# Data Regulation Ontology

Guillaume Delorme, Guilaine Talens, Eric Disson
Laboratoire Magellan
University Jean Moulin Lyon 3, Iaelyon School of Management
Lyon, France
@univ-lyon3.fr

Guillaume Delorme
Group Security
Solvay
Lyon, France
@solvay.com

***Abstract*—*The recent upsurge enactment of regulations seeking to regulate data processing induces a complexification of compliance management for regulated firms. Firms wishing to implement efficient, cost effective and compliant information security and risk management require an increased comprehension of regulatory requirements. Following a previous paper defining Data Regulation Risk, this paper describes an ontology to apprehend the business and operational impacts of regulatory requirements. The ontology is structured to handle various firms' legal context while remaining agnostic of risk management methodologies.*

***Keywords-Ontology; Compliance; Knowledge-based; Privacy***

## I. INTRODUCTION

Over the past decades, the upsurge enactment of regulations seeking to reinforce the protection of individuals' rights and privacy, economic interests and national security has led to the appearance of a new class of risk called Data Regulation Risk (DRR) [1]. We defined Data Regulation as a norm governing data processing and/or ICT governances and processes and/or information technologies and services. Despite addressing similar concepts, such regulations are yet often demanding divergent or particular controls. This article aims to furnish the necessary information to facilitate DR management.

Several authors pointed out the need for ontology in the security domain [2, 3]. Similarly, several conceptualizations of the legal domain have been presented and studies and comparison of legal ontologies can also be found [4, 5]. Despite important contributions, there is a need for methodologies and models to identify multi-disciplinary risks like DR management. We seek to address DR by building an ontology which facilitates its management.

Ontologies are designed to facilitate the sharing, use and re-use of knowledge [6]. Defined as explicit conceptualization of a domain [7], they enable its modulization with the desired level of abstraction depending on the initial objective. We develop an ontology following the Enterprise Model Approach [8] with the ambition of facilitating the apprehension of business and operational impacts of regulatory requirements. It focuses on regulatory controls while leaving the option of mapping the controls with additional threats for a broader or multi-disciplinary risk management. To reach our target, we use to the extent possible the terminologies of the WordNet database developed by Princeton University [9] as well as concepts present in existing ontologies.

This contribution is structured as follows. Section 2 discusses the core ontology and its purpose. Section 3 presents the building and usability of the ontology. Finally, section 4 draws conclusions and discusses some further research directions.

## II. THE CORE ONTOLOGY ARCHITECTURE AND ITS KEY CONCEPTS

The creation of an ontology requires to determine what entities should be considered and studied.

### *A. Methodology*

With the ambition of easing methodology building, [6] surveyed different ontology building methodologies such as TOVE [10], Enterprise Model Approach [8], Methontology [11] and Ontolingua [7]. Recent methodologies have been developed to focus on specific needs such as [12]. As no methodology seems to stand out and all of them have their pros and cons [5], we decided to adopt the Enterprise Model Approach which is a stage-based approach, widely spread, providing sufficient freedom of representation [13]. It is appropriate to a cross disciplinary ontology such as ours and is articulated around four main stages : identify purpose, building the ontology, evaluation and documentation. The second stage incorporates the ontology capture, ontology coding and the integration of existing ontologies [8].

As opposed to the classic bottom-up and top-down approach to identify the main terms of our ontology, we opt for the middle out approach presented in [8]. This approach allows one to identify the primary concepts of the ontology before moving on to specialize or generalize terms [11]. The middle out approach implicitly leads to more stable concepts. In regards to clarity, which is the foundation of the usability and reusability of an ontology, we need a world known, easily accessible, proven and accepted terminology database. Suggested Upper Merged Ontology (SUMO) [14] is a formal public ontology providing definitions for general purpose terms and is intended as a unifying framework for more specific domain level ontologies. As SUMO is designed as an upper ontology, it provides generic terms and therefore fails to address the needs of more specific domains ontologies [15]. We then decide to use when possible the terminologies of the WordNet database developed by Princeton University [9]. WordNet "is a large lexical database of nouns, verbs, adjectives and adverbs grouped into sets of cognitive synonyms (Synsets), each expressing a distinct concept. Synsets are interlinked by

DOI reference number: 10.18293/SEKE2022-024

means of conceptual-semantic and lexical relations." Each concept, relation or attribute in our ontology is mapped with a unique Synset using the Synset ID.

