

Artif Intell Law (2008) 16:89–105  
DOI 10.1007/s10506-007-9055-z

ORIGINAL PAPER

## Artificial institutions: a model of institutional reality for open multiagent systems

Nicoletta Fornara · Francesco Viganò ·  
Mario Verdicchio · Marco Colombetti

Published online: 26 July 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** Software agents' ability to interact within different open systems, designed by different groups, presupposes an agreement on an unambiguous definition of a set of concepts, used to describe the context of the interaction and the communication language the agents can use. Agents' interactions ought to allow for reliable expectations on the possible evolution of the system; however, in open systems interacting agents may not conform to predefined specifications. A possible solution is to define interaction environments including a normative component, with suitable rules to regulate the behaviour of agents. To tackle this problem we propose an application-independent metamodel of artificial institutions that can be used to define open multiagent systems. In our view an artificial institution is made up by an ontology that models the social context of the interaction, a set of authorizations to act on the institutional context, a set of linguistic conventions for the performance of institutional actions and a system of norms that are necessary to constrain the agents' actions.

---

Partially supported by Swiss National Science Foundation project 200021-100260, "An Open Interaction Framework for Communicative Agents".

---

N. Fornara (✉) · F. Viganò · M. Colombetti  
Università della Svizzera italiana, via G. Buffi 13, 6900 Lugano, Switzerland  
e-mail: [nicoletta.fornara@lu.unisi.ch](mailto:nicoletta.fornara@lu.unisi.ch)

F. Viganò  
e-mail: [francesco.vigano@lu.unisi.ch](mailto:francesco.vigano@lu.unisi.ch)

M. Verdicchio  
Università degli Studi di Bergamo, via Salvecchio 19, Bergamo 24100, Italy  
e-mail: [mario.verdicchio@unibg.it](mailto:mario.verdicchio@unibg.it)

M. Colombetti  
Politecnico di Milano, piazza Leonardo Da Vinci 32, Milano, Italy  
e-mail: [marco.colombetti@polimi.it](mailto:marco.colombetti@polimi.it)

**Keywords** Artificial institutions · Open systems · Norms · Conventions · Commitments

## 1 Introduction

Since the advent of the Internet, about ten years ago, the landscape of computer science has substantially changed. For the first time, it has become feasible to develop open distributed systems, that is, software systems whose components can aggregate dynamically and are free to enter and leave an interaction at their will. In an open distributed system, the interacting agents are typically designed and implemented by different parties, and may represent conflicting interests, as it happens for example in e-commerce applications. This fact has two important consequences. The first is that interaction will not even be possible unless agents are designed to comply with well-defined standards. The second consequence is that interaction will not lead to coherent outcomes unless the agents' behaviour is suitably regulated. The first problem (i.e., the problem of standards) has been tackled by a number of organizations, including FIPA, the Foundation for Intelligent Physical Agents (<http://www.fipa.org>). The second problem (i.e., the problem of regulating interactions) is much more elusive, and has become an important object of research at least since Noriega's and Sierra's works on electronic institutions in the late nineteen-nineties (Noriega and Sierra 1996; Noriega 1997). Since then, several authors (see for example Esteva et al. 2001; Artikis et al. 2002; Vasconcelos et al. 2002; Vázquez-Salceda et al. 2005; Esteva et al. 2004a; Cliffe et al. 2006) have contributed to the specification of electronic institutions and artificial institutions (Fornara et al. 2007; Viganò et al. 2006).

An electronic institution is usually viewed as a set of norms that the agents interacting within an open distributed system ought to follow (Noriega 1997). Interaction protocols themselves can be conceived as sets of norms (about what an agent may do and when), and are the most obvious component of an electronic institution (Esteva et al. 2004a). As a whole, the function of an electronic institution is to guarantee that if its norms are followed by all agents, the interaction will produce a desirable outcome. Indeed, it is part of the very concept of an autonomous agent that norms may be violated; the system implementing an electronic institution ought to detect such violations and to manage the situation, for example by applying suitable sanctions. Given that norms are an essential component of social reality (see for example Searle 1995), we can regard electronic institution as a means for imposing a well-defined structure to the social reality within which agents interact. However, norms are just one component of social reality; other components, which also seem to us to be important for the specification of open distributed systems, have been largely neglected by most proposals concerning electronic institutions. For this reason we would like to put forward an extension of the concept of an electronic institution, that we call "artificial institution".

Social reality is that part of the world that exists only because it is collectively accepted as existing by a group of agents. Norms are an obvious example of something that exists only because it is commonly recognized by the members of

some community; other important components are: (i) a socially defined ontology, and (ii), linguistic conventions. Let us briefly analyse these components.

Institutions not only regulate, but also create components of social reality. For example, the institution of ownership does not just regulate a pre-existing piece of reality, but creates the very concept of owning something. In other words, an institution introduces a new ontology, including a set of entities with their properties and relationships. A very important class of entities created by an institution is the set of institutional actions. For example, actions like selling and buying, renting and hiring, lending and borrowing (and, in fact, also stealing) are only meaningful within a suitable institution of ownership.

