IEEE Access
The journal for rapid open access publishing

IEEE Xplore Digital Library
http://ieeexplore.ieee.org/Xplore

An Integrated Knowledge Graph to Automate Cloud Data Compliance

KARUNA PANDE JOSHI[1], LAVANYA ELLURI[2], ANKUR NAGAR[3] AND AMITH HENDRE[4]

***Abstract*— To address data protection concerns, authorities and standards bodies worldwide have released a plethora of regulations, guidelines, and software controls to be applied to Cloud data. As a result, service providers maintaining their end-user's private attributes have seen a surge in compliance requirements. Since most of these regulations are not available in a machine-processable format, it requires significant manual effort to adhere to them. Often many of the laws have overlapping rules, but as they are not referencing each other, providers must duplicate efforts to comply with each regulation. We have done a detailed study of all the data protection regulations that apply to Cloud data. We have developed an integrated, semantically rich knowledge graph that captures these various data compliance regulations. It includes the data threats and security controls that are needed to mitigate the risks. In this paper, we present this knowledge graph in detail, along with the system that we have developed to evaluate it. We have validated our knowledge graph against the privacy policies of various Cloud service providers like Amazon, Google, IBM, and Rackspace. This knowledge graph is available in the public domain and can be used by organizations to automate their compliance processes and set their enterprise Cloud security policies.**

***Index Terms*— Cloud Computing, Cloud Security, Security Domains, Security Compliance Models, Cloud Security Models.**

I. INTRODUCTION

Cloud Services are increasingly maintaining their consumer's confidential attributes, like personal, browsing patterns, and financial payment details, to facilitate seamless user experience. A significant portion of this consumer data is often shared by the Cloud service providers with their subsidiaries and third parties for further analysis to ensure customer retention and increase their purchase volume. Hence, even though Cloud-based services provide cost savings and rapid provisioning/scaling, privacy and security of Cloud data remain a concern for most consumers [42]. Because of this surge in sensitive information on the Cloud, regulatory organizations world over are formulating data protection legislation, like the European Union's General Data Protection Regulation (EU GDPR) [63] and Payment Card Industry Data Security Standard (PCI DSS) [64], etc. Cloud service providers must adhere to that. Simultaneously, various security standards for Cloud data have been proposed, or are being developed, by standard organizations like Cloud Security Alliance (CSA) [51], International Organization for Standards (ISO) [52], and National Institute for Standards and Technology (NIST) [6]. Cloud providers are incorporating these regulations and standards in their solutions to make their system robust and acceptable to consumers. This spurt in data protection regulations and security standards has resulted in overwhelming legal compliance challenges of Cloud services, and businesses often fixate on a single tree or branch in the forest of laws, regulations, standards, and seldom step back to gain an overall view of the compliance forest [42].

Data protection regulations are currently not machine-processable and are available only in a textual format requiring significant manual effort to parse their rules and constraints. Therefore, it is nearly impossible to determine in real-time if a compliance violation has occurred. Another issue is that data protection policies often contain legalese jargon that requires expert interpretation resulting in increased compliance costs. Real-time tracking of data flow on the Cloud would ensure that any operation performed on consumer data, from the acquisition of the data to its manipulation or sharing to its end-state archival in an organization, can be verified and documented for future audits.

We envision that an integrated, semantically rich, machine-processable knowledge graph (or ontology) that captures the various data compliance regulations, as they apply to Cloud data, will significantly help in automating an organization's data compliance processes. In addition to saving organizational resources dedicated to compliance adherence, it will also help in proactively identifying data breaches. Another advantage of building this integrated knowledge graph will be that potential contradictory policies in the organization can be identified and rectified as needed.

As a first step towards this vision of a holistic data compliance knowledge graph (or Ontology), we have created a semantically rich knowledge graph to capture the various compliance regulations, potential data threats with corresponding CSA controls [44]. We have also developed a comprehensive representation of the rules encapsulated in PCI DSS and GDPR [44]. We used Semantic Web technologies, Natural Language Processing (NLP), and text mining techniques to create this ontology, which is machine-processable. Hence, it can also contribute significantly to automating the continuous monitoring of data operation, transfer, and sharing. In this paper, we describe this knowledge graph in detail, along with the methodology we have used to build it. We have validated this Knowledge Graph against the data policies of five key vendors. This Knowledge Graph that is available in the public domain [86] [87] can be used to automate data protection compliance in an organization significantly.

We conducted a comprehensive study of the various compliance models and security controls that apply to Cloud-

Department of Information Systems, University of Maryland Baltimore County, 1000 Hilltop Cir, Baltimore, Maryland, USA.

"This work is supported in part by a DoD supplement to NSF IUCRC Center CARTA."

Digital Object Identifier: 10.1109/ACCESS.2020.3008964

based services. We also reviewed the potential threats faced by Cloud consumers and determined the compliance models and security controls that should be in place to manage these risks [61]. For our study, we analyzed more than 20 compliance models for Cloud computing as well as for IT management. We also reviewed more than 100 Cloud providers for their security standards by examining the security-related whitepapers posted on their websites.

In this paper, we first discuss the related work in section II. In section III, we present our analysis of the various Cloud security compliance models and classify them according to their security domains. The semantic web ontology for Cloud security compliances and security standards are described in section IV. In section V, we describe results & validation. We conclude in section VI and define the future work planned.

II. RELATED WORK

A. CLOUD DATA COMPLIANCE

Data protection standards contain a set of rules or policies formulated by regulatory agencies or standards organizations [58]. Security and privacy compliance models, like ISO 27001, COBIT, etc., have been proposed for Cloud computing security to ensure data protection and user privacy. We have analyzed and categorized the various Cloud compliance models according to security controls implemented. The features of each compliance model relevant to Cloud security are discussed in section III.

Cloud security [56] mainly focuses on the policies and controls used to protect the data present in the Cloud. Both Cloud providers and consumers can face security issues. Cloud providers should ensure that consumers understand data protection requirements while using their services. To enforce security, Cloud providers implement various security controls, which can be categorized as Deterrent, Preventive, Detective, and Corrective controls [57].

While the Cloud services and deployment models have been classified into different types, the security controls they use to protect their environment is the same for all - SaaS, PaaS, and IaaS – service types. Compliance models are applied based on security controls. We have to synchronize these models to ensure adequate Cloud security.

The IT compliance model focuses on electronic data processing, network, and IT infrastructure. The compliance model implements some rules and regulations across the various components of IT to make them work harmoniously. The security model is adopted based on these compliance models. One of our key contributions has been to associate the various compliance models and security controls. This transparency amongst the Cloud model, security control model and the compliance model will help the end-users achieve the data protection in a better way.

Before adopting a Cloud service, consumers should consider all potential threats that might compromise their data. CSA [2] lists threats like data breaches, data loss, account or service hijacking, insecure interfaces and APIs, denial of service, malicious insiders, abuse of Cloud services, insufficient due diligence & shared technology vulnerabilities. Cloud providers understand the importance of these persistent issues and have implemented various security standards. Vendors like Amazon [3], Rackspace [4], and Google [5] specify the security standards that they have incorporated on their platform. According to Spamina [37], there are more than 800 Cloud providers available all over the world. The question is, how many of them are using Cloud security standards and are capable of fighting potential threats [6][2].

In [49], security issues of different Cloud services are defined. It is also mentioned that Cloud providers should mention security issues in their SLA (service level agreements). This will give a clear idea to Cloud consumers about Cloud security issues. In 2013, CSA published CCM v3 (Cloud control matrix version 3)[14], which consists of more than 135 security controls and related compliance models. ISO 27001:2013 document [21][7]. consists of 114 security controls in 14 different groups. However, it does not have security controls like data encryption and media protection. NIST 800-53[25] presented its list of security controls with 18 groups. DoD (Department of Defense) has also published a list of eight information assurance areas and controls. There is a need for identifying common security controls that are easy to comprehend by consumers, and our prototype system attempts to do just that.

A. SEMANTIC WEB ONTOLOGY

The semantic web is a representation of the World Wide Web by providing standards to express relationships between web information and deals primarily with data instead of documents. It enables data to be annotated with machine-understandable meta-data, allowing the automation of their retrieval and their usage of incorrect contexts [1][45].

Semantic Web technologies include languages such as Resource Description Framework (RDF) and Web Ontology Language (OWL) for defining ontologies and describing meta-data using these ontologies as well as tools for reasoning over these descriptions [1][17][84][85]. These technologies can be used to provide standard semantics of privacy information and policies enabling all agents who understand basic Semantic Web technologies to communicate and use each other's data and Services effectively [1][17] [84][85].

