Towards compliance checking in reified I/O logic via SHACL

Livio Robaldo
livio.robaldo@swansea.ac.uk
Legal Innovation Lab Wales - Swansea University
Swansea, Wales, UK

ABSTRACT
Reified Input/Output logic [29] has been recently proposed to handle natural language meaning in Input/Output logic [17]. So far, the research in reified I/O logic has focused only on KR issues, specifically on how to use the formalism for representing contextual meaning of norms (see [28]). This paper is the first attempt to investigate reasoning in reified I/O logic, specifically compliance checking. This paper investigates how to model reified I/O logic formulae in Shapes Constraint Language (SHACL) [2], a recent W3C recommendation for validating and reasoning with RDFs/OWL.

KEYWORDS
reified I/O logic, SHACL, RDFs/OWL

1 INTRODUCTION
Reified Input/Output logic [29] is Input/Output logic [17] enriched with reification. The introduction of reification in I/O logic enhances the expressivity of the I/O logic formulae without substantially affecting the I/O logic constructs that implement deontic reasoning.

Reification is a formal mechanism that associates instantiations of high-order predicates and operators with FOL terms [13], [27], [26]. The latter can be then directly inserted as arguments of other FOL predicates, which may be in turn reified again into new FOL terms. In other words, reified I/O logic associates norms with explicit terms, e.g., constants or variables, and not only with truth-conditional symbols such as predicates or (second-order) deontic operators. These terms can be then inserted as parameter of separated meta-properties.

Reified I/O logic is grounded on a specific reification-based approach for Natural Language Semantics: the framework in [12]. The main insight of [12] is to massively use reification in order to transform every second-order operator, including boolean connectives, into a FOL predicate applied to FOL terms. The final resulting formulae are then flat conjunctions of atomic FOL predicates.

As shown in [28], the formal simplicity and the modular structure of reified I/O logic facilitate the implementation of user-friendly interfaces to encode large knowledge bases of norms in reasonable time. [28] presents the DAPRECO knowledge base (D-KB), a repository of 966 formulae in reified I/O formulae that translates norms from the GDPR. The D-KB was built in four months via a special JavaScript editor implemented to this purpose.

While past research in reified I/O logic has focused on how building formulae associated with norms in natural language, this paper represents the first attempt to investigate how these formulae can be implemented and used for compliance checking, i.e., to infer which obligations have been violated in a given state of affairs and with respect to a given set of norms.

Compliance checking has never been really studied in I/O logic. Most past literature in I/O logic has focused on deontic reasoning, and recently, normative reasoning [15].

Deontic reasoning is to reason about what is obligatory and permitted, while dealing with contrary-to-duty reasoning, deontic paradoxes, ethical/moral conflicts, etc. Reasoning about obligations and permissions is of course orthogonal to what agents really do, i.e., whether they did or did not violate their obligations or whether they did or did not perform what they were permitted to do.

Compliance checking does not involve deontic reasoning. Still, compliance checking could not be so simple to handle, e.g., because norms might include exceptions that lead to defeasible reasoning.

This paper proposes a formalization of non-deontic inferences in reified I/O logic via SHACL [2]. While recent literature offered solutions for compliance checking implemented in RDFs/OWL, e.g., [6], only preliminary works use SHACL to this end, e.g., [21].

2 BACKGROUND - REIFIED I/O LOGIC

2.1 Input/Output logic
I/O logic was originally introduced in [17]. I/O logic is a family of logics, just like modal logic is a family of systems $\mathcal{K}$, $\mathcal{S}_4$, $\mathcal{S}_5$, etc.

However, while modal logic uses possible world semantics, I/O logic uses norm-based semantics, in the sense of [11]: I/O systems are families of if-then rules $(a, b)$, such that when $a$ is given in input, $b$ is returned in output. $a$ and $b$ are formulae in another logic, called “the object logic”. It has been argued that norm-based reasoning features some advantages over reasoning based on possible world semantics, first of all a lower computational complexity [30].