### B. Purpose

Defining an ontology purpose and what its intended uses are, is the fundamental step towards developing one [8], [11].

Our approach is an attempt to design a system capable of representing the various legal modalities (ought, ought not, may, or can) and delivering pragmatic information for the users based on generalist and sometimes abstract body of laws. It must then by default be designed to integrate the fast evolution of the regulations, the divergent or particular controls as well as being able to focus on a firm specific information systems' environment. Finally, this system must furnish the necessary information to apprehend the business and operational impacts of regulatory requirements. Our ontology does not seek to assess the effective compliance nor the threat landscape of a company. Our work is solely to express the requirements and constraints based on the deontic models of the laws.

The complexity of DR management resides in the necessity of translating the regulatory constraints and requirements into technical, organizational and operational terms. Not to mention that DR is context specific and depends on one organization's markets, geographical presence and jurisdictions, it therefore requires an in-depth analysis involving a broad set of skills fragmented across the organization's departments. We then identified three main types of users which are: IT managers, security practitioners and compliance managers. All three of them require different pieces of information extracted from the laws in order to perform their duties while ensuring business continuity and their company compliance. For example, the IT manager will need the deontic modalities and regulatory requirements to build and manage the overall information systems while the security practitioners will focus on the mandatory security controls that need to be implemented.

## III. ONTOLOGY BUILDING

### A. Reused Ontology

During our search we were able to distinguish two main areas of work related to ours.

#### 1) Information Security Management Ontologies

As show in [16], security ontologies can be sorted by: general security ontology, security ontology applied to a specific domain and theoretical work. This work was later reused by [17] who extended the classification to eight categories, namely: beginning security ontologies, security taxonomies, general security ontologies, specific security ontologies, web oriented security ontologies, risk based security ontologies, ontologies for security requirements and security modeling ontologies. They reached the conclusion that the existing security ontologies vary a lot and no ontology covers all of the aspects of the security domain.

A strong basis for information security domain knowledge may be found in [18]. Their Information Security Ontology is composed of three sub-ontologies (security, enterprise and location) and is based on established documentation, industry best practices and controls. In their previous work, [19] also proposed a security ontology as a basis for a low cost risk management solution as well as an ontology focusing on threats to corporate assets. The ontology consists of five sub-ontologies (threat, attribute, infrastructure, role and person). Other works introduce ontologies specific to vulnerability analysis and management [20], risk assessment [3], security annotations of agents and web services [21], dependability requirements that include security [22], secure development [23]. Despite the variety of domain specific ontologies in the different branches of information security, they tend to apply to only very limited scope which prevent us from reusing most of them. We will nonetheless reuse the role and person concepts found in [18] as much as possible.

#### 2) Compliance & Legal Ontologies

Several conceptualizations of the legal domain have been presented or studied and comparison of legal ontologies can also be found [4, 5]. For instance, the McCarty's Language for Legal Discourse [24] is semi-formal conceptualization with the ambition of creating a general language for legal domain knowledge. By dividing the domain in three: norm, act and concept description, the issue of reusability of legal ontologies is presented in [25]. The three concepts are designed to be sufficient to conceptualize the subdomains of the legal domain.

There are also ontologies focusing on a single regulation or a type of regulation such as privacy ontologies. For instance, PrivOnto [26] is a semantic framework to represent annotated privacy policies and provide a linguistic instrument for the privacy domain. Another example is GDPRtEXT [27] which is a list of concepts present in the General Data Protection Regulation (GDPR). Its goal is to provide a way to refer to the concepts and terms found in the GDPR without providing an interpretation of compliance obligations. Similarly, the privacy ontology PrOnto [28], models the GDPR main conceptual cores "to support legal reasoning and compliance checking by employing defeasible logic theory". Similarly to the Frame-Based Ontology, PrOnto manages to model norms through its conceptualization of deontic operators. We will reuse and follow as much as possible these design patterns for our ontology.

Finally, LKIF [15] is a legal core ontology presented as a knowledge representation formalism that enables the translation between different legal bases. Comparably to the role and person concepts found in [26], LKIF presents the organization, role and person concepts which we will be reusing.

