A Cloud Provider Description Schema for Meeting Legal Requirements in Cloud Federation Scenarios

George Kousiouris, George Vafiadis, Marcelo Corrales

A Cloud provider description schema for meeting legal requirements in cloud federation scenarios

George Kousiouris[1] , George Vafiadis[1] and Marcelo Corrales[2]

[1]National Technical University of Athens, Greece
gkousiou@mail.ntua.gr, gvaf@iccs.gr
[2] Leibniz University of Hanover, Germany
corrales@iri.uni-hannover.de

Abstract. The advent of Cloud computing has created numerous significant challenges with regard to manipulation of data and especially personal data in cases of Clouds and federated Clouds. Existing legislation currently creates constraints and boundaries in the free usage of external Cloud providers. The aim of this paper is to provide a schema definition and usage mechanism (CPDS) that includes various levels of legal information that is necessary for automating the process of Cloud provider selection and data outsourcing. Thus the aforementioned constraints may be checked in an automated and machine understandable fashion and fully harvest the potential that is created by advances in Cloud computing like dynamic federation. In this direction, legal gaps and necessary actions are identified so that the automation avoids manual and bureaucratic steps that are necessary at the moment.

Keywords: Cloud computing, legal issues, personal data, data management, Cloud federation

1 Introduction

With the advent of Cloud computing, new challenges have arisen with regard to the handling of personal data in infrastructures that are not under the direct control of either the end users or the application service providers, when using externalized infrastructure services from the IaaS/PaaS layer. This lack of control creates threats with regard to data accountability and legal action, especially with regard to current legislation[1].

On the other hand, the significant advantages of Cloud computing, like high availability, seemingly infinite resources, federation capabilities for exploiting multiple IaaS providers and pay-per-use business models are very intriguing for the business aspects of applications dealing with personal data. However, the “iron curtain” that lays in front of these providers with regard to how they manage their infrastructures (and is dictated by their business needs and confidentiality) hinders the available in-

formation that may be exposed and could alleviate the legal fears when it comes to personal data management.

For this reason, an intermediate solution should be found, that can expose critical legal information of the IaaS provider capabilities in order to meet legal constraints, while on the other hand not force the latter to share sensitive information regarding their infrastructure. This information should then be used in order to automate the negotiation process between the SP (Service Provider-the entity that is responsible for finding a suitable cloud resource and deploying the application service) and the IaaS provider (IP), or between IaaS providers in the case of dynamic cloud federation (case of resource sharing between IaaS providers). The aim of this paper is to provide such an XML-based definition of needed information and its usage in a dynamic multi-cloud utilization scenario, in order to bridge the gap between legal compliance and dynamic business models. The main triggering factor for this work is the OPTIMIS project's deployment and usage scenarios [12], that include federated and multi-cloud operations that may encounter legal boundaries and constraints [2].

The paper proceeds with introducing related work in Chapter 2. In Chapter 3, the main OPTIMIS deployment scenario is portrayed along with the legal constraints that prohibit its full potential. The structure of the necessary XML schema and mechanism is presented in Chapter 4 while its usage in an automated environment is portrayed in Chapter 5. The identified legal gaps that are necessary to complete the automation of the process in the mentioned scenarios are described in Chapter 6 while Chapter 7 concludes the paper.

2 Related Work

Legal implications and fears from the usage of Cloud computing have been highlighted in numerous research efforts. For example, the issue of losing control over data and the risks or concerns that this involves is highlighted in [3], as an aspect that is or will slow down the proliferation of Cloud usage. Jurisdictional issues, accountability and compliance are also critical in [4], and the usage of XML-based descriptions for legal text and rules has been taken under consideration, mainly in the SLA creation but not IP selection. In this interesting work, also the role of an intermediate third party trusted instance is considered, for the validation of auditing processes for example. LegalXML is an interesting attempt towards machine understandable contracts and terms, from which the following more interesting initiateves may be identified. LegalDocML[21] is a modeling language aiming at overcoming differences in XML-defined legal terms between different standards. LegalRuleML[22] is an ongoing effort to standardize legal rules and reasoning, while LegalXML eContracts is a recommendation towards standardized contract structures from a legal point of view. The latter may be used in conjunction to our work for describing for example the Standard Contractual Clauses. This family of recommendations may be valuable from an expression point of view, however it does not focus on provider selection.

