

## Article

# An SQWRL-Based Method for Assessing Regulatory Compliance in the Pharmaceutical Industry

Efthymios N. Lallas <sup>1,\*</sup>, Ilias Santouridis <sup>2</sup>, Georgios Mountzouris <sup>1</sup> , Vassilis C. Gerogiannis <sup>1</sup>   
and Anthony Karageorgos <sup>1</sup>

<sup>1</sup> School of Technology, University of Thessaly, 41500 Larissa, Greece

<sup>2</sup> School of Economics and Business Administration, University of Thessaly, 41500 Larissa, Greece

\* Correspondence: elallas@uth.gr

**Abstract:** Nowadays, data integrity has become a critical issue in the pharmaceutical regulatory landscape, one that requires data to be compliant to ALCOA principles (i.e., data must be Attributable, Legible, Contemporaneous, Original, and Accurate). In this paper, we propose a method which exploits semantic web technologies to represent pharma manufacturing data in a unified manner and evaluate in a systematic manner their ALCOA compliance. To this purpose, in the context of a pharma manufacturing environment, a data integrity ontology (DIOnt) is proposed to be utilized as the basis for the semantic representation of pharma production data and the associated regulatory compliance management processes. We further show that semantic annotations can be used to represent the required ALCOA compliance information, and that semantic reasoning combined with SQWRL queries can be used to evaluate ALCOA compliance. The proposed approach has been implemented in a proof-of-concept prototype and validated with real world pharma manufacturing data, supporting the combined execution of SWRL rules and SQWRL queries with the aim to support the ALCOA compliance assessment and calculate non-compliance percentages per each ALCOA principle.



**Citation:** Lallas, E.N.; Santouridis, I.; Mountzouris, G.; Gerogiannis, V.C.; Karageorgos, A. An SQWRL-Based Method for Assessing Regulatory Compliance in the Pharmaceutical Industry. *Appl. Sci.* **2022**, *12*, 10923. <https://doi.org/10.3390/app122110923>

Received: 13 September 2022

Accepted: 25 October 2022

Published: 28 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** ALCOA; data integrity; regulatory compliance; ontology technology; pharmaceutical industry; SWRL; SQWRL

## 1. Introduction

The pharma industry operates in a condition of unlimited need for the constant innovation and speedy delivery of efficient, safe, and quality pharmaceutical products. Hence, it comes as no surprise that the pharmaceutical sector is heavily regulated. The pharma industry is under the constant and vigorous scrutiny of organizations that enforce regulatory compliance, such as the European Medicines Agency (EMA) and the US Food and Drug Administration (FDA). These organizations establish regulations for pharmaceutical industries, with the overall aim to protect public health. In pharma manufacturing, regulatory compliance refers to the capability of a pharma manufacturing plant's management system to define and comply with a set of regulations, emanating from pharma production processes. Unfortunately, inappropriate practices that lead to regulatory violations are not rare in the pharma industry. The number of notices of concern, statements of non-compliance, and warning letters issued by the World Health Organization (WHO), EMA, and FDA have shown an alarming increase in recent years [1].

According to the WHO, in the pharmaceutical industry context, data integrity is “the degree to which data are complete, consistent, accurate, trustworthy and reliable and these characteristics of the data are maintained throughout the data life cycle. The data should be collected and maintained in a secure manner, such that they are attributable, legible, contemporaneously recorded, original or a true copy and accurate.” [2]. Therefore, when actions take place, either deliberately or accidentally, which result in data not bearing the aforementioned characteristics,

data integrity is considered to be compromised. In 2003, the FDA assembled the compliance requirements for data integrity in 21 CFR Part 11 by devising the acronym ALCOA, which means that data must be Attributable, Legible, Contemporaneous, Original, and Accurate. ALCOA was later extended to ALCOA+ by adding the Complete, Consistent, Enduring, and Available principles. It is common in the relevant literature to use the acronym ALCOA to include all the principles from both ALCOA's initial version and ALCOA+, and this approach is also followed in the current study. Apart from the FDA, ALCOA has also been endorsed by the European Commission in cGMP Annex 11.

There have been many cases relevant to problematic regulatory compliance of the pharma industry. Indicatively, 21 out of 28 warning letters issued by the FDA to pharmaceutical companies between January 2015 and May 2016 were relevant to data integrity violations [1]. Another report indicates that the increase of FDA warning letters pertinent to data integrity has been exponential and that the corresponding number quintupled from 2014 to 2017 [3].

The work presented in this study is part of the SPuMoNI R&D project, which is funded by the European Union's Horizon Europe Framework Programme for Research and Innovation. The SPuMoNI project has been developed for providing data authenticity, data quality, compliance, and audibility within a pharma manufacturing environment [4], via adoption of ontologies, commonly utilized for pharma conceptual modeling [5]. The SPuMoNI technology is offering means to build an explicit ontology model of the pharma production processes domain, which is data storage independent, is both machine and human readable, and supports knowledge management as well as inferential reasoning. The aim of this study is the provision of a proof-of-concept prototype, based on ontology conceptual design, which takes advantage of a combined utilization of generic semantic web rule language (SWRL) rules with semantic query-enhanced web rule language (SQWRL) queries for performing ALCOA assessments, and thus to support the corresponding required regulatory compliance.

This paper is structured as follows. Section 2 includes a review of relevant work by firstly discussing compliance management and regulatory technology issues and then focusing on the utilization of ontology technology in regulatory technology solutions in the pharmaceutical industry. Section 3 describes in brief the data integrity ontology. It starts by providing an overview of the SPuMoNI system architecture and a short reference of the ALCOA data integrity principles, followed by a brief description of the ontology structure. Section 3 closes with the description of a proof-of-concept implementation of the proposed ontology. Section 4 is dedicated to the description of generic SWRL rules, combined with SQWRL queries, required for the accurate provision of assessments for each of the ALCOA principles. Section 5 provides a critical discussion of the work presented in this paper as well as the challenges of adapting a semantic approach to data regulation compliance management in the pharmaceutical manufacturing domain. Finally, Section 6 presents the concluding remarks of the paper.

## 2. Literature Review

### 2.1. Compliance Management and Regulatory Technology

Compliance management processes have gained a constantly greater prominence in most business sectors, including, but not limited to, the pharmaceutical, financial services, information security, manufacturing, construction, software, and environmental sectors. Compliance has been defined by McIntyre (2008) [6], as “A desired outcome, with regard to law and regulations, internal policies and procedures, and commitment to stakeholders that can be consistently achieved through managed investment of time and resources. The compliance management includes the legal and tactical activities in day-to-day business processes.” Compliance processes are commonly administered by the governance, risk, and compliance function of companies [7], and can be employed at the design (pre-execution), run (execution), or auditing (post-execution) of a business process [8].

According to O'Neill (2014) [9], an organization engages in compliance relevant activities with the aim to enforce externally imposed regulations, define and abide to the organization's internal rules and policies and fulfill social requirements stemming from regulatory processes. Moreover, when implemented successfully, compliance management can lead to increased market and investor confidence and business performance [10]. Focusing on the pharma manufacturing industry, regulatory compliance has a dual aim. The first is to ensure that produced medicines are safe and efficient, with the second being the acceleration of the passage of new products through the investigative and regulatory stages of their production [11]. At the same time, it has to be pointed out that compliance management often poses significant challenges to organizations. These result from (i) the large number of stakeholders interested in the associated processes, such as compliance professionals, business partners, customers, and regulatory bodies, and (ii) the frequently changing compliance requirements alongside the variety of their authors and the lack of a shared understanding of the relevant concepts [7,12].

The manual and paper-based management of compliance is often a very difficult, and costly task that can be susceptible to errors due to its complexity and the repeatedly changing regulations in most industries. As a result of these problems, in recent years several academic and commercial information technology systems have been reported in literature, which aim to automate the support for regulatory compliance. On several occasions, such systems have been denoted as belonging to a specific information technology stream, labeled as regulatory technology (RegTech). Butler and O'Brien (2019) investigated RegTech in the financial services sector and provided a definition of the term, which can be extended in other industries as well [13]. They defined RegTech as *"information technology (IT) that (a) helps firms manage regulatory requirements and compliance imperatives by identifying the impacts of regulatory provisions on business models, products and services, functional activities, policies, operational procedures, and controls; (b) enables compliant business systems and data; (c) helps control and manage regulatory, financial and non-financial risks; and (d) performs regulatory compliance reporting."* They also indicated that RegTech can be used (i) to facilitate reporting compliance, (ii) for the better matching of the actual regulatory requirements with their interpretation and implementation within a company, (iii) to provide for the more efficient and faster detection of regulatory breaches, and (iv) for more efficient delivery of the compliance processes.

The complexity of regulatory compliance and the frequently changing requirements pose challenges not only to compliance management but also to RegTech. Butler and O'Brien (2019) [13], and Butler (2017) [14], in their research work on RegTech for financial services, identified two major perils. They named the first one as the *Tower of Babel* problem, to refer to the lack of a common language in the industry and the use of different terms to describe similar objects and processes. They referred to the second problem as the *Translation* problem and this is linked to the differences between the semantics of the regulatory domain, as understood by business professionals and as comprehended and coded by people of information technology. This means that the technology platform on which a RegTech system is built is needed to provide for a common language and semantic consistency. Along this line, Kim et al. (2007) [15] argued that a compliance data model must be evaluative, formal, reusable, and sharable. This is where ontology technology, which is discussed in the next section, comes into play.