Contrary to natural actions, like eating or walking around, institutional actions cannot be executed only by exploiting physical abilities. Rather, they require suitable conventions that allow agents to perform institutional actions by producing certain forms of behaviour, typically linguistic behaviour. The need of specifying linguistic conventions for open distributed systems has been recognised at least since the definition of KQLM (Finin et al. 1997). Since then, Agent Communication Languages (ACLs) have been an important subject of research. Early models of multiagent systems tended to regard ACLs as local to every specific system (or even to every pair of roles in the system, as suggested by the Aalaadin model Ferber and Gutknecht 1998). Later research in agent communication, however, highlighted the importance of defining a universal ACL (Dignum 2004, van Eijk et al. 2005). Moreover, some approaches to the semantics of ACLs show that the very definition of a communication language may be based on elements of social reality, like the concept of commitment (Singh 1999b; Colombetti 2000; Colombetti et al. 2004): if this approach is correct, ACL messages can be understood only within a suitable system of artificial institutions.

In general, producing an instance of the correct conventional behaviour is necessary but not sufficient to perform an institutional action. For example, the president of society can open the annual meeting by saying “The meeting is open”, while a generic member of the society cannot do the same. The issue, here, is power or, as we prefer to call it, authorization: institutional actions are successful only if they are performed by an authorized agent, and authorizations are typically associated to the roles played by the agent in the institution. Even authorized actions, however, cannot always be performed freely. For example, the owner of a car is authorized to sell it, but may be prohibited to do so by his or her spouse. Given the authorizations of the car owner, the action of selling will have its normal institutional effects, but it will also violate the commitment that the owner has with his or her spouse (in this example two institutions, ownership and marriage, interact).

This brief analysis shows that institutions are made up by a number of components: an ontology (including the definition of institutional actions), a set of authorizations to perform institutional actions (typically associated to roles), a set of linguistic conventions for the execution of institutional actions, and a system of norms regulating the agents’ interaction. All these components are so strictly interconnected that they have to be dealt with within a single conceptual framework. The metamodel of artificial institutions that we propose in this paper is an attempt to

define such a framework. Our presentation is structured as follows. In Sect. 2 we introduce our metamodel of artificial institutions, defining their fundamental concepts and relationships. In Sect. 3 an example of the specification of an artificial institution, the Basic Institution, is presented. Using the notion of commitment defined by the Basic Institution, in Sect. 4 we present our operational definition of norms. Finally, in Sect. 5 we conclude with a brief discussion of related work and open questions.

## 2 A metamodel of artificial institutions

Artificial institutions are *models* of institutional reality which specify a class of interaction domains by exploiting a set of common concepts and notations. For this reason we introduce a *metamodel* of artificial institutions, called *OCeAN* (Ontology Commitments Authorizations Norms), to describe an abstract syntax and a semantics for each construct that characterizes artificial institutions. In our view the fundamental components of an artificial institution are: a core ontology for the definition of institutional entities and actions, and for the specification of roles and events, a set of conventions and authorizations for the actual performance of institutional actions, and a set of ECA-rules that are crucial for the definition of norms. In Fig. 1 the abstract syntax of our artificial institution metamodel is presented by means of a notation inspired by the UML metamodel (Object Management Group 2005b), showing the fundamental concepts and the relationships existing between them.

Even if we have adopted a notation inspired by UML, we are not prescribing that artificial institutions and agents acting within them should be implemented by object oriented technologies. In fact, artificial institutions are intended to specify institutional reality for open multiagent systems, which are characterized by the fact that heterogeneous agents realized with different technologies and by different

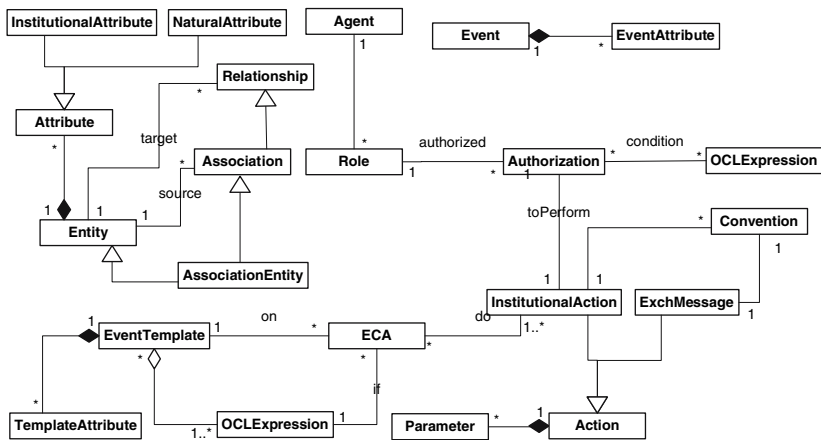


Fig. 1 The artificial institution metamodel

organizations should interact. Therefore, it is important that the specification does not commit to a specific implementation language, platform, or internal architecture. For this reason, we do not exploit most of the features of object oriented technologies, like methods or polymorphism, and to avoid confusion we will not talk of classes. Therefore, we do not adopt the meta-metamodel of UML,<sup>1</sup> which introduces, among others, the concept of MetaClass and MetaOperations.

We use UML as a starting point, reusing its well known graphical notation and the Object Constraint Language (OCL Object Management Group 2005a) as our language to express constraints. An advantage of this approach is that it employs concepts that are close to the intuition and knowledge of practitioners. We believe the metamodel we have developed can be easily understood by software engineers who design and implement open multiagent systems.

Like the metamodel of UML, our metamodel of artificial institution has a declarative semantics and suppresses implementation details. Therefore, we abstract away from methods and implementation issues and provide an abstract syntax to specify agent interaction systems.