On the other hand, IT scientists have devoted a lot of time and effort in order to enable from a technical point of view flexible and dynamic management [5] and sharing

of resources between different Cloud providers in order to meet demand peaks, minimize cost [19] or achieve higher availability (through multi-site placement) [19]. The federated aspects of Cloud computing model are considered capable to create immense potential as they offer significant performance gains as regards to response time and cost saving under dynamic workload scenarios [6].

3 OPTIMIS federation scenarios and legal constraints

In the OPTIMIS ecosystem [12], the Service Provider (SP) is an entity that needs to find a suitable Infrastructure Provider (IP) in order to deploy the application developer's service. In this operation, it must also meet the constraints set by the latter in the service level agreement that relate to the legal issues [13]. Furthermore, during operation, the IP may use external resources from other IPs in federated cloud scenarios, for various reasons such as cost minimization, risk management and avoidance of Service Level Agreement (SLA) failures due to overprovisioning. However, in that case the initial IP is assigned the role of assuring the legal aspects of data management by the federated resources. One of the most critical outcomes [10] is the determination of data location. The one who determines where the data are stored (in general this role is described as the Data Controller) is responsible in the end for meeting the legal requirements for data management manipulation. If the Data Controller decides to move the data to another provider, it must ensure that the proper agreements are in place prior to the federation. These agreements in the legal plane are summarized in the following cases:

- Binding Corporate Rules[7]: these refer to intra-company procedures (so in our case they apply to the case of a company located in Europe that wants to federate to its affiliate data center in e.g. Asia) for reassuring proper technical measures when it comes to data manipulation.

-Standard Contractual Clauses: these refer to inter-company agreements, with regard to how data are treated during the manipulation by the target federated provider. While these clauses have been formalized by the EU [14], their final form (that dictates the agreement between the home European IP and the target foreign IP) is subject to the agreement and/or possible modification from the involved IP parties.

-Intellectual Property rights: these refer to the ownership of the produced data from the utilization of a service offered by the IP.

The aforementioned documents (mainly the BCR and SCC) in order to be valid must be certified by the Data Protection Authority of the EU country of the originating (or home) IP. At the moment there is no automated way for performing the necessary legal checks so that the EU IP can on the fly check the legal compliance of the federated (external) IP. These checks may include the location of the federated IP, its security protocols and certifications, its main legal documents compatibility, such as the BCR, SCC and IPR declarations, with the EU law. If the target provider is located in the EU/EEA or the compatible countries [8], then the federation may be performed with minimal interventions. However if the IP is located in countries outside this list, then the federation cannot proceed if these legal checks are not in place. Thus it is

limiting the applicability of the federated scenarios to providers from the "white list" or includes the need for manual bureaucratic procedures prior to the federation.

Other requirements include the security from data loss, which is mainly reflected in data replication or multi-site placement techniques, and data encryption both at rest and during transfer. However these aspects are generally met in current technical solutions anyhow, due to well-established and pre-existing non-functional requirements of IT systems.

The complete scenario of this process appears in **Fig. 1**. In the OPTIMIS case, with different roles in different scenarios, we can predict that this role will be spread around multiple levels (SPs, IPs, federated IPs). However the main responsible is regarded the IP, since it receives the legal constraints and must act according to them. The home IP located in the UK may select from a variety of providers located in the EU/EEA and compatible countries. However they cannot exploit offerings from third countries, before the legal documents are certified by the respective DPA of the originating EU country. This certification process depends on each DPA and in many cases needs manual submission of the legal texts and validation.

