Legislation-Aware Cloud Computing: An Overview

Beniamino Di Martino, Giuseppina Cretella, and Antonio Esposito

Abstract During the recent few years, Cloud Computing has gained a strategic importance in the development of the ICT sector. Cloud Services have been adopted by companies, public administrations, and private customers to store and process data. As the requirements in terms of information reliability, accessibility, and availability increase, so do the requirements related to confidentiality and security of sensitive and classified data (e-mail, medical records, etc.). These are therefore translated into legal requirements constraints. When data are used by a Cloud application, security and confidentiality issues may arise due to the geographical distribution of the Cloud Providers' data center. Indeed, it is a very common practice for a Cloud Provider to have its registered office in a country, while the several data centers it controls are located in different continents of the world. This business strategy, clearly aimed at maximizing profit by reducing expenses (especially in terms of taxes), is in contrast to the variegated national laws, which may provide specific requirements and security measures for the treatment of different kinds of data. This chapter presents an overview of the current initiatives concerning the legislation awareness in Cloud Computing and reports the main aspects to take into account to approach such a topic.

Keywords Cloud Computing • Law awareness • Law compliance

1 Introduction

Recently, Cloud Computing has gained relevant shares in the IT market and has consequently imposed itself in all the industrial sectors which rely on information technologies to effectively deliver their services. The massive use of Internet and the widespread distribution of mobile devices, which provide a cheap and continuous access to online resources to companies, professionals, public administrations,

B. Di Martino • G. Cretella • A. Esposito (✉)

Department of Industrial and Information Engineering, Second University of Naples, Aversa, Italy

e-mail: beniamino.dimartino@unina2.it; giuseppina.cretella@unina2.it; antonio.esposito@unina2.it

and consumers, have strongly contributed to the development and adoption of Cloud Computing. Also, the volume of data exchanged has dramatically increased, thus amplifying the issues connected to their management. Cloud Computing provides the means to build online service which guarantees access to scalable, always available and transparent services and resources, without the need for customers to own such resources or know their exact location. This is why in both legal-economic and sociological literature, the concept of Access Culture [1] is used to describe a situation in which it is not important to own a resource, but it is fundamental to always be able to access it. Since such resources are offered by third parties, it is necessary that providers guarantee the continuous availability and reliability of their services. In order to enforce the respect of Service Level Agreements (SLAs) and the safe and secure management of the data exchanged between users and service providers, it is necessary to emanate efficient and rigorous laws which regulate access to both services and data. In the last years, particular attention has been devoted to standardization efforts, which aim at defining shared formalisms for the description of Cloud Services [2, 3] and the measurement of their performances. The indexes and parameters used to evaluate a service's performances have also to take into consideration their safety, security, and privacy levels. However, maintaining the desired levels of protection of data and privacy required by current legislation in a Cloud Computing infrastructure is a serious challenge, as is meeting the restrictions on cross-border data transfer [4], because of the geographically distributed nature of the Cloud.

The remainder of this chapter is organized as follows: Sect. 2 reports the main issues in data sharing and privacy involved with Cloud Computing applications; Sect. 3 describes projects and initiatives carried out to define standards for law representation and tagging; Sect. 4 presents few hints on an ontology-based approach which exploits semantics to ensure law alignment among Cloud Services; Sect. 5 reports some conclusion regarding the current issues and some consideration on possible future works.

2 Legislation Awareness and Data Privacy

As we have already stated in the introduction, with the steady diffusion of mobile applications the volume of data processed has increased, and the urge to define means to control and manage them has grown accordingly. When applications rely on Cloud Computing Services, the need to provide a strong legislation to protect customers' data arises. As Cloud Services process users' data on machines that the users do not own or operate themselves, **privacy issues** often arise. Probably, privacy represents the main concern users express on the adoption of Cloud Computing, since many customers fear for the disclosure of private and sensitive data. Unless technological mechanisms to allay users' concerns are introduced, this may prove fatal to many different types of Cloud Services. One typical scenario in which Cloud users express most of their concerns regards the actual possibility for