B. TEXT EXTRACTION

Researchers have used and applied Natural Language Processing technique to extract relevant information from the vast corpus of text documents. In the research, Rusu et al. [10] the authors suggested the technique to extract the information and relevant phrases in the form of subject-predicate-object triples. To do so, Parse Trees were generated from English sentences, and triples were extracted from the parse trees [17][10]. In the research work of Etzioni et al. [11], the author developed the KNOWITALL system, which helped in the automation of extracting extensive collections of facts from the web in an unsupervised, domain-independent, and scalable manner [17]. The author used the approach of Pattern Learning to address this challenge [17]. In another research, other necessary NLP technique approach applied for

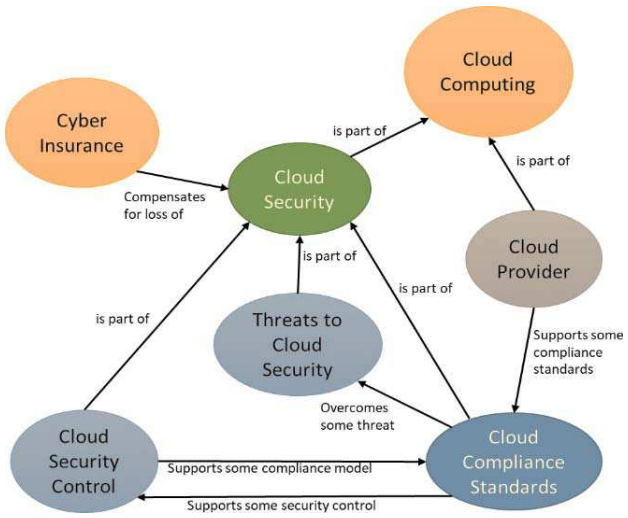

Fig. 1. High-level reference architecture of Cloud Data Security

information extraction from unstructured text is 'Noun Phrase Extraction' [71]. Author Rusu et al. [10] showed the technique of creating triplets by considering 'Noun Phrases' obtained via various part-of-speech taggers. Different automated techniques have been used for extracting the permissions and obligations from legal documents [17]. Techniques such as text mining and semantic techniques have been explored and applied by various authors in the past [17][24][25]. In the research work of Kagal et al. [19][22], the authors proposed an ontology-based policy framework to model conversation specifications and policies using obligations and permissions [17][19][22].

III. DATA PROTECTION REGULATIONS

As a first step towards building our integrated data compliance ontology, we did a detailed study of various security and privacy regulations and guidelines that apply to data managed by Cloud services. Figure 1 illustrates our high-level reference model for Cloud data security that we used to build our methodology. In this section, we list the key Cloud compliance standards along with the security controls that are needed for these regulations. In our prior work [12], we have analyzed the critical security threats faced by Cloud services consumers and related them to the security controls and compliance models that protect from these threats.

The following are the critical security controls that affect Cloud security. We have referenced the NIST and CSA security documents [28] [14] [54]. We also co-relate them with security standards based on the description of controls.

A. DATA ENCRYPTION, KEY MANAGEMENT:

Data encryption is necessary to provide data confidentiality and integrity. Encryption/decryption key management also allows users to access authorized data securely. Data encryption includes application encryption and network encryption. The compliance model for data encryption should be capable of preventing accidental exposure and misuse of the data in public domains. After analyzing several security standards, we found that data encryption standards like FIPS 140-2 and Vaultive fulfill these requirements. CSA guide [14] suggests avoiding old security standards like DES (Data Encryption standards). Key management is also an essential aspect of data encryptions. Key management can also be done using KEK (Key encrypting keys) [40].

B. MEDIA PROTECTION:

Media protection includes the protection of entertainment content like music, movies, and software [61]. It is the responsibility of Cloud providers to protect the entertainment content of users from piracy [61]. It may contain pre-release material from creative arts to the software industry. Strong compliance models should be adhered to, and legal action should be taken against the attackers. If the media protection security control model is implemented correctly, more consumers will store the data in the Cloud. The MPAA compliance model is specially designed for media protection.

C. IDENTIFICATION, AUTHENTICATION, AND AUTHORIZATION:

Identification not only consists of user identification but also device and resource identification. Multi-tenancy requires that consumers share common resources in the public domain. The identification of correct resources to authorized users is an essential aspect of this security control [61]. After identification, authentication of users also plays a crucial part in this model. The users should be identified by key management and passwords. Cloud providers should also provide access controls to users so that they can give rights to other authorized users [61]. This is called authorization. Cloud providers should apply the compliance model that manages these three tasks. This will not only enforce data security but will also help to implement other security control models more effectively. Compliance models like Oauth and NIST 800-63 provide guidelines for valid authentication, identification, and authorization of the Cloud system.

D. VIRTUALIZATION AND RESOURCE ABSTRACTION:

Virtualization in the Cloud can be used to achieve higher density through multi-tenancy and resource utilization, which makes the organization more efficient. Virtualization and Resource abstraction control models (mainly technology, architecture, and service models) should focus on new tools and techniques to improve visibility for security operators. Virtualization brings more specific Cloud security issues like inter-virtual machine attacks, hypervisor security, etc. It is recommended that a virtual machine setup should also include firewall implementation. PCI-DSS standard is not only focused on the payment card industry, but it also supports hypervisor security implementation.

E. PORTABILITY AND INTEROPERABILITY:

Various components in the Cloud system working together for higher performance are called Interoperability. Interoperability is achieved by creating standards for application

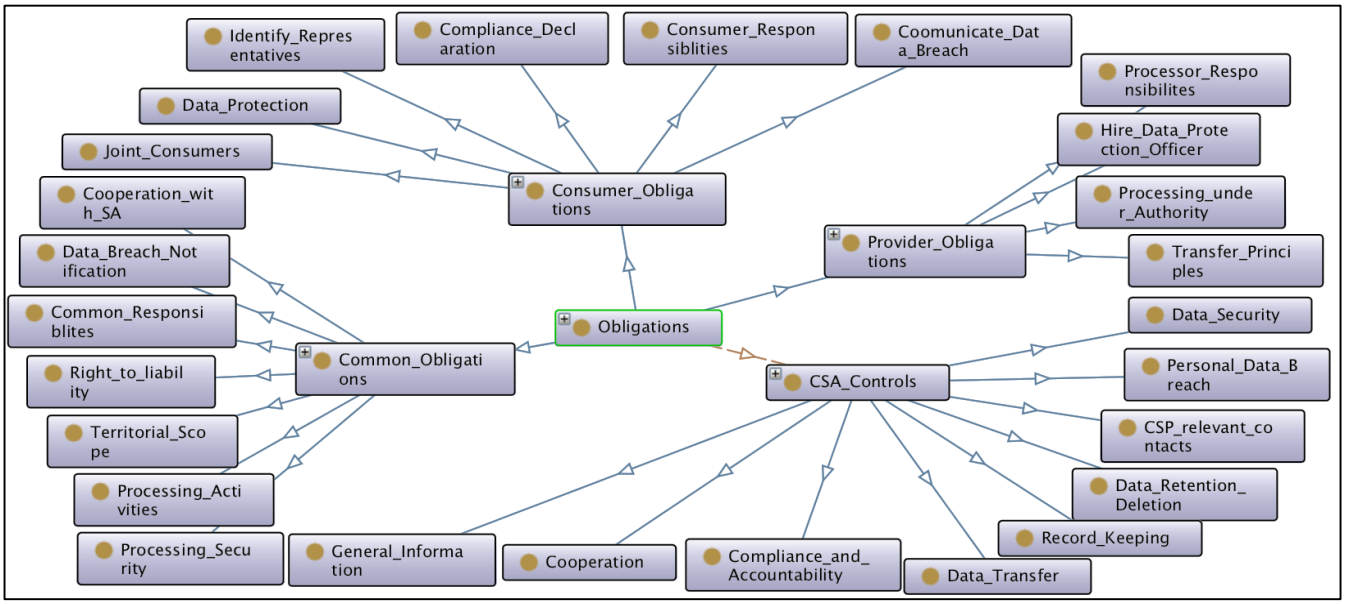

Fig. 2. High-Level Ontology for GDPR Data Protection on Regulations that apply to Cloud Data

interfaces (APIs) for collaborating with all the components. Different platforms have different APIs, so there should be some standard, which will make systems interoperable with each other. It is advised that OCCI (Open Cloud computing interface) [14], libCloud should be applied whenever possible. Portability is reusing the components of the Cloud system. Portability decreases the production cost. However, we have to make sure there should be some mechanism, which will reuse the component between different systems, but data is secured. The security standards implemented on the Cloud system should enable information sharing amongst the other system. Otherwise, it will increase additional expense and reengineering.

F. APPLICATION SECURITY:

Application security is the overall security of the applications running on the Cloud. If we want to achieve application security, we have to take care of the following processes - secured SDLC (software development lifecycle), authentication, and authorization. Secure SDLC can be achieved if we implement the maturity models like system security engineering capability maturity models (SSE-CMM). Application security controls should implement and validate controls for validation and authentication.

G. SECURITY RISK ASSESSMENT AND MANAGEMENT:

When security is added to the Cloud, the risk factors should also be considered. Cloud computing allows the sharing of resources across all the consumers at a low cost. However, Cloud providers should implement the authorization and risk assessment for utilizing shared resources. FedRAMP is a compliance model, which provides guidelines for risk assessment and management. It also differentiates between the shared authorization model and the system-centric authorization model.

H. PRIVACY, ELECTRONIC DISCOVERY, AND OTHER LEGAL ISSUES:

Privacy and electronic discovery focuses on managing the physical location of data and also accessing confidentially. It also implements privacy and confidentiality policies to ensure compliance. For this security control documents, terms of services and privacy policies should be reviewed. EDRM-PSRRM compliance model provides security and risk reduction models for privacy and e-discovery.

I. CONTINGENCY PLANNING:

It is the Cloud consumers' responsibility to understand the Cloud provider's contingency plans and Service Level Agreements (SLAs) to make sure that Cloud providers meet all the requirements. According to NIST 800-34, steps for contingency plan are development of statement, conduct business impact analysis, identify preventive controls, create strategies, develop a contingency plan, ensure testing, and plan for maintenance.

J. DATACENTER OPERATIONS, MAINTENANCE:

Security controls should also have standards for maintaining data centers. Maintenance of data centers includes configuration and personnel security with a background check to enter secured data center location, physical privacy of data center, and authentication [61].

