HOLACONF - Cloud Forward: From Distributed to Complete Computing,
Towards a Legislation-aware Cloud Computing Framework

Beniamino Di Martino, Giuseppina Cretella, Antonio Esposito

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

Abstract

Most of the application frequently processes, stores, or transmits data that are subject to regulatory and compliance requirements. When data falls under regulatory or compliance restrictions, the choice of cloud deployment hinges on an understanding that the provider is fully compliant. Otherwise, there is the risk of violating privacy, regulatory or other legal requirements. Surely using the cloud may raise legal risks, but these need not be an impediment to the adoption of cloud. The important thing is to be fully aware of the risks and the regulations that acts on the data and the security features offered by the providers before deciding to put data into the cloud. This paper presents tentative design and implementation of a framework able to support the automatic checking of cloud application compliance taking into account the purpose of the application, the data nature and the kind of treatment.

© 2014 The Authors. Published by Elsevier B.V.

Key words: Cloud Computing; law-awareness; law compliance

1. Introduction

Law awareness in cloud computing is nowadays a hot topic because the adoption of cloud computing is in an advanced state and the boost to adopt cloud computing by companies, professionals and Public Administrations is particularly felt. Before entrusting data to a cloud provider the customer needs to have a very close read of their terms and conditions to understand deeply how the data will be treated on the cloud. But, even if this is sufficient for the customer scope, this cannot be sufficient for the actual legislation that applies in the data country owner. This problem particularly affected public administration which generally handle sensitive citizens data. As is well known, there is a big difference between the rules applicable to countries belonging to EU and outside the EU; the national and international legislation (Consumer Code, international conventions, private international law and procedural law) leaves no room for interpretation doubts, providing the applicability of the contract law of the country where the consumer has his habitual residence. Moreover, cloud computing necessarily involves the transfer of data and, in many cases, generates flows and cross-border movement of the data. There must also be consideration for what happens

* Giuseppina Cretella. Tel.: +39-0815010380.
E-mail address: giuseppina.cretella@unina2.it
when the provider-customer relationship ends. In most cases, this event will be addressed before an application is deployed to the cloud.

The problem is to understand how to comply with privacy laws and regulations when choosing cloud based services. The most common and significant legal issues that can arise in contracts with vendors for cloud computing services are:

- **Data privacy security and confidentiality**: the confidentiality, availability and integrity of data must be ensured by means of appropriate organizational and technical measures. These also include the protection of systems and data from the risks of unauthorized or arbitrary destruction, arbitrary loss, technical faults, forgery, theft and unlawful use, as well as from unauthorized modification, copying, access or other unauthorized processing.
- **Location of data**: Some vendors form contracts expressly reserve the right to store customer data in any country in which they do business. While dispersed geographical storage is beneficial from a data protection and backup perspective, it can raise law issues for some kind of data.
- **Suspension and termination of the service**: the vendor will return or destroy any copies of data once the cloud service is no more used by the customer, but this cannot be assumed.
- **Ownership of data**: the contract should expressly make clear that all data belongs to the customer and that the vendor acquires no rights or licenses to use the data for its own purposes.

This paper proposes a semantic based framework able to provide support for checking the compatibility of services with the laws according to the requirements of the user application. The verification of the requirements is based on three main information: the kind of data treatment (scientific, statistical, historical or generic treatment), the type of data (sensitive, health, judicial or not subject to protection data) and location of data (the geographic location of the provider data center for the specific service).

The remainder of this paper is organized as follows: section 2 reports the main issues in data sharing and privacy involved with Cloud Computing applications; section 3 describes the existing frameworks and current studies aimed at provide automatic processing of law; section 4 present a semantic-based approach which exploits semantics to ensure legislation checking of Cloud services according to user requirements; section 5 reports some conclusion regarding the current work and some consideration on future work.

2. Law-awareness Cloud Computing

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 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 on-line service which guarantee 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’’ 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 and the measurement of their performances. The indexes and parameters used to evaluate a service’s performances have also to take in 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, because of the geographically distributed nature of the Cloud.

