way and categorised according to the associated concepts. After the input, a report can be generated.

Similar to the Microsoft's tool, Nymity has also developed a spreadsheet based assessment [28]. Obligations are referred to as *Privacy Management Activities*. Unlike Microsoft, the questions are not categorised but follow the order of the GDPR text, whereby each obligation is linked to the corresponding GDPR paragraph. The spreadsheet is designed to work with their commercial software, *Nymity Attestor*[5], through which a report can be generated.

Each of the aforementioned GDPR compliance assessment tools show a list of questions which do not have any contextual connections between them. For instance, even if consent is not the basis for processing, a user still needs to answer all questions for consent as the relations between the questions are missing. As a result, the user has to go through all the questions (162 questions for the Microsoft's tool), even questions which are not applicable, to finish the assessment. Also, surprisingly none of the tools currently consider related national or domain specific legislation.

3 Framework for a compliance assessment system

Due to the shift towards information and knowledge-driven economies, the use of software intensive information systems is increasing. When it comes to legislations such as the GDPR, companies need to ensure that the data processing and sharing carried out by such systems complies with relevant legal obligations. Ensuring compliance is important, otherwise non-compliance can lead to large penalties and reputation damage. As such, companies often rely on compliance assessment tools that can be used to help them to assess if their existing business processes and systems comply with relevant legal obligations.

From a requirements perspective, it is important that compliance tool vendors are able to demonstrate the exhaustiveness of their tool in terms of legal obligations, as wrong conclusions could potentially be drawn from incomplete assessments. Ensuring traceability i.e. providing references to the legislation text is considered to be important for such tools [9,11,29]. References, for instance, allow companies to consult the legislations in case of confusion or if they need to verify an assessment. Also, it is important that such tools are kept up-to-date and are capable of taking into account updated legal interpretation of the relevant regulations [9,11,20,29]. For instance, the GDPR mentions *appropriate measures* for security (Article 32.1) where the measure of appropriateness can change over time.

To address these requirements, we propose a framework for compliance assessment, as depicted in Figure 1, which can be used to support multiple legislations as well as to manage changes in interpretation over time, by decoupling the data component from the compliance system.

[3] https://ico.org.uk/for-organisations/resources-and-support/data-protection-self-assessment/controllers-checklist

[4] https://ico.org.uk/for-organisations/resources-and-support/data-protection-self-assessment/processors-checklist

[5] https://www.nymity.com/solutions/attestor/

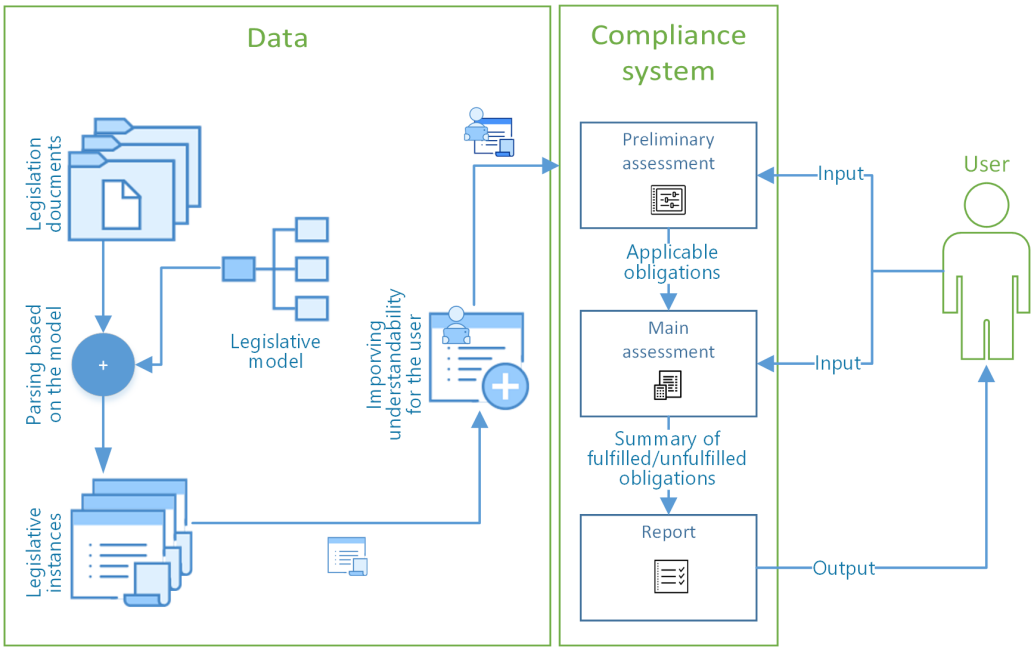

Fig. 1. Framework of the compliance tool

For the data component, a generic legislative model, ODRL, is used to represent legislative obligations and relations. For the *parsing process*, first the text defining obligations is extracted from the legislations. Next, the relations are identified between the extracted obligations and represented according to the legislative model. Following on from this, the modelled obligations are translated into a format that can be read by the compliance system, referred to as *Legislative instance*. Finally, the last step involves making the legislative instance more understandable for the user. Questions are prepared for the obligations such that the tool ask the user for the fulfilment of the obligations. Associated definitions are also added to ensure intelligibility of the questions.