providers to use their data for purposes of which they may not be aware. For instance, a Cloud Provider could use the data exchanged with the user for marketing purposes, without asking for her consent. Therefore, there is an increasing awareness for the need for design for privacy from both companies and governmental organizations. While local laws for data management already exist, expressing precise rules for the exchanging, treatment, and storage of sensitive data within the national boundaries, there are still problems regarding their actual application. As already stated before, the resources offered by Cloud can be provided by servers which may be scattered throughout the world: so, it could be difficult to determine the exact legislation to apply in each situation. Furthermore, if the data are migrated from a data center to another, and such centers reside in different countries, then the legislation could vary and data management should adapt accordingly. This is not an easy task, since it requires a complete knowledge of the legislations of the involved countries, and mechanisms to ensure compliance to such legislation each time data leave a center to move elsewhere.

The privacy challenge for software engineers is to design Cloud Services in such a way as to decrease privacy risk and to ensure legal compliance. Laws placing geographical and other restrictions on the collection, processing, and transfer of personally identifiable and sensitive information limit usage of Cloud Services as currently designed. For example, a UK business storing data about individual customers with the prominent Cloud Service Provider Salesforce.com could find itself in breach of UK data protection law. Customers may be able to sue enterprises if their privacy rights are violated, and in any case the enterprises may face damage to their reputation. There have been a number of high-profile privacy breaches in the news recently. It is also important to allay users' fears about the usage of Cloud Services. Concerns arise when it is not clear to individuals why their personal information is requested or how it will be used or passed on to other parties: this lack of control leads to suspicion and ultimately distrust. There are also security-related concerns about whether the personal data in the cloud will be adequately protected.

3 An Overview of Existing Initiatives and Frameworks

Several sectors are becoming interested in the adoption of Cloud Computing technologies and some of them need to face with regulation compliance issues to exploit the benefit of Cloud Computing because in spite of obvious benefits, the issue of regulatory compliances by Cloud Computing can seriously impede its utilization. In particular, the healthcare industry challenges are continuing to grow in the areas of regularity compliance for protecting patient privacy. The challenge looms large as the healthcare sector is under heavy pressure due to regulatory compliance mainly for protecting the privacy of PHI (protected health information). Godbole and Lamb in [5] discuss the adoption of cloud-based application architectures in the healthcare sector and its implication. Khan and Bai in [6]

propose an approach that enables Cloud Computing clients to verify health regulatory compliance claimed by Cloud Computing providers. In their approach, clients of Cloud Computing could check automatically how the Cloud Provider meets the regulatory compliance for their health records. The approach is based on three processes: (1) mechanisms to represent health regulations in machine-processable form; (2) collection of service-specific compliance-related real-time data from cloud servers; and (3) automatic reasoning about the compliances between the machine-processable regulations and the collected data from servers. However, this approach lacks an automatic technique to collect and represent regulations in a machine-processable form. This can be achieved by exploiting a plethora of standards for law representation. In fact at European level, many initiatives have been carried out to define standards for law representation and tagging:

- **Metalex** [7], which has also been used as an input for the CEN workshop on an Open XML interchange format for legal and legislative resources, which officially started on July 7, 2006. The objective of the Workshop is to develop a **CEN Workshop Agreement** (CWA) on an Open XML interchange format for legal and legislative resources. Several European initiatives are currently collaborating to improve the MetaLex standard in order to retain compliance among the different formats used throughout Europe. Among these, the most relevant ones are represented by LexDania, CHLexML, NormeInRete, and Formex. The **MetaLex** standard has been developed within the **EPOWER** project, with the objective to introduce ICT technologies to support citizens and governments in accessing and managing the growing volume of legal information produced by national, international, European, and local authorities. The standard is based on an XML-based formalism for the markup of legal documents, and it provides a generic and easily extensible base for the complete representation of legal documents and constraints. The current version of the standard, going under the name of CEN MetaLex, can be considered as an interchange format between other, more jurisdiction-specific XML standards. MetaLex is independent of the specific juridical systems and of the languages in which the laws are expressed, and it can be integrated with the XML schemas used by other frameworks, operating at national and international levels. Together with the European initiatives presented in Sect. 3, MetaLex also provides compliance with the **Akoma-Ntoso** standard, which has been realized within the *Strengthening Parliaments' Information Systems in Africa* project.
- The **SDU BWB** [8] standard is an XML-based format used to encode the laws currently stored in the Dutch **Basiswettwenbestand** (BWB) database. The original standard was published by **SDU** and is currently maintained by the Dutch government.
- **LexDania** [9] is a project initiated by the Danish Ministry of Science, Technology and Innovation with the purpose of defining a standard for the creation and interchange of legislative documentation. The project has developed a complex structure for the definition of XML schemas: core elements and types have been defined and organized in a multilayered architecture and are used as building

blocks for the definition, creation, and maintenance of new documents types and applications.

- **CHLexML** [10] is a Swiss standard, based on XML, for the representation of legal texts, with particular attention to multilingual issues, which represent a hot topic for the Swiss Confederation. The project, started in 1998 and directed by the *Coordination Office for the Electronic Publication of Legal Data Federal Office of Justice* (COPIUR), aimed at harmonizing federal, cantonal, and even private sector legal documentation, by employing information technologies in the legislative field to elaborate uniform norms and standards.

- The **E-LAW** [11] project, born in Austria, aims at a complete reform of legal text production, which include documents like government bills, committee reports, legal enactments of the *Nationalrat*, and decisions of the *Bundesrat*. As in other initiatives, XML represents the foundation stone for the creation of a uniform electronic production chain for legal documents. However, in the E-LAW project XML represents only a storage and interchange format, while more classical software and formats (like Word) are to be used during production of the legal texts.

- The **NormeInRete** [12] project can be considered as a precursor for the definition of standards for tagging documents with legislation information. Also, the project aimed at providing instruments to automate the tagging and guarantee interoperability among public administrations adhering to the project.

  The standard produced within the project consists in an XML schema which can be used to represent and describe approved laws, thus enabling users to mark up their documents and data with information relative to the applicable norms. The project has been developed in three steps, which reflected the necessity to uniform the semantics and nomenclature used in laws and to provide a machine-readable standard for their representation. In particular, the three steps can be resumed in:

  – Issues of circulars for the standardization of names used to address legal documents.
  – Production of a standard, based on XML, for the digital representation of laws and norms, in order to support the markup of documents and data.
  – Design and implementation of a federated website, in which public administrations can add and retrieve law defined through the approved standard.

In order to ease the adoption of the NormeInRete standards, a set of tools for editing and textual analysis able to adapt existing and new documents to these standards and to provide specific functionality for drawing (drafting) the enactment of legislation have been developed. XMLeges family tools have been developed by ITTIG [13] (Institute of Legal Information Theory and Techniques) in order to create a unique point of access to legal documents in a distributed environment and a system of permanent links between legal materials, able to guide users toward the participating authorities websites. This application suite for legal drafting includes:

– *xmLeges Linker* that can be used to implement automatic legislative document hyperlinking by identifying the regulatory references in a text and describing them using the standard URN-NIR.
– *xmLeges Marker* that is able to convert a legislative text into a XML-NIR file. Automatic detection and tagging of the documents formal structure provides a fast and safe conversion of the text in a specific XML standard, simplifying the burden of a manual conversion.
– *xmLeges Classifier* that is a tool for the automatic classification of legislative text paragraphs into provision types (e.g., Duty, Permission, Sanction, Abrogation, etc.). It is able to implement the Provision Model provided by the NormeInRete XML standard. It is implemented by a multiclass SVM machine learning technique.
– *xmLegesExtractor* whose purpose is to select relevant text fragments corresponding to specific semantic roles that are relevant for the different types of provisions. xmLegesExtractor is realized as a suite of Natural Language Processing tools for the automatic analysis of Italian texts, specialized to cope with the specific stylistic conventions of the legal parlance.

