

Editorial Manager(tm) for Artificial Intelligence and Law
Manuscript Draft

Manuscript Number: ARTI25R1

Title: From Human Regulations to Regulated Software Agents' Behaviour. Connecting the abstract declarative norms with the concrete operational implementation. A position paper.

Article Type: Special Issue Agents Inst.&Legal Theory

Section/Category:

Keywords: Multi agent systems, electronic institutions, norms, norm enforcement

Corresponding Author: Dr. Javier Vázquez-Salceda, PhD

Corresponding Author's Institution: Universitat Politecnica de Catalunya

First Author: Javier Vázquez-Salceda, PhD

Order of Authors: Javier Vázquez-Salceda, PhD; Huib Aldewereld, M. D.; Davide Grossi, M.D.; Frank Dignum, PhD

Manuscript Region of Origin:

Abstract: In order to design and implement electronic institutions that incorporate norms governing the behavior of the participants of those institutions, some crucial steps should be taken. The first problem is that human norms are (on purpose) specified on an abstract level. This ensures applicability of the norms over long periods of time in many different circumstances. However, for an electronic institution to function according to those norms, they should be concrete enough to be able to check them run time. A second problem is that norms describe which behavior is desirable and permitted, but not how this is achieved in an institution. In the "real world" regulations often indicate procedures for implementing and enforcing the law. Likewise we should devise means to annotate the norms with practical aspects such as enforcement mechanisms, sanctions, etc. in order to get

requirements for an institution that will enforce norms (by either constraining behavior within the norms or reacting to violation of the norms). The choice of which kind of mechanism is chosen is not a normative one, but usually based on criteria of efficiency and/or feasibility of the mechanism. In this paper we present our view on how to approach these problems and other related issues to be solved in order to develop e-institutions capable to operate in complex, highly regulated scenarios.

From Human Regulations to Regulated Software Agents' Behaviour

Connecting the abstract declarative norms with the concrete operational implementation. A position paper

Javier Vázquez-Salceda (jvazquez@lsi.upc.edu)

Knowledge Engineering and Machine Learning Group, Universitat Politècnica de Catalunya, Barcelona, Spain

Huib Aldewereld (huib@cs.uu.nl), Davide Grossi

(davide@cs.uu.nl) and Frank Dignum (dignum@cs.uu.nl)

Institute of Information and Computing Sciences, Utrecht University, The Netherlands

May 30, 2006

Abstract. In order to design and implement electronic institutions that incorporate norms governing the behavior of the participants of those institutions, some crucial steps should be taken. The first problem is that human norms are (on purpose) specified on an abstract level. This ensures applicability of the norms over long periods of time in many different circumstances. However, for an electronic institution to function according to those norms, they should be concrete enough to be able to check them run time. A second problem is that norms describe which behavior is desirable and permitted, but not how this is achieved in an institution. In the "real world" regulations often indicate procedures for implementing and enforcing the law. Likewise we should devise means to annotate the norms with practical aspects such as enforcement mechanisms, sanctions, etc. in order to get requirements for an institution that will enforce norms (by either constraining behavior within the norms or reacting to violation of the norms). The choice of which kind of mechanism is chosen is not a normative one, but usually based on criteria of efficiency and/or feasibility of the mechanism. In this paper we present our view on how to approach these problems and other related issues to be solved in order to develop e-institutions capable to operate in complex, highly regulated scenarios.

Keywords: Multi agent systems, electronic institutions, norms, norm enforcement

1. Introduction

Internet, as an extension of the real world, is affected by the regulations of one or several countries on activities carried out through the web. For instance, Electronic Commerce activities between two parties are regulated by the law of the parties' countries plus international commerce treaties. In the Health Care field, highly regulated, the citizens' rights are precisely defined and regulated by national and international laws. How to make sure that such norms and regulations are met on the activities and information exchanges through Internet? How to create



© 2006 Kluwer Academic Publishers. Printed in the Netherlands.

mechanisms to enforce norm compliance, and therefore, increase trust between individuals and companies?

In most software and agent methodologies, these external regulations, along with the internal norms and regulations of the organization to be modelled, are seen only as extra requirements in the analysis phase of the system. If either the external or the internal regulations change (as they usually do from time to time), it becomes very hard to track all the changes to be done in the implementation, as there is no explicit representation of the norms and regulations, but a chain of design decisions that were guided by the norms' requirements (i.e., if norms are embedded in the agents' design and code, all the design steps have to be checked again and all the code verified to ensure compliance with new regulations). The alternative is to have an explicit representation of the norms.

Research on Distributed Artificial Intelligence has created the concept of Electronic Institutions. As their human counterparts, an Electronic Institution is an entity defining a set of norms over the behavior of individuals inside the institution. Recently research in electronic institutions has focused on the use of Software Agent technology. An *Agent-Mediated Electronic Institution* (**e-institution** for short) belongs to a new and promising field where interactions between a group of (software) agents are regulated by means of a corpora of explicit norms, expressed in a computational language that agents can interpret. An e-institution [12, 14] is a safe environment mediating in the interaction of agents. The expected behavior of agents in such an environment is described by means of an explicit specification of norms, which is a) expressive enough, b) readable by agents, and c) easy to maintain.

1.1. CHALLENGES IN THE FIELD

In our view, there are basically four issues to be solved in order to successfully design e-institutions in complex, highly regulated scenarios:

- I1 **abstractness of human regulations:** human regulations are written in a very abstract way, open to several interpretations to make the legal text stable for a long time. This abstraction poses a problem when trying to implement them on computers, where meanings should be precise and unambiguous.
- I2 **operationalisation of norms:** usually norms are formally specified in languages based in deontic logic, which is expressive enough to cover some of the concepts that appear in regulations (obligations, permissions, prohibitions), has formal semantics and allows

the verification of sets of norms. However, deontic-based languages are declarative in nature and do not fully express the operational impact of the norms in the behavior of the multiagent system.

I3 implementation of norms from the institutional perspective

Agent platforms should be extended with mechanisms to detect attempts from the agents to break the norms. This issue is not only relevant in open scenarios where non-trusted agents may enter into the platform, but also in scenarios with trusted agents, as sometimes these agents may break the norms unwillingly (e.g. not detecting that a norm applied in a given situation).

I4 methodology to design electronic institutions Although there are some toolkits and some frameworks to build electronic institutions, currently there is no methodology which covers all the aspects, from the specification of the abstract norms to the connection with their implementation, both from the institutional and the agent viewpoints.

In our opinion, all four issues above are very important to create agent-mediated electronic institutions able to take into account domain regulations during the design phase and also to enforce compliance of agent behavior to those regulations during run-time. Depending on the characteristics of the application domain (openness, frequency of regulations' change, potential negative impact of breaking a norm) some of these issues will be more or less relevant. For instance, in the Health Care domain, some norms may be breakable without harm to a given patient, while others may endanger the patient's health. An additional issue would be to have agents which can understand a given set of norms and be able to reason the suitability of their behavior against the norms. Although it would be desirable to have such agents it is not crucial for the success of e-institutions, as by issue I3 regulations' compliance is ensured by the institution, and we cannot assume in an open scenario that all agents interacting within an e-institution would be capable of normative reasoning to self-ensure proper behavior. ¹

The rest of the paper is organized as follows. Our view on I1 is explained in section 2. Our approach on I2 and I3 is explained in sections 3 and 4. Our view on I4 is briefly explained in section 5. Finally in section 6 we summarize our approach and point to some ongoing work.

¹ The reader interested in the issue of norm-aware agents can find more details in [7].

2. Explaining abstractness in human regulations

In [8, 31] it has been variously stressed how the design of norm-governed multi-agent systems has to cope with the inherent abstractness of norm formulations. Human regulations are usually written in a quite abstract language and are open to interpretation. The main reason for this is to cover with the same legal text the major number of cases and therefore be stable for longer periods of time. This abstraction and capability of multiple interpretations that are positive for humans pose a problem when trying to implement them on computers, where meanings should be precise and unambiguous. Attempts in the past to build Legal Knowledge-Based Systems capable to automate rules of interpretation for any situation had to confront the *open texture* of norms [22]. In the case of e-institutions the open texture interpretation problem is reduced by fixing a context for interpretation (the system does not need to handle all possible interpretations of a norm, but the interpretation which is valid in the context of the e-Institution). Therefore, when designers try to include the norms in the design process of an e-Institution, at least two steps should be performed:

- an interpretation of the norms in the context of the e-institution should be provided (usually by close collaboration with legal experts on the domain), and
- such interpretation should be then connected to the processes and structure of the implemented e-institution.

The second step is explained in section 3. First step of the problem can be distilled in the question:

How are norms, which are specified by means of abstract terms (e.g. *“personal data which are not strictly relevant for the transplant activities should not be included in the transplant data base”*), connected to norms specified instead via more concrete ones (e.g. *“data about age may be included in the transplant data base”*)?

Our approach: in previous work [11, 13] we have focused on the formal definition of norms by means of some variations of deontic logic that include conditional and temporal aspects [5, 10], and we provided formal semantics. However there are complex ontological issues to be fixed in order to solve the open texture interpretation problem. Our view is that Institutions provide structured interpretations of the concepts in which norms are stated. To put it in a nutshell, institutions do not only consist of norms, but also of *ontologies* of the to-be-regulated domain. For instance, whether something within a given institution

counts as personal data and should be treated as such depends on how that institution interprets the term 'personal data'.