### B. Ontology Capture

The preceding sections presented the requirements for our ontology and some concepts we reuse from existing ontologies.

#### 1) Key Concepts

Capturing our ontology implies the findings of precise unambiguous text definitions and terms' identification for the different concepts and relationships [8]. We group the top level concepts of our ontology in four subontologies: enterprise, security, legal and location.

We reuse the top concepts Individual and Role from [15,18]. The concept Individual (Synset ID: 100007846), (ent: Individual ⊑ ⊤) is used to represent an identifiable natural person. The concept Role (Synset ID: 100722061), (ent: Role ⊑ ⊤) and its corresponding subconcepts are used to represent the normal or customary activity of a person in a particular social setting. Every individual has one or more roles which enables a flexible handling of the concepts in complex scenarios.

The creation of the subontologies enterprise, security and location is derived from [18]. While the whole subontologies do not fit the needs of ours, we reuse and adapt their concepts Control, Asset, Organization, Data and Location to create respectively Security_Measure (Synset ID: 100823316), (sec: Security_Measure ⊑ ⊤), Information_System (Synset ID: 103164344), (ent: Information_System ⊑ ⊤), Legal_Entity (Synset ID: 100001740), (ent: Legal_Entity ⊑ ⊤), Technological_Data (Synset ID: 105816622), (ent: Technological_Data ⊑ ⊤), Country (ent: Loc: Country⊑ ⊤), (Synset ID: 108544813) and Citizenship (loc: Citizenship ⊑ ⊤) (Synset ID: 113953467).

The concept Technological_Data and its corresponding subconcepts are used to represent data in digital format. For this ontology, we model the subconcepts: Business_Data and Personnal_Data. The former (ent: Business_Data ⊑ Technological_Data) corresponds to data involved in the course of conduct of activities of a Legal_Entity while the latter (ent: Personnal_Data ⊑ Technological_Data) are any information relating to an identified or identifiable natural person.

The concept Legal_Entity and its corresponding subconcepts represent a natural or legal person, a public authority body which carries out an activity whatever its legal form. The following subconcepts modeled so far are: Business_Organization, Independant_Organization and Regulatory_Agency.

We use the concept Information_System to describe an organized set of resources (hardware, software, individual, data and procedure) which makes it possible to process data.

Accordingly, we create the concept IT_System (Synset ID: 104377057), (ent: IT_System ⊑ ⊤) to represent a combination of interacting elements (resources) organized to achieve one or more desired objectives. We introduce this concept to provide an agile ontological structure according to the granularity of regulations. To illustrate various data processing, we add the concept Action (Synset ID: 100037396), (ent: Action ⊑ ⊤) and its corresponding subconcepts to represent something done (i.e. action or processing of data).

We then need to create the concept Functionnal_Process (SynSet ID: 101023820), (ent: Functionnal_Process ⊑ ⊤) to describe a set of interrelated or interacting activities that uses inputs to produce an intended result. As an example, an instance of a Functionnal_Process would be the payroll process within an organization.

Security measures are usually gathered within different classes of documents. We then create the concept Documentation (Synset ID: 106588326), (sec: Documentation ⊑ ⊤) to represent the set of documents such as policies, guidelines, procedures or frameworks. The concept is also useful to illustrate external documentation such as standards and frameworks which are often cited in regulations.

We need the concept Norm (Synset ID: 106532330), (leg: Norm ⊑ ⊤) to describe texts of laws. To show an action that is governed by a regulation through legal modalities, we will use the concept Act (Synset ID: 100030358), (leg: Act ⊑ ⊤) as introduced by [23]. Accordingly, a Norm governs an Act which itself governs Individual, Technological_Data, Legal_Entity and Security_Measure.

*2) Key Relationships*

Our next task focuses on determining the relationships between the concepts. Our model consists of two types of relationships: characteristic relationships which are used to represent the links between the different concepts of the model and action relationships when a concept performs a direct action on another concept. Our model is composed of 11 characteristic relationships (govern, has_a, isLocatedIn, belong, involve, protect, define, manage, isOwnedBy, isComposedOf and create) and 3 action relationships (process, isUsedBy, perform).