## 2.2. Ontologies for Regulatory Technology

Over the past few decades, semantic technology based on ontologies has been extensively utilized in information integration and conceptual modeling applications [5]. A commonly cited brief definition of an ontology states that it is a specification of a conceptualization [16]. A more comprehensive definition of the term has been given by Guarino (1998) [17], where an ontology is *"a logical theory accounting for the intended meaning of a formal vocabulary, i.e., its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological com-*

*mitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models."*

Ontologies support the development of models built on a common vocabulary, thus providing a unification-based umbrella for the conceptual representation of all the relevant data in a particular domain, which is stored in various formats and under different labels in often dispersed data storages. The main building blocks of an ontology are classes, which represent concepts having entities with common characteristics. Classes are linked with relations, which define how classes are related to each other. A subclass is one such relation, which inherits the characteristics of the parent class to its child, thus allowing the formation of class hierarchies, very much like the object-oriented paradigm in software engineering. Individuals of a class (i.e., class objects) are the instances belonging to that particular class. Object properties are binary relations between individuals, while data properties are binary relations between an individual and data values. Finally, axioms are constraints on the class level, which enforce restrictions on the class individuals.

Ontologies reach their full potential when combined with formal logic. Such an approach makes them formal ontologies, which according to Roussey et al. (2011) [18], *"require a clear semantics for the language used to define the concept, clear motivations for the adopted distinctions between concepts as well as strict rules about how to define concepts and relationships"*. This is obtained by using an appropriate formal logic (usually first order logics or description logics), where the meaning of the concept is guaranteed by formal semantics. Formal ontologies can provide the foundations on which information systems managing knowledge can be built. Furthermore, this knowledge can be processed by a reasoner to infer new knowledge and support decision making. Since ontologies augment common understanding between users, domain knowledge managed or inferred with the use of a formal ontology must be consensual, in order that it can be shared and reused across different software systems and by various types of users [19].

Data in the pharmaceutical manufacturing industry are massive, heterogeneous, and complex and hard to effectively process. Moreover, the knowledge inference required by a regulatory technology (RegTech) software for regulatory compliance monitoring, has become a very challenging task. For all these reasons, ontology technology has been adopted in the context of the SPuMoNI project, with particular focus on comprehensively offering:

- *Semantic data representation.* Ontologies can inherently support pharma product and process development due to both their openness and richness of semantics for representing, analyzing, interpreting, and managing large amounts of complex and diverse information and their independence from any software or data storage platform. Furthermore, the utilization of ontology features, such as properties, axioms, and rules, can enable the enhancement of data consistency and also can facilitate reasoning based on semantics;
- *Fast and compact decisions during product and process development.* Supporting the decision-making process in pharmaceutical product development requires a systematic and integrated informatics framework based on formal and explicit modeling of related information. The utilized information models should be intelligible by both humans and software thus providing a common understanding for information sharing. Ontologies and the associated semantic technology support the smart discovery of data, the inference of data relationships, and the extraction of knowledge from enormous sets of raw data in various formats and from various sources.

Despite the possible suitability of ontologies for the support of semantic data representations and modeling, there have not been many research attempts of utilizing ontologies in pharma manufacturing environments so far. Venkatasubramanian et al. demonstrated the use of Purdue ontology for pharmaceutical engineering (POPE), for product and process development in a pharma production environment [20–22]. Overall, POPE is a comprehensive ontology, which fulfills to a great extent its aims, but there was no support of the POPE ontology for data integrity and furthermore, for ALCOA compliance.

Sesen et al. (2010) tackled the regulatory compliance issue by presenting OntoReg ontology [5]. OntoReg consists of a process and a regulatory domain, and presents many similarities, concerning concepts, with the OntoCAPE [19] and POPE ontologies [20–22]. The structure of OntoReg was designed mostly based on concepts from European regulatory legislation. Extracts of this legislation were then applied to the aspirin production process to create some representative case studies, thus demonstrating the capabilities of ontological modeling and analysis of regulatory structures. OntoReg further provided the basis for RegCMatic [23], which is a methodology aiming to automate the extraction of regulatory information from relevant documents and the mapping of them to internal processes. OntoReg has been considered a noteworthy research effort towards the development of a solution offering an automated regulatory compliance validation approach in the pharmaceutical domain.

Compliance management systems based on ontologies have also been proposed in domains other than the pharma manufacturing industry. The most prominent sector in the area of information technology solutions to automate regulatory compliance management is that of financial services, where the acronym RegTech was initially devised [13]. A series of major financial scandals that took place during the past few decades have, to a great extent, been attributed to violations of regulatory compliance. Similar to the pharmaceutical industry, apart from the very complex regulatory environment of financial services, there are also significant cost drivers justifying the increased interest in automated systems. Indicatively, governance, risk, and compliance costs are estimated to account for between 15 and 20% of banks' operational costs, while the cost of regulatory compliance is predicted to reach 10% of financial institutions' revenues [13]. Exemplary RegTech research efforts in the financial sector can be found in [13,14,24–26]. Other domains, for which RegTech ontologies have also been proposed in literature, include, but are not limited to, engineering [27–29], business process management [8,30], business process quality [15,31], computer security [32], as well as certification and accreditation of software intensive systems [33]. Finally, Abdullah et al. [7,12] have proposed a generic conceptualization for a shared understanding of a compliance regulation management process, intended to be applied to various sectors.

### 3. Data Integrity Ontology in SPuMoNI

#### 3.1. SPuMoNI RegTech Architecture

Industrial system validation processes involve specific rules and related guidance documents, thus defining a regulatory compliance framework of a specific industrial environment. The SPuMoNI system architecture has been designed to support dynamic data quality assessment, regulatory compliance and auditability of pharmaceutical manufacturing processes and data. Within this framework, the RegTech of the SPuMoNI system, whose architecture is depicted in Figure 1, supports required regulatory validation processes, meeting dynamically changing regulatory compliance requirements. Since the RegTech architecture has been particularly designed with the aim to check ALCOA data integrity rule compliance, an ALCOA assessment module has been included in the RegTech architecture. This ALCOA assessment module is the core module of RegTech architecture and it is responsible for interpreting ALCOA principle rules, for performing ALCOA compliance evaluation, and for producing reports. This module, as seen in Figure 1, is responsible for extracting heterogeneous raw data from corresponding system sources and running validation checks to yield the final positive or negative decision of ALCOA compliance, or non-compliance, respectively.

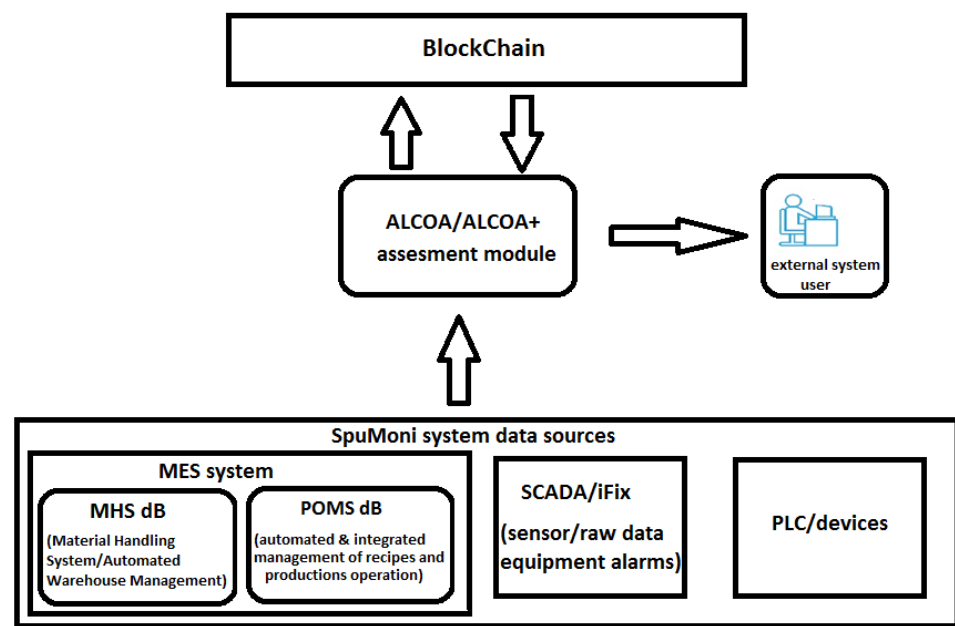


Figure 1. SPuMoNI's RegTech architecture.

### 3.2. ALCOA/ALCOA+ Principles

GxP is an acronym representing “good practice”, guidelines, covering all regulated fields, as denoted by the use of “x”, which set quality standards and enforce regulatory requirements in various industries including the pharmaceutical manufacturing industry. Good manufacturing practices (GMP) is a member of the GxP family of guidelines, aiming to ensure that products are safe and meet their intended use. Regarding the pharma regulatory requirements, compliance to the ALCOA Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, and Available principles reside at the core of GMP.