This perspective on institutions, which emphasizes the semantic dependencies between abstract and concrete norms, goes hand in hand with acknowledged positions in the study of social reality and legal systems. In fact, institutions can be seen as complex systems of norms, which consist of regulative as well as non-regulative components², that is to say, which do not only regulate existing forms of behavior, but they actually specify and constitute -via classification- new forms of behavior. The abstractness of norms is precisely the effect of such a "constitution": non-regulative components of institutions connect concrete concepts, such as 'age' to abstract ones such as 'personal data'. The basic brick of this constitution consists of statements of the following general logical form: "*X counts as Y in context C*" [28, 24]. Two ingredients are displayed in this sentential form: a relation between two concepts *X* and *Y*, (for instance, 'age' and 'personal data'), and a relativization of it to a specific context *C* (for instance, 'national hospitals')³. In [20, 19, 21], we proposed, investigated and applied a framework for formally representing such statements, where the relation between *X* and *Y* is interpreted as a standard concept subsumption, but which holds only in relation to a context *C*: "age is a subconcept of personal data in the context of national transplantation policies". Counts-as statements are therefore studied as contextual subsumption relations: $C : X \subseteq Y$. The key idea of the framework consists in tailoring the semantic machinery of Description Logic [2], with the framework developed in [18] to model context in logic. As a result, it becomes possible to formally represent, in a framework based on computationally appealing logics, how concepts expressed in the abstract language of the norms, can be related to concepts belonging to more specific languages, and in particular, to the language used at the implementation level.

In addition, this approach enables a further noteworthy feature consisting in the possibility of representing a number of fundamental notions for the understanding of contextual ontologies, such as the notions of *core* and *penumbra* of the meaning of a concept which we borrowed from legal theory [23] and which we formally analyzed in [20]:

[Suppose a] legal rule forbids you to take a vehicle into the public park. Plainly this forbids an automobile, but what about bicycles, roller skates, toy automobiles? What about airplanes? Are these, as

² See for example [27], [1], [24], [6] and [28].

³ The importance of the notion of context for specifying and representing institutions has been advocated also in [9, 29].

we say, to be called “vehicles” for the purpose of the rule or not? [...] There must be a *core* of settled meaning, but there will be, as well, a *penumbra* of debatable cases in which words are neither obviously applicable nor obviously ruled out.

The possibility of capturing these notions in a rigorous setting, allows to properly represent the domain by ontologies that provide the connection from the abstract terms in the human regulations to the ones used in the e-institution. As we will see in section 5, the creation of such ontologies should be a key aspect during the formal specification process of the regulations (and the terms appearing in them) that will impact the behavior of agents within the e-institution.

3. From Declarative to Operational

Current work on normative systems’ formalization (mainly focused in Deontic-like formalisms [25]) is declarative in nature, focused on the expressiveness of the norms, the definition of formal semantics and the verification of consistency of a given set. The declarative aspects of norms are important to check whether a given action or situation is permitted, obliged or forbidden (at a certain moment). However, norms also have an operational aspect, as norms should guide (or at least influence) the behavior of the agents. Although norms can be implemented by just putting enough constraints on the behavior of the agents such that the norms will never be violated, one has to define which constraints are to be used exactly, and who (or what) will check that the constraints always hold.

Our approach: in our view, in order to implement norms it is not enough to include a model-checking module by, e.g., implementing a theorem prover that, using the norms semantics, checks whether a given interaction protocol complies with the norms. The implementation of norms should also consider a) how the agents’ behavior is affected by norms, and b) how the institution should ensure the compliance with norms. The former is related to the *implementation of norms from the agent perspective*, by analyzing the impact of norms in the agents’ reasoning cycle (work on this perspective can be found in [4, 3, 7]). The latter is related to the *implementation of norms from the institutional perspective*, by implementing a safe environment (including the enforcing mechanisms) to ensure trust among parties.

As we discussed in [30], we need to specify the operational semantics of the norms to ensure this institutional perspective of norms. In general an operational semantics for norms always comes down to either one

of the following: 1) **Defining constraints on unwanted behavior;** or 2) **Detecting violations and reacting to these violations.**

The choice between these two approaches is highly dependent on the amount of control over the addressee of the norms. Preventing unwanted behavior can only be achieved if there is full control over the addressee (as the constraints are programmed into the agents beliefs and goals); in cases where such control is lacking one should define and handle violations. As we assume the external agents to be black-boxes, i.e. the internal states of the agents cannot be observed nor controlled, we will focus on the latter method of enforcement.

In order to detect and react to violations, norms need to have a declarative as well as an operational meaning. For this purpose we extend normal deontic representations of norms with extra fields to express the operational aspects of the norm. For expressing the declarative part of the norm any kind of deontic logic can be used, however, since the norms need to be useable by agents, a machine-readable format should be used. We proposed in [30] to use the following format for expressing norms that can be conditional or have temporal aspects:

DEFINITION 1 (Norm Condition).

$$\begin{aligned} \text{NORM_CONDITION} &:= N(a, S \langle \text{IF } C \rangle) | \\ &\quad \text{OBLIGED}(a \text{ ENFORCE}(N(a, S \langle \text{IF } C \rangle))) \\ N &:= \text{OBLIGED} | \text{PERMITTED} | \text{FORBIDDEN} \\ S &:= P | \text{DO } A | P \text{ TIME } D | \text{DO } A \text{ TIME } D \\ C &:= \textit{formula} \\ P &:= \textit{predicate} \\ A &:= \textit{action expression} \\ \text{TIME} &:= \text{BEFORE} | \text{AFTER} \end{aligned}$$

This machine-readable declarative representation lacks operational meaning, as it includes no information about violation management, detecting the activation and deactivation of norms, and responses to violations of norms. These elements are necessary for the implementation of norm enforcement as the process of enforcement is composed of the following sub-processes:

- a) the detection of when a norm is active,
- b) the detection of a violation on a norm, and
- c) the handling of the violations.

Before it can be determined whether a norm is violated, it has to be checked whether the norm is actually active. Detecting the *activation of the norm* (when the condition C holds) and the *deactivation of the norm* (when predicate P or action A is fulfilled or C does not hold)

are related to the declarative meaning of the norms. An additional issue is to establish the allowed reaction time between the activation and deactivation of obligations, i.e. the time that is allowed for the completion of the obligation when it becomes active but when the reaction time has passed.

Deadlines represent a special case in the implementation of conditional norms, as they are not that easy to check. Deadlines require a continuous check (second by second) to detect if a deadline is due. If the institution has lots of deadlines to track, it will become computationally expensive. We proposed in [30] to include within the agent platform a **clock trigger mechanism** that sends a signal when a deadline has passed. The idea is to implement the clock mechanism as efficiently as possible to avoid burden on the agents.

The second sub-process of the enforcement is about detecting the actual violation of the norm (for which the norm must be active). In an agent platform with several agents performing different actions at the same time, a question arises on how to implement the detection of the occurrence of actions. The agents enforcing norms may become overloaded on trying to check any action at any time. To solve this problem we proposed in [30] to create two platform mechanisms: 1) a **black list mechanism** of actions to be checked, and 2) an **action alarm mechanism** that triggers an alarm when a given action on the black list attempts to start or is done. This trigger mechanism has to do no further checks, only to make sure the enforcer agent is aware of the occurrence of the action. The action alarm can only be done with actions defined in the e-institutions' ontology.

It is easy to see that a protocol or procedure satisfies a norm when no violations occur during the execution of the protocol. The real problem in norm checking lies, however, in determining when that violation occurs. For instance, in criminal investigations, a police officer should not have more (sensitive or private) information than needed for the investigation. So an officer is doing fine as long as no violation occurs (i.e. he does not have unneeded information). The real problem is that the "need for information" is difficult to ascertain, unless the system is implemented in a way where, e.g., officers should indicate how the requested information contributes to a given case. In some cases this could be formally checked at run-time, while in other cases it might be marked as '*to be checked*' and then checked by people afterwards.

Therefore, the implementation of norm enforcement is dependent on two properties of the checks to be done: a) the checks being *verifiable* (i.e. a condition or action that can be machine-verified from an institutional point of view, given the time and resources needed) and b) the checks being *computational* (i.e. a condition or action that can

be checked on any moment in a fast, low cost way). Using these two properties, we can analyse their impact on the implementation of norm enforcement (explained in more detail in [30]):

- **Norms computationally verifiable:** verification of all predicates and actions can be done easily, all the time.
- **Norms not computationally verifiable directly, but by introducing extra resources:** the condition or action is not directly (easily) verifiable, but can be so by adding extra data structures and/or mechanisms to make it easy to verify. The *action alarm* and *clock trigger* mechanisms are examples of extra resources.
- **Non-computationally verifiable** the check can be machine-verified but it is too time/resource consuming to be verified at any time. Verification should not be done all the time, but can be delayed, done periodically or at random.
- **Observable from the institutional perspective, but not machine-verifiable:** that is, verifiable by other (human) agents that have the resources and/or information needed. Such checks should be delegated appropriately.
- **Indirectly observable from the institutional perspective:** These can be internal conditions, internal actions (like reasoning) or actions which are outside the ability of the system to be observed or detected. To solve this, other observable conditions or actions might be found and used to (indirectly) detect a violation.