The legislative instance is passed as input to the compliance system which assesses compliance based on the user-input and the legislative instance. As the modelled legislations could potentially govern multiple scenarios, it is possible that not all the defined obligations would be relevant for a compliance assessment. For instance, considering the GDPR, obligations related to processing outside the EU would not be applicable if a company does not transfer any personal information outside the EU. Therefore, to ensure that irrelevant obligations are not shown to the user, the assessment process is divided into two steps: (i) preliminary assessment; and (ii) main assessment. In the first step, the legislative instance is read and input from the user is taken. The input relates to the different scenarios which could affect the applicability of the obligations. For example, in case of the GDPR, whether the personal data is processed outside the EU. Based on the input, the system shortlists the applicable obligations and presents the assessment to the user. In the main assessment, the user provides input regarding the fulfilment of the obligations within their company. Once the required input is received, the system generates a report with a list of fulfilled and unfulfilled obligations.

Specific details on our implementation of the data component and the compliance system can be found in Sections 4 and 5 respectively.

4 Data modelling and the GDPR instance

In this section, we provide an overview of the proposed Open Digital Rights Language (ODRL) profile that can be used to model legislative obligations. Following on from this we provide a sequence of steps that are required in order to represent existing legislative text using the proposed model.

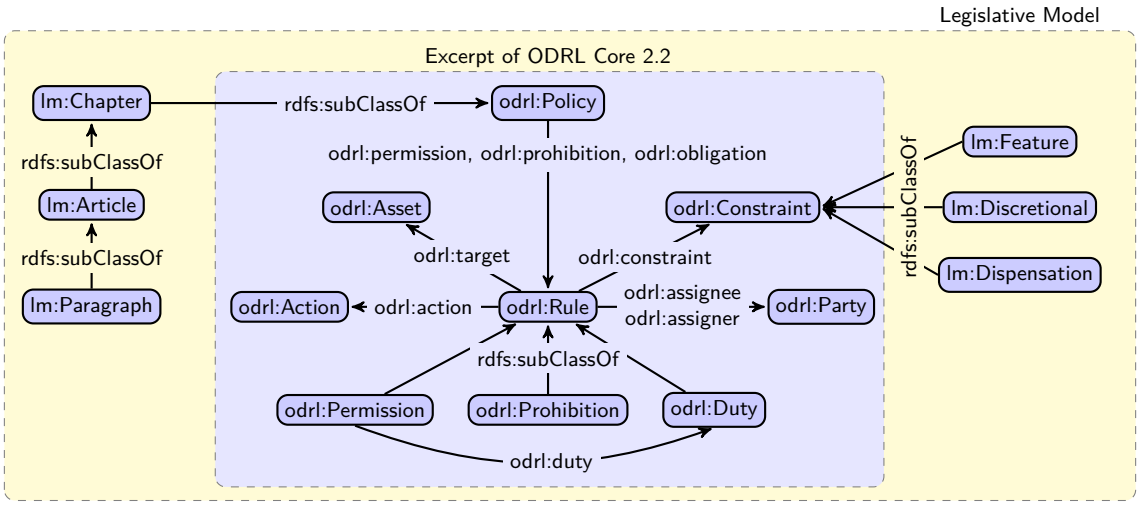

Fig. 2. The Legislative Model: based on an excerpt from ODRL Core 2.2 [34]

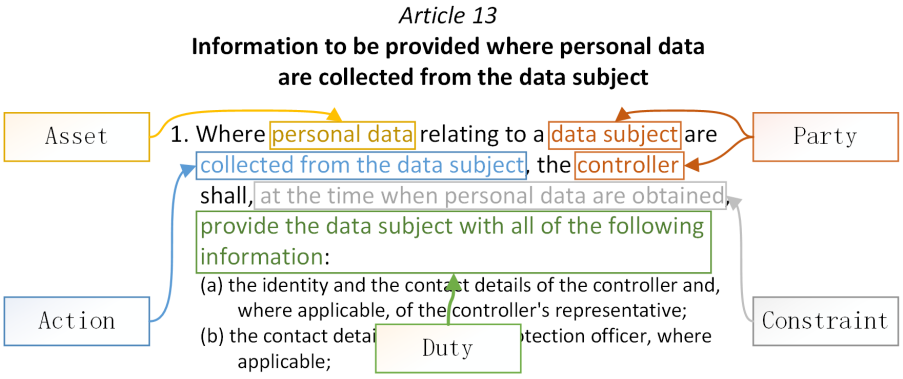

Fig. 3. Breaking down Article 13.1 of the GDPR according to the ODRL model

4.1 Legislative model

Like Korba et al. [22] we chose ODRL [34], which was released as a W3C Recommendation in February 2018, for modelling the regulation. ODRL provides a standard means to define policy expressions and licenses for digital content. The primary motivation for choosing ODRL is the fact that it can easily be extended for other use-cases such as representation of legislations by defining additional profiles[6].

The central entity of the ODRL model, as depicted in Figure 2, is a *Policy* which is used to specify *Rules* that are used to represent *Permissions*, *Prohibitions* and *Duties*. A *Permission* to perform an *Action* is granted if the associated *Duty* is fulfilled. While, an *Action* would not be allowed if any *Prohibition* is associated with it. Finally, a *Party* is an entity which participates in policy related transactions and an *Asset* is something which can be a subject to the policy under consideration.

Legal obligations are conceptually similar to ODRL duties. Consider Article 13 para 1 as depicted in Figure 3. In this example, personal data can be considered as an *Asset*, the controller and the data subjects are the involved *Parties*. While, the collection of personal data from the data subjects would be the *Action* for which the *Duty* is defined. Also, for this *Duty*, a *Constraint* is defined, which indicates that the *Duty* should be fulfilled at the time when personal data is obtained.