I/O logic neatly decouples deontic and non-deontic inferences. I/O logic is indeed a meta-logic wrapped around another logic (e.g., [12], in case of reified I/O logic) called “the object logic”. The meta-logic implements deontic inferences while the object logic implements the non-deontic ones. In I/O systems for legal reasoning, rules $(a, b)$ can be obligations, permissions, and constitutive rules. These are clustered within three distinct sets $O$, $P$, and $C$ such that $\forall(a, b)\in O$ reads as “given $a$, $b$ is obligatory”, $\forall(a, b)\in P$ reads as “given $a$, $b$ is permitted”, and $\forall(a, b)\in C$ reads as “given $a$, $b$ holds”.

Most past research on I/O logic has focused on theoretical investigations in the meta-logic, for modeling deontic reasoning. Since the focus was on studying the meta-logic, the object logic was always kept as simple as possible, i.e., it was always propositional logic. Reified I/O logic is perhaps the most relevant proposal so far in the I/O logic literature that employs an alternative (first-order) object logic: the logical framework in [12].

In I/O logic, inferences in the meta-logic are achieved by imposing axioms and constraints on the sets of if-then rules. Different combinations of axioms and constraints trigger different inferences.
For instance, [17] defines the basic axioms in (1), where the symbol ‘∗’ is the entailment relation of the object logic. Variants of these axioms have been further investigated in [23] and [22].

(1) • SI: from \((a, x) \rightarrow (h, x)\) whenever \(b \vdash a\).
• OR: from \((a, x)\) and \((b, x)\) to \((a \lor b, x)\).
• Wo: from \((a, x) \rightarrow (a, y)\) whenever \(x \vdash y\).
• AND: from \((a, x)\) and \((a, y)\) to \((a, x \land y)\).
• CT: from \((a, x)\) and \((a \land x \land y)\) to \((a, y)\).

By imposing axioms SI, Wo, and AND, we obtain a specific derivation system called \(der1v1\). Adding OR to \(der1v1\) gives \(der1v2\). Adding CT to \(der1v1\) gives \(der1v3\). The five axioms together give \(der1v4\). Each derivation system is sound and complete with respect to a different (norm-based) semantics and can therefore trigger different inferences (see [17] for further discussion and details).

Given a derivation system, we may further constrain its sets of if-then rules, by considering only subsets that do not yield outputs conflicting with given inputs. This is needed to handle contrary-to-duty reasoning, i.e., to determine which obligations are detached in a situation that already violates some among them [16].

This paper is not concerned with the meta-level of I/O logic. Rather, it will focus on the object logic and non-deontic inferences, including defeasible ones to handle exceptions in legal reasoning.

2.2 Adding reification to I/O logic

Reification is a well-known technique used in linguistics and computer science for representing abstract concepts. These are associated with explicit objects, e.g., FOL terms (see below in this section) or RDF resources (see §3 below), on which we can assert properties. These assertions can be recursively reified again into new terms.

Both [12] and RDFs/OWL recursively reify assertions until the knowledge is represented in terms of a flat list of atomic predicates applied to terms. In RDFs/OWL, these flat lists are made of triples “(subject, predicate, object)”, while [12] also allows predicates with higher arity; however, any n-ary predicate can be transformed into an equivalent conjunction of RDF triples.

In [12] and in reified I/O logic, both the antecedent and the consequent of the if-then rule are conjunctions of predicates. Universal and existential quantifiers are added to bound the free variables occurring in the formulae. Universals that osculate the whole if-then rules are used to “carry” individuals from the antecedent to the consequent. Formal details and definitions are available in [29].

A simple example from the D-KB [28] is shown in (2). (2) encodes in reified I/O logic part of Art.5(1)(a) of the GDPR. The if-then rule belongs to the set \(O\) (note “∈ O” in (2)): it is an obligation requiring each personal data processing to be lawful.