### *C. Formalization of the U.S. Export Arm Regulations*

To illustrate the different primary concepts of our model, we will formalize parts of the Export Arm Regulations [29]: EARNorm is_a Norm. EAR Supplement No. 18 to part 734 states the following:

*Transmitting or otherwise transferring "technology" or "software" to a person in the United States who is not a foreign person from another person in the United States.*

Using the concept Act, Supplement No. 18 to part 734 is therefore represented as: EAR734.18Act is_a Act. Using the govern relation: EARNorm governs EAR734.18Act. Data regulated by the EAR Supplement No. 18 to part 734 then correspond to: EARBusiness_Data is_a (Business_Data ⊑ Technological_Data). We then need to create a first person using the concept Individual: PersonReceiveingEARData is_a Individual. Then, this individual must be physically located in the United States: US is_a Country and be an US citizen: USCitizenship is_a Citizenship. PersonReceiveingEARData isLocatedIn US and has_a USCitizenship. We can proceed to create our second individual residing in the US who is the sender of the data: PersonSendingEARData is_a Individual and isLocatedIn US.

Translating the term Release into practical terms would result in granting or receiving access to EAR controlled data. To encapsulate this, the concept Action will be used to represent the transfer of controlled data: TransferEARData is_a (Transfer ⊑ Action). To add an extra layer of granularity, we can come up with additional subconcepts such as granting access and its reverse, receiving access: GrantAccessEARData is_a (GrantAccess ⊑ Transfer) and ReceiveAccessEARData is_a (ReceiveAccess ⊑ Transfer). In the end, transmitting or otherwise transferring would be: An individual that uses an

EARSystem is_a IT_System to perform the action TransferEARData that process EARBusiness_Data.

EAR734.18Act ⊑ governs ((PersonReceiveingEARData ⊓ isLocatedIn.US ⊓ has_a.UScitizenship) and (EARSystem ⊓ perform.ReceiveAccessEARData ⊓ process.EARBusiness_Data))

EAR734.18Act ⊑ governs ((PersonSendingEARData ⊓ isLocatedIn.US) and (EARSystem ⊓ perform.GrantAccessEARData ⊓ process.EARBusiness_Data))

## IV. Conclusion and Further Research Direction

Based on various data sources such as established documentation or industry best practices, existing ontologies [15, 18] and regulations, we present an ontology able to formalize firms' legal context while enabling the sharing and reuse of knowledge to support decision making. We present an ontology with 14 top level concepts grouped in four subontologies (enterprise, security, legal and location) and 14 relationships. With the ambition of facilitating the apprehension of business and operational impacts of regulatory requirements, our ontology is designed for any type of firm. We are currently developing the ontology using Protégé and implementing it at a worldwide chemical company subject multiple regulations.

We also plan to integrate further existing information security and risk management ontologies. We believe that combining them will enable more efficient risk management by combining regulatory risk and information security risk.

## References


[1] Delorme, G., Talens, G., Disson, E., Collard, G., & Gaget, E. (2020, December). On the Definition of Data Regulation Risk. In International Conference on Service-Oriented Computing (pp. 433-443). Springer, Cham.

[2] Donner, M. (2003). Toward a security ontology. IEEE Security & Privacy, 1(03), 6-7.

[3] Tsoumas, B., & Gritzalis, D. (2006, April). Towards an ontology-based security management. In 20th International Conference on Advanced Information Networking and Applications-Volume 1 (AINA'06) (Vol. 1, pp. 985-992). IEEE.

[4] Larmande, P., Arnaud, E., Mougenot, I., Jonquet, C., Rouge, T. L., & Ruiz, M. (2013, May). Proceedings of the 1st International Workshop on Semantics for Biodiversity. In 1. International Workshop on Semantics for Biodiversity (pp. 001-131).

[5] Visser, P. R., & Bench-Capon, T. J. (1998). A comparison of four ontologies for the design of legal knowledge systems. Artificial Intelligence and Law, 6(1), 27-57.

[6] Jones, D., Bench-Capon, T., & Visser, P. (1998). Methodologies for ontology development.

[7] Gruber, T. R. (1992). Ontolingua: A mechanism to support portable ontologies.

[8] Uschold, M., & King, M. (1995). Towards a methodology for building ontologies (pp. 1-13). Edinburgh: Artificial Intelligence Applications Institute, University of Edinburgh.