The *Attributable* principle involves inspection of whether data are associated with the person responsible for data collection. *Legible* refers to the requirement that data records must be readable for any use. This involves inspection of whether text data values are in UTF-8 character encoding and include only valid words and if numeric decimal data values have the same format. The *Contemporaneous* principle checks whether all data items are associated with the date and time of their recording. The *Original* principle is based on the assumption that when a report is submitted for the first time it is considered as original. Any modifications detected in a subsequent re-submission of the same report result in its characterization as being not original. *Accurate* involves the examination of whether all data items have values that lie within the acceptable ranges. The *Complete* principle involves the inspection of whether all mandatory data fields of a report have values and are not empty. *Consistent* refers to whether data items have starting dates preceding their ending dates. The *Enduring* principle refers to data record availability for the entire period for which it might be needed, by checking whether it is associated with a date after which it will not be stored. Finally, *Available* involves the inspection of whether the report has a date associated with it after which it will not be accessible.

### 3.3. Spumoni System Description

The main goal of the SPuMoNI information system is to evaluate data integrity and facilitate data quality assessment of significant amounts of raw data, from various heterogeneous manufacturing line data sources, along with their embedded systems, which control the various processes within the pharma production lines.

Hence, the SPuMoNI pharmaceutical manufacturing process data are mainly raw data, generated directly by heterogeneous manufacturing line data sources, which are further

pre-processed along the batch production processes that occur in the manufacturing plant. The system generates a comprehensive batch record, with plenty of details about product, materials, procedures, and equipment. Furthermore, authorized users can certify and sign any field of batch record, across the production process. There are two basic types of raw data that flow into the SPuMoNI information system. The first is relevant to product type and related recipe data information, while the second is relevant to machinery and equipment type control information. Raw data are represented either as numerical data with single values, as time series numerical measurements, or as non-numerical, categorical data (e.g., chemical/physical properties of material/substances).

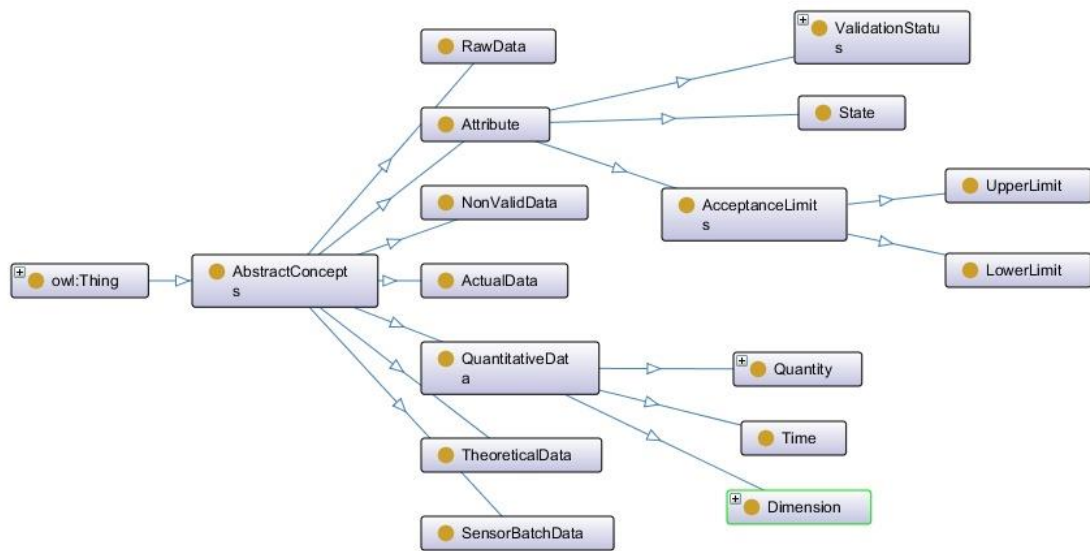
The Manufacturing Execution System (MES), is responsible for the automated and integrated management and execution of the recipes for each product, together with SCADA and the PLCs, for monitoring and controlling the machinery equipment (Figure 1). Specifically, these systems provide data for each batch record that is associated with an order recipe. Each recipe consists of sub-recipes, for the various manufacturing process phases, which are specified by specific plant instructions. Hence, a processing phase is uniquely identified and fully described by a recipe and its sub-recipe instructions.

At a next step, all these raw data types that are retrieved from various heterogeneous data sources are further pre-processed by SWRL rules and refined into a more qualitative information by SQWRL queries, before entering into the ALCOA assessment module (Figure 1). In particular, the most important module within the SPuMoNI's RegTech architecture is the ALCOA assessment module, which is responsible for the ALCOA principle rule interpretation and related compliance evaluation. The functionality of this assessment module involves the retrieval and evaluation of a batch record, consisting of production plant manufacturing information and raw data from various heterogeneous data sources, such as sensors, PLCs, production devices, operator staff, or equipment alarms, so as to decide whether the batch record is ALCOA compliant or not. All this information has been structured as a data model, by means of the data integrity ontology (DIOnt), and preserved in proper knowledge graph databases. Moreover, the ontology data model, and the ontology rules database, have been enriched with a proper ALCOA principle (constraints) knowledge base, as specified by 21 CFR Part 11, and cGMP Annex 11, in order to produce the appropriate rules (in our case, a combination of SWRL rules and SQWRL queries), that would support the ALCOA assessment module to take the final compliance decision. Actually, the SWRL rules are used to preprocess the initial knowledge base in order to infer a renewed knowledge base, for filtering the batch ALCOA compliance, via the use of SQWRL queries, as will be explained in the next subsection of the paper. Furthermore, at the top of the system architecture (Figure 1), there is also a blockchain module. This assures data validity by running data integrity checks on batch order data via hash transaction values, and will be included in our future work extension.

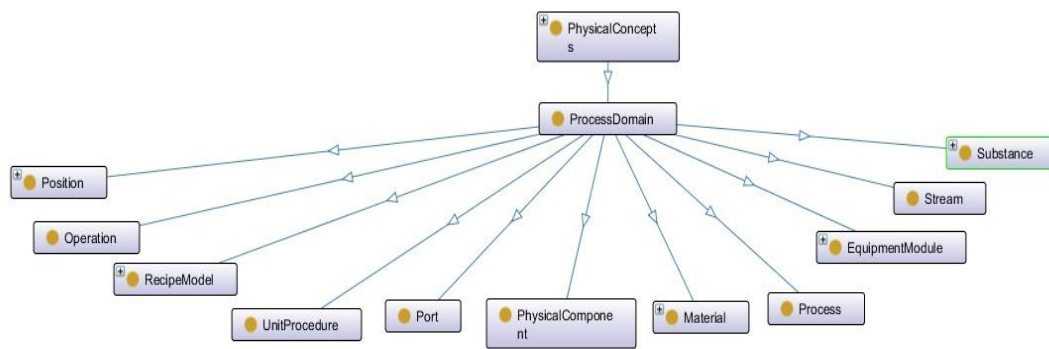
### 3.4. DIOnt Brief Description

Data integrity ontology (DIOnt) was selected for the semantic representation of all involved data and relevant process development, and the modeling of the regulatory requirements, targeting ALCOA data integrity rule compliance, aligned with RegTech system architecture. DIOnt ontology will be described in brief in this subsection, as it has been extensively presented in our previous research work [34].

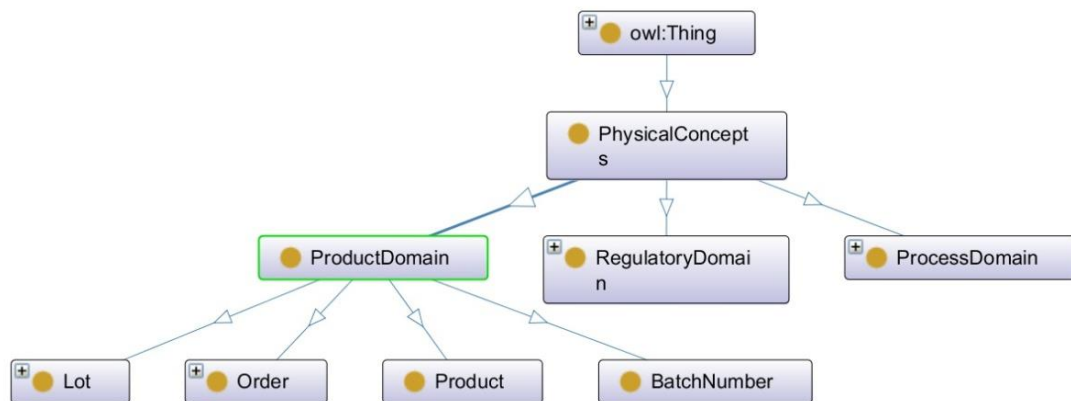
The hierarchical structure of DIOnt, which, to a certain extent, has been motivated by the OntoReg ontology [5], is shown in Figure 2a–d. The higher level of DIOnt consists of two generic classes, *AbstractConcepts* (Figure 2a) and *PhysicalConcepts*, respectively. The class *PhysicalConcepts* is then specialized into its subclasses, namely *ProcessDomain*, *RegulatoryDomain*, and *ProductDomain* (Figure 2b–d), which in turn, are further specialized into various subclasses, thus forming the three basic levels of the DIOnt ontology.