(2) \(\forall e_p(\exists l, z, w, y, x \{ (\text{RevisAtTime } e_p t) \land (\text{DataSubject } w) \land (\text{DataProcessing} e_p x z) \land (\text{Controller } y z) \land (\text{Processer } x) \land (\text{nominates } y x) \land (\text{GiveConsentTo } w e_p) \land \neg \text{ha}((\text{exceptionAgeDS } e_p)) \}) \in C\)

Formulae in reified I/O logic employ two kind of predicates: primed predicates such as \(\text{DataSubject}\) and non-primed predicates such as \(\text{DataSubject}\). The former are obtained by reifying the latter; the first argument of primed predicates is the reification of the non-primed counterpart, i.e., a FOL term.

We should not reify all predicates, but only those we need. For instance, we do need to reify \((\text{PersonalDataProcessing} x z)\) into \((\text{PersonalDataProcessing} e_p x z)\), where \(e_p\) explicitly refers to the action of processing, because we need to assert a property on this action: in the consequent of the obligation, we require it to be lawful, i.e., to satisfy the \(\text{isLawful}\) predicate. Note that in (2), in order to “carry” the variable \(e_p\) from the antecedent to the consequent, a universal quantifier outscoping the if-then rule has been inserted. All other variables are existentially quantified within the antecedent.

The other predicate that \(e_p\) is required to satisfy is \(\text{RevisAtTime}\). This is a special predicates used to assert which reifications “really exist” at a certain time. \(\text{RevisAtTime}\) parallels the well-known predicate \(\text{HoldsAt}\) used in Event Calculus [14].

Thus, formula (2) reads: “for every personal data processing \(e_p\) of some personal data \(z\), owned by a data subject \(w\), controlled by a controller \(y\), and processed by a processor \(x\) (nominated by \(y\)), it is obligatory for \(e_p\) to be lawful.”

2.3 Adding defeasibility to reified I/O logic

It is common in legislation that some rules override others in restricted contexts. These more specific rules are seen as exceptions of the general rules, as penguins may be seen as exceptions of birds with respect to the ability of flying.

In line with the literature, e.g., [10], reified I/O logic models exceptions via special predicates “Ex” that are false by default. This is achieved via negation-as-failure (naf). “\(\neg\text{ha}(\text{Ex})\)” is true if either “\(\text{Ex}\)” is false or it is unknown. On the other hand, when “\(\text{Ex}\)” holds, “\(\neg\text{ha}(\text{Ex})\)” is false, and the general rule is blocked. An example, taken from [28], is given by the following rules:

(a) If the data subject has given consent to processing, then the processing is lawful.

(b) If the age of the data subject is lower than the minimal age for consent of his member state, (a) is not valid.

(c) In case of (b), if the holder of parental responsibility has given consent to processing, then the processing is lawful.

(a)-(c) are formalized as the following constitutive rules:

(3) \(\forall e_p(\exists l, z, w, y, x \{ (\text{RevisAtTime } e_p t) \land (\text{DataSubject } w) \land (\text{DataProcessing} e_p x z) \land (\text{Controller } y z) \land (\text{Processer } x) \land (\text{nominates } y x) \land (\text{GiveConsentTo } w e_p) \land \neg \text{ha}((\text{exceptionAgeDS } e_p)) \}) \in C\)

(4) \(\forall e_p(\exists l, z, w, y, x, s \{ (\text{RevisAtTime } e_p t) \land (\text{DataSubject } w) \land (\text{DataProcessing} e_p x z) \land (\text{Controller } y z) \land (\text{Processer } x) \land (\text{nominates } y x) \land (\text{DataProcessing} e_p x z) \land (\text{StateOf } s w) \land (\text{ageOf}(w) \text{minConsentAgeOf}(s)) \}) \in C\)