The declarative norm representation of definition 1 is extended using these operational aspects to encapsulate the operational meaning of the norm. This gives us the following norm frame:

DEFINITION 2 (Norms).

```

NORM := NORM_CONDITION
      VIOLATION_CONDITION
      DETECTION_MECHANISM
      SANCTION
      REPAIRS
VIOLATION_CONDITION := formula
DETECTION_MECHANISM := {action expressions}
SANCTION := PLAN

```

$$\begin{aligned} \text{REPAIRS} & := \text{PLAN} \\ \text{PLAN} & := \textit{action expression} \mid \\ & \quad \textit{action expression}; \text{PLAN} \end{aligned}$$

In this format, the *norm condition*-field is denoting when the norm becomes active and when it is achieved. The *violation condition* is a formula derived from the norm to express when a violation occurs (e.g. for the norm $\text{OBLIGED}((a, P) \text{ IF } C)$ this is exactly the state when C occurs and P does not, that is, the state where the norm is active, but not acted upon). The *detection mechanism* is a set of actions that can be used to detect the violation. The set of actions contained in the *sanction*-field is actually a plan which should be executed when a violation occurs (which can contain imposing fines, expulsing agents from the system, etc.). Finally, the *repairs* contains a plan of action that should be followed in order to ‘undo’ the violation. An example of an annotated norm is shown in figure 1.

4. Implementing Norms

Given the operational representation of a norm, how does one proceed to implementing that norm in an e-institution (such as ISLANDER, [14]). What elements in the e-institution are needed for this implementation, given that the norm enforcement is supposed to be done by a distributed set of (internal) agents? And how do the norms represented in the formalism of definition 2 translate to these enforcement mechanisms?

Our approach: The ISLANDER formalism [14] provides a formal framework for institutions [26]. This formalism views an agent-based institution as a *dialogical system* where all the interactions inside the institution are a composition of multiple dialogic activities (message exchanges). The messages (or *illocutions*) are structured through agent group meetings called *scenes* that follow well-defined protocols. Furthermore, the AMELI platform [15] allows the execution of e-institutions, based on the rules provided by ISLANDER specifications, wherein external agents may participate.

To implement norm enforcement in the ISLANDER framework, we need to introduce mechanisms for detecting violations and expressing the actions that have to be taken by the internal agents upon the detection of a violation (i.e. the sanctions and repairs). The former is done by specifying *integrity constraints* (derived from [16]), which are used by the system to detect and register the violations of norms. The latter is expressed by *dialogical constraints*, which are in fact obligations

<i>Norm</i>	FORBIDDEN(<i>allocator</i> DO <i>assign</i> (<i>organ</i> , <i>recipient</i>))
<i>condition</i>	IF NOT(<i>hospital</i> DONE <i>ensure_compatibility</i> (<i>organ</i> , <i>recipient</i>)))
<i>Violation condition</i>	NOT(<i>done</i> (<i>ensure_compatibility</i> (<i>organ</i> , <i>recipient</i>))) AND <i>done</i> (<i>assign</i> (<i>organ</i> , <i>recipient</i>)))
<i>Detection mechanism</i>	{ <i>detect_alarm</i> (<i>assign</i> , ' <i>starting</i> '); <i>check</i> (<i>done</i> (<i>ensure_compatibility</i> (<i>organ</i> , <i>recipient</i>))); }
<i>Sanction</i>	<i>inform</i> (<i>board</i> , "NOT(<i>done</i> (<i>ensure_compatibility</i> (<i>organ</i> , <i>recipient</i>))) AND <i>done</i> (<i>assign</i> (<i>organ</i> , <i>recipient</i>)))")
<i>Repairs</i>	{ <i>stop_assignment</i> (<i>organ</i>); <i>record</i> ("NOT(<i>done</i> (<i>ensure_compatibility</i> (<i>organ</i> , <i>recipient</i>))) AND <i>done</i> (<i>assign</i> (<i>organ</i> , <i>recipient</i>)))", <i>incident_Log</i>); <i>detect_alarm</i> (<i>ensure_compatibility</i> , ' <i>done</i> '); <i>check</i> (<i>done</i> (<i>ensure_compatibility</i> (<i>organ</i> , <i>recipient</i>))); <i>resume_assignment</i> (<i>organ</i>); }

Figure 1. Example norm

to the enforcers to execute the sanctions and repairs once a violation has occurred.

DEFINITION 3. *Integrity constraints are first-order formulae of the form*

$$\left(\bigwedge_{i=1}^n \text{uttered}(s_i, w_{k_i}, \mathbf{i}_i) \wedge \bigwedge_{j=0}^m e_j \right) \rightarrow \perp$$

where s_i are scene identifiers, w_{k_i} is a state k_i of scene s_i , \mathbf{i}_i is an illocution scheme l_i of scene s_i and e_j are boolean expressions over variables from illocution schemes \mathbf{i}_i .

Integrity constraints define the set of states that should not occur within the e-institution. They express that a certain combination of illocutions (actions) and expressions lead to a violation of a norm. Integrity constraints are derived from the norm frame of definition 2, as the violation condition of the norm frame expresses exactly the left-hand side of the constraint.

DEFINITION 4. *Dialogical constraints are first-order formulae of the form:*

$$\left(\bigwedge_{i=1}^n \text{uttered}(s_i, w_{k_i}, \mathbf{i}_i^*) \wedge \bigwedge_{j=0}^m e_j \right) \Rightarrow \left(\bigwedge_{i=1}^{n'} \text{uttered}(s'_i, w'_{k_i}, \mathbf{i}'_i) \wedge \bigwedge_{j=0}^{m'} e_j \right)$$

where s_i, s'_i are scene identifiers, w_{k_i}, w'_{k_i} are states of scenes s_i and s'_i respectively, $\mathbf{i}_i^*, \mathbf{i}'_i$ are illocution schemes l_i of scenes s_i and s'_i respectively, and e_j, e'_j are boolean expressions over variables from illocution

schemes \dot{i}_i . These boolean expressions can include functions to check the state of the institution.

The idea of dialogical constraints is that it expresses an obligation to the enforcers. A dialogical constraint expresses that if a certain situation (expressed at the left hand side of the constraint) arises, namely the violation of a norm (i.e. the violation condition of definition 2 holds), the enforcer is obliged to see to it that a series of actions is performed (expressed at the right-hand side of the constraint).

The integrity constraints and dialogical constraints introduced in the previous definitions are the building blocks for operationalising norms in the ISLANDER framework. Integrity constraints are implemented in the infrastructure of the e-institutions, thereby providing the means to detect violations of norms, where dialogical constraints are implemented in enforcing agents which use them to determine the illocutions that should be uttered when a norm has been violated.

In order to connect the annotated norms in section 3, all the observable events and actions appearing in the norms should be mapped into utterances. For instance a norm such as $\text{OBLIGED}((a \text{ DO } A) \text{ IF } C)$ should be translated into $\text{OBLIGED}(\text{utter}(S, W, I) \text{ IF } C)$ taking into account that the state S and world W of the e-institution will correspond to the applicable state meant by the norm, and that I is an illocution performed by a to implement A . More details on this translation process can be found in [17]

5. A methodology for electronic institutions

Comprehensible methodologies to design electronic institutions must be able to describe the characteristics of a normative environment (its regulations, its constraints, its organizational structure and its domain language). These methodologies should also guide the translation process from the human regulations to the final mechanisms implemented to enforce the norms in the agent platform. Furthermore, each step of the process should have properly defined formal semantics, Although there are currently some toolkits and some frameworks to build electronic institutions, currently there is no methodology which covers all the aspects, from the specification of the abstract norms to the connection with their implementation (both from the institutional and the agent viewpoints), including the mechanisms to enforce them.

Our approach: in previous work we have presented OMNI (Organizational Model for Normative Institutions), an integrated framework for modelling agent organizations that allows the balance of global organizational requirements with the autonomy of individual agents. OMNI

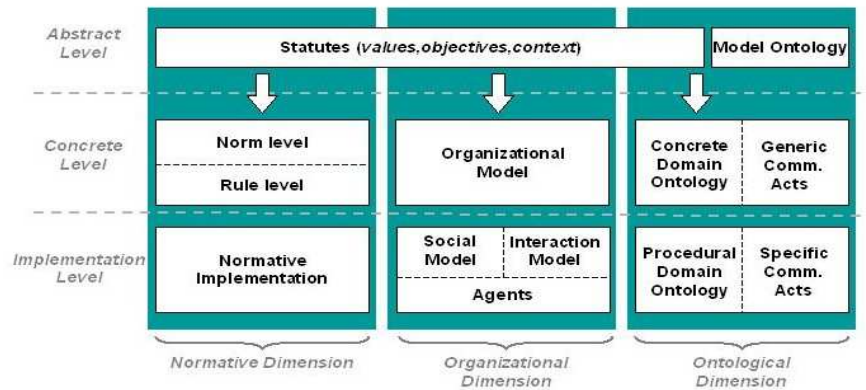


Figure 2. The OMNI framework.

also integrates the norms that regulate interaction between agents, as well as the contextual meaning of those interactions into one framework. OMNI is composed by three dimensions: the Normative Dimension, the Organizational Dimension and the Ontological Dimension.

When applying OMNI for the design of e-institutions, the design process is divided in three abstraction levels (the Abstract Level, the Concrete Level and the Implementation Level), which ease the transition from the very abstract human norms and regulations to the very concrete interaction protocols and enforcement procedures implemented in the system.