(a)



(b)



(c)

Figure 2. Cont.

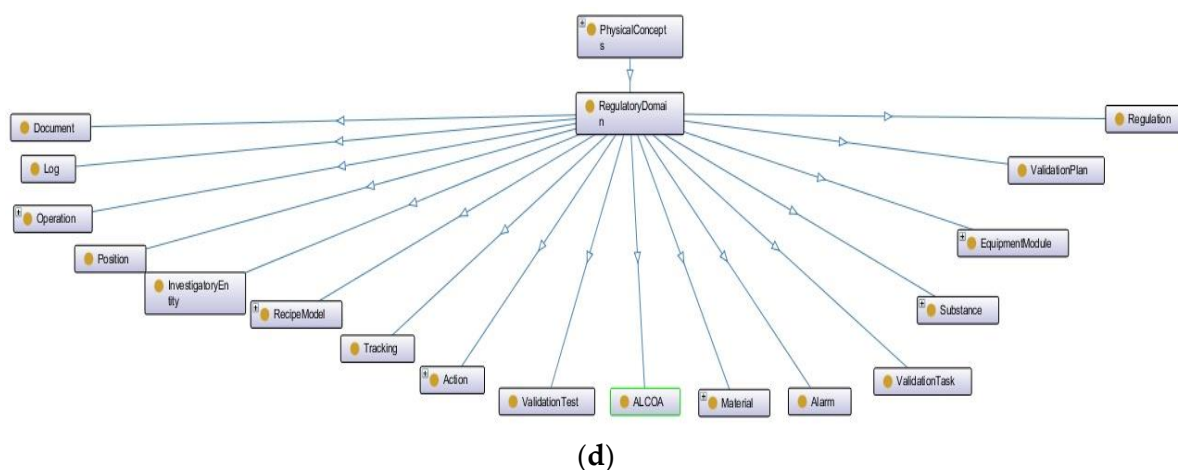


**Figure 2.** (a) Class AbstractConcepts and its subclasses. (b) Class ProcessDomain and its subclasses. (c) Class ProductDomain and its subclasses. (d) Class ProductDomain and its subclasses.

Below that level, there are four hybrid classes, namely Position, Substance, Operation and EquipmentModule, which belong to both Process and Regulatory domains. A pharmaceutical Process consists of UnitProcedures, while every Unit Procedure consists of one or more Operations which are executed by the same Equipment.

The RegulatoryDomain contains all regulatory requirements, derived by the ALCOA principles, and these are represented as relations between class individuals, which are then translated into a set of tasks, executed in a given sequence, via the use of data and object properties (Figure 3). This is achieved via the ValidationTask, ValidationPlan, ValidationTest, and Action classes as well as the Regulation and Position classes. Moreover, the classes Document and TaskReport are responsible for the relevant procedure report documentation. Last, the class ProductDomain, is responsible for accurately describing the manufactured product across the whole pharmaceutical manufacturing line, from its original ingredients to the final batch composition, and its order. This is achieved via Product, Order, Lot, and BatchNumber, class individuals.

Figure 3 shows the inter-domain relationships between the DIOnt entities, represented by object properties (as corresponding numbered arrows), which are presented in Table 1. Due to the fact that the overall DIOnt hierarchical structure consists of a great number of classes (95), as seen in Figure 2a–d, the number of relevant object and data properties required for their accurate relation description is quite high (99 and 52, respectively). Therefore, Figure 3 displays only the Process domain originated entity relationships, which constitute a part of the overall properties required; therefore, many of these relationships have been omitted to simplify the presentation.

In this way, the DIOnt restrictions and constraints can be defined and tailored specifically for corresponding ALCOA principles, and they are specified by respective SWRL rules. These rules enhance the OWL semantics of the DIOnt ontology classes and represent the functionalities required in order to perform the evaluation compliance of the ALCOA principles. Specifically, the SWRL rules enable rule-based reasoning, as they are loaded into the rule engine of the ALCOA assessment module, via the support of a Reasoner mechanism that is capable of inferring logical consequences from existing rules, and thus of enriching the DIOnt ontology with new inferred knowledge. Additionally, an SWRL-based query language (SQWRL) is used for extracting information from the inferred ontology, and its renewed knowledge base, performing ALCOA inconsistency evaluation. Hence, the DIOnt ontology forms the basis for the development of an appropriate regulatory compliance model, on top of the SPuMoNI's RegTech architecture which, as a proof-of-concept prototype, combines generic SWRL rules and SQWRL queries in order to perform ALCOA assessment, with the overall aim to enhance the process data integrity.

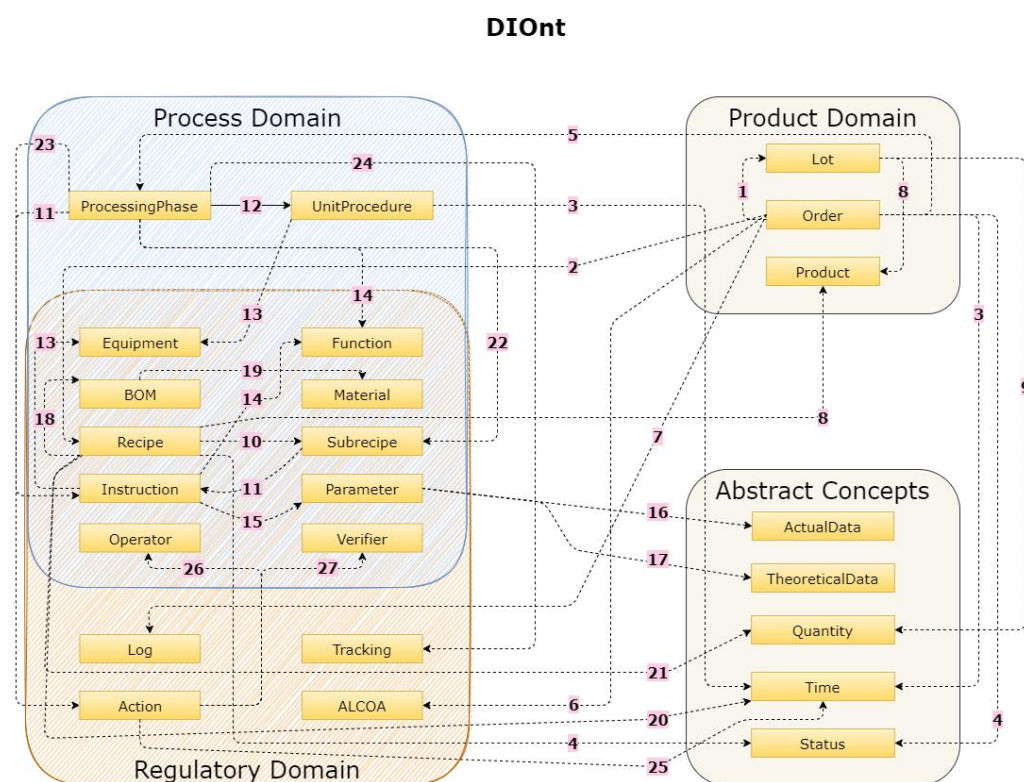


**Figure 3.** DIOnt inter-domain entity relationships.

**Table 1.** DIOnt object properties.

| No | Object Property    | No | Object Property                      |
|----|--------------------|----|--------------------------------------|
| 1  | isOrderOf          | 15 | hasParameter                         |
| 2  | underRecipe        | 16 | hasActualData                        |
| 3  | hasStart/hasEnd    | 17 | hasTheoreticalData                   |
| 4  | hasState           | 18 | hasBOM                               |
| 5  | hasProcessingPhase | 19 | hasMaterial                          |
| 6  | hasALCOA           | 20 | hasEffectivityDate/hasExpirationDate |
| 7  | hasLog             | 21 | hasMinQty/hasStdQty/hasMaxQty        |
| 8  | hasProduct         | 22 | underSubrecipe                       |
| 9  | hasQuantity        | 23 | hasStartAction/hasEndAction          |
| 10 | hasSubrecipe       | 24 | hasTracking                          |
| 11 | hasInstruction     | 25 | hasSignatureTime                     |
| 12 | hasUnitProcedure   | 26 | isPerformedBy                        |
| 13 | hasEquipment       | 27 | hasVerifier                          |
| 14 | hasFunction        |    |                                      |

### 3.5. DIOnt Software Methodology Architecture and Prototype Implementation

DIOnt software methodology consists of three major steps and its architecture layout can be seen in Figure 4. A proof-of-concept prototype has been implemented based on that architecture, by adopting various software modules such as the Protégé ontology editor, a rule engine, along with a reasoner and proper APIs, and finally the SPARQL queries.

The DIOnt development first step, begins with the definition and description of ontology entities (classes, sub-classes, domains), as well as the relationships between class individuals, including definitions of the appropriate object and data properties, within Protégé environment, as described in a previous section.