In this paper we do not propose a specific architecture for the management of institutional states in real systems, because we think that different architectures may be chosen and adapted according to the needs of real systems. Anyway, agents using the interfaces specified through an artificial institution model should not even notice the existence of such design and implementation choices.

In the following sections we shall introduce a notation for the components of artificial institutions and describe their meaning in natural language, whereas the metamodel reported in Fig. 1 provides a set of well-formedness rules for a model of an artificial institution. In Sect. 3 we will exemplify our approach by defining a model of a specific institution, the Basic Institution.

## 2.1 The core ontology

### 2.1.1 Entities

In our view, artificial institutions are a technological extension of human reality and therefore they should represent part of the *state* of the real world. We assume the existence of *external ontologies* defining classes of entities, their relationships and the relevant attributes representing physical properties. For example, an entity may represent a book, which may have a number of pages and a weight. We represent entities, attributes and relations with UML class diagrams.

The *core ontology* is introduced to define new features relative to such entities which exist only thanks to agents' common agreement. For instance, the price of a product on sale exists only because a community of agents recognizes such a property.

For this reason we distinguish between two types of attributes that can be associated to the definition of entities: *natural attributes*, which represent physical

---

<sup>1</sup> For reason of conciseness we do not discuss our meta-metamodel and the relations existing between it and the meta-metamodel defined by UML.

properties and are defined by external ontologies, and *institutional attributes*, which reflect the existence of a common agreement and are introduced by core ontologies.

Sometimes, core ontologies define entities whose attributes are only institutional, like the *commitment entity* defined by the Basic Institution described in Sect. 3. We refer to such entities as *institutional entities*.

### 2.1.2 Institutional actions

In our framework we assume that agents can modify only institutional attributes by performing a particular set of actions, that is, *institutional actions* (Colombetti and Verdicchio 2002). An institutional action describes how institutional attributes change as a consequence of its performance. For example, the effects and preconditions of the act of opening an auction may be described in terms of certain values assumed by the institutional attribute *state*, which represents the current state of an auction.

More precisely, an institutional action is characterized by:

- an *action name*;
- a possibly empty set of *parameters*;
- a possibly empty set of *preconditions*, which specify the values that certain institutional attributes must have for the action to be meaningful (for example, opening an auction is meaningful only if the auction is not already open);
- a nonempty set of *postconditions*, which specify the values of certain institutional attributes after a successful performance of the action.

Preconditions and postconditions of institutional actions are expressed through OCLExpressions (Object Management Group 2005a), which can refer only to the parameters of the institutional action or institutional attributes defined by the core ontology, reflecting the fact that the only effect of an institutional action is to modify institutional attributes.

Because institutional actions change institutional attributes, which exist only thanks to common recognition, it follows that institutional actions have an intrinsic social nature. Therefore, agents cannot perform such actions by exploiting causal links occurring in the natural world, as it would be done to open a door or to move a physical object. Furthermore, a crucial condition for the actual performance of institutional actions is that they must be *public*, that is, made known to the relevant agents by means of some action that can be directly executed by an artificial agent. In the human world such actions vary from certain bodily movements (raising one's arm to vote), to the use of specific physical tools (waving a white flag to surrender), to the use of language (saying “the auction is open” to open an auction). We assume that in a multiagent system all institutional actions are performed by means of a single type of action, namely exchanging a message.

### 2.1.3 Roles

Authorizations, as the metamodel reported in Fig. 1 shows and norms (as we will explain in Sect. 4) are not related to specific individual agents but to sets of agents

identified by means of roles. Roles are therefore an interface between authorizations and norms and concrete agents defined in an artificial institution. For example the president, the secretary, and the participants of a meeting are three roles and in fact the institution of meetings defines different authorizations and norms that apply to the agents that fill such roles. In general, a sound specification of an artificial institution cannot define a role with an empty set of authorizations or ECA rules.

#### 2.1.4 Events

A core ontology also describes a set of events which are relevant for the definition and activation of ECA rules (see Sect. 2.3). UML models four kinds of events (Object Management Group 2005b) : *signals*, *calls*, *passing of time* and *change in state*. Unfortunately, the notation proposed in UML for modeling events is bound to features of State Machine and Statechart Diagrams, while we need a way to describe events in general.

In our framework, an event type can have attributes, providing information about the state transition that caused it, and it is possible to model hierarchies of event types. Moreover we assume that every event has a *time* attribute reporting the time at which an event has occurred. In our formalization we have singled out three main categories of events:

- a *TimeEvent*, which occurs when the system reaches a certain instant of time;
- a *ChangeEvent*, which happens when an institutional entity changes in some way. This kind of event type can be further specialized:
  - an *AttributeChange*, which is registered when an attribute has changed its value;
  - a *RelationChange*, which happens when a new relation is *created* or an existing one between two institutional entities is *dropped*;
  - a *GeneralizationChange*, which occurs when an entity modifies its type in a given taxonomy;
- an *ActionEvent*, that happens when an agent perform an action. In particular, an interesting type of this kind of events is *ExchMsg*, which represents the act of exchanging a message.

The definition of event types allows us to define *event templates*, that is, event types that have some restriction on certain attributes and describe a set of possible event occurrences. Event templates are used in the *on* section of ECA rules to specify what kind of domain dependent events activates a rule.

