Mapping legal requirements to SLAs: an ontology based approach for cloud-based service consumption

Dirk Thatmann¹, Erwin Schuster², and Gökhan Coskun¹

¹ Technische Universität Berlin, Service-centric Networking (SNET), Berlin, Germany
d.thatmann,goekhan.coskun@tu-berlin.de
http://www.snet.tu-berlin.de/
² T-Systems Austria GesmbH, Vienna, Austria
es.ontology@metsky.at
http://www.t-systems.at/

Abstract. This work presents a new approach to ensure compliance to legal regulation in Cloud Computing, especially in Software-as-a-Service. Since high demanding business sectors, such as the health care sector, require high legal certainty, when contracting services offered by external providers. We provide a lightweight ontological representation of the German Federal Data Protection Act (BDSG) and a methodological approach how this work can be extended with additional laws. Furthermore, we integrate the generic ontology into the Linked Unified Service Description Language (Linked-USDL) as Compliance to External Services (Linked-USDL CES) module. This extension enables service customers and providers to negotiate services more fine grained related to legal obligations, which increases legal certainty and thus the acceptance of a cloud-based service consumption. We demonstrate the applicability of the proposed ontology with the concrete use case “physician’s letter” that is part of a running national project TRESOR, which aims at the development of a trusted cloud ecosystem.

Keywords: Linked-USDL, Legal Regulation, Ontology, BDSG, Marketplace, Service Selection, Cloud Computing

1 Introduction

Through increasing awareness for the economic benefits it promises, Cloud Computing approaches gained momentum. Having neither geographical borders nor national limits, it is a global market where providers e.g. in India can have customers in Jamaica. On the one hand, this distribution and flexibility provide benefits for the customers. They are able to choose between offers from all over the world and select the most appropriate one. This process is mostly supported by marketplaces, which provide support for comparing different functional and non-functional aspects of products from different companies. This in turn, increases the competition in this market to the benefits of the customers. On the
other hand, it challenges businesses which are subject of legal restrictions and have to follow national regulators for the used software and the utilized data. The aforementioned marketplaces are currently focusing on functional aspects and have a few non-functional details like pricing information. Due to the complexity of the laws and the expected legal consequences in case of disregard, the legal aspects are omitted so far.

In this paper, we want to attract attention to this issue by presenting concrete use cases from the health care sector along with the legal regulatories from the German privacy law. We advocate that legal aspects of Cloud Computing offers should be semantically described, enabling machine-supported comparability on marketplaces. By this means, the market is opened for businesses dealing with sensitive data. Concretely, we present an extension for Linked-USDL that is a remodeled version of the Unified Service Description Language (USDL). It is described with semantic technologies and published following the Linked Data principles. We exemplify the usage of the proposed extension by applying it to the German privacy law.

The remainder of this paper is structured as follows. In the following section, we elaborate the necessity for describing legal aspects of Cloud Computing offers (focusing on Software-as-a-Service), enabling businesses like the health care sector to benefit from the economic advantages. In Section 3 we present briefly the related work in this field. In Section 4 we describe the core contribution of this work, namely Linked USDL-CES and demonstrate in Section 5 the realization of a concrete use case. After discussing the main critical points of this work in Section 6, we conclude the paper with as summary and outlook in Section 7.

2 The Need for Describing Legal Aspects of SaaS Offers

Today, almost every organization makes use of IT components and software products to some extend. Thus, optimizations of capital as well as operational expenditure for IT solutions is concerning everyone. This explains the payed attention attracted by and the success of Cloud Computing approaches. By bundling resources and allowing shared usage, the different business models like Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-service (SaaS) optimize the exploitation of hardware as well as software components. This is expected to lead to promising significant economic benefits for the customers, which is mostly a convincing argument for the management levels of various organizations.

Being at the highest abstraction layer, SaaS solutions address end-users and have therefore the biggest audience. Whoever uses software is a potential customer for SaaS providers and can leverage its economic advantages. Currently, daily used enterprise software like office products and customer relationship management software as well as very specialized graphic tools for design experts are available “out-of-the-cloud". This allows SaaS providers to offer their products globally and to acquire customers all over the world, where network connectivity is the only requirement.
For supporting the customers in selecting the best offer for their needs, different marketplaces for SaaS offers arise which provide functional as well as non-functional descriptions of existing offers and allow their comparison. However, for particular business sectors SaaS solutions are still not usable. As one concrete example, the health sector is subject of comprehensive legal restrictions and regulations. Although, a very simple SaaS solution for storing and editing doctors’ records about patient treatment would reduce the costs of health care institutions significantly, this sector cannot make use of the economic benefits current Cloud Computing based solutions provide. In the current situation each institution has to have its own IT infrastructure and administration staff. Driven by this motivation, the currently running project Trusted Ecosystem for Standardized and Open cloud-based Resources (TRESOR) aims at opening the SaaS market for the health sector.