A graph database (GDB) of Neo4j NoSQL type, has been utilized for storage purposes. The main reason behind the decision to utilize a graph database instead of a relational one is the native support that the former provides for the storage of relations between

nodes (i.e., in the case of an ontology, between classes). More specifically, a GDB treats relations between data as first-class objects and stores them directly, thus making them equally as important as the data itself. This allows data in the store to be linked together directly and, in many cases, retrieved with one operation. This is in contrast to the indirect representation and storage of relationships in a relational database, where relationships are usually reconstructed at the query execution time. Thus, the choice of a GDB for storing an ontology with many highly interlinked classes provides a natural and efficient solution.

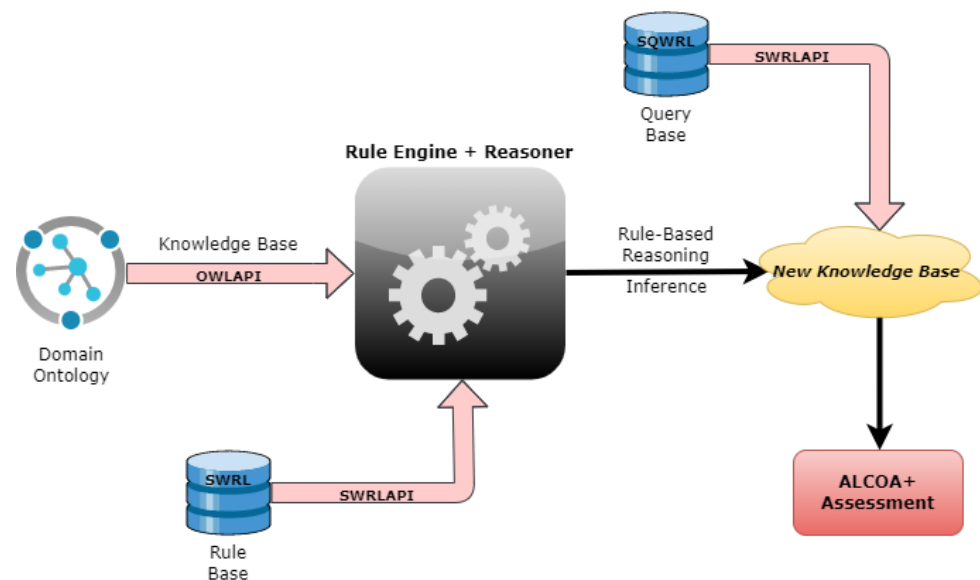


**Figure 4.** DIOnt software methodology architecture.

As a next step, the SWRL rules should be defined in such a way that they are aligned with ALCOA regulations. These rules should be described as OWL concepts (classes, properties, individuals), and be written to a database (RuleBase). As can be seen in Figure 4, all data inputs and SWRL rules are loaded into the rule engine via OWLAPI and SWRLAPI respectively. The rule engine, via the assistance of reasoner, runs and writes any inferred axioms back to the OWL ontology, enriching it with a new inferred knowledge base. These rules are generic in the sense that they are not targeted for the definition of a specific ALCOA principle constraint, but each one could be referred to more than one principle instead. For example, a generic SWRL rule would provide timestamps and, therefore, a time relationship between all entities of DIOnt ontology, via proper object properties. Now all original knowledge has been preprocessed, in terms of time sorting, in which the SQWRL queries will take advantage of that updated inferred knowledge and assist ALCOA module for the final decision regarding ALCOA principles that require time comparison (e.g., Available and Consistent).

Hence, the final step involves the SQWRL (SWRL based query language) queries definitions, that have been particularly designed for the extraction of proper information inferred by the rules, knowledge base regarding ALCOA inconsistency evaluation and data assessment (Figure 4). Each query, or set of them, may be based on more than one generic rule, and they are strictly referred to a specific inconsistency in ALCOA principles, with their corresponding extracted information expressed in percentage terms. Therefore, SQWRL queries and their data retrieval are offering a quantitative approach regarding ALCOA inconsistency, and hence they are aimed for providing ALCOA data assessment. Consequently, SQWRL does not alter or compete SWRL's semantics, on the contrary, it has the advantage of generic rule inferred reasoning, thus enhancing ALCOA assessment performance and accuracy. All generic SWRL rules, along with their corresponding SQWRL queries required for each ALCOA principle assessment, will be described in the next section.

#### 4. ALCOA Assessment

This section describes the ALCOA compliance evaluation mechanism within the SPuMoNI system framework, by introducing, firstly, the seven generic SWRL rules for providing rule-based reasoning and logical inference to OWL ontology, and, secondly, the required queries for providing quantitative data assessment for each ALCOA principle violation. ALCOA principle violations can be seen in terms of percentage (100%) in Figure 5.

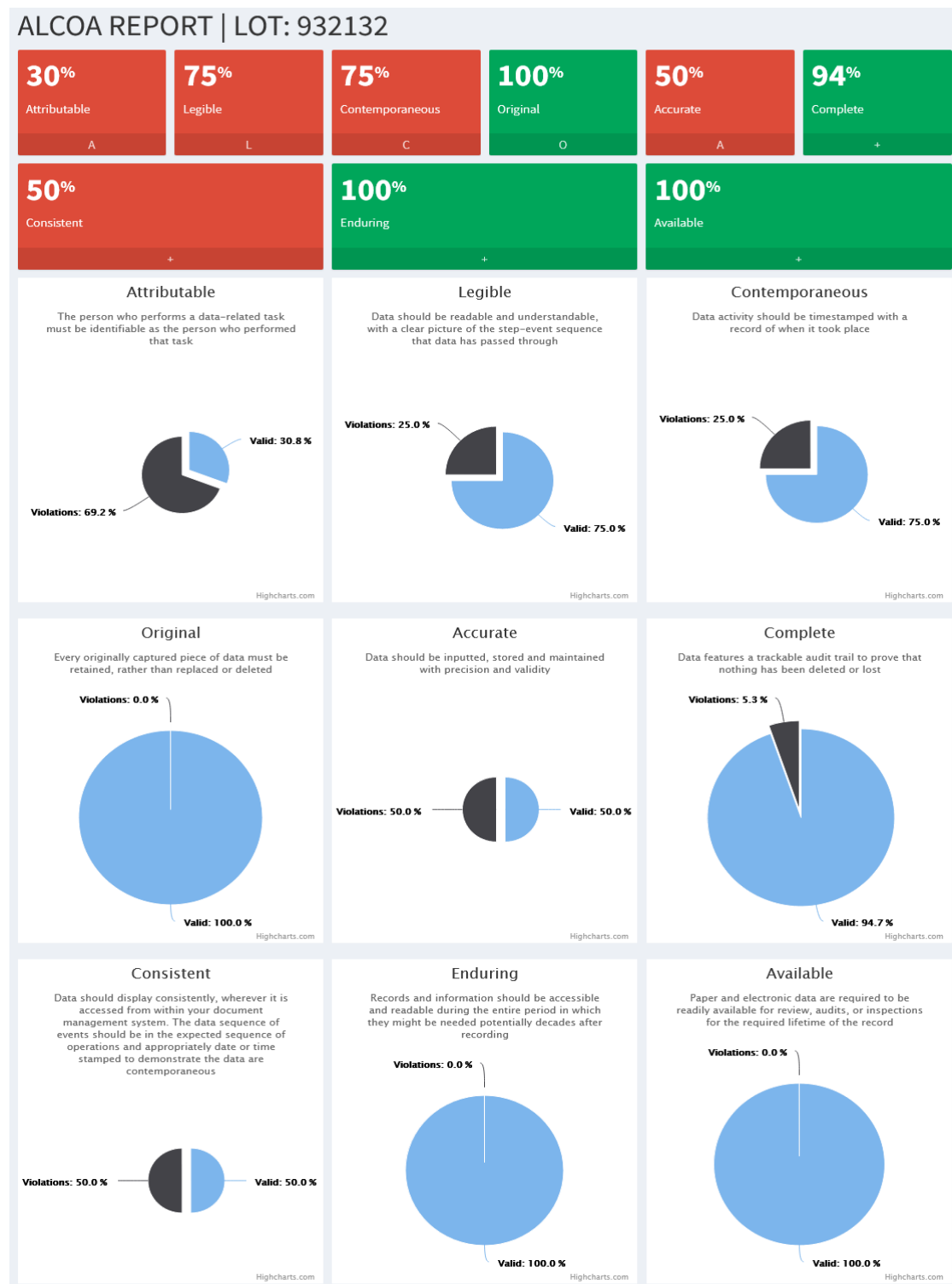


Figure 5. ALCOA violation data assessment.

#### 4.1. SWRL Rule Description

##### 4.1.1. Rule R<sub>1</sub>

Generic rule R<sub>1</sub> seeks the same names between class individuals Personnel and AuthorisedUser and relates them with object property sameAs. Hence, via this rule, only authorized users have been selected among system personnel, and this inferred knowledge will be exploited by the appropriate query in order to detect any unauthorized system user assigned to a task to thus detect Attributable principle violations.