K. INCIDENT RESPONSE:

Cloud providers should develop a response plan in case of any incident like data breaches, data loss, etc. Computer forensics has some different tools and techniques for incident response. The incident response lifecycle consists of the following phases - Preparation for the incident, detection, and analysis of incidents, data sources, forensics, and other

TABLE I

LIST OF SECURITY CONTROLS AND RECOMMENDED SECURITY COMPLIANCE MODELS

Security Controls	NIST Supported [28]	CSA Supported [15]	Security Compliance Models
Data Encryption, Key Management	Y	Y	FIPS 140-2, Valuative
Media Protection	Y	Y	MPAA
Identification and authentication	Y	Y	STIG, FedRAMP, OAuth, NIST 800-63
Virtualization and Resource Abstraction	-	Y	DMTF-OVF, PCI-DSS
Portability and Interoperability	Y	Y	OASIS, PCI-DSS, DMTF-CADF, OCCI, libCloud
Application security	Y	Y	PCI DSS, ISO 27002, SOX, HIPAA
Security risk assessment, authorization	Y	Y	STIG, ISO27002, FedRAMP, DIACAP - FISMA
Privacy, electronic discovery	-	Y	EDRM-PSRRM
Contingency planning	Y	Y	HIPAA, NIST 800-34
Datacenter operation and maintenance	-	Y	PCI DSS, ISO-27002, HIPAA, NIST 800-16, NIST 800-53, ITAR.
Incident response	Y	Y	NIST 800-61, ISO 17799
Compliance, audit, and accountability	Y	Y	DMTF-CADF, ISO 27001, COBIT
Awareness and training	Y	Y	NIST 800-61, ISO 17799
Access control	Y	-	SOX, Safe Harbor
Configuration management	Y	-	NIST 800-100

investigation support for incident analysis and recovery from the incident [61].

L. COMPLIANCE, AUDIT, AND ACCOUNTABILITY:

Cloud computing environments are dynamic and bring new opportunities for additional audit capabilities. These policies require the implementation of robust evaluation criteria. After implementing the compliances, regular audits should be conducted to ensure data security.

M. AWARENESS AND TRAINING:

Cloud awareness and training program should be for those consumers who want to migrate their data to the Cloud but not aware of all the threats and security controls. Cloud providers should develop a response plan in case of any incident like data breaches, data loss, etc. Computer forensics has some different tools and techniques for incident response [61]. The incident response lifecycle consists of the following phases - Preparation for the incident, detection, and analysis of incidents, data sources, forensics, and other investigation support for incident analysis and recovery from the incident.

TABLE II

GDPR KEY TERMS USED TO FORM A KNOWLEDGE GRAPH

Key terms	Frequency
data subject	375
processor	528
controller	1008
profiling	46
data breach	37
personal data	1148
consent	144
notification	28
profiling	46

TABLE III

PCI-DSS KEY TERMS USED TO FORM A KNOWLEDGE GRAPH

Key terms	Frequency
Maintain	10
Control	13
Establish	5
Access	43
unauthorized	6
Ensure	10

N. COMPLIANCE, AUDIT, AND ACCOUNTABILITY:

Cloud computing environments are dynamic and bring new opportunities for additional audit capabilities. These policies require the implementation of robust evaluation criteria. After implementing the compliances, regular audits should be conducted to ensure data security.

O. AWARENESS AND TRAINING:

Cloud awareness and training program should be for those consumers who want to migrate their data to the Cloud but not aware of all the threats and security controls. Based on the security controls definition provided by NIST [28] and CSA [14], we try to relate the security compliance laws to the security controls. In Table 1, the security controls supported by NIST or CSA are listed, followed by the recommended Cloud compliance regulations.

IV. COMPLIANCE KNOWLEDGE GRAPH

In this section, we describe our methodology in detail. We aim to present a rich policy-based knowledge representation of the data compliance regulations with the corresponding CSA controls. Figure 4 illustrates the integrated high-level ontology. The three phases of our methodology are:

1) PREPROCESSING STAGE: For the regulations, we extracted relevant chapters and key terms and then mapped them with corresponding CSA controls. In the first stage of our system, we extracted the repository

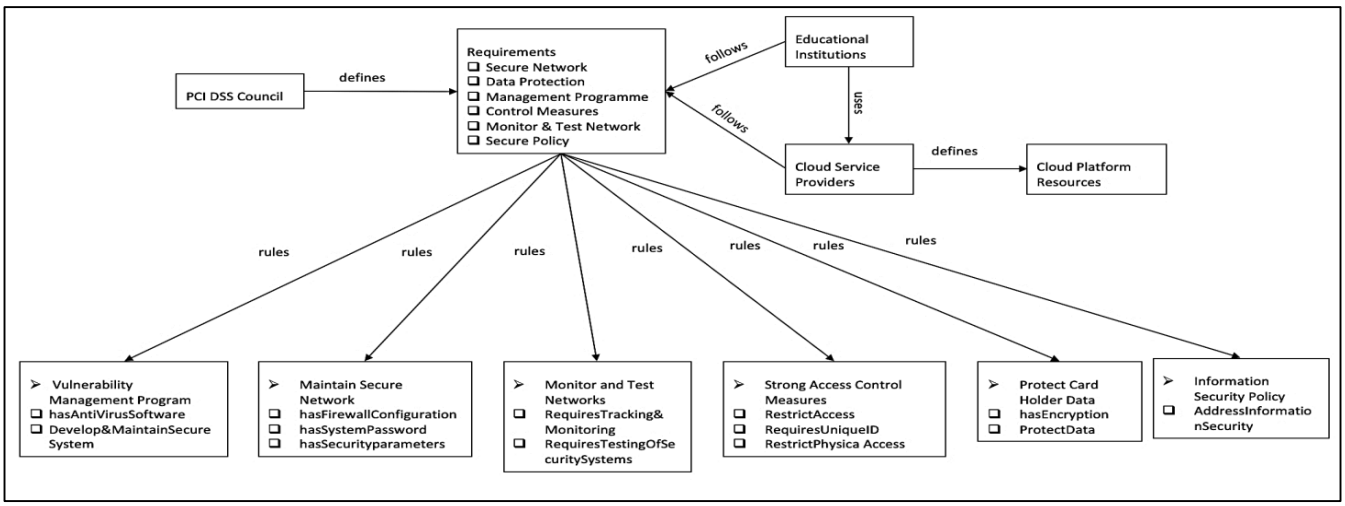

Fig. 3. Integrated High-Level Ontology for PCI DSS Knowledge Graph

& checklist of GDPR [44] and PCI DSS [38], respectively. In our previous work [38] [44], we extracted the relevant key terms from the PCI DSS documents & GDPR and built the knowledge graph accordingly. In the preprocessing stage as part of previous work [44], we extracted chapters 3 and 4 of the GDPR regulation, which are for Consumers and Providers. Like mentioned as part of previous research did [38] [44], we have obtained the key terms which are shown below in the respective Table 2 & Table 3:

2) KNOWLEDGE GRAPH/ONTOLOGY DEVELOPMENT We have developed a comprehensive Data Compliance ontology (Figure 4) that integrates the knowledge representation of various Cloud regulations. For creating the knowledge graph, we utilized the Protégé toolset.

The main classes include-

- The **Stakeholder** class is the main class that represents the key organizations that are affected by the regulations. This class has three main subclasses. These are Consumers, Providers, and Regulators. The **Consumer** class represents the data users and includes properties of end-users. The **Provider** class represents the data providers and includes properties of providing organization Cloud policies. The **Regulators** class represents the regulatory bodies and includes all the details of the council.
- **Regulations** class captures details of the regulation, including its name, description, scope, and country of the regulation. The regulations class is associated with one or more stakeholders. These individual regulations are then captured by different sub-classes. We have also integrated the knowledge graphs that we have already developed for various regulations like GDPR [49], PCI-DSS [38], HIPPA, with this ontology. As part of our ongoing work, we are developing knowledge graphs for other regulations.
- Regulations class is associated with **Cloud Security Controls** and **Cloud Threats** classes.
- **Cloud Security Control**: This class represents the security controls recommended by the Cloud Security Alliance. In this paper, we have related all the regulations that are associated with **Regulations** to Cloud security controls class.
- **Cloud Threats**: The purpose of this class is to associate various Cloud threats to appropriate regulation from the **Regulations** class. This captures the threat name and description as properties.

A. GDPR KNOWLEDGE GRAPH

In our previous work [44], we have identified the keywords that are associated with GDPR. Key terms under GDPR are "data subject", "processor", controller", "profiling",

"data breach", "personal data", "consent", "notification" and "profiling".

To populate our ontology, we have searched for the key terms from the individual organizational policies.

Below are the example statements from the privacy policies for the key term “controller”.

“Microsoft: Identified which Microsoft entities are data controllers under the GDPR, how to contact us, and how to lodge a complaint”.[43]

“WhatsApp: Partners (the data controllers) may submit personal information about their customers to WhatsApp using WhatsApp’s Business Products.” [57]

“Google: Additionally, for products where Google and the customer each act as independent controllers of personal data, we have updated our agreements or made available terms that reflect that status.” [48]

“Facebook: A company is a data controller when it has the responsibility of deciding why and how (the ‘purposes’ and ‘means’) the personal data is processed.” [46]

AWS: ‘the data exporter’ means controller who transfers personal data”.[47]

We then applied deontic logic and divided the whole set rules into either **Permissions or Obligations**. Some of the statements from organizational policies are listed below.

"Facebook: Under the GDPR, data controllers must adopt compliance measures to cover how data is collected, what it's used for, and how long it's retained. They also need to make sure people can access the data about them".[46]

"WhatsApp: The data subject can enforce against the data exporter this Clause, Clause 4(b) to (i), Clause 5(a) to (e), and (g) to (j), Clause 7, Clause 8(2), and Clauses 9 to 12 as third-party beneficiary".[57]

"AWS: This DPA shall continue in force until the termination of the Agreement (the "Termination Date")." [47]

"Facebook: Data processed must be necessary for the Service and defined in the contract with the individual." [46]

We have also identified the key classes of a knowledge graph to represent the GDPR rules. We have referenced the GDPR regulation available at [37] [38] for this.