![Fig. 1. Abstract illustration of the TRESOR project](image-url)

The main objective of TRESOR is the development of a trusted cloud ecosystem that consists of an open platform for offering and consuming cloud services. A broker and marketplace component mediates and combines services, whereas a proxy enables the access to those cloud services by taking enterprise guidelines, regulations by law and security policies into consideration.

A central part of the TRESOR broker is the service description language, that enables describing various aspects of SaaS offers. The Linked-USDL extension that is presented in this paper is a possible add-on of this description language. With this extension, the health care institutions are able to control to which extend the offers are compliant with legal requirements, select the most appropriate one and establish negotiations. Due to the societal aspects of this particular sector, the achieved significant cost reduction is expected to have a remarkable impact on the societal expenses.

3 Related Work

Our goal is to support a fine grained description of compliance to legal regulations in a service description in order to increase legal certainty on both service provider and customer side. Our main task is to realize a Service Description Language (SDL) with a lightweight ontology, able to express the BDSG and
similar acts. Especially, legal obligations for service customers and providers including their relation to operation of a services and any processing of data are of interest. As next, we list and rate the related work in accordance with our tasks. In our rough structuring we differentiate between expressive and lightweight ontologies. Representatives of expressive ontologies are:

There are several ontologies in the field of legal reasoning and argumentation. Edwina Rissland et al. [25] discussed the characteristics of the legal domain and its main points of interest for the application of AI techniques in 1985. In [13, p. 2] seven challenges are listed that AI and the legal domain face. Since then, many representational languages and legal rules have been invented. However, none reached the full expressiveness and complexity of existing legal texts in a consistent manner [17]. Gorden et al. present a formal, mathematical model of argument evaluation which applies proof standard [18]. Prakken’s model [24] is suitable for modeling particular legal procedures, learning about actual legal procedures and to learn about the process of formalizing an actual legal dispute. Brünninghaus’ [10] methods automatically generate legal argumentation and predictions from case texts. The Ontology of Professional Judicial Knowledge (OPJK) [12] focuses on semantic search in the context of question and answer (Q&A) systems. A request, formulated in a natural language, leads to a response with a high level of consilience. [12] but has no added value for our use-case. Gangemi et al. [15] introduce a design pattern for defining legal content ontologies. Whereas Despres et al. [14] focus on how to apply a linguistics-based tool “TERMINAE” to the legal domain and its alignment to core ontologies. However, all these approaches do not support us in our main objective to describe services and its properties related to legal regulations, such as required by the BDSG.

In the field of E-Contracting, the description of legal aspects and obligations mainly focus on general Terms and Conditions. Lamparter et al. [20] introduce a formal model in cooperating common contractual items and rights/obligations and applies the model in a scenario proofing the creditworthiness of customers. The defined ontology allows the use of SWRL-based [19] rules in order to enable automated proofing of the results. However, this work seems to have several similarities and overlaps with our use-case, but does not focus on how to incorporate legal obligations.

The LKIF-core [9] bases on OWL. It covers a standard vocabulary of basic legal terms having the focus on scenarios where the exchange of knowledge between different knowledge-based systems is required. LKIF’s shortcomings are twofold. Firstly, both legal modules “legal-action” and “legal-role” offer terms which support the definition of rules in the context procedural terms and roles (compare action- role process). Furthermore, it is not easily possible to link to a specific law, such as the BDSG. This could lead to the need of extensive expansions. Secondly, the core ontology is defined in English. A mapping layer and legal dictionaries are required, which means that an additional fuzzy layer is present.
Knöpfler’s ontology already concentrates on the BDSG. However, his work is motivated by the Computational Law. His ontology [27] maps rules taken from the BDSG to programmable logic. Due to pragmatic reasons, Knöpfler chooses SWI-PROLOG [29] to proof the fundamental technical feasibility of his idea. He selects just a single section out of a total of 63 which already leads to an impressive amount of rules and objects. A meaningful and appropriate visualization is extremely difficult (comp. [27] p. 298.) to achieve. However, only few segments seem to be reusable for our goal to develop an ontology addressing compliance to legal regulation during a service description and requirement matching.