[6] https://www.w3.org/TR/poe-ucr/

Unfortunately, it is not possible to represent the following concepts using the core ODRL model and vocabulary:

Soft obligations. The term soft-obligation refers to obligations which are non-mandatory. These are similar to recommendations in the sense that they represent best-practices. For instance, consider Example 1 where such a recommendation related to the use of icons is described. Here the text includes "*may be used*", which indicates that the use of icons is optional. As a result, it should not be represented as a *Duty*.

> **Example 1:** Example of an optional constraint from the GDPR
> *Article 12.7*: The information to be provided to data subjects pursuant to Articles 13 and 14 *may be provided* in combination with standardised icons....

Exceptions. Legislations also consist of exceptions, which if present take precedence over the *Duty*. Example 2 illustrates one such exception scenario where obligations defined in certain paragraphs are not applicable if the data subject already has the information.

> **Example 2:** Example of an exception scenario from the GDPR
> *Article 13.4*: Paragraphs 1, 2 and 3 *shall not apply* where and insofar as the data subject already has the information.

Characteristics. There are additional constraints defined in the legislations which describe the features or characteristics of an obligation. Such features should also be fulfilled, along with the corresponding obligations. Example 3 shows constraints such as *conciseness* and *transparency* which should be ensured in order to comply with the *duty* defined in Article 13, depicted in Figure 3).

> **Example 3:** GDPR text defining characteristics
> *Article 12.1*: ...provide any information referred to in Articles 13 ...*in a concise, transparent, intelligible and easily accessible form...*

References to the legislation text. Additionally concepts are also required in order to represent relations with the corresponding legal text, such that it is possible to provide a link to the actual legislative text.

In order to represent these concepts, we define a legislative profile and extend the core ODRL model, as illustrated in Figure 2. We use *Discretional* for the soft-obligations, *Dispensation* for representing exceptions and *Feature* for the characteristics. Also, in order to support referenceability, we define sub-components *Chapter*, *Article* and *Paragraph* under the *Policy* component.

4.2 Instantiation process

Considering the proposed ODRL legislative model, we now discuss the instantiation process that can be used to represent existing legislations in a standard format. The created instance is used as input for the compliance system. The process, as shown in Figure 4 is divided into 5 main steps - (a) filtration of text that relates to obligations; (b) identification of interconnections in the text; (c) normalisation of the text; (d) representation of text in a machine-readable format; and (e) enhancing the readability for the user. In the following, we elaborate on these steps.

(a) Filtration of text that relates to obligations Along with obligations, legislations usually discuss other topics such as the scope of the legislation, relevant definitions and fines for not adhering to the legislation. For a compliance assessment, we focus on the obligations for the stakeholder under consideration, like controllers and processors in the case of the GDPR. Thus, as the first step, the text which is not related to the obligations can be filtered out. For instance, in the GDPR, articles such as Articles 68-76 which define the working of the *European Data Protection Board* can be excluded as these do not introduce any obligations for the controllers or processors.

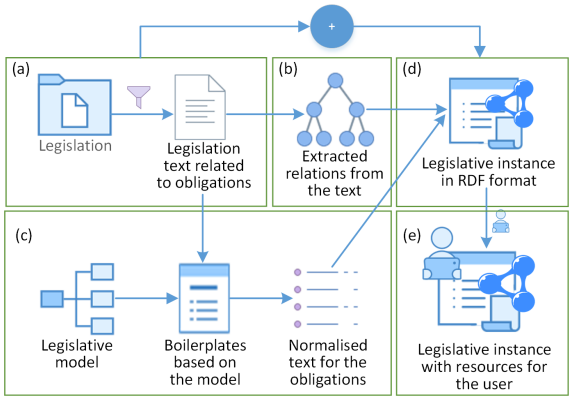

Fig. 4. Steps involved for the instantiation process

(b) Identification of interconnections in the text To represent the filtered legal text as per the legislative model, we have to identify text related to the different components such as *Duty*, *Feature* and *Dispensation*. However, legislations consist of several references within the text to other paragraphs and articles [31]. Example 4 shows text stating connections with Article 13, 14, 15-22 and 34 defined in Article 12 para 1 of the GDPR.

Example 4: Example of the interconnections defined in GDPR
Article 12.1: The controller shall take appropriate measures to provide any information referred to in *Articles 13 and 14* and any communication under *Articles 15 to 22 and 34* relating to processing to the data subject in a concise, transparent....

Thus, connected components are defined in different paragraphs and articles. In order to include all such references for the legislative instance, we extract and document all of the defined relations.

(c) Normalisation of the text Next, we need to represent the legislation text according to the legislative model. To achieve this, it is necessary to manually identify and code parts of the text as components of the legislative model such as *Duty* and *Feature*. However, legislations often represent obligations in different legal styles, which increases the complexity of the coding process. Examples 5 and 6 illustrates two of the many different styles used in the GDPR.

Example 5: Example of the following style: <*processing*> *is lawful if...*<*condition*>
Article 8.1:...*processing* shall be lawful only if and to the extent that *consent is given or authorised by the holder of parental responsibility over the child...*

Example 6: Example of the following style: <processing> is prohibited unless...<condition>
Article 9.1: Processing of personal data revealing racial..origin...shall be prohibited.
Article 9.2: Paragraph 1 shall not apply if...: (a) the data subject..explicit consent...

In the case of Example 5, if <*processing*> would be the *Action* then <*condition*> i.e. authorising consent by the holder of parental responsibility would represent the *Duty*. Similarly, considering Example 6, if <*processing*> would be the *Action* then corresponding *Duty* would be to not perform the *action* as described in Article 9.1. Based on Article 9.2, <*condition*> i.e *explicit consent* would then be the dispensation scenario for the duty. However, this example can also be interpreted in a way similar to Example 5 where for the *Action* of <*processing*>, <*condition*> can also be considered as a *Duty*. Thus, different possibilities may exist for the representation of the text according to the components of the legislative model.