4 Cloud Provider Description Schema Structure

In order to mitigate this effect and be able to have increased flexibility with regard to federation decisions, a suitable declaration procedure must be in place in order to implement this legal framework. In our case, this was decided to be implemented through a suitable XML schema structure, namely the Cloud Provider Description Schema (CPDS), that an IP should complete with its own information and make public, in order to be used during the selection process. Then the SP (or the IP when it is acting as an SP in the federated scenario) at every interaction should request this information, in order to filter the ones that do not meet the requirements. The information that needs to be included in such a description is detailed in the following paragraphs.

As a first step, the IPs must implement a way of declaring the locations of their data centers in terms of country of establishment. This location information may be the most critical, but it is only a part of the information that may be exposed by an IP regarding either legal information or in general capabilities information that is not confidential. These data must be suitably formatted in a machine understandable way so that they can be automatically processed and taken under consideration during a provider selection/ranking. The variety of information may include legal information (location of data, terms of service, legal notices etc.), security information (supported protocols, security services etc.), ecological information (percentage of renewable energy used, energy management certifications), resource and data management support (in terms of SLA guarantees, affinity rules support, available setting of replication etc.).

Not all of the above information is necessary in the legal context, but it was decided to be also included in order to explore also other aspects of IP selection (e.g. eco-efficiency, which indirectly may be linked to legal obligations deriving from the Kyo-

to Protocol [9]). The higher levels of this schema appear in **Fig. 2**. The top level of the structure is the IaaSProviderType. The main type consists of a number of elements of more elaborate types.. The main subtypes are the following

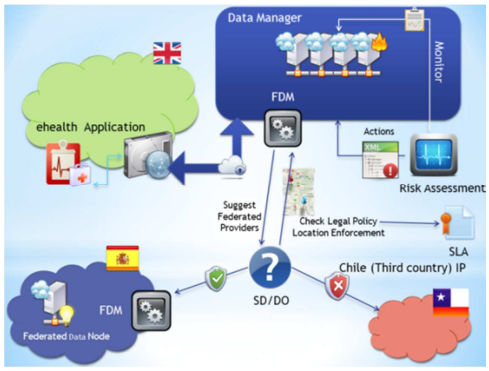

Fig. 1. OPTIMIS federation scenario: the home IP located in the UK may automatically select from a variety of providers located in the EU/EEA and compatible countries but not third countries even if the target IP is compliant but not adequately certified

- **LegalRequirementsType**: this element must be defined by an IP (**Fig. 3**), in order for the SP to be able to validate the legal adequacy of the former. In detail, the provider needs to insert the legal descriptions that govern its policies with regard to the SCCs, the BCR and the IPR with regard to added value from the services. For these textual descriptions it has been foreseen that they will be certified by an independent third party authority with expertise in the legal aspects of cloud computing (either public or private sector, similar to Verisign certification for website security). Another key characteristic refers to the ability of the provider to receive and fulfill requests regarding geographic location of service placement, which is another critical legal requirement, since the location of the data dictates the governing law. The security capabilities that are also critical from a legal point of view have been defined in the SecurityCapabilitiesType, given that they have also a strong technical aspect.
- **SecurityCapabilitiesType**: the provider must specify one element of this type, that contains information regarding its security features. These may include (**Fig. 3**) for example VPN support between the service VMs, Denial of Service attack detection capabilities, different authentication techniques (enumerated list with Boolean values that may be extended), encryption during data transfer or storage with a variety of strength options (bits used for the encryption) and/or a security certification by an external entity that has validated the security strategy of the company. This does not mean that the provider should publish the strategy itself, but just the ability, in order not to reveal sensitive information. Access control refers only to the IaaS level resources. In case of PaaS offerings, a similar field may be in the respective level.
- **DataManagementType**: the provider must specify one element of this type, that contains information regarding its data management features. These may include (**Fig. 5**) for example an element of DataReplicationType, stating the replication strategy of the provider for data redundancy. These features may include a maximum

replication factor, a configurable replication factor capability according to the client demands or a multi-site placement feature, which can lead in more reliable data storage. This is mainly dictated by the legal requirement for protection against data loss [18].