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 an 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 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. Related work

The current literature is reach of works addressed to define models to check the compliance of data geo-location policies for cloud services, such as. However, none of the existing studies take into account the automatic processing of laws directive in conjunction with application requirement specification. At European level, many initiatives have been carried out to define standards for law representation and tagging:

- Metalex, 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 Dutch SDU BWB standard is an XML-based format used to encode the laws currently stored in the Dutch Basiswettenbestand (BWB) database. The original standard was published by SDU and is currently maintained by the Dutch government.
- LexDania is a project initiated by the Danish Ministry of Science, Technology and Innovation with the purpose of define a standard for 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 multi-layered architecture, and are used as building blocks for the definition, creation and maintenance of new documents types and applications.
- CHLexXML 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\textsuperscript{10} 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.

In Italy, the local Ministry of Justice, in conjunction with the “Centro nazionale per linformatica nella pubblica amministrazione ”(CNIPA), known today as “Agenzia per Italia Digitale ”, has promoted since 1999 the national project NormeInRete\textsuperscript{11}. Such a 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 automatize 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. Furthermore, a web-site has been realized to enable users to research for laws and norms described using the NormeInRete (NIR) standard, allowing them to easily search and retrieve information they need to tag the desired documents. The current web-site is reachable at\textsuperscript{12}. The NIR standard is currently used at national level and has been adopted by regional data-centers. The Suprema Corte di Cassazione has applied it to its historical database, producing more than sixty thousand legal documents in XML format, which can be accessed also through the ItalGiureWeb portal\textsuperscript{13}.

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 mark-up 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 other European initiatives, MetaLex also provides compliance with the Akoma-Ntoso standard, which has been realized within the Strengthening Parliaments’ Information Systems in Africa project.

As a consequence, MetaLex is very abstract, and it is considered as a basis to develop the new standards and can be easily integrated and adapted to existing schemas. MetaLex defines a mechanism for schema extension, adding meta-data, cross referencing, constructing compound documents and a basic naming convention.

Several well known web-oriented technologies have been used in the development of the standard: apart from XML, also semantic-based languages like Resource Description Framework (RDF) and Web Ontology Language (OWL) are used, in order to homogenize the terms and the vocabulary used in different legal contexts. Also, the standard can be used together with the Geography Markup Language (GML), in order to provide geographical and spatial information on the applicability of a specific law or normative. The use of such a standard can lead to the development and visualization of entire geographical maps of laws and legal constraints, which can then be used to determine which normative has to be applied in specific situations and geographical areas.

MetaLex XML is based on a strict hierarchical structure, which is divided into:

- **containers**, representing elements that contain other containers or blocks, in a defined sequence
- **blocks**, defining elements occurring within containers, which contain mixed text and in-line elements
- **in-line elements**, occurring within text.

Such elements are required to have unique IDs, which are then used to connect the different parts of the text composing the legislation to other parts, entire documents, semantic descriptions or simple tags, through URIs.
The conceptual model at the base of the XML standard produced by NormeInRete was among the inspirers of a project, carried out with the participation of the University of Bologna and patronized by UNDESA (United Nations - Department of Economic and Social Affairs), called Akoma Ntoso. This 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 re-engineering 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, 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.

4. A Framework to Verify the Law Compliance of Cloud Services

The framework proposed in this paper aims to provide a user oriented tool able to verify the compliance of services provided by different providers respect to the Italian legislation. The user interacts with the framework by using an interface that enables to specify the application requirements, in particular the user can set:

- the kind of data to treat among the defined categories (sensitive, health, judicial or not subject to protection data);
- the aim and kind of treatment among the defined categories (scientific, statistical, historical or generic treatment);
- the service provider the user want to acquire resources from and eventually the specific services he desires to acquire for the application;
- the location of the data center among the possible location defined for the selected provider.

This framework enables different use cases: as instance, a user, fixed the kind of data and the kind of treatment, may use the framework to establish the compliance of a specific provider service running on certain data servers or may use the framework to classify the providers based on their data center location. Moreover the user discovers, fixed the provider and the service, the kind of treatment allowed. The framework was developed to run with the Italian legislation and in particular is based on the formalization of the Italian Legislative Decree 196/2003 and Italian Code for Digital Administration.

The framework is based on a knowledge base that contains two main information sets:

- the semantic rules derived from the legislation analyzed;
- the semantic description of the Terms of Service of some leading providers’ cloud services (such as Amazon, Microsoft, IBM).