1) CONSUMERS AND PROVIDERS:: The regulation splits the tasks and obligations of consumers and providers, obligating consumers and providers that provide "adequate guarantees to implement suitable technical and organizational measures" to meet the regulation's policies and protect data subject's rights [63].

The regulation provides specific counsels for what kinds of security actions should be considered "appropriate to the risk," including [63]:

- The pseudonymization and/or encryption of individual data.
- The capability to certify the ongoing confidentiality, integrity, availability, and resilience of systems and services processing personal data. The aptitude to restore the availability and access to data promptly in the event of a physical or technical occurrence.
- A procedure for regularly testing, assessing, and evaluating the efficiency of technical and organizational measures for ensuring the security of the processing.

2) FINES AND ENFORCEMENT:: Breach of compliance will result in fines of up to 4% of global revenue or €20m, equivalent to roughly $23.4m whichever is greater. It will depend on the severity of the breach and the organization's ability to demonstrate that there were initial measures in place (or not) to protect customer data.

3) BREACH & NOTIFICATION:: In the incident of a personal data breach, data consumers must inform the appropriate supervisory authority without undue delay and, where possible, not later than 72 hours after knowing about a data breach. If notice is not made within 72 hours, the consumer must provide a reasoned justification for the delay [37].

4) DATA PROTECTION OFFICER:: Whoever holds this position will be accountable for managing data protection and data privacy, and free to give approvals or feedback without any fear of negative implications. This only applies if an organization handles huge important volumes of data, typically not applicable to small to medium-sized enterprises.

5) DATA SUBJECT:: Individuals will have more data on how their data is handled, and this information should be available in a clear and reasonable way. Consumers must inform data subjects about the period of (or reasons why) data will be reserved on collection. Data subject consequently wish to have their data removed, and the data is no longer required for the reasons for which it was composed, then it must be erased.

TABLE IV
CONSUMER OBLIGATIONS GDPR AND CSA COC

Obligation	GDPR Article	CSA Controls
Consumer Responsiblities	Article 5 [63] [66]	WWP-1.1, WWP-1.2, WWP-1.3 [65]
Declaration of Compliance	Article [63] [66]	DCA-1.1, DCA-1.2, REC-2.1, REC-2.5 [65]
Data Protection by Design and by Default	Article 25 63] [66]	WWP-1.4, REC 1.7 [65]
Joint Consumers	Article [63] [66]	CAR-1.3, REC-1.2 [65]
Identify Representatives	Article 27 [63] [66]	CAR-1.2, REC-1.2 [65]
Communicate Data Breach	Article 34 [63] [66]	PDB-1.1, PDB-1.2, PDB-1.4, PDB-1.6, PDB-1.7 [65]

To develop the ontology, we have used the mixture of top-down and bottom-up approach by answering the following questions:

- What are the major obligations that will impact an organization?
- What are the specific entities that will be affected?
- Are there any common obligations for consumers and providers?
- Can we come up with a list of obligations that will affect consumers and providers individually?
- Is there a CSA code of conduct control associated with each obligation?

Upon answering the above questions, we could identify our classes, subclasses, and relationships for the ontology, as illustrated in Figure 2. We have identified the associated CSA Code of Conduct controls for the GDPR articles. Table 4, 5, and 6 represents the association between GDPR obligations vs. CSA controls. In our knowledge graph, we have included the associated CSA Code of Conduct controls [80] for the GDPR articles.

B. PCI DSS KNOWLEDGE GRAPH

In our previous paper, we have described the PCI DSS ontology developed by us based on the requirements defined by the PCI DSS council. The security controls and processes required by PCI DSS are vital for protecting cardholder account data, including the PAN – the primary account number printed on the front of a payment card [38]. This includes sensitive data that is printed on a card or stored on a card's magnetic stripe or chip – and personal identification numbers entered by the cardholder [38]. In general, if an organization deals in card transactions, then it must follow the policies listed below [38].

TABLE V
PROVIDER OBLIGATIONS GDPR VS. CSA COC

Obligation	GDPR Article	CSA Controls
Processor Responsibilites	Article 28 (2-4), 10 [63] [66]	WWP-3.2, WWP-3.3, WWP-3.4, WWP-5.1, WWP-5.2, WWP-5.3, WWP-5.4, WWP-5.5 [65]
Processing under Authority	Article 29 [63] [66]	WWP-1.14 [65]
Hire Data Protection Officer	Article 37 [63] [66]	REC-2.2 [65]
Transfer Principles	Article 44 [63] [66]	WWP-5.2, REC 2.4, DTR-1-1, DTR-1-2 [65]

TABLE VI
COMMON OBLIGATIONS VS. CSA COC

Obligation	GDPR	CSA Controls
Territorial Scope	Article 3(1) [63] [66]	DCA-1.2, DCA-1.3 [65]
Common Responsiblities	Article 28(1) [63] [66]	WWP-1.15, RRD-3.1, RRD-4.1, RRD-4.2, CPC-1.1, CPC-1.2 [65]
Records of Processing Activities	Article 30 [63] [66]	WWP-2.1, DCA-1.4, MON-1.1 [65]
Cooperation with the Supervisory Authority	Article 31 [63] [66]	WWP-5.7, PDB-1.6 [65]
Processing Security	Article 32 [63] [66]	SEC-1.2, MON-1.1, CAR-1.5 [65]
Data Breach Notification	Article 33 [63] [66]	PDB-1.1, PDB-1.2, PDB-1.3, PDB-1.4, PDB-1.5 [65]
Right to compensation and liability	Article 82 [63] [66]	SEC-1.1 [65]

1) BUILD AND MAINTAIN A SECURE NETWORK: 'Install and maintain a firewall configuration to protect cardholder data [1] [4]'. The network configuration and its

security requirements should be shared by the IT team and Cloud service providers [38][39]. 'Define the system password and its security parameters' [38][39]. This means that all the default passwords supplied by the providers

should be changed when a system is getting installed in the configured network [38][39].

2) PROTECT CARDHOLDER DAT: 'Protect stored cardholder data' [38][39]. This means that only the necessary data should be stored, and at least every quarter, any unnecessary data should be purged. PAN details should be masked, the first six and last four digits are the maximum number of digits you may display [38][39]. Also, PAN details must be made unreadable wherever it is being stored [38][39]. 'Encrypt transmission of cardholder data across open, public networks' [38][39]. This rule of PCI DSS policy asks the organization to make use of strong cryptography and encryption technologies like SL/TLS, SSH, or IPSec, etc. to safeguard sensitive cardholder data during transmission over any networks [38][39].

3) MAINTAIN A VULNERABILITY MANAGEMENT PROGRAM: 'Use and regularly update the anti-virus software or programs' [8][39]. All the systems and servers should have

anti-virus software's to prevent malicious activity. At the same time, anti-virus services should be running in the background and generating auditing logs [38][39]. 'Develop and maintain secure systems and applications' [38][39]. This policy ensures that all the patches must be installed on time whenever any new patches are published by the vendors [38][39]. Any changes to the system components, coding of applications must be done through proper change and control procedures [38][39]. Also, firewall protection should be ensured for any public-facing web applications [38][39].

4) IMPLEMENT STRONG ACCESS CONTROL MEASURES: 'Restrict access to cardholder data by business need to know' [38][39]. This policy ensures that access is limited to system components and cardholder's data. Also, access control protocol for systems components should be in place for multiple users, and it must restrict access based on a user's needs and should be set to "deny all" unless specifically authorized [38][39]. 'Assign a unique ID to each person with computer access. These policies ensure that any person who is accessing the data should have a unique ID [4]. This will help in tracing an individual's activity in case of any violation or misuse [4]. Also, there should be two-factor authentication for remotely logging into the network for, such as making use of RSA token or other technologies that facilitate two-factor authentication [38][39] 'Restrict physical access to cardholder data' [38][39]. This ensures that proper facility controls should be applied to the cardholder data environment, and individuals only with proper authorization should be allowed to access cardholder data [38][39]. For visitors, the proper token should be given with expiry, and a visitor log must be maintained for tracking purposes [1] [4].

5) REGULARLY MONITOR AND TEST NETWORKS: 'Track and monitor all access to network resources and cardholder data' [38][39]. This ensures that an established process should be implemented to link access of

individuals to system components [38][39]. Log activities of the system components must be reviewed daily, and audit trail history must be retained for at least one year so that three months of activity is available immediately [38][39]. 'Regularly test security systems and processes' [38][39]. This ensures that all the test procedures should be in place to detect access points and unauthorized users [38][39]. Also, external and internal penetration testing should be performed, including network and application-layer penetration tests at least annually [38][39].

6) MAINTAIN AN INFORMATION SECURITY POLICY: This ensures that the PCI DSS policies that have been established, published, and maintained have clear, descriptive definitions of the procedures that everyone in the system knows thoroughly, and such policy must be reviewed at least once a year [38][39].