$[R_1] \text{ Personnel}(?p) \wedge \text{hasName}(?p, ?pname) \wedge \text{AuthorisedUser}(?au) \wedge \text{hasName}(?au, ?auname) \wedge \text{swrlb:equal}(?pname, ?auname) \rightarrow \text{sameAs}(?p, ?au)$

##### 4.1.2. Rule R<sub>2</sub>

Generic rule R<sub>2</sub> implements a look-up match of each incoming data value, with those stored in a dictionary (class individual DictionaryEntry). In the event of an unsuccessful match, rule R<sub>2</sub> correlates via an indexing those values with class individual NonValidData and object property hasDataObject. Hence, via this rule, any incoming pharma parameter value is checked as to whether they have text values of proper character encoding or numeric decimal data values of proper format, or if they have specific text words present in language dictionaries. In such cases of incompatibility, they are characterized as invalid data. The updated inferred knowledge of invalid incoming parameter data values is exploited by proper query in order to detect and count Legible principle violations.

$[R_2] \text{ abox:dpaa}(?ind, ?prop, ?data) \wedge \text{DictionaryEntry}(?entry) \wedge \text{hasEntryValue}(?entry, ?val) \wedge \text{swrlb:notEqual}(?data, ?val) \rightarrow \text{NonValidData}(\text{UnknownData}) \wedge \text{hasDataObject}(?ind, \text{UnknownData})$

##### 4.1.3. Rule R<sub>3</sub>

Generic rule R<sub>3</sub> seeks parameters that show no value at all ("" ) at a specific timestamp. Then it puts default value "1970-01-01T00:00:00" instead. In other words, this rule filters parameters that have no timestamp value. Such inferred information will be exploited by proper query for detecting Contemporaneous and Enduring principle violations.

$[R_3] \text{ abox:dpaa}(?i, \text{hasTimestamp}, "") \rightarrow \text{hasTimestamp}(?i, "1970-01-01T00:00:00" \wedge \text{xsd:dateTime})$

##### 4.1.4. Rule R<sub>4</sub>

Generic rule R<sub>4</sub> checks for order originality by writing down each order's revision number (object property hasRevisionNumber). This rule checks for each order's revision number, classifying them in terms of those that have been submitted for the first time and those that have not, and hence whether they have been modified. Such inferred information is exploited by proper query for the detection of Original principle violations.

$[R_4] \text{ Order}(?order1) \wedge \text{Order}(?order2) \wedge \text{Lot}(?batch) \wedge \text{isOrderOf}(?order1, ?batch) \wedge \text{isOrderOf}(?order2, ?batch) \wedge \text{hasRevisionNumber}(?order1, ?r1) \wedge \text{hasRevisionNumber}(?order2, ?r2) \wedge \text{swrlb:greaterThan}(?r2, ?r1) \rightarrow \text{isRevisedVersionOf}(?order2, ?order1)$

##### 4.1.5. Rule R<sub>5</sub>

At first, generic rule R<sub>5</sub> attempts to correlate default processing phase parameters (theoretical default data) with current running recipe parameters (actual measurement data) via their ID attribute (object property hasID). It then creates a new object (makeOWLThing), to link those parameters whose actual values fall within parameter range (defined by AcceptanceUpperLimit and AcceptanceLowerLimit). By this method, this rule checks whether each parameter's value lies within acceptable range according to pharma recipe instructions. This inferred information is exploited by proper query for detecting Accurate principle violations.

[R<sub>5</sub>] Order(?order) ^ Lot(?lot) ^ ProcessingPhase(?pp) ^ Parameter(?phaseParam) ^ Recipe(?recipe) ^ Subrecipe(?step) ^ Instruction(?instr) ^ Parameter(?recipeParam) ^ isOrderOf(?order, ?lot) ^ hasProcessingPhase(?order, ?pp) ^ hasParameter(?pp, ?phaseParam) ^ hasActualData(?phaseParam, ?actual) ^ hasNumericalValue(?actual, ?actualval) ^ underRecipe(?lot, ?recipe) ^ hasSubrecipe(?recipe, ?step) ^ hasInstruction(?step, ?instr) ^ hasParameter(?instr, ?recipeParam) ^ hasID(?phaseParam, ?phaseParamId) ^ hasID(?recipeParam, ?recipeParamId) ^ swrlb:equal(?phaseParamId, ?recipeParamId) ^ hasAcceptanceLowerLimit(?recipeParam, ?low) ^ hasNumericalValue(?low, ?lowval) ^ hasAcceptanceUpperLimit(?recipeParam, ?up) ^ hasNumericalValue(?up, ?upval) ^ swrlb:greaterThanOrEqual(?actualval, ?lowval) ^ swrlb:lowerThanOrEqual(?actualval, ?upval) ^ swrlx:makeOWLThing(?range, ?phaseParam) -> Range(?range) ^ hasAcceptanceLowerLimit(?range, ?low) ^ hasAcceptanceUpperLimit(?range, ?up) ^ inRange(?param, ?range)

#### 4.1.6. Rule R<sub>6</sub>

Generic rule R<sub>6</sub> seeks for every parameter with a null value, via Null data property and NonValidData class individual. Therefore, this rule filters any parameter that has no value at all. This inferred information is exploited by a proper query for the detection of Complete principle violations, according to which all mandatory data fields of a report should have values and, therefore, not be empty.

[R<sub>6</sub>] abox:dpaa(?i, ?p, "") -> NonValidData(Null) ^ hasDataObject(?i, Null)

#### 4.1.7. Rule R<sub>7</sub>

Generic rule R<sub>7</sub> provides a time relationship between all time individuals of DIOnt ontology, along with their timestamps, via the usage of object property isAfter (and its inferred isBefore). Once all time individuals are related to each other in terms of time, the queries of ALCOA principles that require a time comparison (e.g., Available and Consistent) take advantage of that updated inferred knowledge.

[R<sub>7</sub>] Time(?t1) ^ Time(?t2) ^ hasTimestamp(?t1, ?ts1) ^ hasTimestamp(?t2, ?ts2) ^ temporal:after(?ts1, ?ts2) -> isAfter(?t1, ?t2)

### 4.2. SQWRL Query Description per Principle Violation

On account of the ALCOA assessment module, the SQWRL queries perform a final filtering of batch records associated with those order recipes that have comprehensive information about product, materials, procedures, and equipment that originate from the production plant manufacturing heterogeneous data sources, so as to decide whether it is ALCOA compliant or not.

#### 4.2.1. Attributable Principle Queries

Query Q1 takes advantage of inferred knowledge derived from Rule R<sub>1</sub>, and particularly from object property sameAs, and creates two sets, one with all personnel that perform a task (Personnel class individual), and one with authorized users that perform a task (AuthorisedUser class individual). It then selects the difference between these two sets, which are the unauthorized users (unauthorised class individual). Hence, Q1 query detects any unauthorized user on any task assignment among overall pharma system personnel, thus detecting violations in the Attributable principle, a principle from which any task assignment must only be supported by authorized person.

[Q1] Personnel(?all) ^ performs(?all, ?alltask) ^ Personnel(?user) ^ performs(?user, ?task) ^ AuthorisedUser(?authorised\_user) ^ sameAs(?user, ?authorised\_user) ^ sqwrl:makeSet(?s1, ?task) ^ sqwrl:makeSet(?s2, ?alltask) ^ sqwrl:difference(?s3, ?s2, ?s1) ^ sqwrl:element(?unauthorised, ?s3) -> sqwrl:select(?unauthorised) { Attributable principle violation }

#### 4.2.2. Legible Principle Queries

Query Q2 based on inferred knowledge of generic rule  $R_2$  selects unmatched data values via an indexing of the data property `UnknownData`, as defined in rule  $R_2$ . Hence, via this query, any incoming pharma parameter value with improper character encoding or format, or text words that are not included in language dictionaries, are detected and counted, as are any violations of the Legible principle, according to which, data must be readable, of a certain type of encoding or format, or with text attributes of words presented in language dictionaries.

[Q2] `abox:opaa(?i, ?p, UnknownData) ->sqwrl:select(?i) { Legible principle violation }`

#### 4.2.3. Contemporaneous Principle Queries

Query Q3 is based on inferred knowledge of generic rule  $R_3$  and selects all class individuals with timestamp "1970-01-01T00:00:00", which means it detects individuals (parameters) that have no value at a specific timestamp. This query detects any parameter with no timestamp value, and therefore also any violation of the Contemporaneous principle, according to which any data activity should be timestamped with a record of when it took place.

[Q3] `abox:opaa(?i, ?p, ?time) ^ Time(?time) ^ hasTimestamp(?time, ?ts) ^ temporal:equals(?ts, "1970-01-01T00:00:00"^^xsd:dateTime) ->sqwrl:select(?i) { Contemporaneous principle violation }`

#### 4.2.4. Original Principle Queries

Query Q4 is based on inferred knowledge of the identification of revision numbers per order by generic rule  $R_4$  and selects all orders with revised versions, that is, the non-original ones. Hence, this query detects any violations of the Original principle, according to which any report that has been submitted for the first time is considered to be original.

[Q4] `isRevisedVersionOf(?i, ?j) ->sqwrl:select(?i) { Original principle violation }`

#### 4.2.5. Accurate Principle Queries

Query Q5 is based on inferred knowledge of parameter value ranges by generic rule  $R_5$  and finally selects all violated parameters with actual values that fall outside those ranges (outlier). Hence, this query detects any violation of Accurate principle violation, according to which all data items should have values that lie within the acceptable ranges

[Q5] `Parameter(?all) ^ Parameter(?param) ^ abox:opaa(?param, inRange, ?r) ^ sqwrl:makeSet(?s1, ?param) ^ sqwrl:makeSet(?s2, ?all) ^ sqwrl:difference(?s3, ?s2, ?s1) ^ sqwrl:element(?outlier, ?s3) ->sqwrl:select(?param) { Accurate principle violation }`

#### 4.2.6. Complete Principle Queries

Query Q6 is based on inferred knowledge of the detection and identification of null valued parameters obtained from generic rule  $R_6$  and eventually selects them (class individuals) by indexing those data properties with a Null data property. Hence, this query detects any violation of the Complete principle, according to which all mandatory data fields of a report should have values and are not empty.

