Regulating Virtual Interactions

Pablo Noriega

Artificial Intelligence Research Institute
IIIA - CSIC.
Bellaterra, Barcelona, Spain

Abstract. The evolution of internet has brought about many forms of virtual interactions, like e-learning or e-goverment, whose aim is to achieve an objective that is shared by those parties that interact. It is not unusual that in order for such interactions to be effective, they may require the explicit definition of conventions that participants should abide by, not unlike traditional social interactions. In this paper I explore one way of organizing virtual interactions that is known as an electronic institution. My purpose is to break ground for a discussion with legal thinkers by showing how these artifacts address some of the concerns that have traditionally been in the domain of legal reasoning in as much as they constitute a particular type of regulated social environment where participants may be humans as well as software agents.

Keywords : Electronic Institutions, Multi-agent Systems, Norms.

1   Introduction

Internet has adopted many conventional practices and adapted them to become new forms of those practices, like *e-commerce* or *e-learning*. This process of adoption and innovation has shown that many traditional mechanisms work well in the new medium but it has also shown, both, that some adjustments are needed to make those practices truly applicable in the internet and that unforeseen practices may now be devised. One can look into the process of establishing such virtual practices from an abstract perspective and frame it as the problem of designing coordination artifacts. Artifacts designed to facilitate the achievement of some explicit purpose that is shared by those parties whose interactions are being coordinated.

This paper explores one particular way of making that design problem operational through artifacts that are known as *electronic institutions* (EIs for short). These artifacts organize interactions by establishing a restricted virtual environment where all interactions take place. Because EIs define a virtual space where a set of conventions is meant to hold, we may think of it as a *regulated environment* and invoke some legal constructs to discuss the design and application of the artifact. Thus in this paper I will discuss the notion of electronic institution paying attention to those normative aspects that I think are characteristic to the notion. I trust that such exploration will be interesting to legal thinkers in as much as these EIs may pose new questions and possibly suggest new answers to those traditional legal concerns that take new forms in the emergent realities of the web.
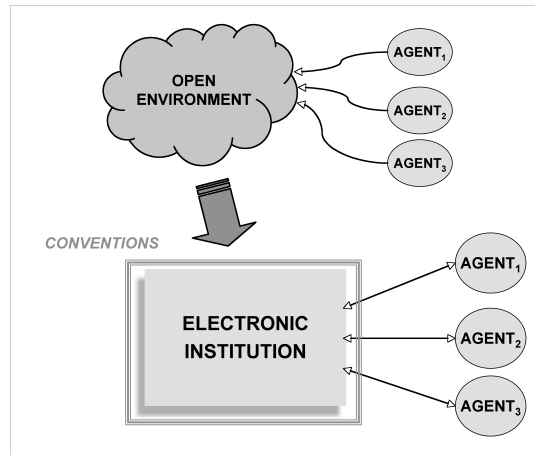
Fig. 1. Electronic Institutions create a virtual environment where interactions among agents in the real world correspond with illocutions exchanged by agents within the restricted environment. Legitimate illocutions —uttered within the institution— change the state of the institution and count as actions in the world

The next section gives an intuitive description of electronic institutions while Section 3 outlines their formal definition and their current implementation. These sections serve as background to a final section where I explore some normative concerns we have found in the process of developing these artifacts.

2 Institutional intuitions

2.1 An example

Picture the traditional way fish is sold in small fishing villages in Europe and particularly in Spain. Every day the local fleet unloads its catch in an large building where buyers gather. Fish is displayed for everyone to see but it is not sold directly by each fisherman to his or her clients. Fish is sold through a third party, an auction house (the *Lonja*), that is the one who offers the fish for sale, sells the fish, receives payments from buyers, charges a commission and pays the fishermen. The actual procedure is exquisitely regulated: fish is displayed in boxes that are orderly presented following the sequence in which the boats entered the port. Boxes are labeled by the lonja with precise information on weight, variety and quality of fish. Each box is offered for sale under a ritual by which an employee of the lonja —the auctioneer— cries aloud the price of the box starting with a reasonably high price and rapidly calling lower prices until someone stops the calling by saying "mine", as shown in Figure 2. This downward bidding process is repeated until the last box is sold. Buyers pay the lonja for the fish they have bought before they can take it away and the lonja pays the fishermen back.