Based on the PCI DSS repository, we created the knowledge graph. Our knowledgebase consists of six different class which incorporate the 12 requirements. Figure 3 illustrates our ontology. The main stakeholder entities are PCI DSS Council, Educational Institutions, and Cloud Service Providers. In our ontology, we have six classes having two or more subclasses

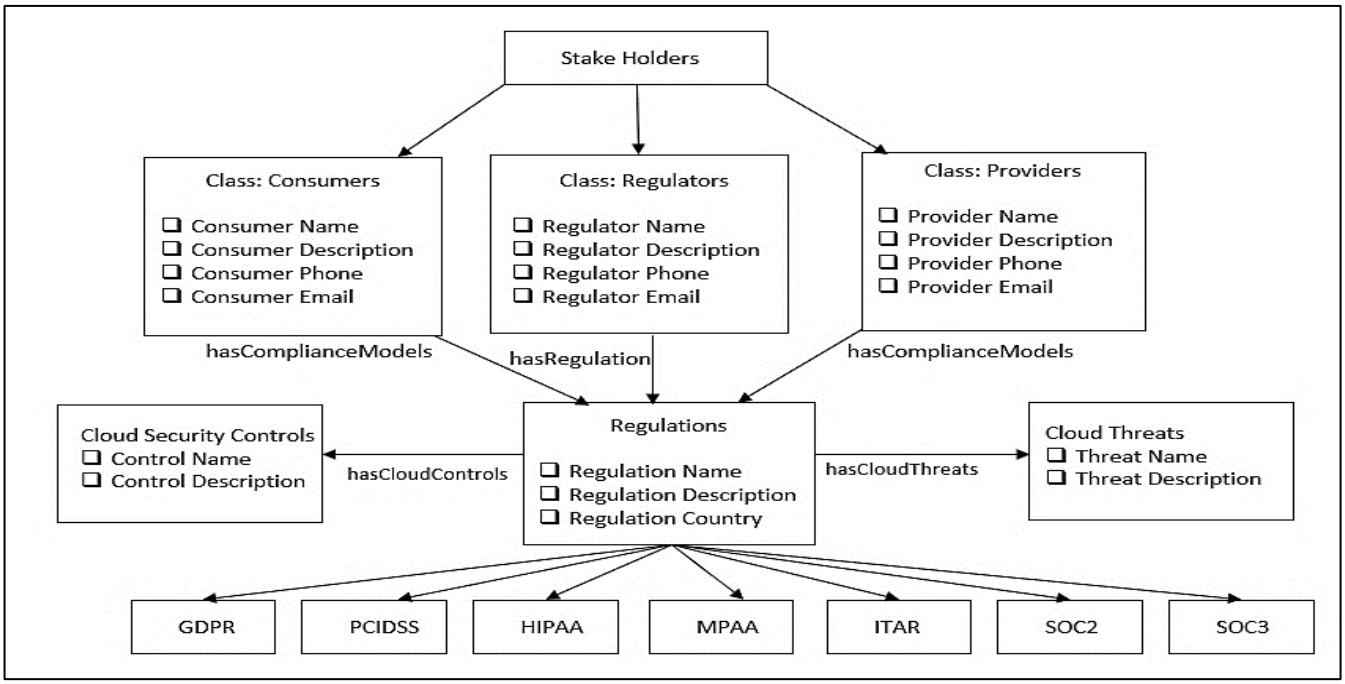

Fig. 4. Integrated High-Level Ontology for all Data Protection Regulations that apply to Cloud Data

in it. Each class are disjoint from other classes which means that an individual (or object) cannot be an instance of more than one of these six classes

Based on the security controls definition provided by NIST [28] and CSA [14], we try to relate the security compliance model to the security controls. In Table 1, the security control supported by NIST or CSA is listed, followed by the recommended Cloud compliance system.

In our previous work [44], we have identified the keywords that are associated with PCI-DSS. Key terms under PCI-DSS are "maintain", "control", "establish", "access", "unauthorized," and "ensure". To populate our ontology, we have searched for these key terms from the individual organizational policies. Below are the example statements from the privacy policies for the key term "control".

"AWS: Service providers now are required to detect and report on failures of critical security control systems". [47]

"eBay: You will maintain such compliance at all times during the term of the Terms. This requirement will survive the duration of the Terms until you return, destroy, or cause

C. CLOUD SECURITY ONTOLOGY

The ontology for Cloud computing security is illustrated in Figure 4. The Cloud computing security class is divided into Cloud security compliance models, Cloud security controls, and Cloud security threats. The relations between all the classes are described in the ontology. The ontology is further developed with individual class and its subclasses. The Cloud security control class and its subclasses are illustrated in Figure 5.

Some of the compliances and security standards are displayed for understanding the relationship between two classes.

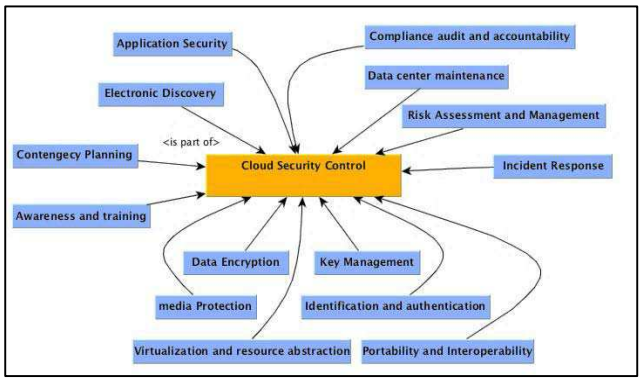

Fig. 5. Ontology for Cloud security controls and its subclasses.

As discussed in section III, each Cloud security standard supports a type of compliance. For example, the security standard MPAA (Motion Picture Association of America) is used for protecting the original content from piracy. It fulfills all the requirements stated in media protection compliance. Hence, we can show that MPAA supports Media protection compliance. Similarly, we can show the relation between security standards and Cloud security compliances mentioned in section III, Table 1.

Figure 6 describes the class Cloud security compliances and its relationship with the security control class. The types of Cloud security compliances, explained in the Appendix, are represented in the ontology. Figure 7 illustrates the relation between security standards and security threats. The security standards overcome the threats if they are correctly used in Cloud security. For example, a data breach is a security threat to the Cloud, but it can be overcome if we apply the

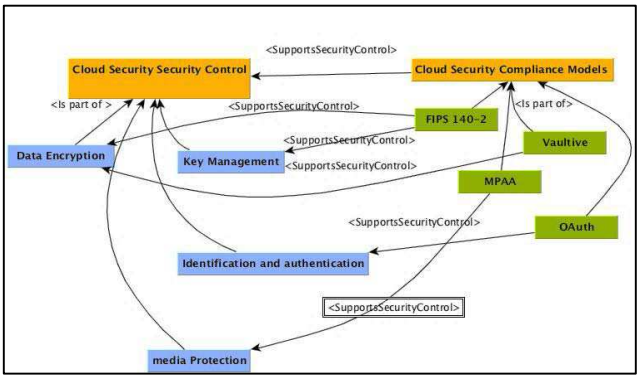

Fig. 6. Sample Ontology for the relationship between Security Controls and Security Compliances classes.

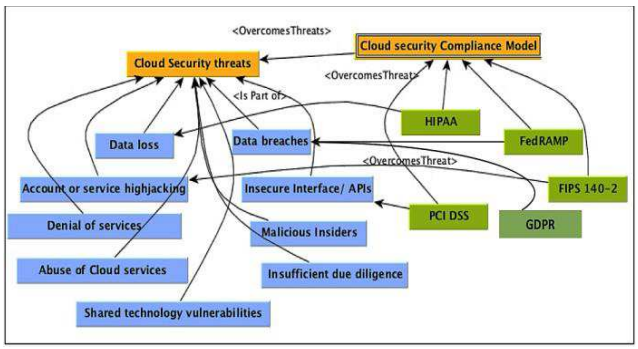

Fig. 7. Ontology of Relation between Security standards and threats.

compliance standard FedRAMP that is specifically used for data security.

V. VALIDATION RESULTS

We have validated our knowledge graph with the privacy policies of various providers like amazon, google, IBM, and Rackspace. Figure 8 illustrates the PCI-DSS regulation instance with all the policies associated with it. Each regulation is associated with threats and controls instances as well. Likewise, we have integrated all the regulations and built relationships with Cloud standards, controls, and threat classes. Figure 9 shows the results of amazon instance from the provider class. End-user can quickly glance if all the regulations are followed by their organization and act by finding out the missing policies. We have listed the SWRL rules in Figure 10.

Based on the key terms extracted from the above sections, we populated the statements in corresponding classes of our ontology. We then check the regulations followed by organizations using the SPARQL queries [70]. Below are the sample queries to check for the consumer and provider obligations under GDPR/PCI-DSS that are followed by an organization.

SPARQL query to check for GDPR provider obligations Amazon:

PREFIX owl: http://www.w3.org/2002/07/owl#
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

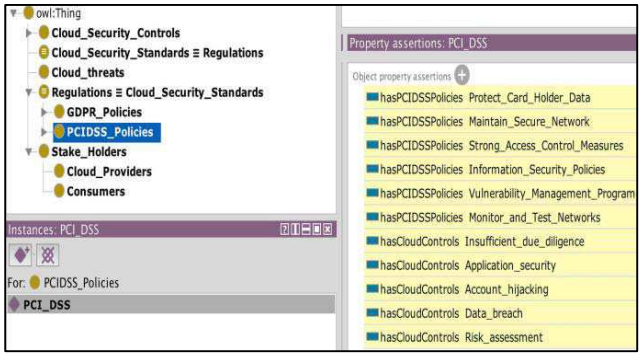

Fig. 8. PCI-DSS regulation rules

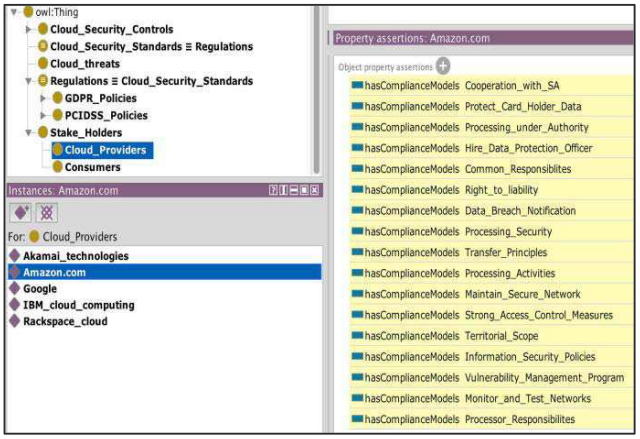

Fig. 9. Amazon Policies

PREFIX cc: <http://www.semanticweb.org/ontologies/2019/6/Cloud_Compliance_Final#>
SELECT ∗
WHERE {cc:Amazon.com cc:ProviderPolicies ?Rules }