[9] WordNet, https://wordnet.princeton.edu/ last accessed 2022/02/20.

[10] Fox, M.S., Chionglo, J., Fadel, F. A Common-Sense Model of the Enterprise, Proceedings of the Industrial Engineering Research Conference 1993

[11] Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering.

[12] Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., & García-Castro, R. (2022). LOT: An industrial oriented ontology engineering framework. Engineering Applications of Artificial Intelligence.

[13] Pinto, H. S., & Martins, J. P. (2004). Ontologies: How can they be built?. Knowledge and information systems, 6(4), 441-464.

[14] Niles, I., & Pease, A. (2003). Mapping WordNet to the SUMO ontology. In Proceedings of the ieee international knowledge engineering conference (pp. 23-26).

[15] Alexander, B. O. E. R. (2009). LKIF core: Principled ontology development for the legal domain. Law, ontologies and the semantic web: channelling the legal information flood, 188, 21.

[16] Blanco, C., Lasheras, J., Valencia-García, R., Fernández-Medina, E., Toval, A., & Piattini, M. (2008, March). A systematic review and comparison of security ontologies. In 2008 Third International Conference on Availability, Reliability and Security (pp. 813-820). Ieee.

[17] Souag, A., Salinesi, C., & Comyn-Wattiau, I. (2012, June). Ontologies for security requirements: A literature survey and classification. In International conference on advanced information systems engineering (pp. 61-69). Springer, Berlin, Heidelberg.

[18] Fenz, S., & Ekelhart, A. (2009, March). Formalizing information security knowledge. In Proceedings of the 4th international Symposium on information, Computer, and Communications Security (pp. 183-194).

[19] Ekelhart, A., Fenz, S., Klemen, M. D., & Weippl, E. R. (2006, December). Security ontology: Simulating threats to corporate assets. In International Conference on Information Systems Security (pp. 249-259). Springer, Berlin, Heidelberg.

[20] Wang, J. A., & Guo, M. (2009, April). OVM: an ontology for vulnerability management. In Proceedings of the 5th Annual Workshop on Cyber Security and Information Intelligence Research: Cyber Security and Information Intelligence Challenges and Strategies (pp. 1-4).

[21] Denker, G., Kagal, L., Finin, T., Paolucci, M., & Sycara, K. (2003, October). Security for daml web services: Annotation and matchmaking. In International Semantic Web Conference (pp. 335-350). Springer, Berlin, Heidelberg.

[22] Dobson, G., & Sawyer, P. (2006, November). Revisiting ontology-based requirements engineering in the age of the semantic web. In Proceedings of the International Seminar on Dependable Requirements Engineering of Computerised Systems at NPPs (pp. 27-29).

[23] Karyda, M., Balopoulos, T., Dritsas, S., Gymnopoulos, L., Kokolakis, S., Lambrinoudakis, C., & Gritzalis, S. (2006, April). An ontology for secure e-government applications. In First International Conference on Availability, Reliability and Security (ARES'06) (pp. 5-pp). IEEE.

[24] McCarty, L. T. (1989, May). A language for legal discourse i. basic features. In Proceedings of the 2nd international conference on Artificial intelligence and law (pp. 180-189).

[25] Van Kralingen, R. (1997, June). A conceptual frame-based ontology for the law. In Proceedings of the first international workshop on legal ontologies (pp. 6-17).

[26] Oltramari, A., Piraviperumal, D., Schaub, F., et al., (2018). PrivOnto: A semantic framework for the analysis of privacy policies. Semantic Web, 9(2), 185-203.

[27] Pandit, H. J., Fatema, K., O'Sullivan, D., & Lewis, D. (2018, June). GDPRtEXT-GDPR as a linked data resource. In European Semantic Web Conference (pp. 481-495). Springer, Cham.

[28] Palmirani, M., Martoni, M., Rossi, A., Bartolini, C., & Robaldo, L. (2018, October). Pronto: Privacy ontology for legal compliance. In Proc. 18th Eur. Conf. Digital Government (ECDG) (pp. 142-151).

[29] Export Administration Regulation (EAR), 15 C.F.R. § 730 et seq, https://www.bis.doc.gov/index.php/regulations/export-administration-regulations-ear, last accessed 2022/02/20.