## 2.2 Count-as relation, authorizations, and conventions

Given that institutional actions modify institutional attributes that exist only thanks to agent common agreement, agents cannot directly perform such actions. Instead, we assume that all institutional actions are performed thanks to the *counts-as* relation which binds the exchange of a certain message to the performance of a

certain institutional action. In particular the counts-as relation represents a constitutive aspect of the construction of institutional reality (Searle 1995).<sup>2</sup>

To model the count-as relation we introduce the notion of *convention*, that is, an agreement about what type of message, sent to certain agents, is bound to a given type of institutional action. In particular conventions are necessary to specify how an institutional action can be concretely carried out by an agent.

In our model the definition of a convention (i.e., statements of the form “=<sub>conv</sub>”) has the following generic form:

$$\text{ExchMsg}(\text{message-type}, \text{sender}, \text{receivers}, \text{content}) =_{\text{conv}} \text{iaction}(\text{parameters})$$

Where the fundamental *message-type* that can be used to perform institutional actions is *declare*; the *sender* is the agent that sends the message, the *receivers* attribute expresses (when it is necessary) the set of agents that should receive the message given that they are affected by the performance of the act; the *content* attribute is used to determine the value of certain parameters of the institutional action or, in case the *message-type* is *declare*, is used to single out the type and the parameters of the related institutional action.

For example the following convention binds a certain message exchange with the performance of the *openAuction* institutional action on condition that all the participants of an auction must receive the message:

$$\begin{aligned} &\text{ExchMsg}(\text{declare}, \text{sender}, \text{participant}, \text{openAuction}(\text{parameters})) \\ &=_{\text{conv}} \text{openAuction}(\text{parameters}) \end{aligned}$$

By itself, a convention is not sufficient to guarantee the successful performance of an institutional action by the exchange of the appropriate message: indeed, some additional conditions must be satisfied about the agent that sends the message and about the state of the system in relation to the content of the message.

### 2.2.1 Conditions on the sender of the message

The sender of the message must be authorized to perform an institutional action. In the specification of an interaction system authorizations are expressed in term of roles; for example, only the auctioneer can open an auction by sending a suitable message to the participants. Moreover an authorization can be given only if certain conditions about the state of the system, expressed by suitable Boolean expressions, are satisfied. For example, it may be established that an auction may be validly opened only if there are at least two participants. Therefore, we abstractly define the authorization to perform a specific institutional action (with given parameters) associating it to a role as follows:

$$\text{Auth}(\text{role}, \text{iaction}(\text{parameters}), \text{conditions})$$

<sup>2</sup> Following Searle (Searle 1995), the construction of social reality in the human world is possible thanks to *constitutive rules* of the form *X counts as Y in C*.



It is worth highlighting that authorizations are a necessary condition for the successful performance of an institutional action and that designers should specify a set of authorizations whenever they introduce a new institutional action. As we will see in Sect. 4, this fact allows us to assume that every authorized institutional action that is not prohibited is permitted.

We decided to separate the notion of convention and authorization because we find it very useful. For example it is part of a possible convention of voting that, under appropriate conditions that have to be specified, raising one's hand counts as voting "aye." Which agents are authorized to exploit such a convention is a different matter, which we find natural to regard as external with respect to the convention. Note that one can modify the set of agents that are authorized to vote without changing the voting conventions.

### 2.2.2 Conditions about the state of the system

All the preconditions of the institutional action associated to the performance of the exchange of the message must be satisfied; for instance, an auction cannot be closed if it has not been opened yet.

To conclude the exchange of a message counts as the performance of an institutional action if and only if: the type of the message is conventionally associated to the type of institutional action and all contextual conditions hold.

## 2.3 Event-Condition-Action rules

The last component of our metamodel is represented by Event-Condition-Action rules, which allow us to model the fact that institutional reality may change as a consequence of the occurrence of certain events. For example, a competition may start at a fixed time instant and be considered closed at another time without an agent having declared it. ECA rules are also useful to model how the performance an institutional action executed within an institution may affect another institution (see for example Viganò et al. 2005).

Inspired by Active Database models, we define an ECA rule as composed by three elements:

1. an *event template*, which represents the class of events which may activate the rule;
2. a *condition*, expressed through a Boolean OCLExpression, which can refer to the variable  $e$  which contains a description of the event that has activated the rule;
3. a nonempty set of *institutional actions* which are executed by the rules.

The semantics of ECA rules is given as usual: when an event matching the description given by the event template occurs in the system, the variable  $e$  is filled with the event instance and the condition is evaluated; if the condition is verified, the set of institutional actions are executed and their effects are brought about in the system. We assume that the system is authorized by default to perform every institutional action. In our model, ECA rules are specified according to the following notation:

**on**  $e$  : *event–template*

**if condition then**

**do** *institutionalAction(parameters)*<sup>+</sup>

As we will see in Sect. 4, we define norms as ECA rules that perform institutional actions creating, cancelling or modifying commitments. Therefore, before introducing norms, we need to define the Basic Institutions, that is, the artificial institution that defines what social commitments are.

### 3 The Basic Institution

The *Basic Institution* is the institution that defines a fundamental concept: the notion of social *commitment*. The importance of commitment is due to the fact that, from our perspective, it is essential to express the meaning of most types of communicative acts (see Fornara and Colombetti 2004, Fornara et al. 2007) and because it is used to define norms as will be shown in Sect. 4. Therefore we assume that the Basic Institution has to be used, together with other special institutions, in the specification of every open interaction framework.