SPARQL query to check for GDPR obligations for Territorial Scope:

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIXcc: <http://www.semanticweb.org/ontologies/2019/6/Cloud_Compliance_Final#>
SELECT ∗
WHERE {cc:Territorial_Scope cc:GDPR_Territorial_Scope ?Rules}

For the process of validation, we referred to Cloud data policies of major Cloud data providers. We wanted to verify if key terms and obligations specified in these data policies and can be populated as instances of our data compliance knowledge graph. First, we populate the ontology by utilizing the original GDPR and PCI-DSS policy documents. We then run SPARQL queries to identify original policy statements under each class of our ontology. Results from these SPARQL queries are exported to compare with the results of organiza-

Name	Query
S1	GDPR_Policies(?s) ^ Cloud_Providers(?t) ^ hasCommonPolicies(?s, ?z) -> hasComplianceModels(?t, ?z)
S2	GDPR_Policies(?s) ^ Cloud_Providers(?t) ^ hasProviderPolicies(?s, ?z) -> hasComplianceModels(?t, ?z)
S3	PCIDSS_Policies(?s) ^ Cloud_Providers(?t) ^ hasPCIDSSPolicies(?s, ?z) -> hasComplianceModels(?t, ?z)
S4	GDPR_Policies(?s) ^ Consumers(?t) ^ hasConsumerPolicies(?s, ?z) -> hasComplianceModels(?t, ?z)
S5	PCIDSS_Provider_Obligations(?s) ^ Cloud_Providers(?t) ^ PCI_DSS_Provider_Rules(?s, ?z) -> ProviderPolicies(?t, ?z)
S6	GDPR_Provider_Obligations(?s) ^ Cloud_Providers(?t) ^ GDPR_Provider_Rules(?s, ?z) -> ProviderPolicies(?t, ?z)
S7	GDPR_Common_Obligations(?s) ^ Cloud_Providers(?t) ^ GDPR_Common_Rules(?s, ?z) -> ProviderPolicies(?t, ?z)

Fig. 10. SWRL Rules

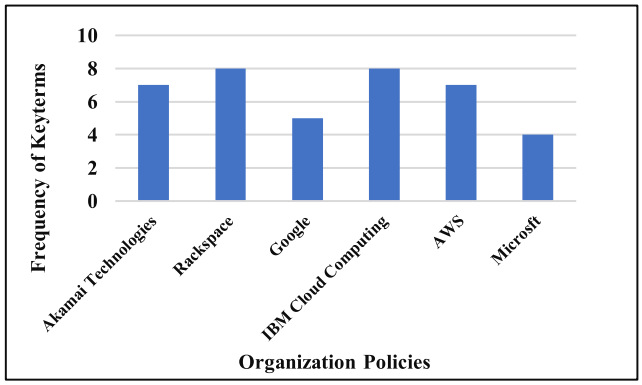

Fig. 11. Validation Results

tional policies. Classes that do not have any induvial means that an organization is not in compliance with that regulation either under GDPR/PCI-DSS.

This analysis will help an organization to verify the results with the original regulation document quickly. We found similar key terms in the organizational policies along with the number of times that term has occurred. The graph in Figure 11 is a snapshot of key terms and the count for various organizations. With the help of these terms, each organization's policies were populated as instances of our knowledge graph. The data policies are now available as an RDF graph and are machine-processable. It will now be possible to automate the compliance validation by using policy reasoning engines that can alert any potential compliance violation.

VI. CONCLUSION

We have developed the Cloud security comparator system for consumers who are planning to move their data to the Cloud but are uncertain due to security concerns as they may not be aware of various compliance models. This study also helped us determine the Cloud security controls and policies and quantify them in a comprehensive manner. As part of our ongoing work, we will further analyze other IT compliance models to improve our recommendation system.

As we discussed, the analysis will clarify the importance of security controls and compliance models. Also, the prototype will help Cloud consumers choose Cloud providers based on the security compliance model. In the future, we plan on refining the recommendation system by adding the cost of Cloud providers. The cost factor will give us the cognitive result to choose the best Cloud provider. Similarly, this prototype model can be implemented on the IT compliance models other than security. We can also integrate this tool with e-commerce providers to find an optimized solution for B2B services.

VII. APPENDIX: ANALYSIS OF VARIOUS CLOUD DATA COMPLIANCE MODELS

After a detailed study of existing data protection regulations, we have identified the following standards that apply to Cloud-based services and applications.

1. ISO 27002

ISO standard for information security controls [20]. It was initially published as ISO 17799. This standard advises how to implement various controls in an organization, but it does not focus on a particular compliance model.

Key features: Network security, incident management, security compliance review

2. ISO 27001

ISO 27001[21] is an auditable international standard for information security management system (ISMS) and focuses on selecting adequate and appropriate security controls. Generally, a full assessment is done every three years, and a surveillance audit is performed every six months.

Key features: Compliance Audit, risk assessment, IT security management.

3. PAYMENT CARD INDUSTRY DATA SECURITY STANDARD (PCI-DSS)

This standard by Payment Card Industry Security Standards Council (PCI-SSC) [39] aims to reduce credit card frauds. It applies to organizations that store, process, and transmit cardholder's information. Note that even though a Cloud provider is PCI-DSS compliant, the Cloud consumer does not necessarily become PCI-DSS compliant.

Key features: Protect Credit, Debit cardholder-related information, Strong access control, Maintain a firewall, Anti-virus software maintenance.

4. STATEMENT ON STANDARDS FOR ATTESTATION ENGAGEMENTS (SSAE16)

This standard [8] was developed by American Institutes of Certified Public Accountants (AICPA) for reporting on Controls at a Service Organization, the Statement on Auditing Standards SAS70. SSAE16 has three kinds of Service Organization Controls(SOC) reports. SOC1 report is required when audits conducted over internal controls over financial reporting, management of the user organization, and management of the service organization. SOC2 report is required when auditing the organization's security, availability, privacy, confidentiality, and processing. SOC3 report is given to Cloud provider organization when there are restrictions on providing information about current and potential customers in auditing,

Key features: Security auditing standards

5. GENERAL DATA PROTECTION REGULATION (GDPR)

GDPR standard is mandated by the European Union for protecting data of European citizens. As part of our previous work [49], we have created a high-level ontology to represent the GDPR rules, and it is described in section IV A.

Key features: Data privacy protection for EU citizens.

6. CONTENT PROTECTION AND SECURITY (CPS)

This standard was created by the Content Delivery & Security Association (CDSA)[23]. This standard mainly focuses on

managing IT security and piracy risks.

Key features: Auditing of system, Risk assessment.

7. HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA)

HIPAA[24] is a security standard for health-related data. If Cloud providers store data related to health, they have to adhere to HIPAA standards to protect it.

Key features: Electronically protected health information, Risk management

8. FEDERAL RISK AND AUTHORIZATION MANAGEMENT PROGRAM (FEDRAMP)

FedRAMP[9] is a federal security authorization process. It enhances the transparency between government and Cloud providers. It reuses the current security assessments, which save high cost, time, and resources. FedRAMP provides a uniform approach to risk-based management.

Key features: Security assessment and management, security authorization

9. DIACAP AND FISMA

DIACAP [68] or DoD (Department of defense) Information Assurance Certification and Accreditation Process is a compliance standard developed by DoD, which is closely aligned with FISMA (Federal Information Security Management Act). DIACAP standard leverages the controls with DoD 8500.2, and these DoD 8500.2 controls are applied based on MAC (machine). DIACAP ensures that risk management is applied to information systems. It also maintains information assurance throughout the system.

Key features: Defense data, Risk assessment, contingency planning.

10. INTERNATIONAL TRAFFIC IN ARMS REGULATIONS (ITAR)

ITAR[26] regulates and controls the import and export of defense-related articles and services over the network.

Key features: Defense-related electronic data

11. Federal Information processing standard (FIPS) FIPS-140-2

Federal Information processing standard (FIPS) 140-2 [27] is a US government standard that specifies the cryptographic modules for data protection. There are four levels of security defined in FIPS 140-2.

Key features: security standards for cryptographic modules, data encryption

12. CONTROL OBJECTIVES FOR INFORMATION AND RELATED TECHNOLOGY (COBIT)

COBIT[10] is a Business Framework for the Governance and Management of Enterprise IT for companies that are under the Sarbanes-Oxley Act.

Key features: IT management guide, maturity models

13. NIST SP 800-53

SP 800-53[25] explains Security and Privacy Controls for Federal Information Systems and Organizations. It covers a risk management framework that addresses security control, according to FIPS.

Key features: Risk assessment and management

14. VAULTIVE

Vaultive is a Cloud data encryption standard used with many regulations such as HIPAA-HITECH, GLBA, PCI. The Vaultive compliant data cannot be directly accessed by the US Government without consumers' authorization.

Key features: Cloud data encryption standard

15. SECURITY TECHNICAL IMPLEMENTATION GUIDE(STIG)

STIG[17] provides security guidance throughout the application development lifecycle. It includes development, design, testing, conversions, and upgrades for existing applications, maintenance, software configuration management, education, and training.