The **Abstract Level** is the first step in the process. In order to make a requirement analysis of the problem the *statutes* of the institution to be modelled are defined. Such statutes are composed by:

- a set of the overall *objectives* of the institution (in the form of goal statements),
- a set of *values* (in the form of abstract normative statements) that direct the fulfilling of the objectives, and
- the *context* where the organization performs its activities.

The analysis of the context is important in those scenarios where the behavior of the agents in the electronic institution should comply with regulations and restrictions that are imposed by the context. In these cases the ontologies already in use in the context should be identified, along with any set of regulations that may affect the behavior of the electronic institution. The Abstract level includes a model ontology, which is a meta-ontology defining all the concepts of the framework

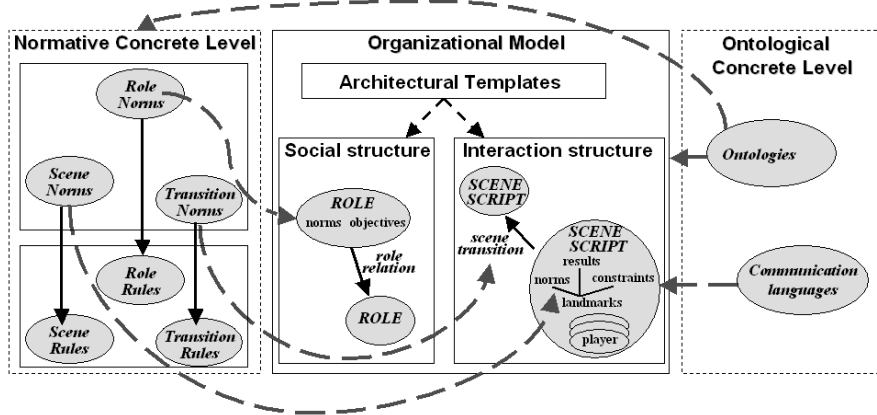


Figure 3. Connection between the multi-level ontology, the norms and the organization structure

itself, such as norms, rules, roles, groups, violations, sanctions and landmarks.

The **Concrete Level** specifies the analysis and design process, starting from the abstract values and objectives defined in the previous level, refining their meaning in terms of norms and rules, roles and concrete ontological concepts. In this level the three dimensions are highly inter-connected: norms and rules in the normative dimension influence the roles, groups and scenes in the organizational model, while all terms are defined in the ontological dimension (see figure 3).

It is in this level where the full list of norms is created, following the annotated format presented in section 3. In parallel the Concrete Domain Ontology is constructed, handling ontological abstractness by means of the counts-as relationship as described in section 2. It is important to note the key role of this ontology, as it gives formal semantics to the terms that appear not only in the norms and rules but also in the definition of the roles, the groups and the actions available within the e-institution.

The **Implementation Level** describes the implementation of the design in a given multi-agent architecture, including the mechanisms for role enactment, the mechanisms for norm enforcement and the implemented ontologies. As we explained in section 4, we use the AMELI platform for execution of e-institutions, so in this level 1) all actions and events in the norms should be translated into utterances and events that can be handled by the AMELI platform, 2) the Social and Interaction Structures should be mapped into roles, scenes and scene scripts in the ISLANDER framework, 3) the Concrete Domain Ontology should

be mapped into the Procedural Domain Ontology and the Specific Communication Acts Ontology (the ontologies to be used at run-time), including the aforementioned mapping of actions into utterances.

6. Conclusions & Ongoing Work

In this paper we have presented our view about the issues to be solved in order to develop successful electronic institutions, able to operate in highly regulated environments, and we have described our approach on each of them. In short, the solution to the open texture of norms consists on the definition of the context of use of the norms (the e-institution), which fixes the interpretation of the norms. The interpretation step is not trivial, but as it is bounded to a given context, an ontology precisely defining an unambiguous interpretation can be defined with the help of legal experts in the domain. Then the declarative aspects of norms should be extended with operational descriptions specifying how norms are checked, and how detected violations to these norms should be handled. Finally norms should be implemented from the institutional perspective by providing operational norm enforcement mechanisms connected to the norms' operational descriptions. The complexity in all these steps in the e-institutions design process requires also the creation of methodologies guiding the designer and covering all the aspects, from the abstract norm specification to the connection with their implementation.

Most of the work presented in this paper is on-going, and we can summarize it in the following research lines: 1) to define a formal language to specify norms for agents, a language should be expressive enough for complex, highly regulated scenarios and machine-parseable; 2) to formally connect a specification of a set of norms (in the above-mentioned language) with an operational specification of the accepted behavior inside an e-Institution; 3) to formally define ontologies which can cope with the abstractness present in human regulations, and which can properly connect such abstract concepts with the concepts used in the e-institution implementation; 4) to refine the OMNI framework to create a methodology and the tools to support designers in the specification, analysis, design and implementation of e-Institutions for highly-regulated environments; and 5) to extend e-Institutions platforms (such as AMELI) with the enforcement mechanisms needed to check compliance of norms by the agents interacting in the platform.

Currently there is a close collaboration with the eInstitutions group at IIIA to extend EIDE by introducing our norm model into ISLANDER, and adding enforcement mechanisms to the AMELI run-time platform.

Acknowledgements

The authors would like to acknowledge the close collaboration of Virginia Dignum, John-Jules Ch. Meyer, Andres García Camino, Juan Antonio Rodríguez Aguilar, Pablo Noriega and Carles Sierra in different stages of this work. Also our work includes ideas coming from valuable discussions with Ulises Cortés, Julian Padget and Owen Cliffe.

References

1. Alchourrón, C. E. and E. Bulygin: 1986, *Normative Systems*. Wien: Springer Verlag.
2. Baader, F., D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider: 2002, *The Description Logic Handbook*. Cambridge: Cambridge University Press.
3. Boella, G. and L. der Torre: 2004, 'Normative multiagent systems'. In: *Proceedings of Trust in Agent Societies Workshop at AAMAS'04*. New York.
4. Boella, G. and L. van der Torre: 2004, 'Fulfilling or Violating Norms in Normative multiagent systems'. In: *Proceedings of IAT 2004*.
5. Broersen, J., F. Dignum, V. Dignum, and J.-J. Ch. Meyer: May 2004, 'Designing a Deontic Logic of Deadlines'. In: *Proceedings of the 7th International Workshop on Deontic Logic in Computer Science (DEON'04)*. Portugal.
6. Bulygin, E.: 1992, 'On Norms of Competence'. *Law and Philosophy* 11 pp. 201–216.
7. Castelfranchi, C., F. Dignum, C. Jonker, and J. Treur: 2000, 'Deliberative Normative Agents: Principles and Architectures'. In: N. Jennings and Y. Lesperance (eds.): *ATAL '99*, Vol. 1757 of *LNAI*. Berlin Heidelberg, pp. 364–378.
8. Dignum, F.: 2001, 'Agents, Markets, Institutions, and Protocols'. In: *Agent Mediated Electronic Commerce, The European AgentLink Perspective*. pp. 98–114.
9. Dignum, F.: 2002, 'Abstract Norms and Electronic Institutions'. In: *Proceedings of the International Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA '02)*, Bologna. pp. 93–104.
10. Dignum, F., J. Broersen, V. Dignum, and J.-J. Ch. Meyer: April 2004, 'Meeting the Deadline: Why, When and How'. In: *3rd Goddard Workshop on Formal Approaches to Agent-Based Systems (FAABS)*. Maryland.
11. Dignum, F., D. Kinny, and L. Sonenberg: 2002a, 'From Desires, Obligations and Norms to Goals'. *Cognitive Science Quarterly* 2(3-4), 407–430.
12. Dignum, V. and F. Dignum: 2001, 'Modelling Agent Societies: Coordination Frameworks and Institutions'. In: P. Brazdil and A. Jorge (eds.): *Progress in Artificial Intelligence*, Vol. 2258 of *LNAI*. pp. 191–204.
13. Dignum, V., J.-J. Meyer, F. Dignum, and H. Weigand: Oct. 2002b, 'Formal Specification of Interaction in Agent Societies'. In: *2nd Goddard Workshop on Formal Approaches to Agent-Based Systems (FAABS)*. Maryland.
14. Esteva, M., J. Padget, and C. Sierra: 2001, 'Formalizing a language for institutions and norms'. In: J.-J. Meyer and M. Tambe (eds.): *Intelligent Agents VIII*, Vol. 2333 of *LNAI*. pp. 348–366.