The Basic Institution is an artificial institution that defines the ontology of commitment, the institutional actions necessary to operate on it, and a set of authorizations for the performance of these institutional actions. In general, institutions also define sets of norms to regulate the behaviour of agents, but in our current view, the Basic Institution does not specify norms. In this section we report a fragment of the Basic Institution for a complete specification see (Fornara et al. 2005).

Other institutions, that we call *special institutions*, can then be defined to model the aspects of institutional reality typical of certain application domains. For instance, for electronic commerce applications it will be necessary to model the institutions of ownership, money, business transactions, auctions, and so on. A detailed specification of the English auction, of the Dutch Auction, and of the Auction House as special institutions, can be found in (Fornara et al. 2007; Viganò et al. 2006).

#### 3.1 Commitments and temporal propositions

In this paper we give only a short description of our model of *commitment* and of *temporal proposition* (Colombetti et al. 2004), that are used to represent the content of commitments. A commitment is characterized by the following attributes: a *debtor*, a *creditor*, a *content*, and a *state*. Figure 2 represents the class diagram of the ontology of the Basic Institution. In the rest of the paper we will refer to commitments using the following notation:

$$Comm(state, debtor, creditor, content)$$

The content of commitments is expressed using temporal propositions. A temporal proposition is characterized by a *statement* about a state of affairs or about an action.

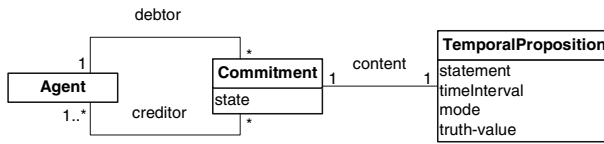


Fig. 2 Class diagram of the ontology of the Basic Institution

The statement is referred to a *time interval* with two possible different *modes*: *exist* ( $\exists$ ) or *for all* ( $\forall$ ). The *truth-value* of a temporal proposition is initially *undefined* ( $\perp$ ). It becomes *true*, thanks to an event driven routine, if the mode is *for all* and the statement is true for every instant of the associated time interval or if the mode is *exist* and the statement is true for some instant in the associated time interval, otherwise it becomes *false*. Temporal propositions are represented with the following notation:

$$TP(statement, [t_{start}, t_{end}], mode, truth-value)$$

A commitment undergoes the life cycle described in (Fornara and Colombetti 2004) by reacting either to institutional actions performed by agents or to domain-dependent events, which modify the truth value of the temporal proposition in its content. If its temporal proposition becomes true, the commitment becomes *fulfilled*; if it becomes false the commitment becomes *violated*. For instance the commitment of agent  $a_1$  to agent  $a_2$  to open the *auction<sub>i</sub>* within 10 min from *now* (or between *now* and *now* + 10) is represented as:

$$Comm(pending, a_1, a_2, TP(open(a_1, auction_i), [now, now + 10m], \exists, \perp))$$

In our framework every agent is authorized to create a commitment by performing the *makeCommitment* institutional action, whose successful performance creates an *unset* commitment. The debtor of an *unset* commitment may refuse it by executing *setCancel*, or it may undertake the proposed commitment by executing *setPending*. We represent a refused commitment by means of the *cancelled* state, whereas an accepted commitment is depicted with the *pending* state. The creditor of a *pending* or *unset* commitment can always set it to *cancelled*. Here, due to space limitation, we report only the definition of one institutional action and of one authorization. The institutional action *makePendingComm*, used in Sect. 4, creates a *pending* commitment and its execution coincides with the sequential performance of *makeCommitment* and *setPending*:

*name* :    **makePendingComm**(*debtor, creditor, content*)  
*pre* :       *not Comm.allInstances()*(*c*|*c.debtor = debtor*  
           *and c.creditor = creditor and c.content = content*)  
*post* :       *Comm.allInstances()*(*c*|*c.state = pending and*  
           *c.debtor = debtor and c.creditor = creditor and*  
           *c.content = content*)

Every registered agent *RegAgt* is authorized to create a commitment having itself as *debtor*:

$$\text{Auth}(\text{RegAgt}, \text{makePendingComm}(\text{debtor}, \text{creditor}, \text{content}), \text{RegAgt} = \text{debtor})$$

## 4 Norms

In Sect. 2 we have defined the counts-as relation among actions that represents a constitutive aspect of the construction of institutional reality (Searle 1995). In this section we will tackle the problem of how to describe the regulative counterpart, that is, the definition of norms.

In the literature, it is possible to find two different approaches to the specification of norms for open systems. The more formal approach uses deontic logic to express the semantics of norms and is more suitable when the goal is to verify the consistency of a set of norms (Dignum et al. 2004). The other approach focuses on the operational aspects of norms, and is more suitable when the goal is the actual implementation of norms in open systems (Vázquez-Salceda et al. 2004; Garcia-Camino et al. 2005). In this paper we study the formalization of norms from the *institutional and operational perspective* (opposite to the agent perspective Vázquez-Salceda et al. 2004), that is, our goal is to define norms that make the evolution of the state of an interaction system partially predictable without violating the autonomy of the interacting agents. In order to reach this purpose, norms describe, on the basis of the state of the system, the expected behaviour of the agents playing different roles, and make it possible to detect deviations from such a behaviour, that is, the violation of obligations and prohibitions defined in the system.

As already said in Sect. 3 in our model norms are a special type of ECA rules characterized by the fact that when they are fired by events happening in the system, they create or cancel commitments to perform or to not perform an action within a certain interval of time, thus affecting *each* agent that satisfies a suitable selection expression. Usually the collection of liable agents corresponds to the set of agents that play a given role in the institution.