In contrast to aforementioned expressive legal ontologies, we identified several lightweight ontologies in the legal domain. This includes ontologies used for representation of legal documents, such as MetaLex [9], a structured and nearly complete representation, and the Akoma Ntosa XML standard presented by Barabucci et al. [8]. The European Union published the multi-language thesauri EuroVoc [4], which contains a subsection for legal terms. EuroVoc bases on a SKOS extension [3] and new definitions are taken from the dublin-core [1]. However, these technologies and achievements can only supplement our solution. In the field of SDL, we identified the Linked-USDL [11], since it combines the Linked Data [22], [2] principles and the Web of Data by remodeling the existing USDL specification as RDF(S) vocabulary in order to enable a better support for machines when trading services on the Web of Data. Linked-USDL currently contains three modules: USDL-Core, USDL-Pricing and USDL-SLA [21]. The USDL comes with a legal module which is designed only to express copyright and license information. Thus, it is not usable to support legal compliance as defined in our scope.

Summing up, we have to propose our own lightweight ontology for describing services and their compliance to legal regulation for service discovery and selection scenarios. We choose Linked-USDL since its wide scope, its Linked-Data alignment and its focus on services descriptions seem to provide the most promising basement for adding appropriate ontologies able to describe acts and their obligations for service providers and consumers.

4 Ontological Description of Legal Compliance

The essential criteria for reliable decisions based upon service descriptions are twofold. The first criteria is a matter of knowledge and trust. The service to be described has to be known in-depth. By letting the service creator and service provider create the description, who are expected to possess the mentioned knowledge, this can be regarded merely as a question of trust. As such, it can be tackled e.g. by the introduction of a trusted 3rd party or some certification procedures. The second criteria is the quality of the service description language as well as its usability. On the one hand, the expressivity needs to allow the correct description of various facets. On the other hand, it needs to be easy to understand and to use.
Aiming at the description of legal aspects of SaaS services, we advocate making use of semantic technologies and following the Linked Data principles in reusing existing vocabularies and interlinking newly created ones with existing Web ontologies. To be more concrete, the Resource Description Framework (RDF), RDF Schema (RDFS) and the Web Ontology Language (OWL) emerged from the Semantic Web vision and defined as standards by the World Wide Consortium (W3C) represent a profound basis for expressiveness in defining a language. Along with manifold tools for editing and reasoning, the usage of these standards is very promising. Additionally, the ever increasing number of online available vocabularies and ontologies for various domains, including the service description domain as well as the legal domain, encourage their reuse by pragmatically following the Linked Data principles. Reusing existing and broadly used ontologies, which can be seen as de-facto-standards, is expected to simplify the understanding of the new language constructs.

4.1 From the Law to the Ontology

The endeavor to develop an ontology, which provides language constructs to describe legal aspects of SaaS offers, requires the analysis of valid and relevant laws for the particular context. Due to the circumstance of the TRESOR project, we focused on the Germany Federal Data Protection Act (Bundesdatenschutzgesetz - BDSG)[16]. Because of the complexity of the domain as well as the hierarchical relations between national, European and global regulations, we decided to start with the concrete and develop the ontology in a bottom up approach. For that purpose, we analyzed the text of the Germany privacy law and modeled the domain ontology.

We extracted nine concrete characteristics applicable to SaaS services, which we call Compliance Criteria. For each we defined a set of possible Criteria Values. Figure 2-illustrates this in a simple notation.

The extraction process we applied consist of the following steps:

1.0 Rough structuring of the BDSG [16] and scope reduction.
2.0 Define and structure the requirements.
   2.1 Compliance BDSG examples
   2.2 Compliance BDSG and USDL-CES
3.0 Defining an ontology - The USDL-CES

Through the first step (rough structuring), we identified the following six sections. (1) General and common provisions, (2) Data processing by public bodies, which includes subsections, such as Legal basis for data processing, Rights of the data subject and Federal Commissioner for Data Protection and Freedom of Information. (3) section Data processing by private bodies and commercial enterprises under public law contains of three subsections. The subsection Legal basis for data processing is followed by Rights of the data subject and Supervisory authority. In the end the structuring conclude with (4) Special-, (5) Final- and (6) Transitional-provisions. Based on this structure we can now concentrate on further questions, such as:
Q: Which sections contain relevant information for legal compliance? Which we answer by: all sections containing obligations.

Q: Which sections contain rights for 3rd parties? For the opposite case this can be seen as obligations for service-customers and service-providers which includes e.g. rights, such as “provision of information and granting permission to consult records to the persons concerned”.