15. Esteva, M., J. Rodríguez-Aguilar, B. Rosell, and J. Arcos: July 2004a, 'AMELI: An Agent-based Middleware for Electronic Institutions'. In: *Third International Joint Conference on Autonomous Agents and Multi-agent Systems*. New York, US.
16. Esteva, M., W. Vasconcelos, C. Sierra, and J. Rodríguez-Aguilar: 2004b, 'Verifying Norm Consistency in Electronic Institutions'. In: *Proceedings of the AAAI-04 Workshop on Agent Organizations: Theory and Practice (AOTP)*. San Jose, California.
17. García-Camino, A., P. Noriega, and J. Rodríguez-Aguilar: 2005, 'Implementing Norms in Electronic Institutions'. In: E. Schweighofer (ed.): *The 4th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS-05)*. pp. 667–673.
18. Ghidini, C. and F. Giunchiglia: 2001, 'Local models semantics, or contextual reasoning = locality + compatibility.'. *Artificial Intelligence* **127**(2), 221–259.
19. Grossi, D., H. Aldewereld, J. Vázquez-Salceda, and F. Dignum: 2005a, 'Ontological Aspects of the Implementation of Norms in Agent-Based Electronic Institutions'. In: *Proceedings of NorMAS'05*. Hatfield, England.
20. Grossi, D., F. Dignum, and J.-J. C. Meyer: 2005b, 'Contextual Taxonomies'. In: J. Leite and P. Toroni (eds.): *Post-proceedings of CLIMA V, 5th International Workshop on Computational Logic in Multi-Agent Systems*, Vol. 3487 of *LNAI*. pp. 33–51.
21. Grossi, D., F. Dignum, and J.-J. C. Meyer: 2006, 'Contextual Terminologies'. In: F. Toni and P. Torroni (eds.): *Post-proceedings of CLIMA VI, 6th International Workshop on Computational Logic in Multi-Agent Systems*, Vol. 3900 of *LNAI*. pp. 284–302.
22. Hart, H.: 1994, *The Concept of Law*. Oxford 1961, 2nd. ed.
23. Hart, H. L. A.: 1958, 'Positivism and the Separation of Law and Morality'. *Harvard Law Review* **71**, 593–629.
24. Jones, A. J. I. and M. Sergot: 1993, 'On the Characterization of Law and Computer Systems'. *Deontic Logic in Computer Science* pp. 275–307.
25. Lomuscio, A. and D. Nute (eds.): 2004, *Proceedings of the Seventh International Workshop on Deontic Logic in Computer Science (DEON04)*, Vol. 3065 of *LNCS*. Springer Verlag.
26. Rodriguez, J.: 2001, 'On the Design and Construction of Agent-mediated Electronic Institutions'. Ph.D. thesis, Inst. d'Investigació en Intel·ligència Artificial.
27. Ross, A.: 1968, *Directives and Norms*. London: Routledge & Kegan Paul.
28. Searle, J.: 1995, *The Construction of Social Reality*. Free Press.
29. Vázquez-Salceda, J.: 2004, *The role of Norms and Electronic Institutions in Multi-Agent Systems*. Birkhuser Verlag AG.
30. Vázquez-Salceda, J., H. Aldewereld, and F. Dignum: 2004, 'Implementing Norms in Multiagent Systems'. In: G. Lindemann, J. Denzinger, I. Timm, and R. Unland (eds.): *Multiagent System Technologies*, Vol. 3187 of *LNAI*. pp. 313–327.
31. Vázquez-Salceda, J. and F. Dignum: 2003, 'Modelling electronic organizations'. In: J. M. V. Marik and M. Pechoucek (eds.): *Proceedings CEEMAS'03*, Vol. 2691 of *LNAI*. Berlin.

Response to reviewers' comments for ARTI 25

“From Human Regulations to Regulated Software Agents' Behaviour.

Connecting the abstract declarative norms with the concrete operational implementation. A position paper”

Reviewer #1: The article is mainly a report on previous work of the same authors and it does not seem to offer any new result. Moreover, its very sketchy style of presentation makes it very difficult to appraise its technical accuracy as well as other relevant aspects.

The paper is one of the papers coming from the 1st Round Table on Electronic Institutions and Law. Just after the Round Table **we were invited to do a Position Paper**, which should explain the current approach of the authors, the research lines pursued, and also present a list of the most important issues related to norms and institutions, from the authors' point of view. With these directives we re-shaped our contribution to make it fit those directives.

That is why the article is mainly “*a report on previous work*”, as in our view, that is exactly what a "Position Paper" is: an statement of the "position" we take on Institutions' research, our hypotheses, our objectives, our approach. Following the directives we were given, on this paper we explain the work we do, where we come from, the issues we think are important and our past or present work in those issues...

After exchanging some messages with Giovanni Sartor, we have followed his suggestion: in the subtitle, in the abstract and in the conclusions we state clearly that this is a position paper where we state our position with regard to the domain of Electronic Institutions.

Reviewer #2: The paper moves from previous published work.

It discusses how to represent norms equipped with an operational flavour to handle violations and sanctions. Norms implementation and norm enforcement is achieved by adding constraints into ISLANDER.

Finally the OMNI framework is proposed as a methodology for building electronic institutions.

The paper discusses several things (norms, the need for ontologies in normative languages and operational support for them, the need for verifying the compliance of a system of agents to norms, just to recall the main issues), may be too many. The drawback is - in a certain sense - that each issue is discussed at a very superficial level.

I suggest to reduce the set of arguments introduced, and better explain (e.g. ontologies, rather than norms or viceversa).

Please see response to previous reviewer. As one of the directives coming from the Round Table, we were suggested to list the issues we considered the most challenging, the most relevant to the field. Of course, for each of these issues, one could write full chapters.

As this was precisely one of the points we were told to focus on when re-shaping the paper for the special issue, following the suggestions from the editors we have used the extra space that editors have provided us to better clarify those issues, and make proper reference to other works where the interested reader can find more details.

One reference is missing (question mark in text).

This has been fixed.

Reviewer #3: In my opinion this is an interesting and challenging paper, but I do have some questions concerning the assumptions in this paper. The basic assumption of this paper seems to be that an e-institution should completely reflect an off line institution and be governed by the same, abstract norms. It can be highly questioned whether this is feasible, but also whether this is desirable. If literature on AI and Law is considered we see that exactly the open texturedness and vagueness of norms offers problems for building legal knowledge based systems that are almost impossible to overcome. It looks like the authors are falling in this same pitfall. They do not clarify why this is necessary and why not other directions can be taken.

The work presented in this section actually has standard work on open texture [H.L.A. Hart, *The Concept of Law* (Oxford 1961; 2nd edn. 1994)] as starting point. Main difference between the open texturedness problem in other legal systems and in Electronic Institutions is that in Law systems people want to **automate rules of interpretation for any specific situation** in which a norm (with open texture) might be applied. In our application we actually can **define the interpretation of the open texture for the specific context in which the institution functions**. So, we don't need to dynamically adapt this interpretation based on the context anymore. That is fixed by the institution. The interpretation step is not trivial, but can be made (as is done by the judges in court), by having legal experts helping the designer to introduce the proper interpretation in the system.

So, in short, our solution to the open texture interpretation problem is to define the context of the use of a norm through the definition of an (electronic) institution in such a way that the interpretation can be made unambiguous. The specification of the institution (including an ontology that defines an unambiguous contextual interpretation of the abstract norms) is exactly our proposed solution for the problem.

We have introduced all these clarifications in the text, especially in section 2.

For instance:

```
p.3 I3: `...Agent platforms should be extended with mechanisms to detect attempts from agents to break the norms': of course, but how and when and which norms?
```

The *how* and *when* questions are answered in section 4 of the paper.

There are other solutions possible. One could consider only to accept authorized or safe agents on the platform.

Accepting only authorized and/or safe agents would require a very strong governance of the interactions within the institution in order to ensure that no deviations occur. This is what was originally proposed in *Islander*, but which proves not feasible in all cases and very

inefficient in other cases. Even if we have trusted agents, the platform should be able to detect (unintentional) norm violations from these agents (e.g. an agent not detecting that a given norm applied to the current situation, and therefore unwillingly breaking it).

So, we do not reject other solutions, but provide an alternative for situations where those solutions are not usable.

The paragraph on p. 3 "In our opinion, all four issues..... to self insure proper behavior" is unclear. Do the institutions have to take these issues into account during design phase? What is meant?

The main idea is that this depends on the application domain: on its openness, on whether the impact of breaking a norm may be catastrophic or not, etc. We have added some clarification on this in the text (see page 3).

p. 4 "What counts as personal data... of information as personal data" : Hospital and police register can (and probably should) be separated and not be consulted by the same agents. Especially using agents it is very well possible not to mix these registers and build agents that are defined and customized for one particular task. Safety issues can be taken into account. Further: it can be questioned whether the normative questions attached to the classification of information as personal data can be shared. I would be willing to discuss this, since at least part of it depends on the interpretation of concepts.

The way that example was explained in the text led to confusion. We have fixed this. Following our approach, the normative questions attached to the classification of information as personal data can be shared within the e-institution, as within this context there is a shared interpretation of the term that intervening agents should accept: the one defined by the institution itself. This is linked to our previous response on the open texture problem.

p. 5 "As A result implementation level" Does this not specifically ask for a clear(er) definition of the norms? This would contradict the assumptions of the authors.

As the reviewer states, yes, a clearer definition of the norms is created during the process of designing the e-institution. But this does not contradict our assumptions; on the contrary, it is the basic idea: as norms are vague and abstract, one should create a clearer interpretation of them in the context of the e-institution, and also provide an operational description of such norms that can be then checked by the institutional platform. In our paper we present our proposals for each step in that process.

p. 8 'the real problem lies in determining when the officer has too much information' I think the real problem is not so much that he has too much information, but how he obtained the information.

Fixed. We have introduced a better explanation of the example.

p. 13: what are the statutes of the institution?

Fixed. We have added an explanation in the text that basically explains that statutes are composed by:

- a set of the overall *objectives* of the institution (in the form of goal statements),
- a set of *values* (in the form of abstract normative statements) that direct the fulfilling of the objectives, and
- the *context* where the organization performs its activities.