From our point of view, commitments are not a specialization of norms as in (López y López et al. 2004) and norms are not themselves a special kind of commitments as in (Castelfranchi 1995) and (Singh 1999a). We perceive norms as rules that manipulate commitments of the agents engaged in an interaction. This because in the abstract formalization of a system it is important to model norms associated to roles rather than to individual agents, whereas during the actual

evolution of the system it is fundamental to create commitments associated to individual agents and to detect their fulfillment or violation. Obviously such violations can be interpreted also as the violation of the corresponding norm.

In our model, when an agent fills a role in a software system implementing an institution, the norms of that institution will create commitments binding the agent to the system, which will be itself regarded as an agent. The identity of the creditor agent allows keeping trace of the commitments created by a system and thus, in case of need depending on the application domain, distinguishing them from the ones created by other agents.

The general structure of a norm can be described as follows:

```

on e : event-template
if condition then
  foreach agent in selection-expression
    do commitment-Ops

```

where *agent* is an identifier varying on the set of agents that satisfy the selection expression; *selection-expression* is a list of agent identifiers or a *role* defined in a certain artificial institution; *commitment-operations* is a sequence of commitment operations defined using BNF notation as follows:

```

commitment-Ops := comm-Op | comm-Op, commitment-Ops
comm-Op := makePendComm(agent, instAgent, content) |
             setCancel(agent, instAgent, content)
content := TP(action, [t, t],  $\exists$ ) | TP( $\neg$ action, [t, t],  $\exists$ )
t := now | event.time | instant | now + number | event.time + number

```

where *action* is an institutional action as described in Sect. 2; *now* is the time of the system when the temporal proposition is created and *instant* is an instant of time.

In (Fornara et al. 2007) we give a formalization of the English Auction Institution. Here we report only an example of a norm that creates an obligation for the auctioneer to open the auction when the *start\_time* has elapsed, if at least two agents have been registered as participants:

```

on e : TimeEvent(UnsetEnglishAuction.startTime)
if UnsetEnglishAuction.participant.sizeOf()  $\geq$  2 then
foreach agent in UnsetEnglishAuction.auctioneer
  do makePendingComm(agent, instAgent,
    TP(openAuction(UnsetEnglishAuction.id), [now, now +  $\delta$ ],  $\exists$ ))

```

where  $\delta$  is the time allowed to the agent to fulfill its obligation.

#### 4.1 Deontic relations

Using our metamodel of artificial institutions and the definition of norms it is possible to represent fundamental deontic relations between agents:

- *Obligations* are represented by commitments, created by norms, to perform an action of a given type. It may happen that agents do not perform obligatory actions; in such a case the commitment becomes violated and the system does not evolve to the expected new state.
- An action is *prohibited* when a norm has created a commitment not to perform it. The main difference between prohibition and the absence of authorization resides in the effects of the action: if an agent is prohibited to perform action  $a_i$  but performs it anyway, the effects of the action take place and the commitment to not perform the action is violated; differently if an institutional action  $a_i$  is not authorized neither the counts-as relation will hold, nor the effects of  $a_i$  will take place.

When commitments that represent obligations or prohibitions are violated is up to the system to impose sanctions to the misbehaving agent, and possibly recover the system to a safe state. The treatment of sanctions, either direct (fine, expulsion) or indirect (trust, reputation), and the study of plans to recover the system from unsafe states are a wide and interesting research problem we plan to tackle in the future.

Regarding permission, if an action is not prohibited (and in case it is an institutional actions it is also authorized) it is permitted. This definition is close to the notion of weak permission discussed in deontic logic (von Wright 1959).

## 5 Discussion and conclusions

In this paper we have described in short our operational model of artificial institutions that can be used for the specification of open interaction systems. In particular, we have focused on the definition of a syntax and an intuitive semantics of the concepts that an agent, or its designer, has to take into account and to reason on, in order to be able to interact properly with different systems.

Existing proposals for the definition of electronic institutions like the one proposed in the EIDE platform (Esteva et al. 2001 2001; Esteva et al. 2004b), the one presented by Cliffe and Padget (Cliffe et al. 2006), or the OMNI framework (Vázquez-Salceda et al. 2005) are mainly focused on the specification of the normative component. Unlike them, our model includes also the concepts for defining the institutional reality of the system.

As far as the formalization of normative aspects is concerned, likewise the other mentioned proposals, we distinguish among permission, obligation, and prohibition. We also distinguish between normative aspects and facts dealing with institutional power (Jones and Sergot 1996), which separate meaningful/empowered actions from meaningless actions that do not have an effect. In our opinion, this last type of fact, like in (Artikis et al. 2002), should be considered as part of the declarative apparatus of an electronic institution, and not of the normative one. Indeed we think that norms regulate agent interaction by indicating, among all meaningful actions at a certain stage, those that are obliged or forbidden. Moreover an interesting aspect of our proposal is its uniformity; in fact, our operational formalization of norms uses

the same concepts used for the definition of the semantics of ACL, that is, social commitments. Therefore, differently from (Esteva et al. 2001; Vasconcelos 2003; Esteva et al. 2004b), where the semantics of the communicative acts is given in terms of predefined state transitions, in our model it is possible to define an application independent semantics of a Communicative Act Library, as presented in (Fornara et al. 2007), with a syntax compatible with FIPA-ACL (Foundation for Intelligent Physical Agents 2002).