Finally, *xmLeges Editor* integrates all the previous modules into a visual tool. By using this tool, it is possible to produce documents according to NIR DTD (XMLSchema), URN, and Metadata legislative standards. The environment has the aspect of a common text editor; it doesn't require any technical knowledge about standards, since it hides all the xml details to the final user. This editor uses a set of grids to aid the user in drafting documents but is expressly designed for the Italian languages; thus, it does not support other than Italian language.

- **Akoma-Ntoso** [14] project aimed at providing concepts, formats, and tools for the development of IT support to parliamentary activities for countries in sub-Saharan Africa. The project Akoma-Ntoso started indeed from an internationalization and a complete reengineering of the XML of Normeinrete, which represent its backbone.

  The Akoma-Ntoso standard can be applied to the entire legislative chain, from law proposal to the final approval of the legislative decree, also including the reports of commissions, chambers, magistrates, and so on. Having a unique pattern to follow in the entire proposal-approval chain allows users to efficiently and effectively query a native XML database, in order to retrieve all information available on a specific legislation: which parliamentary acts have altered a certain regulation, when such changes have been enacted, and who has proposed an amendment to a law and when.

  Together with the development of a set of interconnected standards for the machine-readable representation of laws and actors, the project also aims at defining supporting languages and guidelines for their use, making it possible to accurately describe, produce, store, and manage judicial, parliamentary, and legislative documents.

  The standard is based on the following elements:

- A common model for the representation of documents, based on XML for the definition of their structure and syntax.
- A shared model for the exchange of documents, based on the similarities which characterize the different legislative processes, despite the influence of languages and cultural and historical backgrounds. In particular, such a model focuses on resolving issues like:

  1. The creation of documents using a single shared tool, regardless of its nature.
  2. Provide a graphical representation of the document using again a unique shared tool, always independent of nature, language, etc.
  3. Guarantee accessibility to the documents, intended as the possibility to navigate and explore them through references, in a fashion similar to text hyperlinks.

- A shared scheme for data representation, which has to be explicitly supported by the documents adhering to the standard.
- A shared scheme for the representation of references ontologies and metadata. Metadata provide auxiliary information which enriches the documents, such as the publication date, or the name the last modifier. However, in order to correctly interpret them, ontologies are used to add semantics to the documents.
- A common scheme for quotations and cross-references: the entire set of parliamentary, legislative, and judiciary documents can be seen as a network in which each document is connected to another by means of expressions in natural language. The adoption of a common convention for the names and references, among common reference mechanisms, as proposed by Akoma Ntoso, improves documents' accessibility and navigability.

The Italian Senate, from 2013, has already begun to mark the bills in Akoma-Ntoso, and tests are in course within the initiative OpenParlamento.

4   A Semantic-Based Approach

In the previous sections, we have briefly introduced some of the main attempts to provide a homogeneous description of legislative documents and some initiatives aimed to propose tools to automatically check the compliance of Cloud Services to regulations. Apart from Akoma-Ntoso, which was born to support African governments, the described formalisms have been developed in European countries. Other initiatives have been carried out worldwide: in the USA both the government [15] and private organizations (**Legal-RDF** [16]) have contributed to define formalisms for laws description; in Australia the government has financed the project **EnAct** [17], which involves Tasmania, Canada, some federal states of the USA, and New Zealand; in Japan several initiatives are being carried out by academia [18], legal experts [19, 20], and government [21].

Most of the cited approaches have a few elements in common:

- The use of XML for the definition of schemas to support some or all the activities connected to the production, storage, and transmission of legal documents.
- The exploitation of semantic-web technologies (RDF, OWL) to enrich the annotated documents and provide useful semantics.