To overcome the confusion which arises due to different writing styles, in the field of requirements engineering, the use of boilerplates has been recommended which help in representing the text in

Table 1. Boilerplates used for expressing obligations in a standard style

Type	Boilerplate
Main	`Party` to perform `Action` on a given `Asset` should fulfil `Duty` in order to ensure compliance
Feature	`Duty` has additional requirement of `Feature` which must also be ensured
Dispensation	If `Dispensation` scenario for a `Duty` is true then that `Duty` is not applicable
Discretional	If `Discretional` for a `Duty` or `Feature` is true then that `Duty` or `Feature` is not compulsory

a standard form [6, 17, 24]. A boilerplate is defined as a natural language pattern that restricts the syntax of the sentences to pre-defined linguistic structures [6]. Example 7 illustrates a boilerplate to represent the previous examples in a standard format.

Example 7: Illustration of a boilerplate to represent Example 5 and 6 in a standard form
Boilerplate: <`Party`> to perform <`Action`> on a given <`Asset`> should fulfil <`Duty`>
- *Controller* to perform *Processing* on *Minors' data* should *Obtain consent by their parents*
- *Controller* to perform *Processing* on *Sensitive data* should *Obtain explicit consent for it*

This way, based on a boilerplate, we first represent the text in a standardised format. As we are interested in identification of components like *Action*, *Duty* and *Feature*, the boilerplates are based on the components of the legislative model and are listed in Table 1.

(d) Representation of text in a machine-readable format After the use of boilerplates, the obligations need to be expressed in a format which can be easily read by the compliance system and is standardised such that the data model can be reused for other systems as well. We chose, the Resource Description Framework (RDF) format [7] for the representation, which is also currently used for the exchange of legislation data in Europe[8]. To represent the obligations as RDF, Protege (an open-source ontology editor)[9] was used as it provides a simple GUI for accomplishing the task. Listing 1 shows a snippet of the text related to Article 13.1 of the GDPR in the RDF format. Using RDF, each triple, which is composed of a *subject-predicate-object* expression, asserts a binary relationship between two pieces of information. These triples are placed in common *namespaces*, referenced via *prefixes*. The prefix `odrl` represents the components from the ODRL model `<http://www.w3.org/ns/odrl/2/>`. The prefix `rdf` is used for the RDF built-in vocabulary, `lm` to denote the legislative vocabulary `<http://privacylab.at/vocabs/lm/>`, and `gdpr` for the GDPR instantiation `<http://privacylab.at/vocabs/gdpr/>`.

Listing 1: Snippet of the GDPR instance based on the duty from Article 13.1

```
gdpr:P13_1 rdf:type lm:Paragraph .
gdpr:P13_1 odrl:duty gdpr:ProvideInfo .
gdpr:ProvideInfo rdf:type odrl:Duty .
gdpr:ProvideInfo odrl:action gdpr:DirectCollection .
gdpr:ProvideInfo lm:dispensation gdpr:DataSubjecthasInfo .
gdpr:ProvideInfo lm:feature gdpr:Transparency .
gdpr:ProvideInfo lm:feature gdpr:Conciseness .
gdpr:ProvideInfo lm:discretional gdpr:Icons .
```

[7] https://www.w3.org/TR/rdf11-concepts/
[8] http://www.eli.fr/en/
[9] https://protege.stanford.edu/

In Example 4 we had illustrated an interconnection between Article 13 and 12. In Listing 1, along with representing the duty from Article 13.1, we also include connections to other articles and paragraphs. For instance, line 6 and 7 of the listing represent connections to *transparency* and *conciseness* from Article 12.1 as illustrated in Example 4. Similarly, line 5 of the listing represents the connection to the dispensation defined in Article 13.4 (see Example 2). Also, line 8 represents the discretional task of using privacy icons, illustrated in Example 1 from Article 12.7. Thus, the duty based on Article 13.1 is related to other parts of the text such as to Article 12.1, 12.7 and 13.4. These relations were established with the help of identified interconnections in step (b).

(e) Enhancing readability for the users In the RDF model, additional information such as legal definitions can be added by defining new data fields for the components. For instance, in the GDPR, Article 4 is dedicated for such definitions which can be added to a GDPR instance. Along with the resources such as definitions, in order to take input from the user, questions need to be added to the instance. This way, the compliance system can present the data model in form of a questionnaire. Example 8 illustrates some templates used for creating such questions. Using, the template, the *Duty* for providing the required information to the data subject (Article 13.1) would correspond to a question: "*Does your organisation ensure that the required information is provided to the data subject?*".

Example 8: Example for the structure of the questions
Action: Does your organisation (perform) <`Action`>?
Duty: Does your organisation (ensure) <`Duty`>?
Feature: Does your organisation (ensure) <`Feature`>?

Listing 2 illustrates how questions can be added to the instance. While, Listings 3 and 4 illustrate *Action* and *Feature* questions respectively.