(5) \(\forall e_p(\exists l, z, w, y, x, s, h \{ (\text{RevisAtTime } e_p t) \land (\text{DataSubject } w) \land (\text{DataProcessing} e_p x z) \land (\text{Controller } y z) \land (\text{Processer } x) \land (\text{nominates } y x) \land (\text{StateOf } s w) \land (\text{ageOf}(w) \text{minConsentAgeOf}(s)) \land (\text{isLawful } e_p) \}) \in C\)
3 COMPLIANCE CHECKING IN RDFs/OWL

RDFs/OWL is nowadays the W3C standard language for the Semantic Web [1]. RDFs/OWL represents knowledge via flat sets of triples 

“subject, predicate, object”, in which the predicate is 

an rdfs:Property while the subject and the object can be any 
rdfs:Resource, including other rdfs:Property(s). In other words, 

RDFs/OWL allows to treat rdfs:Property(s) as first-order terms on 

which separately asserting other (meta-)properties.

It is then evident that reification is, in essence, the very same 

mechanism used to represent knowledge in RDFs/OWL, thus the 

idea of implementing reified I/O logic in the W3C standard.

Some proposals have been done to implement compliance check- 

ing in RDFs/OWL, e.g., [9] and [6]. In these approaches compliance 

checking is achieved by enriching the ontology with classes refer- 

ring to sets of individuals compliant with the norms and by 

enforcing “is-a” inferences on these classes.

For instance, the OWL ontology used in [9] includes a class 

Supplier including individuals that supply consumers with some 

goods. Since suppliers are obliged to communicate their contractual 

conditions to their consumers (rule R1), the corresponding class in- 

cludes a boolean datatype property hasCommunicatedConditions 

which is true for those suppliers that has complied with their obli- 

gation and false otherwise. The ontology includes then a class 

SupplierR1compliant defined as to include only individuals in 

Supplier for which hasCommunicatedConditions is true. Com- 

pliance checking is enforced by applying simple “is-a” inferences.

In the same spirit, [6] encodes in a fragment of OWL2 selected 

norms from Artt. 6, 7, 15, 23, and 30 of the GDPR, which concern 

data usage policies. Compliance on these policies is again imple- 

mented via “is-a” inferences.

While [9] and [6] are of course important contributions towards 

the same direction of research advocated here, it is not clear how to 

model exceptions in those frameworks. Furthermore, adding explicit 
classes specifically devoted to “collect” the individuals compliant 

with the norms, as well as introducing new ones to properly handle exceptions, does not appear to be an easy and intuitive solution.

The rest of the paper proposes to use SHACL as an alternative 

to the accounts in [9] and [6].

4 COMPLIANCE CHECKING IN SHACL

This paper proposes and makes initial investigations to encode 

legal rules in a formal language different from RDFs/OWL. This 

formal language is SHACL [2], proposed by W3C precisely for 

validation and inferences on RDFs/OWL graphs. The use of SHACL 

is currently a matter of ongoing research in the Semantic Web 

community (see [7], [24], among others).

SHACL appears to be the right formal language for modelling 

compliance checking, although so far it has been scarcely investi- 

gated to this end, preliminary works being [20], [21], and [8].

SHACL was originally proposed to define special conditions on 

RDFS/OWL graphs, called “SHACL shapes”, more expressive than 

standard OWL cardinality and quantifier restrictions. RDFs/OWL 

graphs can be then validated against a set of such SHACL shapes.

However, SHACL “may be used for a variety of purposes beside 

validation, including user interface building, code generation and data integration” (cit. [2]). This paper adds a new use cases for SHACL 

in that it proposes to use it for serializing reified I/O logic formulae 

fit to check compliance.

In order to enhance the expressivity and the flexibility of the 

standard, a current W3C Working Group Note proposes to enrich 

SHACL shapes with advanced features such as “SHACL rules” to 

derive inferred triples from asserted ones, prior to validation.