Using XML and semantic-web technologies, it is indeed possible to develop a methodology for the description of Cloud Services and their composition [22]: so, we think it is feasible to leverage the existing frameworks and technologies to annotate Cloud Services with legal information and support users in choosing the ones satisfying their requirements.

Figure 1 reports a schematic representation of a possible framework architecture which associates law definitions to Cloud Services' representations. This schema includes three semantic databases in which are stored three different kinds of information:

- The *Cloud Services Database* will include information about the functionalities and the service levels of the Cloud Providers offer, in particular, this database's
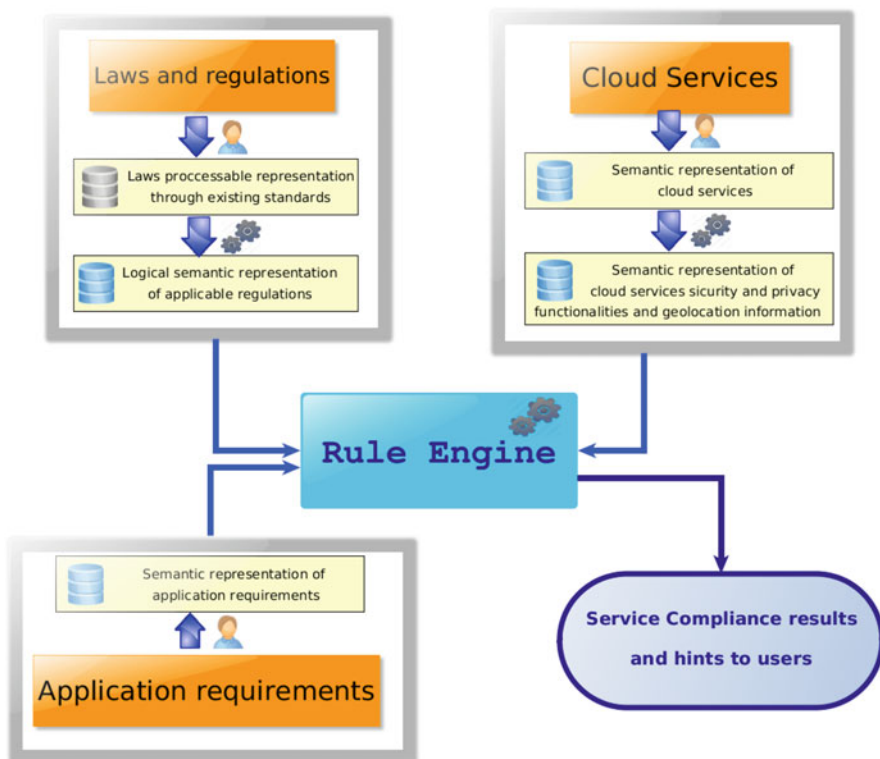