Listing 2: Snippet of the GDPR instance from Listing 1 with the added question

```
1 gdpr:ProvideInfo rdf:type odrl:Duty .
2 gdpr:ProvideInfo odrl:action gdpr:DirectCollection .
3 gdpr:ProvideInfo lm:dispensation gdpr:DataSubjecthasInfo .
4 gdpr:ProvideInfo lm:feature gdpr:Transparency .
5 gdpr:ProvideInfo lm:feature gdpr:Conciseness .
6 gdpr:ProvideInfo lm:discretional gdpr:Icons .
7 gdpr:ProvideInfo lm:hasquestion "Does your organisation ensure that the
8                  required information is provided to the data subject?" .
```

Listing 3: Illustration of an *Action* with added question

```
1 gdpr:DirectCollection rdf:type odrl:Action .
2 gdpr:DirectCollection lm:hasquestion "Does your organisation collect
3               personal information directly from the data subjects?" .
```

Listing 4: Illustration of a *Feature* related to the duty from Listing 2

```
1 gdpr:Transparency rdf:type lm:Feature .
2 gdpr:Transparency lm:hasquestion "Does your organisation ensure
3                          transparency for the provided information?" .
```

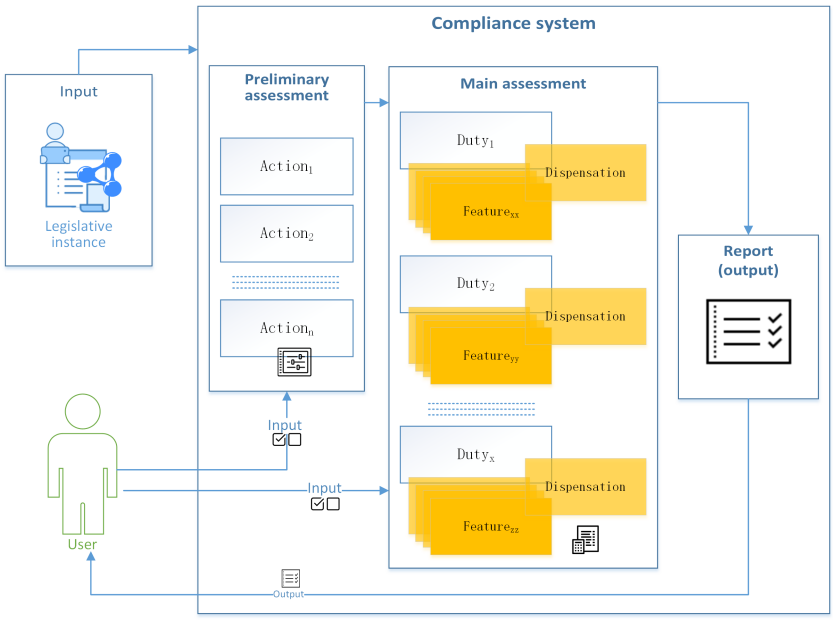

Fig. 5. Detailed process for the assessment of compliance

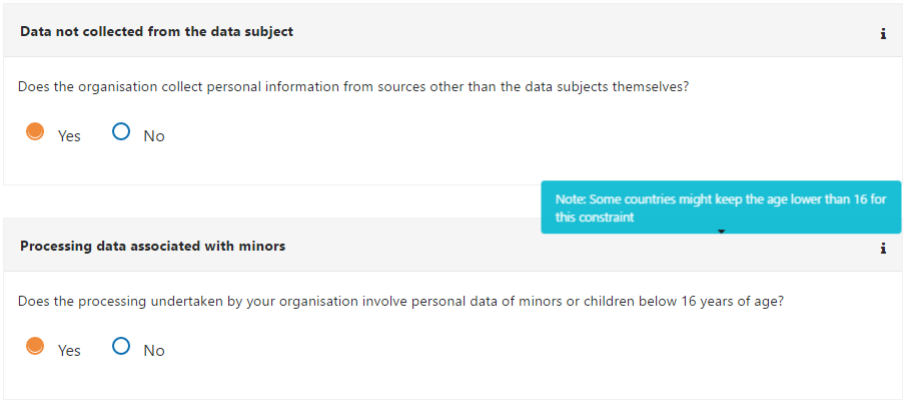

Fig. 6. A screenshot showing some questions from the preliminary analysis. The blue bubble shows additional information related to the question

5 The compliance system

After the definitions and questions are added to the legislative instance, it can be passed as input for the compliance system as shown in Figure 5. We now elaborate on the compliance system and discuss how it can be used for GDPR compliance assessment. For the assessment, we split the process into three parts: (i) preliminary assessment; (ii) main assessment; and (iii) report.

5.1 Preliminary assessment

The aim for the preliminary assessment is to find out the applicable obligations such that user does not have to identify and mark the non-applicable obligations similar to the existing tools [18, 26]. Based on the legislative model, as depicted in Figure 2, in order to perform *Action*, the associated *Duty* must be fulfilled Hence, the component *Action* can be used for the preliminary analysis to filter the applicable obligations. For instance, consider the *Action* illustrated in Listing 3. The *Duty*