The first set of information is obtained from the laws of reference on privacy, by formalizing the aspects of the law that are useful and applicable for cloud services in order to generate a knowledge base useful to be processed by the framework. In particular, starting from the text, we have identified four types of information to be formalized: implicit (concepts assumed to be already known and thus not further explained), explicit (which defines a particular concept), complementary (concepts not explicitly defined in the analyzed decree but necessary to interpret the law) and prescriptive (the key concept that represent the law disposition) information. The prescriptive sentences has been translated in logical rules. An example of these rules is reported in figure 1. This rule states that if the data are statistical or historical or scientific, when the treatment ends the data must be destroyed. In this case the rule ensures
to verify that the service implements the auto-scrub functionality. If not a warning, that indicates the actions to be undertaken manually in case you wish to use the service analyzed, will be notified to the user.

\[
\begin{align*}
\text{DEFINE} & \quad \text{service } \rightarrow \text{featureDataAutoScrub} \\
\text{IF} & \quad \text{NOT} \quad \text{scope } \rightarrow \text{historical OR statistic OR scientific} \quad \text{AND NOT} \quad \text{service } \rightarrow \text{featureDataAutoScrub} \\
\text{THEN WARNING} & \quad \text{[reference to the article]} \quad \text{“It’s necessary to delete data manually when the treatment ends.”}
\end{align*}
\]

Fig. 1. A logical rule derived from the law for data treatment

The description of the terms of the services of the provider has been provided by using ontologies, in particular figure 2 reports the main classes and relationships used to describe the cloud services features.

Fig. 2. Class and relationships of the Cloud Service Ontology

By means of the input provided by the user that describes the application requirements, and the description of the cloud services functionalities, the rules that represent the law will be examined in order to verify the compliance of the services. A prototypical application has been implemented in order to test these rules. In particular figure 3 reports the architecture of the implemented prototype. There are two main components: the back end that is composed by the
Ontology Cache, the OWL Parser and the SWIPL Façade and the front end. The Owl Parser extracts information from ontologies coded in OWL and convert them into Prolog facts that are then questioned using the rules by the SWIPL Façade component.

The communication between the back end and the front end takes place through AJAX calls and JSON messages. The front end is represented by the user interface implemented in HTML. Suppose that a user need to use a cloud service to store medical data for a scientific treatment. Suppose the user want to use the Amazon Simple Storage Service within a data center located in EU. Figure 4 illustrates how the user can specify the application requirement by using the web page of the application that enable to specify the kind of data, the kind of treatment, the service
provider and the specific service, the location of the data center among the ones possible for the specific provider. The result of this specific request is illustrated in figure 5. Due to the specific kind of data and the particular kind of treatment the selected service need to be enriched with some feature that can be complemented by using one of the complementary services suggested. Furthermore the system lists the rules that are satisfied by the particular services and some warnings that represent tips to the user in order to advise him of some normative obligations, as instance to verify if the data owner have already signed a consent form.

5. Conclusion

The adoption of cloud computing may bring along common and significant legal issues that can arise in contracts with vendors. Most of these issues are basically business ones, requiring business decisions, but most of them are related to legal aspects, in the sense that they may conflict with the legislation in force. Maintaining the levels of protection of data and privacy, confidentiality and security required by current legislation in cloud computing infrastructure is a new challenge, as is meeting the restrictions on cross-border data transfer (the problem of data location) and holding the ownership of data. Unfortunately, among the many initiatives that have been carried out to define standards for law representation formalism definition there aren’t works that approach the topic from the point of view of automatize the checking of law compliance. In particular if we think at cloud computing issues and, as a significant example, at the specific scenario in which a public administration will store citizens data on the cloud, a very desired feature will be to verify the compliance of the vendors contracts with legislation of the country. This can be achieved by exploiting the existing standards already developed for law representation and annotation.
and by enriching them with semantic formalization of information related to the technical aspect of the normative disposition which can be useful to match the vendors contracts with the customers requirements. In particular this can be implemented by representing the regulations that emerges from the legislation as semantic assertions and inferences rules, the characteristics of the cloud providers offers and the requirements of the customers as semantic assertions and by using a shared semantic vocabulary. In this paper we propose prototypical implementation of a semantic based application that can be used as a starting point to implement a framework that will support the automatic checking of cloud application compliance with legislation.

Acknowledgments.

This research has been supported by the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n 256910 (mOSAIC Project), by PRIST 2009, “Fruizione assistita e context aware di siti archeologici complessi mediante dispositivi mobili” and CoSSMic (Collaborating Smart Solar-powered Micro-grids - FP7-SMARTCITIES-2013).

References