Key features: Configuration management, incident response

16. ELECTRONIC DISCOVERY REFERENCE MODEL (EDRM) -PRIVACY AND SECURITY RISK REDUCTION MODEL (PSRRM)

PSRRM [29] is a process for reducing the volume of private, protected, and risky data by using a series of steps applied in sequence. The steps are to Define Risk, Identify available data, Create Filters, Run filters, Verify the output, and Quarantine.

Key features: Privacy, electronic discover

17. SARBANES-OXLEY ACT (SOX)

SOX [30] is an act to protect investors by improving the accuracy and reliability of corporate disclosures made according to the securities laws, and for other purposes.

Key features: IT security management

18. ISO 17799

ISO/IEC 17799[32] is a code of practice. It contains guidelines for information security management. It is meant to provide a high level, general description of the areas currently considered necessary when initiating, implementing, or maintaining information security in an organization.

Key features: IT security management, incident response, compliance

19. DISTRIBUTED MANAGEMENT TASK FORCE (DMTF) - CLOUD AUDITING DATA FEDERATION (CADF) AND OVF(OPEN VIRTUALIZATION FORMAT)

The CADF[33] Working Group determined to develop and publish granular use cases around Cloud auditing and data federation that will be used as input for the development of their data format and interface specification.

OVF [13] provides an open, secure, portable, and efficient standard for virtual applications. This standard does not depend on any hypervisors.

Key features: audit management, Virtualization standards

20. NIST 800-16

NIST 800-16[34] is a Role-Based Model for Federal Information Technology/ Cyber Security Training. Its primary focus is to develop a methodology for cybersecurity training.

Key features: security awareness and training

21. NIST 800-50

The purpose of this security standard[35] is to build awareness and training programs in the IT security system.

Key features: security awareness and training.

22. MOTION PICTURE ASSOCIATION OF AMERICA (MPAA) COMPLIANCE

The MPAA[36] protects the right for those who create entertainment content like creative arts in the software industry. The main objective of MPAA is to protect the pre-release content and report piracy.

Key features: Media protection, anti-piracy

23. ORGANIZATION FOR THE ADVANCEMENT OF STRUCTURED INFORMATION STANDARDS (OASIS) SAML

The Security Assertion Markup Language (SAML) [31] standard defines a framework for exchanging security information between online business partners. It was developed by the Security Services Technical Committee (SSTC) of the standards organization OASIS [31].

Key features: XML-based framework, authentication

24. NIST 800-61 [59]

NIST 800-61 standard is used for Incident handling. This compliance model is beneficial for understanding incident response.

Key features: incident response

25. NIST 800-63 [53]

NIST 800-63 standard is used as the guidelines for electronic authentications. The levels of authentications are explained in the document to ensure that the user is authenticated correctly.

Key features: electronic authentication

26. NIST 800-100 [69]

It is the handbook of information security to assist the managers in the implementation of information security in the organization.

Key features: Information security management

27. OAUTH [59]

Oauth is an open standard for authorization. Oauth 2.0 framework enables the third party to obtain access over HTTP service from the resource owner.

Key features: Authorization

ACKNOWLEDGMENT

This research was supported by a DoD supplement to the NSF award# 1747724, Phase I IUCRC UMBC: Center for Accelerated Real-time Analytics (CARTA).

REFERENCES

[1] Karuna Pande Joshi, Yelena Yesha, and Tim Finin, "Automating Cloud Services Lifecycle through Semantic technologies", IEEE Computer Society Press, pp.109-122, Jan 2014.

[2] Alliance, C. S. "Cloud Security Alliance Warns Providers of' The Notorious Nine' Cloud Computing Top Threats in 2013." Top Threats Working Group The Notorious Nine Cloud Computing Top Threats in 2013 (2013): 8.

[3] Amazon Web Services,Nov 2013, Security at Scale: Governance in AWS, Analysis of AWS features that can alleviate on-premise challenges, p6-8

[4] Rackspace,Rackspace security management, https://www.rackspace.com/security/management/,

[5] Google,June 6 2012, Google Apps to offer additional compliance options for EU data protection, http://googleenterprise.blogspot.com/2012/06/google-apps-to-offer-additional.html

[6] Fang Liu; Jin Tong; Jian Mao; Robert B. Bohn; John V. Messina; Mark L. Badger; Dawn M," NIST, sept 2011, NIST Cloud Computing Reference Architecture", Leaf,p 15-16

[7] Watkins, Steve G. "An Introduction to Information Security and ISO 27001: 2013 A Pocket Guide. It Governance Ltd, 2013".

[8] SSAE16, dec 14, 2013, The SSAE16 Auditing Standard, http://www.ssae-16.com/

[9] FedRAMP, About FedRAMP http://www.gsa.gov/portal/category/102375

[10] D. Rusu, L. Dali, B. Fortuna, M. Grobelnik, and D. Mladenic, "Triplet extraction from sentences," in Proceedings of the 10th International Multiconference" Information Society-IS, 2007, pp. 8– 12.

[11] Etzioni, O., Cafarella, M., Downey, D., Popescu, A. M., Shaked, T., Soderland, S., …& Yates, A. (2005). Unsupervised named-entity extraction from the web: An experimental study. Artificial intelligence, 165(1), 91-134.

[12] Amit Hendre, Tim Finin, and Karuna Pande Joshi, "Cloud Security and ComplianceOntology", July 2014, http://ebiq.org/r/361/

[13] DMTF, OVF. "Specification V1. 1. 0." (2010).

[14] Metsch, Thijs, and Andy Edmonds. "Open Cloud Computing Interface–Infrastructure,"." Standards Track, no. GFD-R in The Open Grid Forum Document Series, Open Cloud Computing Interface (OCCI) Working Group, Muncie (IN). 2010.

[15] CSA, Nov 14 2014, CSA Security Guidance v3

[16] Mell, P. & Grance, t. (2011) "The NIST Definition of Cloud Computing (Special Publication 800-145)". Gaithersburg MD: National Institute of Standards and Technology.

[17] K. P. Joshi and C. Pearce, "Automating Cloud Service Level Agreements Using Semantic Technologies," 2015 IEEE International Conference on Cloud Engineering, Tempe, AZ,2015, pp. 416-421, doi: 10.1109/IC2E.2015.63

[18] STIG, Apr 25 2014, Application Security and Development STIG, V3R7,

[19] L. Kagal and T. Finin, Agent Communication: International Workshop on Agent Communication, AC 2004, New York, NY, USA, July 19, 2004, Revised Selected and Invited Papers. Springer Berlin Heidelberg, 2005, ch. "Modeling Communicative Behavior Using Permissions and Obligations"

[20] ISO 27002, Introduction to ISO 27002, http://www.standards.bz/iso-27002.html

[21] ISO 27001, ISO/IEC 27001 - Information security management, http://www.iso.org/iso/home/standards/management-standards/iso27001.htm

[22] Kagal, L. & Finin, T., "Modeling conversation policies using permissions and obligations," Auton Agent Multi-Agent Syst (2007) 14: 187. doi:10.1007/s10458-006-0013-z

[23] Content protection and Security (CPS), CDSA, http://www.cdsaonline.org/content-protection-and-security-standards-and-procedures/

[24] HIPAA, may 2003,Summary of HIPAA privacy rules. http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacysummary.pdf

[25] NIST SP 800-53, Information security, http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final_updated-errata_05-01-2010.pdf

[26] ITAR, Licence for export of defense articles, http://www.pmddtc.state.gov/regulations_laws/documents/official_itar/2013/ITAR_Part_123.pdf

[27] FIPS 140-2, Jan 2011, security requirement for cryptographic models, http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf

[28] NIST, May 2010, "Recommended Security Controls for Federal Information Systems and Organizations", http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-53r4.pdf

[29] EDRM-PSRRM, "Privacy and Security risk reduction model", http://www.edrm.net/resources/psrrm

[30] SOX, Sarbanes-Oxley Act, http://en.wikipedia.org/wiki/Sarbanes%E2%80%93Oxley_Act

[31] OASIS SAML, Overview of OASIS SAML v2, https://www.oasis-open.org/committees/download.php/27819/sstc-saml-tech-overview-2.0-cd-02.pdf

[32] ISO 17799, International Standard ISO/IEC 17799:2000 Code of Practice for Information Security Managementhttp://csrc.nist.gov/publications/secpubs/otherpubs/reviso-faq.pdf

[33] DMTF CADF, June 2012, Cloud Auditing Data Federation http://www.dmtf.org/sites/default/files/standards/documents/DSP2028_1.0.0a.pdf

[34] NIST 800-16, for security training, oct 2013, http://csrc.nist.gov/publications/drafts/800-16-rev1/draft_sp800_16_rev1_2nd-draft.pdf

[35] NIST 800-50, for awareness and training,oct 2003, program,,http://csrc.nist.gov/publications/nistpubs/800-50/NIST-SP800-50.pdf

[36] MPAA, 2012, entertainment content security and protection, http://www.fightfilmtheft.org/docs/2012_Annual_Trending_Report_Final.pdf

[37] Spamina, List of Cloud providers as of 2014, http://www.spamina.com/eng/Cloud_hosting_providers_list.php

[38] Nagar, Ankur, and Karuna Pande Joshi. “A Semantically Rich Knowledge Representation of PCI DSS for Cloud Services.” 6th International IBM Cloud Academy Conference ICACON 2018, Japan. 2018.

[39] https://www.pcisecuritystandards.org/documents/pci_dss_v2.pdf

[40] https://csrc.nist.gov/glossary/term/Key_Encryption_Key

[41] Tim Mather, Subra Kumaraswamy, Shahed Latif, O’Reilly Media, Inc,, “An Enterprise Perspective on Risks and Compliance”, Cloud Security and Privacy: 2009

[42] Abdul Raouf Khan, “Access control in Cloud computing environment”, VOL. 7, NO. 5, MAY 2012, ARPN Journal of Engineering and Applied Sciences

[43] Privacy.microsoft.com. (2018). Change history for Microsoft Privacy Statement – Microsoft privacy. [online] Available at: https://privacy.microsoft.com/en-us/updates [Accessed 17 Aug. 2018].