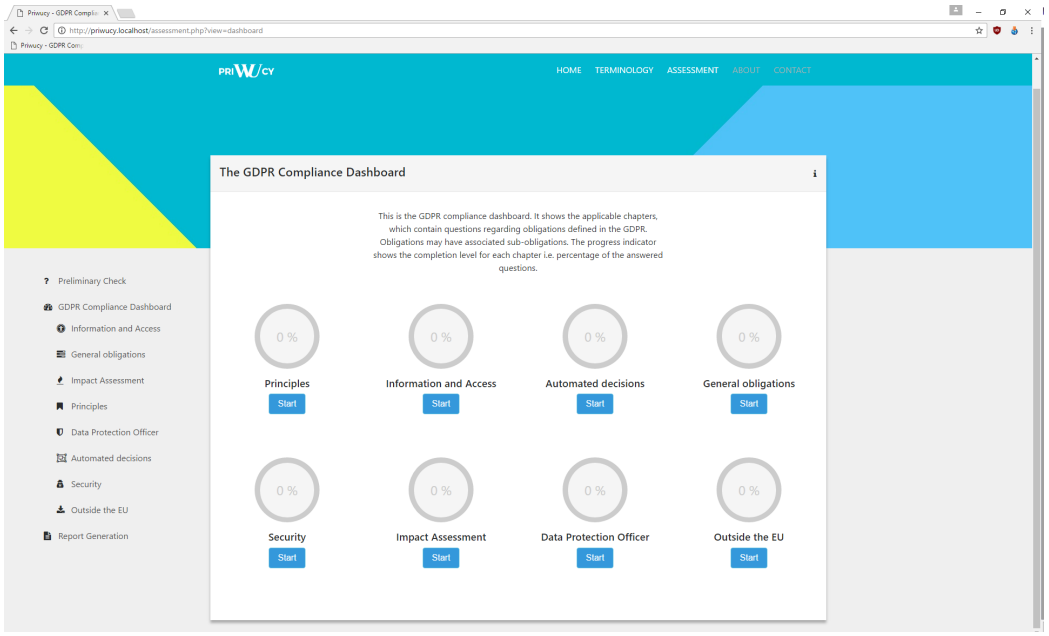

Fig. 7. Dashboard based on the GDPR chapters for the main assessment

shown in Listing 2, based on the connection with the considered *Action*, would only be applicable if that *Action* is performed. As shown in Figure 6, a list of questions are presented to the user which can be answered as *Yes* or *No*. For every question, there exists a title to give some context for the question. In addition, on the top right corner of every question, "i" button has been provided to display the additional resources such as definitions or external links for further reference. Once the user submits all the answers, the system then uses this information to select the applicable parts which are associated with the actions where the user responds with a *Yes*.

5.2 Main assessment

Based on the selected *Actions*, all the associated *duties* are extracted from the instance. These *duties* are the basis for the main assessment. Referring back to Figure 2, the *Duty* component is connected to the constraints: *Feature*, *Dispensation* and *Discretional*. Thus, along with the *Duty*, other connected components are also presented to the user. Considering Listing 2, the assessment would also show the question for the *Duty* as well as for the connected components such as transparency, shown in Listing 4. Even after eliminating the non-applicable parts, the number of duties can be overwhelming to show as a flat list. Thus, in an attempt not to overwhelm the user with 100+ questions on a page, we group the questions, by clustering the questions according to the *chapters* as shown in Figure 7.

The user can start the assessment with any of the displayed chapters. Based on the preliminary assessment, the number of *chapters* shown may vary as the dashboard is dynamically created based on the applicable obligations. After the user selects a chapter, a list of questions is shown which is based on *duties* belonging to the selected *chapter*. Like the questions for the preliminary analysis, all questions for the main analysis have a short title and one "i" button on the top right corner. Initially, only the questions based on the *Duty* are shown. If the user selects *No* then nothing happens. However, if *Yes* is selected, a cascaded list of questions is displayed. These questions are based on the connected *Dispensation* and *Features*. By putting questions in a cascaded format, the user only sees the relevant parts. For instance, for duty illustrated in Listing 2, in case the user selects *No* for the question related to the *Duty* then the questions for the associated *features* like transparency, depicted in Listing 4 are not relevant and are not shown to the user. Only when the user selects *Yes* for the *Duty*, the related questions are shown. The user has the option to go back to

the dashboard even when the all the questions have not been answered. The progress is saved and reflected as percentage complete on the dashboard.

5.3 Report

The last part for the compliance system is the report which provides a list of all the fulfilled and unfulfilled obligations. An obligation is considered to be fulfilled if a *Duty* is fulfilled along with all of the associated *Features*. *Duties* and *Features* represented as *Discretional* are also documented in the report. Along with the fulfilment status, references to the source (based on the *Articles* and *Paragraphs* which are defined in the legislative instance) are provided, such that users can refer to the legislation for additional information. Furthermore, fulfilled components (*Duty* and *Feature*) are shown in green boxes, *Discretional* components in orange and unfulfilled components are shown in red boxes.

6 Discussion

Our legislative model overcomes several of the challenges discussed in Section 2. It can represent both actions and purposes using the *Action* component of the model, which is one of the shortcoming for Bath et al's approach [7]. Also, as compared to May et al's approach [25] it can represent specifications for the obligations by using the *Feature* component. We have also considered soft-obligations and exceptions, which we refer to in our model as *Discretional* and *Dispensation* respectively.

To compare the capabilities of the compliance tools, we analyse 3 different capabilities: support for exceptions, management of evolving law and traceability. For the compliance tools, similar to legal modelling, **support for exceptions** is also important. For instance, in the GDPR, paragraphs like 17.3 define scenarios where obligation related to "right to be forgotten" is not applicable. Secondly, as law is considered to be dynamic where the interpretation involves based on amendments as well as on important judicial decisions [9, 11, 20, 29], the GDPR tools should support **management of evolving law** by ensuring provisions for updating the obligations accordingly. Lastly, **traceability** i.e. ensuring traceable references between the legal text and obligations is considered to be important [9, 11, 29]. References provide an overview of the articles and the paragraphs which a tool covers for the evaluation. With such traceable links, changes in the law can also be easily traced to the corresponding obligations defined for the tool.
Based on these criteria, in the following, we compare the GDPR compliance tools. The capabilities have been summarised in Table 2.

ICO The checklist for data protection self assessment provided by ICO [18] does not consider the exceptions. However, the questions can be answered as *not applicable* for cases where a user is aware of the exceptions. Also, as the checklist is web-based the updation of obligations can only be managed by the ICO. In terms of traceability, references to the GDPR text are missing which makes it difficult to analyse how much of the GDPR is covered by their tool.