From Human Regulations to Regulated Software Agents' Behaviour

Connecting the abstract declarative norms with the concrete operational implementation. A position paper

Javier Vázquez-Salceda (jvazquez@lsi.upc.edu)

Knowledge Engineering and Machine Learning Group, Universitat Politècnica de Catalunya, Barcelona, Spain

Huib Aldewereld (huib@cs.uu.nl), Davide Grossi

(davide@cs.uu.nl) and Frank Dignum (dignum@cs.uu.nl)

Institute of Information and Computing Sciences, Utrecht University, The Netherlands

May 30, 2006

Abstract. In order to design and implement electronic institutions that incorporate norms governing the behavior of the participants of those institutions, some crucial steps should be taken. The first problem is that human norms are (on purpose) specified on an abstract level. This ensures applicability of the norms over long periods of time in many different circumstances. However, for an electronic institution to function according to those norms, they should be concrete enough to be able to check them run time. A second problem is that norms describe which behavior is desirable and permitted, but not how this is achieved in an institution. In the "real world" regulations often indicate procedures for implementing and enforcing the law. Likewise we should devise means to annotate the norms with practical aspects such as enforcement mechanisms, sanctions, etc. in order to get requirements for an institution that will enforce norms (by either constraining behavior within the norms or reacting to violation of the norms). The choice of which kind of mechanism is chosen is not a normative one, but usually based on criteria of efficiency and/or feasibility of the mechanism. In this paper we present our view on how to approach these problems and other related issues to be solved in order to develop e-institutions capable to operate in complex, highly regulated scenarios.

Keywords: Multi agent systems, electronic institutions, norms, norm enforcement

1. Introduction

Internet, as an extension of the real world, is affected by the regulations of one or several countries on activities carried out through the web. For instance, Electronic Commerce activities between two parties are regulated by the law of the parties' countries plus international commerce treaties. In the Health Care field, highly regulated, the citizens' rights are precisely defined and regulated by national and international laws. How to make sure that such norms and regulations are met on the activities and information exchanges through Internet? How to create

© 2006 Kluwer Academic Publishers. Printed in the Netherlands.

mechanisms to enforce norm compliance, and therefore, increase trust between individuals and companies?

In most software and agent methodologies, these external regulations, along with the internal norms and regulations of the organization to be modelled, are seen only as extra requirements in the analysis phase of the system. If either the external or the internal regulations change (as they usually do from time to time), it becomes very hard to track all the changes to be done in the implementation, as there is no explicit representation of the norms and regulations, but a chain of design decisions that were guided by the norms' requirements (i.e., if norms are embedded in the agents' design and code, all the design steps have to be checked again and all the code verified to ensure compliance with new regulations). The alternative is to have an explicit representation of the norms.

Research on Distributed Artificial Intelligence has created the concept of Electronic Institutions. As their human counterparts, an Electronic Institution is an entity defining a set of norms over the behavior of individuals inside the institution. Recently research in electronic institutions has focused on the use of Software Agent technology. An *Agent-Mediated Electronic Institution* (**e-institution** for short) belongs to a new and promising field where interactions between a group of (software) agents are regulated by means of a corpora of explicit norms, expressed in a computational language that agents can interpret. An e-institution [?, ?] is a safe environment mediating in the interaction of agents. The expected behavior of agents in such an environment is described by means of an explicit specification of norms, which is a) expressive enough, b) readable by agents, and c) easy to maintain.

1.1. CHALLENGES IN THE FIELD

In our view, there are basically four issues to be solved in order to successfully design e-institutions in complex, highly regulated scenarios:

- I1 **abstractness of human regulations:** human regulations are written in a very abstract way, open to several interpretations to make the legal text stable for a long time. This abstraction poses a problem when trying to implement them on computers, where meanings should be precise and unambiguous.
- I2 **operationalisation of norms:** usually norms are formally specified in languages based in deontic logic, which is expressive enough to cover some of the concepts that appear in regulations (obligations, permissions, prohibitions), has formal semantics and allows

the verification of sets of norms. However, deontic-based languages are declarative in nature and do not fully express the operational impact of the norms in the behavior of the multiagent system.

- I3 implementation of norms from the institutional perspective** Agent platforms should be extended with mechanisms to detect attempts from the agents to break the norms. This issue is not only relevant in open scenarios where non-trusted agents may enter into the platform, but also in scenarios with trusted agents, as sometimes these agents may break the norms unwillingly (e.g. not detecting that a norm applied in a given situation).
- I4 methodology to design electronic institutions** Although there are some toolkits and some frameworks to build electronic institutions, currently there is no methodology which covers all the aspects, from the specification of the abstract norms to the connection with their implementation, both from the institutional and the agent viewpoints.

In our opinion, all four issues above are very important to create agent-mediated electronic institutions able to take into account domain regulations during the design phase and also to enforce compliance of agent behavior to those regulations during run-time. Depending on the characteristics of the application domain (openness, frequency of regulations' change, potential negative impact of breaking a norm) some of these issues will be more or less relevant. For instance, in the Health Care domain, some norms may be breakable without harm to a given patient, while others may endanger the patient's health. An additional issue would be to have agents which can understand a given set of norms and be able to reason the suitability of their behavior against the norms. Although it would be desirable to have such agents it is not crucial for the success of e-institutions, as by issue I3 regulations' compliance is ensured by the institution, and we cannot assume in an open scenario that all agents interacting within an e-institution would be capable of normative reasoning to self-ensure proper behavior.¹

The rest of the paper is organized as follows. Our view on I1 is explained in section 2. Our approach on I2 and I3 is explained in sections 3 and 4. Our view on I4 is briefly explained in section 5. Finally in section 6 we summarize our approach and point to some ongoing work.

¹ The reader interested in the issue of norm-aware agents can find more details in [?].

2. Explaining abstractness in human regulations

In [?, ?] it has been variously stressed how the design of norm-governed multi-agent systems has to cope with the inherent abstractness of norm formulations. Human regulations are usually written in a quite abstract language and are open to interpretation. The main reason for this is to cover with the same legal text the major number of cases and therefore be stable for longer periods of time. This abstraction and capability of multiple interpretations that are positive for humans pose a problem when trying to implement them on computers, where meanings should be precise and unambiguous. Attempts in the past to build Legal Knowledge-Based Systems capable to automate rules of interpretation for any situation had to confront the *open texture* of norms [?]. In the case of e-institutions the open texture interpretation problem is reduced by fixing a context for interpretation (the system does not need to handle all possible interpretations of a norm, but the interpretation which is valid in the context of the e-Institution). Therefore, when designers try to include the norms in the design process of an e-Institution, at least two steps should be performed:

- an interpretation of the norms in the context of the e-institution should be provided (usually by close collaboration with legal experts on the domain), and
- such interpretation should be then connected to the processes and structure of the implemented e-institution.

The second step is explained in section 3. First step of the problem can be distilled in the question:

How are norms, which are specified by means of abstract terms (e.g. *“personal data which are not strictly relevant for the transplant activities should not be included in the transplant data base”*), connected to norms specified instead via more concrete ones (e.g. *“data about age may be included in the transplant data base”*)?

Our approach: in previous work [?, ?] we have focused on the formal definition of norms by means of some variations of deontic logic that include conditional and temporal aspects [?, ?], and we provided formal semantics. However there are complex ontological issues to be fixed in order to solve the open texture interpretation problem. Our view is that Institutions provide structured interpretations of the concepts in which norms are stated. To put it in a nutshell, institutions do not only consist of norms, but also of *ontologies* of the to-be-regulated domain. For instance, whether something within a given institution

counts as personal data and should be treated as such depends on how that institution interprets the term 'personal data'.

This perspective on institutions, which emphasizes the semantic dependencies between abstract and concrete norms, goes hand in hand with acknowledged positions in the study of social reality and legal systems. In fact, institutions can be seen as complex systems of norms, which consist of regulative as well as non-regulative components², that is to say, which do not only regulate existing forms of behavior, but they actually specify and constitute -via classification- new forms of behavior. The abstractness of norms is precisely the effect of such a "constitution": non-regulative components of institutions connect concrete concepts, such as 'age' to abstract ones such as 'personal data'. The basic brick of this constitution consists of statements of the following general logical form: "X counts as Y in context C" [?, ?]. Two ingredients are displayed in this sentential form: a relation between two concepts X and Y, (for instance, 'age' and 'personal data'), and a relativization of it to a specific context C (for instance, 'national hospitals')³. In [?, ?, ?], we proposed, investigated and applied a framework for formally representing such statements, where the relation between X and Y is interpreted as a standard concept subsumption, but which holds only in relation to a context C: "age is a subconcept of personal data in the context of national transplantation policies". Counts-as statements are therefore studied as contextual subsumption relations: $C : X \subseteq Y$. The key idea of the framework consists in tailoring the semantic machinery of Description Logic [?], with the framework developed in [?] to model context in logic. As a result, it becomes possible to formally represent, in a framework based on computationally appealing logics, how concepts expressed in the abstract language of the norms, can be related to concepts belonging to more specific languages, and in particular, to the language used at the implementation level.