In order to reduce the scope, answering the opposite questions help to exclude non-relevant legal text parts.

<table>
<thead>
<tr>
<th>Recipient (Empfänger)</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contract data processor</td>
<td>1</td>
<td>Bodies collecting, processing or using personal data on behalf of others, in accordance with §11 and §46 (3)</td>
</tr>
<tr>
<td>Identifiable natural person (“subject”)</td>
<td>2</td>
<td>A natural person who can be identified and to whom the data belongs, in accordance with §3 (1)</td>
</tr>
<tr>
<td>3rd Party</td>
<td>3</td>
<td>Each recipient of data or any person or body other than the controller, in accordance with §3 (8), §46 (3)</td>
</tr>
<tr>
<td>Responsible Processing Point (Controller)</td>
<td>4</td>
<td>Any person or body which collects, processes or uses personal data on his, her or its own behalf, or which commissions others to do the same. In accord. with §3 (7)</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of Compliance Criteria “Recipient (Empfänger)”

In our case, based on the recently introduced BDSG sections, our result can be summed up as: the core of the relevant legal sections are part of Section 2 and 3. Here we find concrete rules and guidelines for handling data within different scopes and purposes, distinguished by public and private institutions.
This information is important, since service providers do not know service-customers' obligations. Furthermore, Section 4 “Lawfulness of data collection, processing and use” is important, since (1) stated: “The collection, processing and use of personal data shall be lawful only if permitted or ordered by this Act or other law, or if the data subject has provided consent”. In summary, it can be ascertained that after rough structuring and scope reduction some Sections, such as Section 1, Section 2 including Subsection 3, Section 5 and 6 (Final- and Transitional-provisions) can be skipped.

Gathering criteria is time consuming (and can be improved or automated for sure), since it requires to understand the legal text in details and derive necessary conclusions. The complete result of our work is spread over approximated 20 printed pages. Therefore, we are going to list just some for Section 5 relevant use-case criteria in Table 1 and 2. However, additional information is available online [7]:

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family-related</td>
<td>1</td>
<td>Personal or family related in accordance with §1 (2) No.3</td>
</tr>
<tr>
<td>EU countries</td>
<td>2</td>
<td>EU/EWR country without germen establishment in accordance with §1 (5)</td>
</tr>
<tr>
<td>Public</td>
<td>3</td>
<td>Processing Points in accordance with §2 (1)- (3)</td>
</tr>
<tr>
<td>Public competition</td>
<td>4</td>
<td>Public undertaking, participation in the economic contest in accordance with §27 (1) No. 2 (commercial enterprises which are, though in public ownership, exposed to competition)</td>
</tr>
<tr>
<td>Non-public</td>
<td>5</td>
<td>Non-public Processing Point in accordance with §2 (4)</td>
</tr>
</tbody>
</table>

Table 2. Characteristic of Compliance Criteria “Public and Private Bodies (Stelle)”

4.2 Abstracting from the Law Ontology

Following the design of Linked-USDL, we introduce USDL-CES module in order to address common Compliance for External Services (CES). The goal of CES is to create a structure which can express on the one hand afore mentioned BDSG taxonomy and on the other laws, structured in a similar way. Figure 3-❄ depicts the ontology. The three levels between Figure 2-❃ and Figure 3-❄ are congruent and show how to instantiate the BDSG-Taxonomy. “BDSG” maps to “Statute or Act”, “Recipients” maps to “Compliance criteria” and “3rd Party” maps to “Criteria Value”. In case of replacing the taxonomy by an
Fig. 3. Combining the generic BDSG taxonomy with the Linked-USDL ontology

appropriate taxonomy expressing other acts, different namespaces should be introduced and applied.

4.3 LinkedUSDL-CES (Integration into LinkedUSDL)

In order to use this ontology for service description, we combine the USDL-CES with Linked-USDL, as depicted in Figure 3.