[Q6] `abox:opaa(?i, ?p, Null) ->sqwrl:select(?i) { Complete principle violation }`

#### 4.2.7. Consistent Principle Queries

Query Q7 is based on inferred knowledge of time comparison based on timestamps provided by rule  $R_7$  and ensures the proper recording of the time sequence of operations (class individuals), as required by default by the Consistent principle. It finally selects all violated operations with improper time sequence. Simply, this query checks whether all batch report data items have starting dates that precede their ending dates, therefore detecting any Consistent principle violation.

[Q7] *abox:opaa(?operation, hasStart, ?start) ^ abox:opaa(?operation, hasEnd, ?end) ^ isAfter(?start, ?end) ->sqwrl:select(?operation) { Consistent principle violation }*

#### 4.2.8. Enduring Principle Queries

Query Q8 is based on inferred knowledge of generic rule  $R_3$ , as well as query Q3, and selects all lots with the ExpectedLifetime timestamp “1970-01-01T00:00:00”, which means it detects lots that have no ExpectedLifetime value at a specific timestamp, which are therefore no longer available and violate the Enduring principle.

[Q8] *Lot(?lot) ^ hasExpectedLifetime(?lot, ?time) ^ hasTimestamp(?time, ?ts) ^ temporal:equals(?ts, “1970-01-01T00:00:00”^^xsd:dateTime) ->sqwrl:select(?lot) { Enduring principle violation }*

#### 4.2.9. Available Principle Queries

Query Q9 is based on inferred knowledge of time comparison based on timestamps provided by rule  $R_7$ , in particular it checks current timestamp time relation with the ExpectedLifetime of a lot (batch). It finally selects all lots with expired ExpectedLifetime (prior to current timestamp), and therefore those that have violated the Available principle.

[Q9] *Lot(?lot) ^ hasExpectedLifetime(?lot, ?exp) ^ hasTimestamp(?now, “now”) ^ isBefore(?exp, ?now) ->sqwrl:select(?lot) { Available principle violation }*

### 5. Discussion and Evaluation

In this paper, an SQWRL-based assessment prototype has been designed and developed for enhancing data integrity regulatory compliance in alignment with ALCOA principles, in the pharmaceutical domain, and based on DIOnt ontology. This prototype utilizes the combined action of generic SWRL rules, with SQWRL queries, on top of DIOnt software architecture, to accurately provide ALCOA assessment. In particular, it checks any batch record of manufacturing plant production, and effectively detects any non-compliant batches regarding ALCOA violations, and, finally, writes down the exact type of failure per each violation. This work has been tested with real world pharma manufacturing data, generated by the pharmaceutical process plant of an industrial partner of SPuMoNI project.

An innovative ontological conceptual approach has been chosen due to its inherent semantic representation of heterogeneous pharma production data and processes so as to fully conform with rapidly changing regulatory frameworks and relevant documentation. DIOnt ontology has been specifically designed and developed for ALCOA data integrity regulatory principles. The DIOnt ontology hierarchical structure, with its main ProcessDomain, RegulatoryDomain, and ProductDomain high level classes, along with their multiple lower-level subclasses, as described, allows for a thorough and exhaustive analysis of ALCOA compliance audit mechanisms, described as relations, indicating tasks, roles, and assignments between ontology class individuals. The joint description of these three domain branches can completely cover any pharma regulatory framework, throughout the overall pharma manufacturing production, starting from the batch product composition (ProductDomain), via the appropriate processes required (ProcessDomain), and ending at ALCOA principle compliance via SWRL rule validation with SQWRL queries (RegulatoryDomain).

The SQWRL based assessment prototype has been exploited to check ALCOA inconsistency expressed as a 100% violation percentage per principle (Figure 5). Generic SWRL rules have been employed to specify the constraints and restrictions, along with the support of SQWRL-based queries that accurately provide, in numbers, the ALCOA percentage violation, as described in detail in previous sections. In fact, the assessment prototype measures the non-violation success percentage per principle, as seen analytically in JSON printout of Figure 6. The system detects the non-compliant batches and writes down the exact type of failure per each violation, e.g., for the Contemporaneous principle. It then simply calculates the percentage of succeeded, non-violated batches, given a total number of batches.



**Figure 6.** ALCOA assessment report for Contemporaneous principle.

SWRL rule semantic representation capability enhances OWL description logic limitations in cases such as DIOnt, where there are a number of class individuals, each with complex relationships. Specifically, the if-then-else SWRL inherent expression capability offers experienced users the potential of setting their own user defined constraints and subsequently building the new inferred knowledge derived by involved class individuals and their relationships. The updated inferred ontology is developing and is definitely richer in terms of expressivity and completeness. On the other hand, these generic rule-based constraints, by their nature, do not necessarily need to provide memory consuming customized cut and tailored technical details, but generic information and inferred results that could be efficiently reused again and again, thereby improving knowledge usability and reducing memory usage and computation speeds. Then, the SQWRL queries, functioning in a complementary way by extracting information from inferred OWL ontology and its new knowledge base, thus provide ALCOA inconsistency evaluation, as described in the example of Section 3.4. The SQWRL queries would request and refine the precise information level required, which in our case, is the ALCOA violation percentage.

Another advantage of DIOnt, upon which our prototype is based, is the integration of SPuMoNI pharma system heterogeneous data sources and related data flows from the ALCOA assessment module, and the blockchain, via the support of a neo4j graph database. DIOnt ontology, by default, offers semantic data integration by using a conceptual representation of the data and of their relationships in order to eliminate heterogeneities. A neo4j graph database, on the other hand, is a database that typically considers nodes as the entities in a graph and can hold any number of attributes (called properties) or relationships and thus provide directed, semantically relevant connections between any two node entities. Therefore, a graph database offers an efficient, real-time way of accessing

nodes and relationships and thus efficiently integrate highly connected heterogeneous data without the need for complex queries that relational databases typically require in conventional heterogeneous data integration schemes.

Moreover, DIOnt, along with the expression potential of a graph database, does not need to infer connections between entities using foreign keys or map–reduce processing, but to directly access entity nodes and relationships, independently of the total dataset size, and consequently, to drastically reduce the speed with which reasoners need to fulfill the required ALCOA compliance analysis tasks. Currently, the DIOnt ontology has been tested on a small amount of batch data with the aim of analyzing its computation performance. Due to the fact that the ontology consists of a great number of classes and properties, which are required for full description of the pharma industry domain (Figure 2a–d), the run time of the computations required to perform ALCOA compliance analysis would not always be polynomial in the size of the data input. Our future intention is to improve DIOnt computation performance by considering an enhanced version of the ontology which will be applicable to larger data inputs.

Last, the inherent simplicity of neo4jgraph structure results in broad DIOnt data representation structures, in the sense that it could be applied for a wide range of use cases or application domains. Our assessment prototype shares this attribute, and without major modeling modifications. Our future aim is to extend DIOnt ontology with a semantic data fusion capability, of various heterogeneous data source streams, for different manufacturing lines and factories.

## 6. Conclusions

An SQWRL-based assessment prototype based on the presented DIOnt ontology can be proved to be a simple and promising solution for effectively tackling ALCOA data integrity issues in drastically evolving and complex pharma production regulation environments. The presented proof of concept of the system prototype takes advantage of a combined utilization of generic semantic web rule language (SWRL) rules with semantic query-enhanced web rule language (SQWRL) queries for the performance of ALCOA assessment, and thus the corresponding required regulatory compliance.

The current work will be further extended in the future by enhancing the DIOnt ontology with semantic data fusion capabilities in order to fuse various heterogeneous data source streams that transcend different manufacturing lines and factories. Therefore, we aim to suggest a unified data model, merging various heterogeneous pharma production lines. Moreover, the next version of the DIOnt ontology will be capable of dealing with larger batch data amounts and achieving improved computation and analysis performance.