Fig. 2. Traditional voice bidding in the Blanes *Llotja*. The auctioneer (left) calls prices for boxes of fish in a descending order until a buyer says "mine"

Auctioning of this sort has been used for centuries to sell different types of goods and because it is such an effective way of trading some types of goods it was rather natural to adapt it for electronic trading. Actually, a rather direct transcription from the lonja process in Blanes (Catalonia), *Fishmarket*, has been implemented [16, 11]. The design and implementation of *Fishmarket* brought to light the need of different sorts of adaptations. Some adaptations were easy to foresee, *e.g.*, software agents may react very quickly, hence the auctioning clock needs to be adapted. Other adaptations were more subtle: *e.g.*, if two software agents are bidding for the same good, they may have buying strategies that produce the same bidding and hence deadlock the process unless the virtual lonja introduces some tie-breaking mechanism. The more interesting adaptations, though, are those adaptations that profit from the combined novelties of agents and internet to create trading conventions that were hitherto unfeasible.

A nice example of previously unattainable trading artifacts is the MASFIT system. MASFIT is a production system that grew out of the *Fishmarket* efforts. It is a system that permits on-line (remote) bidding in actual fish auctions that are being carried out simultaneously in different ports. MASFIT consists of a virtual *federation* of lonjas such that: (i) Each traditional lonja functions exactly as before except that buyers are aware that there may be other buyers that are not physically present but may nonetheless participate in any bidding round under the same conditions as regular buyers do. (ii) For each traditional lonja there is a virtual lonja where those remote buyers bid through software agents that represent them. (iii) A remote buyer may have a buyer interface that allows him or her to place bids in any lonja, receiving the same information that regular buyers receive, and (iv) A remote buyer may have one or many software agents that buy on his or

her behalf in *real-time* in any of the federated lonjas (Figure 3). Two features of MASFIT are worth underlining. First, the system that has been implemented allows buyers to deploy "teams" of agents with a shared buying plan, each agent participating in the rounds of a different auction house and reporting back when necessary to the owner or to the other agents in the team. The salient point is that software agents may contain decision-making models that take advantage —in real time—- of all the information that the market is generating and also from historical market information and communication with fellow agents, possibilities that are beyond human power. The second significant feature is the challenging design problem of adapting both the auctioning conventions and their enforcement in the virtual environment in order to guarantee that the same levels of fairness, transparency and efficiency of the classical lonjas exist in the new set-up [2].
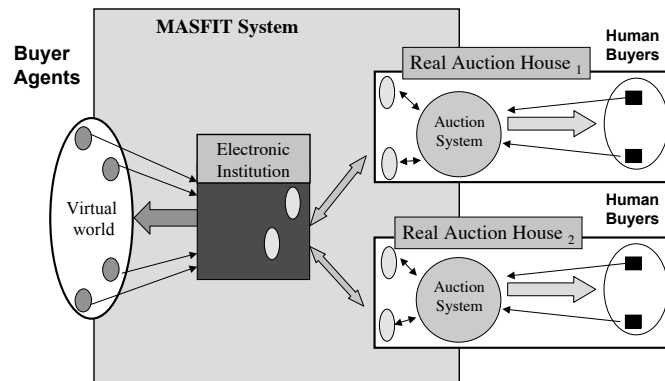

Fig. 3. Functional architecture of the MASFIT system. Regular buyers bid in the traditional lonjas under the familiar conditions. Remote bidding is made by buyers who have access to a buyer interface or to agents that bid through the local auction system of each lonja. An electronic institution handles information flow and remote and local bidding of all the lonjas simultaneously.

What can we learn from the lonja example that can be used for designing practices for the new social media enabled by internet? Lonja auctioning is a highly structured process that is enacted repeatedly. The process involves participants who interact according to well-established conventions —on how the price is set, for instance— that apply to individuals according to the *role* they play —buyers, seller or auctioneers— in the daily ritual. Simple gestures —like saying "mine"— have a precise meaning inside the lonja and entail concrete commitments —to acquire a box of fish and pay for it— that are honored because the lonja has seen to it that the buyer had enough credit to pay before allowing him / her to bid. Furthermore, as mentioned before, the actual auctioning of a traditional lonja can be reproduced in a virtual one.