Fig. 1 Schema of a possible framework for Cloud Service law awareness

own information regarding the geographic areas in which the infrastructures running them are located and the functionality offered in terms of security and privacy by the services. This database will be obtained by processing existing semantic Cloud Services representations and deriving from them the information that is relevant to perform the compliance checking. Figure 2 provides a draft ontology that includes a set of concepts that might be useful to formalize the
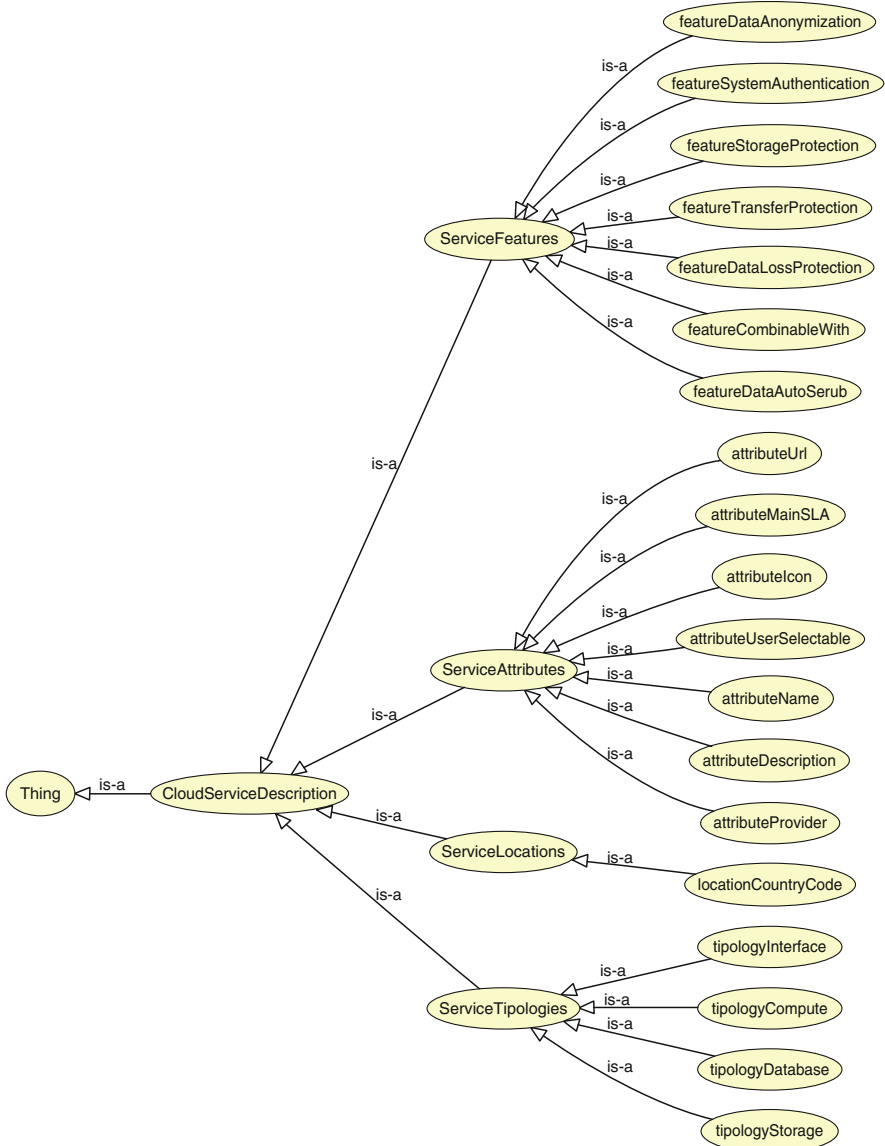

Fig. 2 Tentative Cloud ontology for law awareness

service description. These concepts include the service location, service features such as data anonymization, data loss protection, transfer protection, and so on.

- The *Laws and Regulations Database* will include the formalization in logical predicates of the law prescriptions. This information will be obtained by processing the representation of regulations performed by existing standards such as the ones reported in Sect. 3.
- The *Application Requirements Database* will include a processable representation of the application requirements in terms of kind of data the application handles and kind of treatment that will be performed on the data.

The law and regulations database contains all the annotated law and supplies inferences rules to the engine that running these rules on the assertions that represent the description of the Cloud Service is able to determine if the service is compliant with the law of interest. In particular, the engine takes into account the more appropriate regulations and laws by discerning based on the information provided by the application requirements.

5   Conclusion

Cloud Computing represents a revolution in Information Technology (IT) that offers various benefits to computing users. However, Cloud Computing usage is generally associated with security and compliance issues such as laws, regulations, data privacy, data protection, and contractual agreements. While most of the Cloud Providers can guarantee some measurable nonfunctional performance metrics, e.g., service availability or throughput, there is a lack of adequate mechanisms for guaranteeing that the provider is compliant with the actual legislation in terms of security, trust, and privacy. This lack represents an obstacle for moving most business relevant applications into the Cloud.