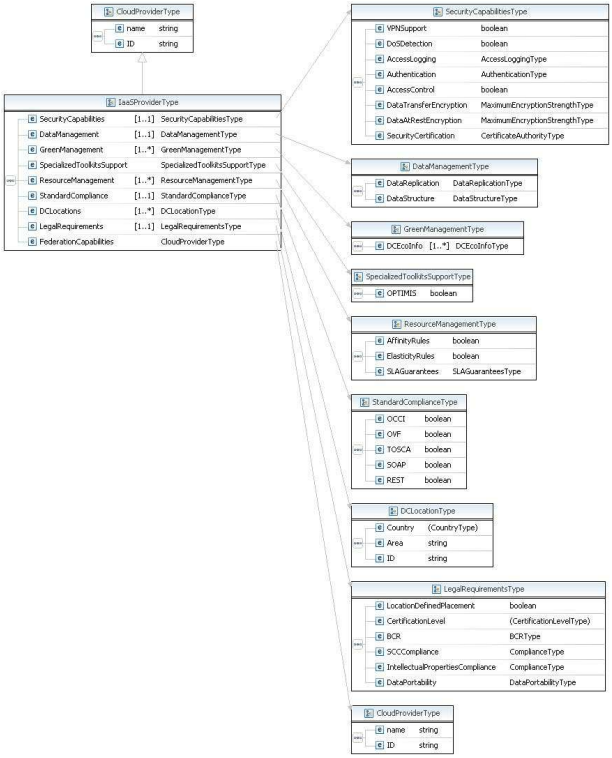

Fig. 2. Overall Structure of CPDS

- **GreenManagementType**: the provider must specify at least one element of this type, that contains information regarding its green management features per data center. The provider may declare different capabilities per data center (**Fig. 6**), given that not all of its infrastructures may be subject to green management, for example be equipped with energy generation from renewable sources. The basic subtype (**Fig. 7**) is the DCEcoInfoType, which contains information per DC, regarding the location (country, area and ID), the available green certificates (enumerated type consisting of the various available certificates) and the percentage of renewable energy produced (possibly on an annual average basis). The available certificates (as in all certificate types) may be signed by a CertificateAuthorityType, which is a generic type consisting of the authority's name and digital signature.
- **ResourceManagementType**: the provider must specify one element of this type, that contains information regarding its resource provision features per data cen-

ter. These may include the possibility to have affinity rules (a feature that leads to optimized placement of interacting VMs), elasticity rules (a feature that leads to better self management capabilities) and the type of SLA guarantees (e.g. minimum availability)

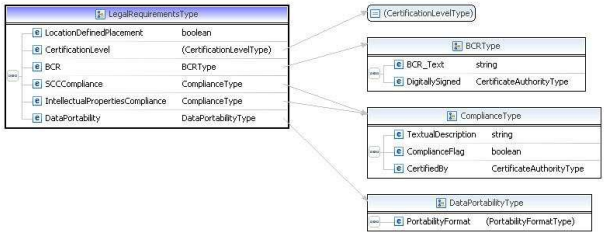

Fig. 3. Legal Requirements Type

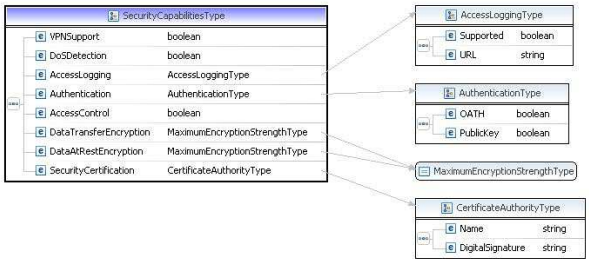

Fig. 4. Security Capabilities Type

• **DCLocations**: this element is used by the provider to declare at least one location of their data centers. This information is also used at the SP level, for the SP to filter out non-eligible locations (legal requirement). This type may include for example different locations for each DC, others in legal locations and others not. The final selection may take this under consideration, given that providers APIs usually offer a way of specifying desired target locations.