In (Boella et al. 2006), the authors propose to model a control mechanism as a *normative agent*, which is an autonomous agent whose goals and beliefs represent norms and constitutive rules. Such model is suitable for an agent that wants to reason about the norms of the systems or to check, at design time, the effectiveness of a control mechanism. In contrast with their model, we do not think that the system of norms constitutes an agent which can autonomously decide whether and how to enforce norms. Instead, in our view, interactions that take place in an open system, modelled and regulated by an institution, are constantly monitored and norms are automatically applied whenever their conditions are met. For this reason, we propose to model norms as rules that create commitments.

A few proposals have been put forward to implement institutions. Automatic translations of the language presented in (Esteva et al. 2001) to specify electronic institutions into executable programs have been presented in (Esteva et al. 2004b, Vasconcelos 2003). In particular, in (Esteva et al. 2004b) the authors define an architecture to execute and monitor specifications of electronic institutions, while in (Vasconcelos 2003) such specifications are translated into Prolog programs, which can be used to simulate the behaviour of a system and to check simple properties. Also, Vasconcelos (2003) defines a precise semantics for labels associated to transitions of electronic institutions. Rubino et al. (2006) present the concept of *computational institution*, which is “regarded as a virtual organization ruled by norms”. In particular, they propose a possible mapping of normative and constitutive rules over the abstraction provided by TuCSoN (Omicini and Zambonelli 1999), an existing infrastructure to implement open systems. In Rubino et al. (2006) the authors focus their attention more on the proposal of a specific architecture than on the definition of the concepts needed to specify an institution. Instead, the attention of this paper has been devoted more on the conceptualization of the institutional notions needed to specify an institution, and we think that a considerable advantage of our approach is its independence of a specific architecture or implementation language, which is an important feature of open systems.

Several research questions are still open, and will be tackled in future works. We will investigate the development of methods for discovering inconsistencies among the specification of one or more artificial institutions. In particular, we are interested in verifying during the specification phase whether certain norms may create obligations to perform unauthorized actions, or under what conditions two norms may generate conflicting commitments. We plan to investigate how to model the notion of strong permission and moreover we intend to devise an explicit representation of the sanctions and repair procedures connected to the violation of commitments.

**Acknowledgements** We would like to thank the anonymous reviewers for their useful comments.

## References

- Artikis A, Sergot M, Pitt J (2002) Animated specifications of computational societies. In: Castelfranchi C, Johnson WL (eds) *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*. ACM Press, pp 535–542
- Boella G, Hulstijn J, Tan Y-H, der Torre LV (2006) Modeling control mechanisms with normative multiagent systems: the case of the renewables obligation. In: Boissier O, Padget J, Dignum V, Lindemann G, Matson E, Ossowski S, Simão Sichman J, Vázquez-Salceda J (eds) *Coordination, organization, institutions and norms in multi-agent systems*, vol 3913 of LNAI. Springer, pp 114–126
- Castelfranchi C (1995) Commitments: from individual intentions to groups and organizations. In: Lesser V (ed) *Proceedings of the 1st International Conference on Multi-Agent Systems*. AAAI-Press and MIT Press, San Francisco, USA, pp 528–535
- Cliffe O, Vos MD, Padget J (2006) Specifying and analysing agent-based social institutions using answer set programming agents. In: Boissier O, Padget J, Dignum V, Lindemann G, Matson E, Ossowski S, Simão Sichman J, Vázquez-Salceda J (eds) *Coordination, organization, institutions and norms in multi-agent systems*, vol 3913 of LNAI. Springer, pp 99–113
- Colombetti M (2000) A commitment-based approach to agent speech acts and conversations. In: *Proceedings of the Workshop on Agent Languages and Communication Policies, 4th International Conference on Autonomous Agents (Agents 2000)*. Barcelona, Spain, pp 21–29
- Colombetti M, Fornara N, Verdicchio M (2004) A social approach to communication in multiagent systems. In: Leite JA, Omicini A, Sterling L, Torroni P (eds) *Declarative agent languages and technologies*, vol 2990 of LNAI. Springer, pp 191–220
- Colombetti M, Verdicchio M (2002) An analysis of agent speech acts as institutional actions. In: Castelfranchi C, Johnson WL (eds) *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*. pp 1157–1166
- Dignum F (ed) (2004) *Advances in agent communication, international workshop on agent communication languages, ACL 2003*, Springer Melbourne, Australia, July 14, 2003, vol 2922 of LNAI.
- Dignum F, Broersen J, Dignum V, Meyer JC (2004) Meeting the deadline: why, when and how. In: *Proceedings of the 3rd International Workshop on Formal Approaches to Agent-Based Systems (FAABS)*. pp 30–40
- Esteva M, Rodríguez-Aguilar J, Rosell B, Arcos J (2004a) AMELI: An Agent-Based Middleware for Electronic Institutions. In: Jennings NR, Sierra C, Sonenberg L, Tambe M (eds) *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)* ACM Press, pp 236–243
- Esteva M, Rodríguez-Aguilar J, Sierra C, Vasconcelos W (2004b) Verifying Norm Consistency in Electronic Institutions. In: Dignum V, Corkill D, Jonker C, Dignum F (eds) *Proceedings of the AAAI-04 Workshop on Agent Organizations: Theory and Practice*. pp 8–15
- Esteva M, Rodríguez-Aguilar JA, Sierra C, Garcia P, Arcos JL (2001) On the Formal Specification of Electronic Institutions. In: Dignum F, Sierra C (eds) *Agent mediated electronic commerce, The European agentlink perspective*, vol 1991 of LNAI. Springer, pp 126–147
- Finin T, Labrou Y, Mayfield J (1997) KQML as an agent communication language. In: Bradshaw JM (ed) *Software Agents*. AAAI Press /The MIT Press, Chapt. 14, pp 291–316
- Fornara N, Colombetti M (2004) A commitment-based approach to agent communication. *Appl Artif Intelligence Int J* 18(9–10):853–866
- Fornara N, Viganò F, Colombetti M (2005) Agent communication and institutional reality. In: van Eijk RM, Huget MP, Dignum F (eds) *Agent communication: international workshop on agent communication, AC 2004*, New York, NY, USA, July 19, 2004, Revised Selected and Invited Papers. pp. 1–17, Springer Verlag. (For an extended version see N. Fornara, F. Viganò, and M. Colombetti, 'Agent Communication and Artificial Institutions'. *Journal of Autonomous Agents and Multi-Agent Systems*, to appear)
- Fornara N, Viganò F, Colombetti M (2007) Agent communication and artificial institutions. *Auton Agents Multi-Agent Syst* 14(2):121–142