As explained in [25], SHACL rules can trigger ontological or 

non-ontological inferences. Ontological inferences derive facts that 

can be added to the model. On the other hand, non-ontological 

inferences have the sole purpose of aggregating data, without neces- 

sarily asserting them in the model, in order to facilitate validation.

5 SERIALIZING REIFIED I/O LOGIC IN SHACL

This paper represents the first attempt to investigate how to serialize 

reified I/O formula modeling obligations as SHACL shapes and 

reified I/O formula modeling constitutive rules as SHACL rules.

(6) shows the SHACL shape that serializes (2) above. Both require 

every personal data processing to be lawful.

(6) CheckLawfulness 
rdf:type sh:NodeShape; 
sh:targetClass PersonalDataProcessing; 
sh:property [ sh:path is-lawful; 
sh:hasValue "true"^^xsd:boolean ]; 

In (6), “sh:” is SHACL namespace prefix. (6) is a sh:NodeShape 

requiring each individual of the sh:targetClass to satisfy the sh:property. The latter constrains the individuals reached from the sh:targetClass through the sh:path to satisfy sh:hasValue.

On the other hand, PersonalDataProcessing.is-lawful, and 

and all other RDFs/OWL resources used in this paper are associated 

1:1 with the predicates used in the reified I/O logic formulae such as (2), in the same way as the predicates occurring in the D-KB [28] are associated with RDFs/OWL resources from the PrOnto ontology [19]. an OWL ontology proposed to conceptualize the data protection domain. Space constraints avoid to provide further details about the 1:1 mapping between reified I/O logic predicates and RDFs/OWL resources.

SHACL shapes refer to constraints, a solution that appears to be 

more intuitive and economical than overpopulating the ontology 

with extra classes as suggested in [9] and [6].

The validation facts, as well as new individuals, derived through 

SHACL are not mandatorily inserted in the ontology. The SHACL 

rules to model the reified I/O logic formulae in (3), (4), and (5) 

represent non-ontological inferences, in the sense explained in [25]; 

these rules are only functional to infer the truth value of is-lawful 

before the SHACL shape in (6) is validated.

(3), (4), and (5) are serialized in the SHACL rules in (7), (8), (9) 

and, below, (10).

(7) sh:rule [ rdf:type sh:TripleRule; sh:order 0; 
sh:subject sh:; 
sh:object [sh:path 
(has-theme has-personal-data 
is-personal-data-of has-member-state 
has-min-consent-age); ]; ]; 

1See https://www.w3.org/TR/shacl-sf

2021-04-23 08:42. Page 3 of 1–5.
(8) sh:rule [rdf:type sh:TripleRule; sh:order 1;
    sh:condition [
        sh:property [sh:path has-min-consent-age;
        sh:minCount 1;];
    sh:property [sh:path has-agent;
        sh:lessThan has-min-consent-age;];]
    sh:subject [sh:path has-theme;]
    sh:predicate rdf:type;
    sh:object exceptionAgeDS; ];

(9) sh:rule [rdf:type sh:TripleRule; sh:order 2;
    sh:condition [
        sh:property [sh:path has-theme;
            sh:class exceptionAgeDS;];
        sh:subject [sh:path has-theme;]
        sh:predicate is-lawful;
        sh:object "true"^^xsd:boolean; ];

The sh:targetClass of all these SHACL rules is GiveConsent. Rules are executed according to the sh:order, from the lowest to the highest value. Each rule in (7)-(9) makes a new assertion: the rdf:Property specified in the sh:predicate of the rule is asserted between the two RDFs/OWL resources in the sh:subject and the sh:object. The sh:subject and the sh:object may be the sh:targetClass itself (keyword "sh: this"), a resource reachable from the sh:targetClass through a path specified in sh:path, any other resource in the ontology, or a literal.