In addition, this approach enables a further noteworthy feature consisting in the possibility of representing a number of fundamental notions for the understanding of contextual ontologies, such as the notions of *core* and *penumbra* of the meaning of a concept which we borrowed from legal theory [?] and which we formally analyzed in [?]:

[Suppose a] legal rule forbids you to take a vehicle into the public park. Plainly this forbids an automobile, but what about bicycles, roller skates, toy automobiles? What about airplanes? Are these, as we say, to be called "vehicles" for the purpose of the rule or not?

² See for example [?], [?], [?], [?] and [?].

³ The importance of the notion of context for specifying and representing institutions has been advocated also in [?, ?].

[...] There must be a *core* of settled meaning, but there will be, as well, a *penumbra* of debatable cases in which words are neither obviously applicable nor obviously ruled out.

The possibility of capturing these notions in a rigorous setting, allows to properly represent the domain by ontologies that provide the connection from the abstract terms in the human regulations to the ones used in the e-institution. As we will see in section 5, the creation of such ontologies should be a key aspect during the formal specification process of the regulations (and the terms appearing in them) that will impact the behavior of agents within the e-institution.

3. From Declarative to Operational

Current work on normative systems' formalization (mainly focused in Deontic-like formalisms [?]) is declarative in nature, focused on the expressiveness of the norms, the definition of formal semantics and the verification of consistency of a given set. The declarative aspects of norms are important to check whether a given action or situation is permitted, obliged or forbidden (at a certain moment). However, norms also have an operational aspect, as norms should guide (or at least influence) the behavior of the agents. Although norms can be implemented by just putting enough constraints on the behavior of the agents such that the norms will never be violated, one has to define which constraints are to be used exactly, and who (or what) will check that the constraints always hold.

Our approach: in our view, in order to implement norms it is not enough to include a model-checking module by, e.g., implementing a theorem prover that, using the norms semantics, checks whether a given interaction protocol complies with the norms. The implementation of norms should also consider a) how the agents' behavior is affected by norms, and b) how the institution should ensure the compliance with norms. The former is related to the *implementation of norms from the agent perspective*, by analyzing the impact of norms in the agents' reasoning cycle (work on this perspective can be found in [?, ?, ?]). The latter is related to the *implementation of norms from the institutional perspective*, by implementing a safe environment (including the enforcing mechanisms) to ensure trust among parties.

As we discussed in [?], we need to specify the operational semantics of the norms to ensure this institutional perspective of norms. In general an operational semantics for norms always comes down to either one of the following: 1) **Defining constraints on unwanted behavior**; or 2) **Detecting violations and reacting to these violations**.

The choice between these two approaches is highly dependent on the amount of control over the addressee of the norms. Preventing unwanted behavior can only be achieved if there is full control over the addressee (as the constraints are programmed into the agents beliefs and goals); in cases where such control is lacking one should define and handle violations. As we assume the external agents to be black-boxes, i.e. the internal states of the agents cannot be observed nor controlled, we will focus on the latter method of enforcement.

In order to detect and react to violations, norms need to have a declarative as well as an operational meaning. For this purpose we extend normal deontic representations of norms with extra fields to express the operational aspects of the norm. For expressing the declarative part of the norm any kind of deontic logic can be used, however, since the norms need to be useable by agents, a machine-readable format should be used. We proposed in [?] to use the following format for expressing norms that can be conditional or have temporal aspects:

DEFINITION 1 (Norm Condition).

$$\begin{aligned} \text{NORM_CONDITION} &:= N(a, S \langle \text{IF } C \rangle) \mid \\ &\quad \text{OBLIGED}(a \text{ ENFORCE}(N(a, S \langle \text{IF } C \rangle))) \\ N &:= \text{OBLIGED} \mid \text{PERMITTED} \mid \text{FORBIDDEN} \\ S &:= P \mid \text{DO } A \mid P \text{ TIME } D \mid \text{DO } A \text{ TIME } D \\ C &:= \textit{formula} \\ P &:= \textit{predicate} \\ A &:= \textit{action expression} \\ \text{TIME} &:= \text{BEFORE} \mid \text{AFTER} \end{aligned}$$

This machine-readable declarative representation lacks operational meaning, as it includes no information about violation management, detecting the activation and deactivation of norms, and responses to violations of norms. These elements are necessary for the implementation of norm enforcement as the process of enforcement is composed of the following sub-processes:

- a) the detection of when a norm is active,
- b) the detection of a violation on a norm, and
- c) the handling of the violations.

Before it can be determined whether a norm is violated, it has to be checked whether the norm is actually active. Detecting the *activation of the norm* (when the condition C holds) and the *deactivation of the norm* (when predicate P or action A is fulfilled or C does not hold) are related to the declarative meaning of the norms. An additional issue is to establish the allowed reaction time between the activation

and deactivation of obligations, i.e. the time that is allowed for the completion of the obligation when it becomes active but when the reaction time has passed.

Deadlines represent a special case in the implementation of conditional norms, as they are not that easy to check. Deadlines require a continuous check (second by second) to detect if a deadline is due. If the institution has lots of deadlines to track, it will become computationally expensive. We proposed in [?] to include within the agent platform a **clock trigger mechanism** that sends a signal when a deadline has passed. The idea is to implement the clock mechanism as efficiently as possible to avoid burden on the agents.

The second sub-process of the enforcement is about detecting the actual violation of the norm (for which the norm must be active). In an agent platform with several agents performing different actions at the same time, a question arises on how to implement the detection of the occurrence of actions. The agents enforcing norms may become overloaded on trying to check any action at any time. To solve this problem we proposed in [?] to create two platform mechanisms: 1) a **black list mechanism** of actions to be checked, and 2) an **action alarm mechanism** that triggers an alarm when a given action on the black list attempts to start or is done. This trigger mechanism has to do no further checks, only to make sure the enforcer agent is aware of the occurrence of the action. The action alarm can only be done with actions defined in the e-institutions' ontology.

It is easy to see that a protocol or procedure satisfies a norm when no violations occur during the execution of the protocol. The real problem in norm checking lies, however, in determining when that violation occurs. For instance, in criminal investigations, a police officer should not have more (sensitive or private) information than needed for the investigation. So an officer is doing fine as long as no violation occurs (i.e. he does not have unneeded information). The real problem is that the "need for information" is difficult to ascertain, unless the system is implemented in a way where, e.g., officers should indicate how the requested information contributes to a given case. In some cases this could be formally checked at run-time, while in other cases it might be marked as *'to be checked'* and then checked by people afterwards.

Therefore, the implementation of norm enforcement is dependent on two properties of the checks to be done: a) the checks being *verifiable* (i.e. a condition or action that can be machine-verified from an institutional point of view, given the time and resources needed) and b) the checks being *computational* (i.e. a condition or action that can be checked on any moment in a fast, low cost way). Using these two

properties, we can analyse their impact on the implementation of norm enforcement (explained in more detail in [?]):

- **Norms computationally verifiable:** verification of all predicates and actions can be done easily, all the time.
- **Norms not computationally verifiable directly, but by introducing extra resources:** the condition or action is not directly (easily) verifiable, but can be so by adding extra data structures and/or mechanisms to make it easy to verify. The *action alarm* and *clock trigger* mechanisms are examples of extra resources.
- **Non-computationally verifiable** the check can be machine-verified but it is too time/resource consuming to be verified at any time. Verification should not be done all the time, but can be delayed, done periodically or at random.
- **Observable from the institutional perspective, but not machine-verifiable:** that is, verifiable by other (human) agents that have the resources and/or information needed. Such checks should be delegated appropriately.
- **Indirectly observable from the institutional perspective:** These can be internal conditions, internal actions (like reasoning) or actions which are outside the ability of the system to be observed or detected. To solve this, other observable conditions or actions might be found and used to (indirectly) detect a violation.

The declarative norm representation of definition 1 is extended using these operational aspects to encapsulate the operational meaning of the norm. This gives us the following norm frame:

DEFINITION 2 (Norms).

```

NORM := NORM_CONDITION
      VIOLATION_CONDITION
      DETECTION_MECHANISM
      SANCTION
      REPAIRS
VIOLATION_CONDITION := formula
DETECTION_MECHANISM := {action expressions}
SANCTION := PLAN
REPAIRS := PLAN

```

$$\text{PLAN} := \text{action expression} \mid \\ \text{action expression} ; \text{PLAN}$$

In this format, the *norm condition*-field is denoting when the norm becomes active and when it is achieved. The *violation condition* is a formula derived from the norm to express when a violation occurs (e.g. for the norm $\text{OBLIGED}((a, P) \text{ IF } C)$ this is exactly the state when C occurs and P does not, that is, the state where the norm is active, but not acted upon). The *detection mechanism* is a set of actions that can be used to detect the violation. The set of actions contained in the *sanction*-field is actually a plan which should be executed when a violation occurs (which can contain imposing fines, expulsing agents from the system, etc.). Finally, the *repairs* contains a plan of action that should be followed in order to ‘undo’ the violation. An example of an annotated norm is shown in figure 1.

4. Implementing Norms

Given the operational representation of a norm, how does one proceed to implementing that norm in an e-institution (such as ISLANDER, [?]). What elements in the e-institution are needed for this implementation, given that the norm enforcement is supposed to be done by a distributed set of (internal) agents? And how do the norms represented in the formalism of definition 2 translate to these enforcement mechanisms?