[44] Elluri, Lavanya, Ankur Nagar, and Karuna Pande Joshi. “An Integrated Knowledge Graph to Automate GDPR and PCI DSS Compliance.” 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018.

[45] Semantic web definition, http://semanticweb.org/wiki/Semantic_Web

[46] Facebook Business. (2018). General Data Protection Regulation. [online] Available at: https://www.facebook.com/business/gdpr [Accessed 17 Aug. 2018].

[47] Anon, (2018). [ebook] Available at: https://d1.awsstatic.com/legal/aws-gdpr/AWS_GDPR_DPA.pdf [Accessed 17 Aug. 2018].

[48] Privacy.google.com. (2018). Compliance | How Google complies with data protection laws. [online] Available at: https://privacy.google.com/businesses/compliance/#?modal_active=none [Accessed 17 Aug. 2018].

[49] Elluri, Lavanya, and Karuna Pande Joshi. “A knowledge representation of Cloud data controls for EU GDPR compliance.” 2018 IEEE World Congress on Services (SERVICES). IEEE, 2018.

[50] Kandukuri, B.R., V.R. Paturi, and A. Rakshit. 2009. "Cloud security issues."

[51] Cloud Security alliance, https://Cloudsecurityalliance.org/

[52] International Organizations of standards, http: //www. iso.org

[53] NIST 800-63 Guideline, NIST Electronic Authentication. “NIST Special Publication 800-63 Version 1.0. 2.” (2006).

[54] Mell, Peter, and Tim Grance. “The NIST definition of Cloud computing.” (2011).

[55] Krutz, Ronald L., and Russell Dean Vines. “Cloud Computing Security Architecture.” Cloud Security: A Comprehensive Guide to Secure Cloud Computing. Indianapolis, IN: Wiley, 2010. 179-80. Print.

[56] Carlin, Sean, and Kevin Curran. “Cloud computing security.” International Journal of Ambient Computing and Intelligence (IJACI) 3.1 (2011): 14-19.

[57] WhatsApp.com. (2018). WhatsApp Legal Info. [online] Available at: https://www.whatsapp.com/legal/#privacy-policy [Accessed 17 Aug. 2018]

[58] “Security Guidance for Critical Areas of Focus in Cloud Computing V2.1.” Cloud Security Alliance. Cloud Security Alliance, December 2009. Web. 19 Apr 2010. <http://www.Cloudsecurityalliance.org/csaguide.pdf>.

[59] Hardt, Dick. “The OAuth 2.0 authorization framework.” (2012).

[60] NIST, SP. “800-61.” Computer Security Incident Handling Guide (2004): 800-61.

[61] Hendre, Amit, and Karuna Pande Joshi. “A semantic approach to Cloud security and compliance.” 2015 IEEE 8th International Conference on Cloud Computing. IEEE, 2015.

[62] https://developer.ebay.com/cms/files/api_license_2018-10-26.pdf

[63] General Data Protection Regulation (GDPR) – Final text neatly arranged.” General Data Protection Regulation (GDPR), gdpr-info.eu/.

[64] Payment Card Industry (PCI) Data Security Standard, Version 3.2April2016 https://www.pcisecuritystandards.org/document_library

[65] GDPR Resource Center. (n.d.). Retrieved March 07, 2018, from https://gdpr.Cloudsecurityalliance.org/

[66] Additional Resources about the GDPR. (n.d.). Retrieved March 07, 2018, from https://www.eugdpr.org/more-resources-1.html

[67] COBIT, Strategic Planning Using COBIT 5, http://www.isaca.org/Knowledge-Center/cobit/Documents/CF-Vol-2-2014-Strategic-Planning-Using-COBIT-5_nlt_Eng_0414.pdf

[68] DIACAP/FISMA,http://www.securestate.com/Federal/Certification%20and%20%20Accreditation/Pages/DIACAP-D0D8500.aspx

[69] Bowen, Pauline, Joan Hash, and Mark Wilson. “SP 800-100. Information Security Handbook: A Guide for Managers.” (2006).

[70] ApacheJenaFuseki Server, https://jena.apache.org/documentation/fuseki2/Getting started with RDF SPARQL queries and inference using Apache Jena Fuseki, Christine Draper, https://christinemdraper.wordpress.com/2017/04/09/getting-started-with-rdf-sparql-jena-fuseki/

[71] K. Barker and N. Cornacchia, “Using noun phrase heads to extract document keyphrases,” in Advances in Artificial Intelligence. Springer, 2000, pp. 40–52.

[72] T. D. Breaux, M. W. Vail, and A. I. Anton, “Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations,” in RE’06: Proceedings of the 14th IEEE International Requirements Engineering Conference (RE’06), IEEE Society Press, September 2006.

[73] N. Kiyavitskaya, N. Zeni, T. D. Breaux, A. I. Anton, J. Cordy, L. Mich, and J. Mylopoulos, “Automating the extraction of rights and obligations for regulatory compliance,” in ER’08: Proceedings of the 27th International Conference on Conceptual Modeling (ER’08), Springer-V erlag, October 2008.

[74] L. Kagal and T. Finin, Agent Communication: International Workshop on Agent Communication, AC 2004, New York, NY, USA, July 19, 2004, Revised Selected and Invited Papers. Springer Berlin Heidelberg, 2005, ch. Modeling Communicative Behavior Using Permissions and Obligations.

[75] Kagal, L. & Finin, T., “Modeling conversation policies using permissions and obligations,” Auton Agent Multi-Agent Syst (2007) 14: 187. doi:10.1007/s10458-006-0013-z

[76] GDPR and PCI DSS: How They Differ, How They’re Similar and…(2018, July 10). Retrieved from http://paymentsjournal.com/gdpr-and-pci-dss/

[77] Calver, N. (2018). “How the PCI DSS can help you meet the requirements of the GDPR”. [online] IT Governance Blog. Available at: https://www.itgovernance.co.uk/blog/how-the-pci-dss-can-help-you-meet-the-requirements-of-the-gdpr/

[78] Jones, A. and IS Partners, L. (2018). “4 Ways to Use PCI DSS to Achieve GDPR Compliance” | IS Partners. [online] IS Partners. Available at: https://www.ispartnersllc.com/blog/4-ways-to-use-pci-dss-to-achieve-gdpr-compliance/

[79] “General Data Protection Regulation (GDPR) – Final text neatly arranged.” General Data Protection Regulation (GDPR), gdpr-info.eu/.

[80] Cloud Security Alliance Releases Code of Conduct for GDPR Compliance.(n.d.).from https://www.morganlewis.com/blogs/sourcingatmorganlewis/2017/11/Cloud-security-alliance-releases-code-of-conduct-for-gdpr-compliance

[81] Lam, Jessica. “Who is the Data Processor and what are its responsibilities under the GDPR?” Law Infographic, 7 Nov. 2017, www.lawinfographic.com/data-processor-responsibilities-gdpr/

[82] Modal logic: http://plato.stanford.edu/entries/logic-modal/

[83] Michael R. Overly, Legal compliance challenges of Big Data: “Seeing the forest for the trees”, https://www.csoonline.com/article/2883796/big-data-security/legal-compliance-challenges-of-big-data-seeing-the-forest-for-the-trees.html, last retrieved 8/19/2018

[84] Resource description framework (rdf). [Online]. Available: http://www.w3.org/RDF/

[85] “Owl web ontology language.” [Online]. Available: http://www.w3.org/TR/owl-features/

[86] http://purl.org/csc/ontologyfiles

[87] http://purl.org/csc/policydocuments

KARUNA PANDE JOSHI is an Associate Professor of Information Systems at UMBC and UMBC Site Director of Center for Accelerated Real-Time Analytics (CARTA). She also directs the Knowledge Analytics Cognitive and Cloud (KnACC) Lab. Her research focus is in the areas of Data Science, Cloud Computing, Data Security and Privacy, and Healthcare IT systems. She has published over 50 papers, and her research is supported by ONR, NSF, DoD, GE Research, and Cisco. She teaches courses in Big Data, Database Systems Design and Software Engineering. She received her MS and Ph.D. in Computer Science from UMBC, where she was twice awarded the IBM Ph.D. Fellowship, and her bachelor’s in computer engineering from the University of Mumbai, India. Dr. Joshi also has extensive experience of working in the industry primarily as an IT Program/Project Manager at the International Monetary Fund.

LAVANYA ELLURI is currently a third-year Ph.D. student in Information Systems (IS) at the University of Maryland Baltimore County. In parallel, she is working as a Senior Database Engineer at REI Systems, Sterling, VA. She received a master's degree in Management Information Systems at the University of Houston Clear Lake. Her research interests include cloud computing, Semantic Web, and Text Mining.

ANKUR NAGAR worked on the Automated Legal Document Analytics (ALDA) project, where he developed novel techniques for automating rules for Mobile Wallet processing. His research interests lied in the field of Cloud Computing and Financial Services. He successfully defended his Master's thesis in Aug 2019 and is currently working for UBS on Electronic trading business, client hub connectivity.

AMIT HENDRE is currently working as a technical lead in Xoriant Solutions. He received his MS in computer science from University of Maryland Baltimore county. He successfully defended his masters thesis on the subject "Cloud securities and policies comparator" in July 2014. He is currently working on projects for financial and healthcare sectors using cloud technologies like oracle cloud and AWS. His research interest lied in the field of cloud computing.