5 Realizing a Use-Case with Linked USDL-CES

We apply our approach to a sample use-case motivated by the TRESPOR project. In this example a hospital, which is a Public Enterprise and acts as a Public Body wants to use a Physician’s Letter-service provided by an external service provider (Recipient: 3rd party). The hospital’s requirements on legal regulation related to this use-case are listed in Table ???. Since the Business Objective is set to Internal Usage, the hospital may - under specific requirements - use (compare BDSG Section 14) the Special Personal Data for other internal purposes. Special Personal Data refers to especially Sensitive Personal Information (compare BDSG Section 3, Paragraph 9) to be processed, which includes medical data of patients. For instance, the service provider retrieves the knowledge by both statements, that the usage of Special Personal Data under the terms and definition of BDSG Section 14, Paragraph 5 is permitted even if the purpose is not listed. Since Physician’s Letter is stated as purpose the service provider can derive that it is prohibited to transfer the data to 3rd parties, such as laboratories. This example shows how restrictive the criteria can be handled and enforced with our approach. In addition, we achieve a high level of compliance to legal regulation which means a higher level of legal certainty for all parties, service customer (SC) and service provider (SP).
D. Thatmann, E. Schuster, G. Coskun

<table>
<thead>
<tr>
<th>Compliance Criteria (Name)</th>
<th>Criteria Value (Ausprägung)</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>German</td>
</tr>
<tr>
<td>Pub./Priv. Bodies</td>
<td>Stelle (Stelle)</td>
</tr>
<tr>
<td>Recipient</td>
<td>Empfänger (Empfänger)</td>
</tr>
<tr>
<td>Business Objectiv</td>
<td>Geschäftszweck (Geschäftszweck)</td>
</tr>
<tr>
<td>Protection Class</td>
<td>Schutzklasse (Schutzklasse)</td>
</tr>
<tr>
<td>Purpose and Scope</td>
<td>Zweckbestimmung (Zweckbestimmung)</td>
</tr>
<tr>
<td>Processing</td>
<td>Verarbeitung (Verarbeitung)</td>
</tr>
<tr>
<td>Purpose</td>
<td>Zweck (Zweck)</td>
</tr>
<tr>
<td>Enterprise Class</td>
<td>UnternehmenArt (UnternehmenArt)</td>
</tr>
<tr>
<td>Protection Class</td>
<td>Schutzart (Schutzart)</td>
</tr>
</tbody>
</table>

Table 3. Description of BDSG legal compliance criteria and values - Physician’s Letter

6 Discussion

Due to the challenging objective of this work, it has inherently some critical points, we want to discuss briefly in this section. The first one is the semantic complexity and expressiveness of the proposed ontology. However, the creation of an ontological representation of laws is a difficult task. Even if the ontology engineer focuses only on a small part. On the one hand, understanding the meaning of legal text in-depth without being an expert for the concrete legal text at hand is very challenging. In order to understand all aspects and to comprehend the interrelation with other laws, it is also necessary to know how judges interpreted the text and how they made their decisions in concrete examples. On the other hand, creating a comprehensive ontology, that represents such a complex knowledge exhaustively seems to be unfeasible. This is due to the fact, that reliable decisions in this context can only be made within margin of discretion by humans, respective judges.

Therefore, an ontology for such a domain is not expected to represent the knowledge enabling automatic decision making, but the basic terms in order to allow communicating the legal aspects between providers and customers. In this concrete work, we aim at providing a basic set of terms, representing an extraction from the German law, namely BDSG, and allowing SaaS providers to communicate legal compliance. We are convinced that such a description is essential for the success of SaaS for sensitive business sectors and want to attract attention to this issue and make a first proposal.

A second critical aspect is its focus on the German law. However, as one of the most comprehensive privacy data protection laws available, we think that it provides a good starting point and represents the basis for fruitful discussions for the next steps, towards an international standard.

The last critical aspect, we want to mention is the following. For a really legally valid description language, an internationally accepted standard has to be created with the authorities in this area. Until then, we think it is of high
value to work on basics towards this challenging goal and expect some light-weight ontologies become de-facto-standards. These in turn, can simplify the definition of a real standard.

7 Summary and Outlook

In order to support a better legal compliance when negotiating contracts between SaaS consumers and providers, we propose a generic methodology for deriving a taxonomy for specific laws/acts, such as the German Federal Data Protection Act (BDSG). Based on the taxonomy we described how to instantiate the taxonomy in our generic Linked-USDL CES module, which we propose as new extension for Linked-USDL. As proof of concept, we applied our approach in a sample use-case named Physician’s Letter in the context of the Cloud Ecosystem TRESOR [6]. Since we finished our research before Pedrinaci et al. presented a Linked-USDL vocabulary [23], we have to check whether an adaptation is required. A next step could be to create taxonomies for other German or European acts.

Acknowledgments. This work was performed in the context of the TRESOR project and was funded by the German Federal Ministry of Economic Affairs and Energy.

References

7. Linked-Data CES (March 2013), http://cloud-tresor.de/linked-usdl-ces/