Microsoft Microsoft's GDPR detailed assessment toolbox [26] also does not support exceptions but like ICO's tool provide an option to answer a question as *n/a*. As the tool is spreadsheet based, the users have an option to modify or update questions if any interpretation changes. The tool also provides references to the GDPR text. However, the references are not defined per obligation but rather for a group of obligations which makes it difficult to identify the reference of a single obligation.

Nymity Nymity's GDPR readiness spreadsheet [28] also does not support exceptions but the questions are framed in a way to exclude the exception scenarios. For instance, for obligation related to "right to be forgotten" the question includes "*where required by law*". The references are then provided to the corresponding article and paragraph and a user can then refer to the GDPR text to check if that obligation is applicable or not. Also, as this tool is also based a spreadsheet the user has the option to modify or update obligations if required.

Table 2. Comparison of the compliance tools

Tool	Support for exceptions	Manage evolving law	Traceability
ICO	**No** manual selection as N/A	**Limited** controlled by ICO	**No** references are absent
Microsoft	**No** manual selection as N/A	**Yes** editing the spreadsheet	**Limited** not defined individually
Nymity	**Limited** has conditional questions	**Yes** editing the spreadsheet	**Yes** references to paragraphs
PriWUcy	**Yes** represented as dispensation	**Limited** requires self-hosting	**Yes** references to paragraphs

PriWUcy In the data model as we defined a component *Dispensation* the exceptions are supported by the tool. For an obligation, if the dispensation is answered as *Yes* then that obligation would not be considered for the analysis. Like ICO's tool, PriWUcy is also web-based and users would not be able to change the obligations unless they self-host the tool. However, as the data component is decoupled from the user interface, updating the obligations based on the changes in the law would not be difficult. Also, by introducing *Chapter*, *Article* and *Paragraph* to the model, we were able to represent the references for all the obligations.

Currently, for the questions used for PriWUcy, we have used the terms as defined in the GDPR. For instance, consider the term transparency defined in Article 12.1 where the corresponding question in the tool is "Does your organisation ensure transparency with respect to the processing of the information provided?" The use of the term *transparency* in the question introduces certain limitations regarding ambiguities. The question does not have a precise interpretation and for the user it is difficult to measure if transparency is ensured. Questions with such ambiguities can be confusing to answer. As a result, removing ambiguities is described as an important prerequisite for defining requirements for a system in the field of Requirements Engineering [2,13,33]. However, on the other hand, according to the legal literature, ambiguity in the legal texts can be intentional and should not be removed or resolved from the legal texts [29]. Moreover, resolving ambiguities can possibly result in wrong specification of the obligations [19]. So, in case if we do not resolve ambiguities then users may have different interpretations and might answer incorrectly. Also, if we resolve ambiguities, for instance describing transparency is some measurable form then we face of risk of misrepresentation of the GDPR text. This can lead to including a wrong question for the assessment which would lead to a wrong report. Either way, we risk ending up with a wrong assessment of compliance. Therefore, it is crucial to find a right balance for ambiguity in order to ensure correctness of the assessment.

7 Conclusions

In this paper, we described a flexible and modular compliance assessment framework, where changes to the legislative instances are automatically reflected in the compliance assessment tool. In addition we proposed a general legislative model and vocabulary based on the Open Digital Rights Language. In order to assess the effectiveness of the proposed framework and model we discuss how it can be used to model the General Data Protection Regulation. Additionally, we compare our compliance assessment tool with those provided by the Information Commissioner's Office (ICO) in the UK, Software vendor Microsoft, and a company called Nymity who provide tools and consultancy to privacy officers worldwide. Learning from one of the main shortcoming of the P3P [30] i.e. high complexity, we know that companies would also not adopt a compliance tool unless the complexity is kept to the minimum. Thus as a next step, we would work on the ambiguity issue such that the questions can be simplified without affecting the correctness of the questions from a legal perspective.

Also, although in this paper we focus on modelling the GDPR, in future work we plan to demonstrate how our legislative model can be used to express related legislative obligations, such as those found in the e-Privacy regulation or the Payment Services Directive. Additionally, we plan to explore

automation techniques such as those investigated by Kiyavitskaya et al. [21], which are designed to automatically extract obligations from legal texts. Such techniques could potentially help in reducing the manual efforts required for the modelling process.

Acknowledgments. Partially supported by the European Unions Horizon 2020 research and innovation programme under grant 731601 and the Austrian Federal Ministry of Transport, Innovation and Technology (BMVIT) DALICC. For Figure 1, 4 and 6, icons have been taken from *icons8* (https://icons8.com/).