Our approach: The ISLANDER formalism [?] provides a formal framework for institutions [?]. This formalism views an agent-based institution as a *dialogical system* where all the interactions inside the institution are a composition of multiple dialogic activities (message exchanges). The messages (or *illocutions*) are structured through agent group meetings called *scenes* that follow well-defined protocols. Furthermore, the AMELI platform [?] allows the execution of e-institutions, based on the rules provided by ISLANDER specifications, wherein external agents may participate.

To implement norm enforcement in the ISLANDER framework, we need to introduce mechanisms for detecting violations and expressing the actions that have to be taken by the internal agents upon the detection of a violation (i.e. the sanctions and repairs). The former is done by specifying *integrity constraints* (derived from [?]), which are used by the system to detect and register the violations of norms. The latter is expressed by *dialogical constraints*, which are in fact obligations to the enforcers to execute the sanctions and repairs once a violation has occurred.

Figure 1. Example norm

DEFINITION 3. *Integrity constraints are first-order formulae of the form*

$$\left(\bigwedge_{i=1}^n \text{uttered}(s_i, w_{k_i}, \dot{\mathbf{u}}_i) \wedge \bigwedge_{j=0}^m e_j \right) \rightarrow \perp$$

where s_i are scene identifiers, w_{k_i} is a state k_i of scene s_i , $\dot{\mathbf{u}}_i$ is an illocution scheme l_i of scene s_i and e_j are boolean expressions over variables from illocution schemes $\dot{\mathbf{u}}_i$.

Integrity constraints define the set of states that should not occur within the e-institution. They express that a certain combination of illocutions (actions) and expressions lead to a violation of a norm. Integrity constraints are derived from the norm frame of definition 2, as the violation condition of the norm frame expresses exactly the left-hand side of the constraint.

DEFINITION 4. *Dialogical constraints are first-order formulae of the form:*

$$\left(\bigwedge_{i=1}^n \text{uttered}(s_i, w_{k_i}, \dot{\mathbf{u}}_i^*) \wedge \bigwedge_{j=0}^m e_j \right) \Rightarrow \left(\bigwedge_{i=1}^{n'} \text{uttered}(s'_i, w'_{k_i}, \dot{\mathbf{u}}_i'^*) \wedge \bigwedge_{j=0}^{m'} e_j \right)$$

where s_i, s'_i are scene identifiers, w_{k_i}, w'_{k_i} are states of scenes s_i and s'_i respectively, $\dot{\mathbf{u}}_i^*, \dot{\mathbf{u}}_i'^*$ are illocution schemes l_i of scenes s_i and s'_i respectively, and e_j, e'_j are boolean expressions over variables from illocution schemes $\dot{\mathbf{u}}_i$. These boolean expressions can include functions to check the state of the institution.

The idea of dialogical constraints is that it expresses an obligation to the enforcers. A dialogical constraint expresses that if a certain situation (expressed at the left hand side of the constraint) arises, namely the violation of a norm (i.e. the violation condition of definition 2 holds), the enforcer is obliged to see to it that a series of actions is performed (expressed at the right-hand side of the constraint).

The integrity constraints and dialogical constraints introduced in the previous definitions are the building blocks for operationalising norms in the ISLANDER framework. Integrity constraints are implemented in the infrastructure of the e-institutions, thereby providing the means to detect violations of norms, where dialogical constraints are implemented in enforcing agents which use them to determine the illocutions that should be uttered when a norm has been violated.

In order to connect the annotated norms in section 3, all the observable events and actions appearing in the norms should be mapped into utterances. For instance a norm such as $\text{OBLIGED}((a \text{ DO } A) \text{ IF } C)$ should be translated into $\text{OBLIGED}(\text{utter}(S, W, I) \text{ IF } C)$ taking into account that the state S and world W of the e-institution will correspond to the applicable state meant by the norm, and that I is an illocution performed by a to implement A . More details on this translation process can be found in [?]

5. A methodology for electronic institutions

Comprehensible methodologies to design electronic institutions must be able to describe the characteristics of a normative environment (its regulations, its constraints, its organizational structure and its domain language). These methodologies should also guide the translation process from the human regulations to the final mechanisms implemented to enforce the norms in the agent platform. Furthermore, each step of the process should have properly defined formal semantics. Although there are currently some toolkits and some frameworks to build electronic institutions, currently there is no methodology which covers all the aspects, from the specification of the abstract norms to the connection with their implementation (both from the institutional and the agent viewpoints), including the mechanisms to enforce them.

Our approach: in previous work we have presented OMNI (Organizational Model for Normative Institutions), an integrated framework for modelling agent organizations that allows the balance of global organizational requirements with the autonomy of individual agents. OMNI also integrates the norms that regulate interaction between agents, as well as the contextual meaning of those interactions into one framework.

Figure 2. The OMNI framework.

OMNI is composed by three dimensions: the Normative Dimension, the Organizational Dimension and the Ontological Dimension.

When applying OMNI for the design of e-institutions, the design process is divided in three abstraction levels (the Abstract Level, the Concrete Level and the Implementation Level), which ease the transition from the very abstract human norms and regulations to the very concrete interaction protocols and enforcement procedures implemented in the system.

The **Abstract Level** is the first step in the process. In order to make a requirement analysis of the problem the *statutes* of the institution to be modelled are defined. Such statutes are composed by:

- a set of the overall *objectives* of the institution (in the form of goal statements),
- a set of *values* (in the form of abstract normative statements) that direct the fulfilling of the objectives, and
- the *context* where the organization performs its activities.

The analysis of the context is important in those scenarios where the behavior of the agents in the electronic institution should comply with regulations and restrictions that are imposed by the context. In these cases the ontologies already in use in the context should be identified, along with any set of regulations that may affect the behavior of the electronic institution. The Abstract level includes a model ontology, which is a meta-ontology defining all the concepts of the framework itself, such as norms, rules, roles, groups, violations, sanctions and landmarks.

Figure 3. Connection between the multi-level ontology, the norms and the organization structure

The **Concrete Level** specifies the analysis and design process, starting from the abstract values and objectives defined in the previous level, refining their meaning in terms of norms and rules, roles and concrete ontological concepts. In this level the three dimensions are highly inter-connected: norms and rules in the normative dimension influence the roles, groups and scenes in the organizational model, while all terms are defined in the ontological dimension (see figure 3).

It is in this level where the full list of norms is created, following the annotated format presented in section 3. In parallel the Concrete Domain Ontology is constructed, handling ontological abstractness by means of the counts-as relationship as described in section 2. It is important to note the key role of this ontology, as it gives formal semantics to the terms that appear not only in the norms and rules but also in the definition of the roles, the groups and the actions available within the e-institution.

The **Implementation Level** describes the implementation of the design in a given multi-agent architecture, including the mechanisms for role enactment, the mechanisms for norm enforcement and the implemented ontologies. As we explained in section 4, we use the AMELI platform for execution of e-institutions, so in this level 1) all actions and events in the norms should be translated into utterances and events that can be handled by the AMELI platform, 2) the Social and Interaction Structures should be mapped into roles, scenes and scene scripts in the ISLANDER framework, 3) the Concrete Domain Ontology should be mapped into the Procedural Domain Ontology and the Specific

Communication Acts Ontology (the ontologies to be used at run-time), including the aforementioned mapping of actions into utterances.

6. Conclusions & Ongoing Work

In this paper we have presented our view about the issues to be solved in order to develop successful electronic institutions, able to operate in highly regulated environments, and we have described our approach on each of them. In short, the solution to the open texture of norms consists on the definition of the context of use of the norms (the e-institution), which fixes the interpretation of the norms. The interpretation step is not trivial, but as it is bounded to a given context, an ontology precisely defining an unambiguous interpretation can be defined with the help of legal experts in the domain. Then the declarative aspects of norms should be extended with operational descriptions specifying how norms are checked, and how detected violations to these norms should be handled. Finally norms should be implemented from the institutional perspective by providing operational norm enforcement mechanisms connected to the norms' operational descriptions. The complexity in all these steps in the e-institutions design process requires also the creation of methodologies guiding the designer and covering all the aspects, from the abstract norm specification to the connection with their implementation.

Most of the work presented in this paper is on-going, and we can summarize it in the following research lines: 1) to define a formal language to specify norms for agents, a language should be expressive enough for complex, highly regulated scenarios and machine-parseable; 2) to formally connect a specification of a set of norms (in the above-mentioned language) with an operational specification of the accepted behavior inside an e-Institution; 3) to formally define ontologies which can cope with the abstractness present in human regulations, and which can properly connect such abstract concepts with the concepts used in the e-institution implementation; 4) to refine the OMNI framework to create a methodology and the tools to support designers in the specification, analysis, design and implementation of e-Institutions for highly-regulated environments; and 5) to extend e-Institutions platforms (such as AMELI) with the enforcement mechanisms needed to check compliance of norms by the agents interacting in the platform.

Currently there is a close collaboration with the eInstitutions group at IIIA to extend EIDE by introducing our norm model into ISLANDER, and adding enforcement mechanisms to the AMELI run-time platform.

Acknowledgements

The authors would like to acknowledge the close collaboration of Virginia Dignum, John-Jules Ch. Meyer, Andres García Camino, Juan Antonio Rodríguez Aguilar, Pablo Noriega and Carles Sierra in different stages of this work. Also our work includes ideas coming from valuable discussions with Ulises Cortés, Julian Padget and Owen Cliffe.