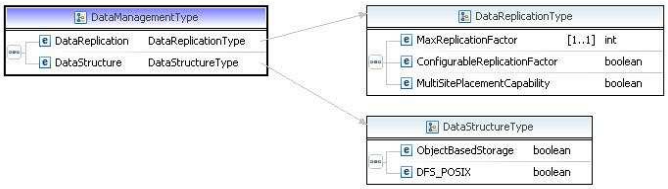

Fig. 5. Data Management Type

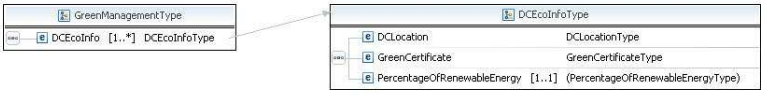

Fig. 6. Green Management Type

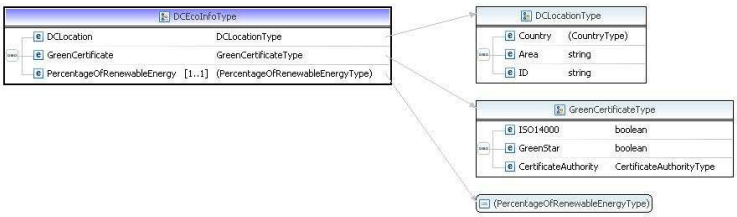

Fig. 7. Ecological Info Type

5 Usage during runtime and selection

The XML description of a provider following the aforementioned template is offered in the OPTIMIS platform through a restful GET interface of the Data Manager [11] component (getCPdescription). The description can be retrieved by any interested party (e.g. SP in initial deployment or the home IP in a federated scenario). Then the content of the XML may be directly checked through the checkLegal method in the OPTIMIS platform. This method compares the published information with the user requirements coming from the OPTIMIS SLA instance (e.g. location of provider, encryption capability and strength, eco-efficiency etc.) and concludes if the specific provider is eligible for the specific service.

In order to ensure that the level of information is valid, each field that requires an external verification from a certification authority needs to be digitally signed by the specific entity. Furthermore, the XML description itself should also be digitally signed, so that the requester of the document (in our case the SP or the home IP) may validate the content of the description. Furthermore, the caller should also validate that the digital signatures internally in the XML refer to the same entity as the provider that declares these capabilities and that the text itself is not altered and it is the same as the one submitted offline for validation in the DPA. This process appears in **Fig. 8** and is identical in concept to the validation process of websites through certification authorities such as Verisign etc.

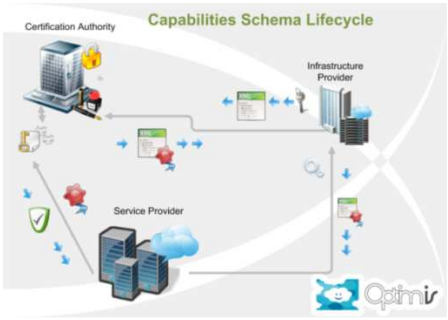

Fig. 8. Runtime Validation of An Infrastructure Provider's Description Schema

6 Identified Legal Gaps and Automation steps

6.1 Validation of arbitrary terms and conditions

Besides the improved selection of a provider based on this schema, the most important part is the legal section. In this, specific sections have been included like the BCR and SCC. These documents are arbitrary for each target company/provider and their conditions must be certified by the respective authority of each individual country (Data Protection Authority for the case of BCR and SCC) before a provider of this country federates to the target provider. This process is also manual in many countries, implying that before each federation decision for example, an employee of the home provider must go through a bureaucratic certification procedure of the target provider, bringing the text to the DPA and waiting for the resulting decision, that may be issued days afterwards. This of course creates an unrealistic legal framework for dynamic on the fly federation scenarios.