**Author Contributions:** Conceptualization, E.N.L., I.S., G.M., V.C.G. and A.K.; Methodology, E.N.L., I.S., V.C.G. and A.K.; writing—original draft preparation, E.N.L., I.S., V.C.G. and A.K.; writing—review and editing, E.N.L., V.C.G. and A.K.; Project administration, A.K.; Software, G.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by from 2019 to 2022 by CHIST-ERA, the Horizon 2020 Future and Emerging Technologies programme of the European Union through the ERA-NET Cofund funding scheme (CHIST-ERA BDSI Call 2017), and General Secretariat for Research and Innovation (GSRI) of Ministry of Development and Investments of the Hellenic Republic, Greece.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This work has been developed under the auspice of the “Smart Pharmaceutical Manufacturing (SPuMoNI)” research project ([www.spumoni.eu](http://www.spumoni.eu), accessed on 10 July 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Perez, J.R. Maintaining data integrity. *Qual. Prog.* **2017**, *50*, 14–15.
- World Health Organization. *WHO Guidance on Good Practices and Record Management Practices*; WHO TRS 996; WHO: Geneva, Switzerland, 2016.
- Unger, B. An Analysis of 2017 FDA Warning Letters on Data Integrity. *Pharmaceutical Online*. Available online: <https://www.pharmaceuticalonline.com/doc/an-analysis-of-fda-warning-letters-on-data-integrity-0003> (accessed on 21 April 2021).
- Leal, F.; Chis, A.; Caton, S.; Gonzalez-Velez, H.; Garcia-Gomez, J.M.; Dura, M.; Sanchez-Garcia, A.; Saez, C.; Karageorgos, A.; Gerogiannis, V.C.; et al. Smart Pharmaceutical Manufacturing: Ensuring End-to-End Traceability and Data Integrity in Medicine Production. *Big Data Res.* **2021**, *24*, 100172. [\[CrossRef\]](#)
- Sesen, M.B.; Suresh, P.; Benares-Alcantara, R.; Venkatasubramanian, V. An ontological framework for automated regulatory compliance in pharmaceutical manufacturing. *Comput. Chem. Eng.* **2010**, *34*, 1155–1169. [\[CrossRef\]](#)
- McIntyre, S.R. *Integrated Governance, Risk and Compliance: Improve Performance and Enhance Productivity in Federal Agencies*; Technical Reports; PricewaterhouseCoopers: London, UK, 2008.
- Abdullah, N.S.; Indulska, M.; Sadiq, S. Compliance management ontology—A shared conceptualization for research and practice in compliance management. *Inf. Syst. Front.* **2016**, *18*, 995–1020. [\[CrossRef\]](#)
- Hashmi, M.; Governatori, G.; Lam, H.P.; Wynn, M.T. Are we done with business process compliance: State of the art and challenges ahead. *Knowl. Inf. Syst.* **2018**, *57*, 79–133. [\[CrossRef\]](#)
- O'Neill, A. An action framework for compliance and governance. *Clin. Gov. Int. J.* **2014**, *19*, 342–359. [\[CrossRef\]](#)
- KPMG. *The Compliance Journey: Making Compliance Sustainable*; KPMG International: Amstelveen, The Netherlands, 2005.
- Deloitte. *A Bold Future for Life Sciences Regulation; Predictions 2025*; Deloitte Centre for Health Solutions: London, UK, 2018.
- Abdullah, N.S.; Sadiq, S.; Indulska, M. A compliance management ontology: Developing shared understanding through models. In Proceedings of the 24th International Conference on Advanced Information Systems Engineering (CAiSE), Gdansk, Poland, 25–29 June 2012; pp. 429–444. [\[CrossRef\]](#)
- Butler, T.; O'Brien, L. Understanding RegTech for digital regulatory compliance. In *Disrupting Finance: FinTech and Strategy the 21st Century*; Lynn, T., Mooney, J.G., Eds.; Palgrave Macmillan: London, UK, 2019; pp. 85–102.
- Butler, T. Towards a standards-based technology architecture for RegTech. *CAPCO Inst. J. Financ. Transform.* **2017**, *45*, 49–59.
- Kim, H.M.; Fox, M.S.; Sengupta, A. How to build enterprise data models to achieve compliance to standards or regulatory requirements (and share data). *J. Assoc. Inf. Syst.* **2007**, *8*, 5. [\[CrossRef\]](#)
- Gruber, T.R. A translation approach to portable ontologies. *Knowl. Acquis.* **1993**, *5*, 199–220. [\[CrossRef\]](#)
- Guarino, N. Formal ontology and information systems. In Proceedings of the Formal Ontology in Information Systems (FOIS'98), Frontiers in Artificial intelligence and Applications, Trento, Italy, 6–8 June 1998; Guarino, N., Ed.; IOS Press: Amsterdam, The Netherlands, 1998; pp. 3–15, ISBN 978-90-5199-399-8.
- Roussey, C.; Pinet, F.; Kang, M.A.; Corcho, O. An Introduction to Ontologies and Ontology Engineering. In *Ontologies in Urban Development Projects. Advanced Information and Knowledge Processing*; Springer: London, UK, 2011; Volume 1. [\[CrossRef\]](#)
- Morbach, J.; Wiesner, A.; Marquardt, W. OntoCAPE: A (re)usable ontology for computer-aided process engineering. *Comput. Chem. Eng.* **2009**, *33*, 1546–1556. [\[CrossRef\]](#)
- Hailemariam, L.; Venkatasubramanian, V. Purdue Ontology for Pharmaceutical Engineering: Part I. Conceptual Framework. *J. Pharm. Innov.* **2010**, *5*, 88–99. [\[CrossRef\]](#)
- Venkatasubramanian, V.; Zhao, C.; Joglekar, G.; Jain, A.; Hailemariam, L.; Suresh, P.; Akkisetty, P.; Morris, K.; Reklaitis, G.V. Ontological informatics infrastructure for pharmaceutical product development and manufacturing. *Comput. Chem. Eng.* **2006**, *30*, 1482–1496. [\[CrossRef\]](#)
- Hailemariam, L.; Venkatasubramanian, V. Purdue Ontology for Pharmaceutical Engineering: Part II. Applications. *J. Pharm. Innov.* **2010**, *5*, 139–146. [\[CrossRef\]](#)
- Sapkota, K.; Aldea, A.; Duce, D.A.; Younas, M.; Bañares Alcántara, R. Towards semantic methodologies for automatic regulatory compliance support. In Proceedings of the 2011 ACM Workshop for Ph.D. Students in Information and Knowledge Management (PIKM'11), Glasgow, UK, 28 October 2011; pp. 83–86. [\[CrossRef\]](#)
- Elgammal, A.; Butler, T. Towards a framework for semantically-enabled compliance management in financial services. In Proceedings of the 1st International Workshop on Knowledge Aware Service Oriented Applications (KASA'15), co-located with ICSOC, Paris, France, 3–6 November 2014; pp. 171–184. [\[CrossRef\]](#)
- Espinoza, A.; Abi-Lahoud, E.; Butler, T. Ontology-driven financial regulatory change management: An iterative development process. In Proceedings of the 2nd Semantic Web and Linked Open Data Workshop (SW-LOD), Oaxaca, Mexico, 5 November 2014.
- Ford, R.; Denker, G.; Elenius, D.; Moore, W.; Abi-Lahoud, E. Automating financial regulatory compliance using ontology+rules and Sunflower. In Proceedings of the 12th International Conference on Semantic Systems (SEMANTiCS 2016), Leipzig, Germany, 12–15 September 2016; pp. 113–120. [\[CrossRef\]](#)
- Beach, T.H.; Rezguy, Y.; Li, H.; Kasim, T. A rule-based semantic approach for automated regulatory compliance in the construction sector. *Expert Syst. Appl.* **2015**, *42*, 5219–5231. [\[CrossRef\]](#)
- Hagedorn, T.J.; Smith, B.; Krishnamurthy, S.; Grosse, I. Interoperability of disparate engineering domain ontologies using basic formal ontology. *J. Eng. Des.* **2019**, *30*, 625–654. [\[CrossRef\]](#)

29. Zhong, B.; Gan, C.; Luo, H.; Xing, X. Ontology-based framework for building environmental monitoring and compliance checking under BIM environment. *Build. Environ.* **2018**, *141*, 127–142. [[CrossRef](#)]
30. Pham, T.A.; Le Thanh, N. An ontology-based approach for business process compliance checking. In Proceedings of the 10th International ACM Conference on Ubiquitous Information Management and Communication, New York, NY, USA, 4–6 January 2016. [[CrossRef](#)]
31. Schmidt, R.; Bartsch, C.; Oberhauser, R. Ontology-based representation of compliance requirements for service processes. In *Semantic Business Process and Product Lifecycle Management, Proceedings of the Workshop SBPM 2007, Innsbruck, Austria, 7 April 2007*; Hepp, M., Hinkelmann, K., Karagiannis, D., Klein, R., Stojanovic, N., Eds.; CEUR: Innsbruck, Austria, 2007.
32. Humberg, T.; Wessel, C.; Poggenpohl, D.; Wenzel, S.; Ruhroth, T.; Jürjens, J. Ontology-based Analysis of Compliance and Regulatory Requirements of Business Processes. In Proceedings of the 3rd International Conference on Cloud Computing and Services Science, Aachen, Germany, 8–10 May 2013; pp. 553–561.
33. Lee, S.W.; Gandhi, R.A.; Ahn, G.J. Certification process artifacts defined as measurable units for software assurance. *Softw. Process Improv. Pract.* **2007**, *12*, 165–189. [[CrossRef](#)]
34. Lallas, E.; Santouridis, I.; Mountzouris, G.; Gerogiannis, V.C.; Karageorgos, A. An ontology based conceptualization of data integrity regulatory compliance in pharmaceutical industry: The spumoni case. In Proceedings of the 25th Pan-Hellenic Conference on Informatics, Volos, Greece, 26–28 November 2021; pp. 460–465. [[CrossRef](#)]