References

1. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. OJ **L 281**, 0031–0050 (1995-10-24), http://data.europa.eu/eli/dir/1995/46/oj
2. IEEE recommended practice for software requirements specifications: Approved 25 June 1998, IEEE Std, vol. 830-1998. IEEE, New York, NY (1998)
3. Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC. OJ **L 337**, 35–127 (2015-12-23), http://data.europa.eu/eli/dir/2015/2366/oj
4. 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 **L 119**, 1–88 (2016-05-04), http://data.europa.eu/eli/reg/2016/679/oj
5. Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications). COM(2017) **2017/03 (COD)** (2017-10-1)
6. Arora, C., Sabetzadeh, M., Briand, L.C., Zimmer, F.: Requirement boilerplates: Transition from manually-enforced to automatically-verifiable natural language patterns. In: 2014 IEEE 4th International Workshop on Requirements Patterns (RePa). pp. 1–8. IEEE (2014)
7. Barth, A., Datta, A., Mitchell, J.C., Nissenbaum, H.: Privacy and contextual integrity: Framework and applications. In: 2006 IEEE Symposium on Security and Privacy. p. 15. IEEE (2006)
8. Biasiotti, M., Francesconi, E., Palmirani, M., Sartor, G., Vitali, F.: Legal informatics and management of legislative documents. Global Center for ICT in Parliament Working Paper **2** (2008)
9. Boella, G., Humphreys, L., Muthuri, R., Rossi, P., van der Torre, L.: A critical analysis of legal requirements engineering from the perspective of legal practice. In: Requirements Engineering and Law (RELAW), 2014 IEEE 7th International Workshop on. pp. 14–21. IEEE (2014)
10. Breaux, T.D., Vail, M.W., Anton, A.I.: Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations. In: 14th IEEE International Requirements Engineering Conference (RE'06). pp. 49–58 (2006)
11. Breaux, T.D.: Legal requirements acquisition for the specification of legally compliant information systems. North Carolina State University (2009), http://www.lib.ncsu.edu/resolver/1840.16/3376
12. Cranor, L.F.: P3p: Making privacy policies more useful. IEEE Security & Privacy **99**(6), 50–55 (2003)
13. Génova, G., Fuentes, J.M., Llorens, J., Hurtado, O., Moreno, V.: A framework to measure and improve the quality of textual requirements. Requirements Engineering **18**(1), 25–41 (2013)
14. Ghanavati, S., Amyot, D., Peyton, L.: Towards a framework for tracking legal compliance in healthcare. In: International Conference on Advanced Information Systems Engineering. pp. 218–232. Springer (2007)
15. Grimm, R., Rossnagel, A.: P3P and the privacy legislation in Germany: can P3P help to protect privacy worldwide? In: Proc. ACM Multimedia, Nov (2000)
16. Holzmann, G.J.: Design and validation of protocols: A tutorial. Computer Networks and ISDN Systems **25**(9), 981–1017 (1993)
17. Hull, E., Jackson, K., Dick, J.: Requirements engineering. Practitioner series, Springer, London, 2nd ed. edn. (2005)

18. Information Commissioner's Office (ICO) UK: Getting ready for the GDPR (2017), https://ico.org.uk/for-organisations/resources-and-support/data-protection-self-assessment/getting-ready-for-the-gdpr/
19. Kamsties, E., Berry, D.M., Paech, B.: Detecting ambiguities in requirements documents using inspections. In: Proceedings of the first workshop on inspection in software engineering (WISE01). pp. 68–80. Citeseer (2001)
20. Kiyavitskaya, N., Krausová, A., Zannone, N.: Why eliciting and managing legal requirements is hard. In: Requirements Engineering and Law, 2008. RELAW'08. pp. 26–30. IEEE (2008)
21. Kiyavitskaya, N., Zeni, N., Breaux, T.D., Antón, A.I., Cordy, J.R., Mich, L., Mylopoulos, J.: Automating the extraction of rights and obligations for regulatory compliance. In: International Conference on Conceptual Modeling. pp. 154–168. Springer (2008)
22. Korba, L., Kenny, S.: Towards meeting the privacy challenge: Adapting drm. In: ACM Workshop on Digital Rights Management. pp. 118–136. Springer (2002)
23. Massacci, F., Prest, M., Zannone, N.: Using a security requirements engineering methodology in practice: The compliance with the Italian data protection legislation. Computer Standards & Interfaces **27**(5), 445–455 (2005)
24. Mavin, A., Wilkinson, P., Harwood, A., Novak, M.: Easy approach to requirements syntax (EARS). In: 17th IEEE International Requirements Engineering Conference. pp. 317–322. IEEE (2009)
25. May, M.J., Gunter, C.A., Lee, I.: Privacy APIs: Access control techniques to analyze and verify legal privacy policies. In: 19th IEEE Computer Security Foundations Workshop. p. 13. IEEE (2006)
26. Microsoft Trust Center: Detailed GDPR Assessment (2017), http://aka.ms/gdprdetailedassessment
27. Nissenbaum, H.: Privacy as Contextual Integrity Symposium - Technology, Values, and the Justice System. Washington Law Review **79** (2004)
28. Nymity: GDPR Compliance Toolkit, https://www.nymity.com/gdpr-toolkit.aspx
29. Otto, P.N., Antón, A.I.: Addressing legal requirements in requirements engineering. In: 15th IEEE International Requirements Engineering Conference (RE 2007). pp. 5–14. IEEE (2007)
30. Schwartz, A.: Looking back at P3P: Lessons for the future. Center for Democracy & Technology (2009), https://www.cdt.org/files/pdfs/P3P_Retro_Final_0.pdf
31. Sushant Agarwal, Sabrina Kirrane, Johannes Scharf: Modelling the General Data Protection Regulation. In: 20. Internationales Rechtsinformatik Symposion (IRIS) 2017, Feb 23, 2017 - Feb 25, 2017, Salzburg (2017)
32. Toval, A., Olmos, A., Piattini, M.: Legal requirements reuse: a critical success factor for requirements quality and personal data protection. In: Proceedings IEEE Joint International Conference on Requirements Engineering. pp. 95–103. IEEE (2002)
33. van Lamsweerde, A.: Requirements engineering: From system goals to UML models to software specifications, vol. 10. Chichester, UK: John Wiley & Sons and Wiley and Chichester : John Wiley [distributor], Hoboken, N.J. (2009)
34. W3C ODRL Community Group: ODRL Information Model 2.2 (2018), https://www.w3.org/TR/odrl-model/