Automation step

Thus a necessary automation step is that an EU country's DPA should act also as a certification authority for external providers that wish to be included in such federation operations. The process of certifying that an external (to EU) provider's BCR, and SCC framework is compatible to EU law should be handled offline between the this provider and the DPA and the latter should then act as a digital certification authority (similar to Verisign) and digitally sign the textual descriptions. This key then may be included in the specific provider's CPDI (Cloud Provider Description Instance). A provider interested in using the specific target provider for federation may then acquire the description and key, validate the key's correctness and thus acknowledge automatically that this provider is certified with relation to its policies.

6.2 Variability of IPR strategies

The IPR strategies of each IP (or even at higher levels such as PaaS and SaaS) are determined by their strategy and potential exploitation/business schemes. However, for the SP and the end user this is a very significant issue, if it affects their own ex-

ploitation or legal framework of operation, when choosing the Cloud provider. An example of this is the Google Docs Terms of Service, which specifies that: "When you upload or otherwise submit content to our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. The rights you grant in this license are for the limited purpose of operating, promoting, and improving our Services, and to develop new ones."[16].. This may be against an exploitation scheme by the end user (and document producer) that includes confidentiality (such as a product design/documentation) and/or royalties for the material (e.g. book). Given that the terms and conditions for using services is in many cases arbitrary and based on company policy, their compliance validation in an automated framework means that they must be somehow grouped in classes and categories.

Automation Step

The categorization of IPR policies in predefined classes and the standardization of the latter create a suitable framework for automating the compatibility process between an end user and the provider. This would in essence be similar to the licensing schemes for software (GPL, BSD, Apache etc. licenses) and could also be accompanied with a rule-based decision framework that may examine cases of different classes compatibility (like in the case of [13] for licensing or even simpler approaches like an inter-license compatibility list) instead of strict "same class" comparison .

7 Conclusions

Dynamic business strategies (like Cloud federation, combination of private/public Clouds etc) that have emerged during the last years due to the technological breakthroughs in Cloud computing have created a number of issues with regard to the legal implications of data management for the application services.

In this paper, an effort to model these requirements in the form of a suitable description schema that the Cloud providers must complete and publish is portrayed, that may be used in an automated fashion during a legal check prior to the selection of a specific Cloud provider. The type of content regards information that normally is publicly available and/or can be found through relevant documentation. The content

Towards this direction, relevant legal and procedural gaps have been identified and corrective actions have been proposed (DPAs as certification authorities, IPR classes standardization need) in order to mitigate this problem and thus enable the usage of Cloud computing even in cases that involve personal data manipulation in dynamic and complex federation scenarios. DPAs acting as CAs may seem difficult, however this is a standard practice when it comes to internet certification and security.

For the future, we aim at enriching the XML schema with more details (e.g. by incorporating it in existing modeling approaches covering functional characteristics of providers like PIM4CLOUD [17]) but also extending the legal checks implementation with the digital signature validation process.

ACKNOWLEDGEMENT

This work has been supported by the OPTIMIS project and has been partly funded by the European Commission's IST activity of the 7th Framework Programme under contract number 257115.

8 References

[1] COM (2012) 11 final. Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), pp. 1-2 available at: http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf

[2] OPTIMIS Project Deliverable D.7.2.1.3 "Cloud Legal Guidelines: Technical Implementation of Legal Requirements, Exploitation of the Toolkit in Use Cases and Component Licenses", pp. 23-33.

[3] R. Chow, P. Golle, M. Jakobsson, E. Shi, J. Staddon, R. Masuoka, and J. Molina. 2009. Controlling data in the cloud: outsourcing computation without outsourcing control. In Proceedings of the 2009 ACM workshop on Cloud computing security (CCSW '09). ACM, New York, NY, USA, 85-90.