- Foundation for Intelligent Physical Agents (2002) FIPA communicative act library specification. <http://www.fipa.org>
- García-Camino A, Noriega P, Rodríguez-Aguilar JA (2005) Implementing norms in electronic institutions. In: Proceedings of the 4th International Joint Conference on Autonomous agents and Multi-Agent Systems (AAMAS 2005). ACM Press, pp 667–673
- Ferber J, Gutknecht O (1998) A meta-model for the analysis and design of organizations in multi-agent systems. In: Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS98). Paris, France, pp 128–135
- Jones A, Sergot MJ (1996) A formal characterisation of institutionalised power. *J IGPL* 4(3):429–445
- López y López F, Luck M, d'Inverno M (2004) Normative agent reasoning in dynamic societies. In: Jennings NR, Sierra C, Sonenberg L, Tambe M (eds): Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004). ACM Press, pp 535–542
- Noriega P (1997) Agent mediated auctions: the Fishmarket Metaphor. Ph.D. thesis, Universitat Autònoma de Barcelona
- Noriega P, Sierra C (1996) Towards layered dialogical agents. In: Mueller JP, Wooldridge M, Jennings N (eds) Proceedings ATAL 96. Intelligent Agents III. Agent theories, architectures, and languages, vol 1193 of LNAI. Springer-Verlag
- Object Management Group (2005a) UML 2.0 OCL Specification. <http://www.omg.org/>
- Object Management Group (2005b) Unified modeling language specification. <http://www.omg.org/>
- Omicini A, Zambonelli F (1999) Coordination for internet application development. *Auton Agent Multi Agent Syst* 2(3):251–269
- Rubino R, Omicini A, Denti E (2006) Computational institutions for modelling non-regulated MAS: an 625 approach based on coordination artifacts. In: Boissier O, Padget J, Dignum V, Lindemann G, Matson E, Ossowski S, Simão Sichman J, Vázquez-Salceda J (eds) Coordination, organization, institutions and norms in multi-agent systems, vol 3913 of LNAI. Springer, pp 127–141
- Searle JR (1995) The construction of social reality. Free Press, New York
- Singh MP (1999a) An ontology for commitments in multiagent systems: toward a unification of normative concepts. *Artif Intell Law* 7:97–113
- Singh MP (1999b) A social semantics for agent communication languages. In: Proceedings of IJCAI-99 workshop on agent communication languages. pp 75–88
- van Eijk R, Huget M, Dignum F (eds) (2005) Agent communication, international workshop on agent communication, AC 2004, New York, NY, USA, July 19, 2004, Revised selected and invited papers, vol. 3396 of LNAI. Springer
- Vasconcelos W (2003) Logic-based electronic institutions. In: Leite J, Omicini A, Sterling L, Torroni P (eds) Workshop on Declarative Agent Languages and Technologies (DALT 2003), vol 2990 of LNCS. Springer, pp 221–242
- Vasconcelos W, Sabater J, Sierra C, Querol J (2002) Skeleton-based agent development for electronic institutions. In: Castelfranchi C, Johnson WL (eds) Proceedings of the 1st international joint conference on Autonomous Agents and Multiagent Systems (AAMAS '02). ACM Press, pp 696–703
- Vázquez-Salceda J, Aldewereld H, Dignum F (2004) Implementing norms in multiagent systems. In: Lindemann IG, Denzinger J, Timm IJ, Unland R (eds) Multiagent System Technologies (MATES 2004), vol 3187 of LNAI. Springer Verlag, Berlin, Germany, pp 313–327
- Vázquez-Salceda J, Dignum V, Dignum F (2005) Organizing multiagent systems. *Auton Agent Multi Agent Syst* 11(3):307–360
- Viganò, F, Fornara N, Colombetti M (2006) An event driven approach to norms in artificial institutions. In: Boissier O, Padget J, Dignum V, Lindemann G, Matson E, Ossowski S, Simão Sichman J, Vázquez-Salceda J (eds) Coordination, organization, institutions and norms in multi-agent systems, vol 3913 of LNAI. Springer, pp 142–154
- von Wright GH (1959) On the logic of negation. *Societas Scientiarum Fennica*, Helsinki