(7) is executed as first because its sh:order is "0". This rule sets the value of the property has-min-consent-age for each individual in the class GiveConsent. This value is set to the integer value reachable from the sh:path defined on sh:object in (7). Specifically, this is the minimal consent age (has-min-consent-age) of the Member State (has-member-state) of the data subject owning the personal data (has-personal-data is-personal-data-of) involved in the personal data processing occurring as the theme of the GiveConsent instances (has-theme).

It is important to understand that has-min-consent-age will not be asserted on the individuals of GiveConsent also in the reference ontology, but only in the derived one. In other words, (7) is a non-ontological inference rule that collects/aggregates this value in GiveConsent for validation purposes only. After the validation, these values will be discharged.

Rule (8) compares the minimal consent age of the agents’ Member State, just asserted by (7) on GiveConsent’s instances, with the agents’ age. The two rules are then executed in a pipeline, thanks to SHACL command sh: order. Mirroring these inferences in native RDFs/OWL seems to be more difficult in that the formalism does not allow to specify a priority between the inference rules. When the agents’ age has been specified (sh:minCount 1) and it is lower than (sh:lessThan) the minimal consent age of the Member State previously asserted by (7), rule (8) asserts the individual of PersonalDataProcessing in the has-theme property of the individual of GiveConsent as member of the class exceptionAgeDS (see rdf:type in sh:predicate).

Finally, (9) sets as true the property is-lawful of the instances of PersonalDataProcessing that do not (sh:not) belong to class exceptionAgeDS. (9) implements the reified I/O logic formula shown above in (3) and the SHACL operator sh:not the negation-as-failure (predicate naf) occurring therein. sh: not is in fact true when the ontology does not include any specific assertion of the personal data processing as member of the class exceptionAgeDS. In other words, since the close world assumption hold for both RDFs/OWL and SHACL, sh: not is true when it is either false or unknown whether the personal data processing belongs to this class.

Finally, (10) implements the reified I/O logic formula (5) above:

(10) sh:rule [rdf:type sh:TripleRule; sh:order 2;
    sh:condition [
        sh:property [sh:path has-theme;
            sh:property [sh:path has-agent;
                sh:lessThan has-min-consent-age;];
            sh:lessThan has-min-consent-age;];
        sh:subject [sh:path has-theme;]
        sh:predicate is-lawful;
        sh:object "true"^^xsd:boolean; ];

If the age of the data subject (has-age) who owns the personal data of the processing (has-personal-data is-personal-data) that is the theme of a GiveConsent individual (has-theme) is lower than (sh:lessThan) the minimal consent age of his/her Member State and the agent of this GiveConsent individual is the holder of the data subject’s parental responsibility (has-holder-of-pr), then the boolean is-lawful is again set to true.

6 CONCLUSIONS
Reified I/O logic is a recent deontic logical framework explicitly designed to handle natural language semantics, i.e., to represent norms occurring in existing legislation such as the GDPR.

So far, the research in reified I/O logic has focused only on knowledge representation issues, specifically on how to use the formalism for representing contextual meaning of norms [3].

On the other hand, this paper is the first attempt to investigate computational issues in reified I/O logic, specifically how to represent the reified I/O logic if-then rules in a computable machine-readable format fit to enforce compliance checking.

This paper proposed to model regulatory rules as SHACL shapes and constitutive rules as SHACL rules. SHACL shapes and rules are applied to RDFs/OWL models that describe states of affairs.

The solution proposed here is alternative to some recent approaches that model compliance checking on RDFs/OWL ontologies, e.g., [9] and [6].

On the other hand, the present work only represents the first step of a research endeavour aiming at developing a full inference engine for reified I/O logic that implements and integrates all components involved in normative reasoning. Much further work needs to be done in order to obtain a formally well-defined framework, tested on existing industrial use cases.

Further directions of research include the automatic or semi-automatic generation of RDFs/OWL or SHACL assertions from legal texts, possibly via NLP (cf. [4], [5], [18]).
REFERENCES