[4] Pearson, S., Charlesworth, A.: Accountability as a Way Forward for Privacy Protection in the Cloud. In: Jaatun, M.G., Zhao, G., Rong, C. (eds.) Cloud Computing. LNCS, vol. 5931, pp. 131–144. Springer, Heidelberg (2009)

[5] Rochwerger, B.; Breitgand, D.; Levy, E.; Galis, A.; Nagin, K.; Llorente, I. M.; Montero, R.; Wolfsthal, Y.; Elmroth, E.; Caceres, J.; Ben-Yehuda, M.; Emmerich, W.; Galan, F.; , "The Reservoir model and architecture for open federated cloud computing," *IBM Journal of Research and Development* , vol.53, no.4, pp.4:1-4:11, July 2009

[6] Buyya R, Ranjan R, Calheiros RN. InterCloud: Utility-oriented federation of cloud computing environments for scaling of application services. Proceedings of the 10th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP 2010), Busan, South Korea. Springer: Germany, 21–23 May 2010; 328–336.

[7] Article 29 Working Party, WP 74, Transfer of Personal Data to Third Countries: Applying Article 26 (2) of the EU Data Protection Directive to Binding Corporate Rules for International Data Transfers.

[8] European Commission, Commission Decisions on the Adequacy of the Protection of Personal Data in Third Countries, available at: http://ec.europa.eu/justice/data-protection/document/international-transfers/adequacy/index_en.htm

[9] http://www.kyotoprotocol.com/

[10] OPTIMIS Project Deliverable 7.2.1.1 Cloud Legal Guidelines, pp. 91-93.

[11] G. Kousiouris,G. Vafiadis and T. Varvarigou, "A Front-end Hadoop based Data Management Service for Efficient Federated Clouds", Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on , vol., no., pp.511-516, Nov. 29 2011-Dec. 1 2011

[12] A. Ferrer et al, OPTIMIS: A holistic approach to cloud service provisioning, Future Generation Computer Systems, Volume 28, Issue 1, January 2012, Pages 66-77, ISSN 0167-739X, 10.1016/j.future.2011.05.022.
[13] Barnitzke, B. et al, (2011). Legal Restraints and Security Requirements on Personal Data and Their Technical Implementation in Clouds. In Workshop for E-contracting for Clouds, eChallenges.
[14] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:039:0005:0018:EN:PDF
[15] Thomas F. Gordon. 2011. Analyzing open source license compatibility issues with Carneades. In Proceedings of the 13th International Conference on Artificial Intelligence and Law (ICAIL '11). ACM, New York, NY, USA, 51-55.
[16] http://www.google.com/intl/en/policies/terms/
[17] Brandtzg, E., Parastoo, M., Mosser, S.: Towards a Domain-Specic Language to Deploy Applications in the Clouds. In: Third International Conference on Cloud Computing, GRIDs, and Virtualization (CLOUD COMPUTING'12). pp. 1{6. Nice, France (Jul 2012)
[18] Djemame et al. “Legal issues in clouds: towards a risk inventory”, p. 5. December 10, 2012 doi:10.1098/rsta.2012.0075 *Phil. Trans. R. Soc. A 28 January 2013 vol. 371 no. 1983 20120075*
[19] Toosi, A.N.; Calheiros, R.N.; Thulasiram, R.K.; Buyya, R.; , "Resource Provisioning Policies to Increase IaaS Provider's Profit in a Federated Cloud Environment," *High Performance Computing and Communications (HPCC), 2011 IEEE 13th International Conference on* , vol., no., pp.279-287, 2-4 Sept. 2011
[20] Zhang, Zehua; Zhang, Xuejie; , "A load balancing mechanism based on ant colony and complex network theory in open cloud computing federation," *Industrial Mechatronics and Automation (ICIMA), 2010 2nd International Conference on* , vol.2, no., pp.240-243, 30-31 May 2010
[21] OASIS LegalDocumentML TC: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legaldocml
[22] OASIS LegalRuleML TC: https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legalruleml#technical
[23] OASIS LegalXML eContracts TC https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=legalxml-econtracts#technical