The approach proposed in [6] is valid, but it is limited to the healthcare field. The approach should be extended to all areas (such as juridical, statistical, etc.) and should allow the user to define the requirements of his/her application, for instance by specifying whether the application is related to legal data, medical data, or sensitive data. Moreover, given the plethora of existing frameworks developed for law representation and tagging, it would be worthwhile to investigate how and if it's possible to extract useful information from these representations to perform automatic checking of law compliance.

References

1. Valtysson, B. (2010). Access culture: Web 2.0 and cultural participation. *International Journal of Cultural Policy, 16*(2), 200–214.
2. Oasis topology and orchestration specification for cloud applications (tosca) tc. https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=tosca
3. Davis, D., & Pilz, G. (2012). Cloud infrastructure management interface (cimi) model and rest interface over http.
4. Pearson, S. (2009). Taking account of privacy when designing cloud computing services. In *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing* (pp. 44–52). IEEE Computer Society.
5. Godbole, N. S., & Lamb, J. (2013). *The triple challenge for the healthcare industry: Sustainability, privacy, and cloud-centric regulatory compliance*. 10th International Conference and Expo on Emerging Technologies for a Smarter World (CEWIT 2013), pp. 1–6, IEEE.
6. Khan, K. M., & Bai, Y. (2013). *Automatic verification of health regulatory compliance in cloud computing*. 2013 I.E. 15th International Conference on e-Health Networking, Applications & Services (Healthcom), pp. 719–721, IEEE.
7. Boer, A., Winkels, R., & Vitali, F. (2008). *Metalex xml and the legal knowledge interchange format*. Berlin: Springer.
8. World encyclopedia of law—sdu-bwb. http://lawin.org/sdu-bwb/
9. Tucker, H. (2004). Lex dania-white paper. *A system of XML schemas for Danish legislative documentation*.
10. Association e-justice.ch—chlexml. http://www.svri.ch/de/CHLexML.html
11. Schefbeck, P. A. E. (2006). *The e-law project in Austria*.
12. Francesconi, E. (2006). Norme in rete project: Standards and tools for italian legislation. *International Journal of Legal Information, 34*, 358.
13. Institute of legal information theory and techniques. www.ittig.cnr.it/
14. Vitali, F., & Zeni, F. (2007). Towards a country-independent data format: The akoma ntoso experience. In *Proceedings of the V legislative XML workshop* (pp. 67–86). Florence, Italy: European Press Academic Publishing.
15. Gullickson, K. (2008). *Drafting legislation using xml in the us house of representatives*. World e-Parliament Conference, pp. 25–26.
16. McClure, J. (2006). Legal-rdf vocabularies, requirements & design rationale. In *Proceedings of the V Legislative XML Workshop* (pp. 149–159).
17. Arnold-Moore, T., Clemes, J., & Tadd, M. (2000). Connected to the law: Tasmanian legislation using enact. *Journal of Information, Law and Technology, 1*, 00–1.
18. Ogawa, Y., Inagaki, S., & Toyama, K. (2008). Automatic consolidation of Japanese statutes based on formalization of amendment sentences. In Y. Ogawa, S. Inagaki, & K. Toyama (Eds.), *New frontiers in artificial intelligence* (pp. 363–376). Berlin: Springer.
19. Maeda, M. (2003). *Workbook hosei shitsumu* (Rev. ed.). Gyosei, Tokyo.
20. Ogawa, Y., Yamada, M., Kato, R., & Toyama, K. (2011). Design and compilation of syntactically tagged corpus of Japanese statutory sentences. In Y. Ogawa, M. Yamada, R. Kato, & K. Toyama (Eds.), *New frontiers in artificial intelligence* (pp. 141–152). Berlin: Springer.
21. Ministry of internal affairs and communications, horei data teikyo system. http://law.e-gov.go.jp/cgi-bin/idxsearch.cgi
22. Di Martino, B., Esposito, A., & Cretella, G. (2015). Semantic representation of cloud patterns and services with automated reasoning to support cloud application portability. *IEEE Transactions on Cloud Computing, PP*(99), 1–1.