Other common practices exhibit these same traits. Think for instance of the procurement process of a local government, betting at he race-track or processing the complaint of an airline passenger. Note also that although these examples have existed as regular practices, they may exist as virtual practices as well and may involve not only human participants but also autonomous software agents. Because of that, it seems natural from an IT perspective to think of designing a tool that may serve to deploy in a practical way the computational analogues of these practices. The purpose of an electronic institution is to make such notion a reality: to generalize the example of the lonja so that one may build computational entities that capture other forms of electronic commerce, due process, conflict resolution or coordination mechanisms in general that share the same type of features and concerns that are addressed in the lonja.

2.2 Institutional conventions and concerns

Before describing what electronic institutions actually are, let us back-up for a moment and try to figure out what features and concerns of the lonja would be shared by those electronic institutions we intend to build.

My using the term "institution" is not accidental. Electronic institutions are designed as artifacts that articulate actions by imposing artificial constraints on participants, or more bluntly: conventions that restrict the behavior of participants in order to make their interaction less uncertain and more effective. Economists use the term institution to refer to such sets of constraints (cf. North [13], p. 5) and we shall borrow their term. Intuitively, as illustrated in Figure 1, we can picture an institution (traditional or electronic) as that restricted environment where the rules of the game hold: a sort of leveled playing ground where interactions that are regulated by a set of conventions, take place. The actions that take place in that precinct are only those actions that are compatible with the institutional conventions, and the effects of those *institutional interactions* are those that were intended to be when the conventions were stated.

Although economist make clear that (traditional) institutions are the conventions that a group of agents follows in order to accomplish some socially agreed upon objective, it is not unusual to abuse language and identify an institution —the set of conventions— with an entity —a firm, company, organization— which is the guarantor of those conventions.[1]

Institutions, in traditional terms, are established to facilitate effective interactions, and in order to do so they are devised to deal with a few complementary concerns, the most salient are:

– *Establish the institutional conventions.* By so doing, conventions have an objective reference that participants may invoke to understand the conventions, to be able to follow them, be accountable for their satisfaction and deal with the wrongdoing of other participants.

---

[1] In keeping with this abuse of language we will refer to the *implementation* of an electronic institution as an electronic institution.

– *Assure permanence and stability of the conventions.* In order that participants may hold sufficient certainty of the requirements and outcomes of their interactions and that they may expect and choose to participate in different opportunities without undue adjustment of their participation requirements.
– *Enforce satisfaction of institutional commitments.* So that all participants may rely and be held responsible for their institutional actions as far as the institutional conventions state.
– *Guarantee accountability of institutional interactions.* An institution should be able to allocate risk and blame in an objective and effective manner. When establishing commitments within an institution, participants may be liable and the institution should be able to blame them when they misbehave. Furthermore, when participants are unreliable or even malevolent, there is risk involved that the institutional conventions should be able to allocate properly and limit potential damage effectively.
– *Manage access and identity of participants.* Validate that they satisfy the requirements of capability, resources or entitlement as long as they act within the institution, in order to be held accountable for their institutional actions.

We want to address these same types of concerns with electronic institutions albeit in a virtual environment. The lonja example provides us with pertinent elements to discern the features involved in addressing those concerns in a virtual institution. In fact, thanks to the lonja example we can distinguish two types of features: those that belong to the interactions that take place among the agents involved and those features that have to do with the effectiveness of those interactions in accomplishing their intended goals. Both types of features are reflected —as one would expect— in a number of conventions that are easy to identify in the lonja and by analogy in other institutions. These conventions may be classified in three groups:

1. *Interaction conventions* that regulate the way participants interact within the institution.
   In the lonja, interaction conventions regulate the behavior of buyers, sellers and lonja staff. For example, the order in which boxes are auctioned, how prices are called and when the calling stops, how to break a bidding tie, what commission is paid by participants, when is a buyer allowed to enter the auction hall, and so on.
2. *Grounding conventions* that establish a *correspondence* between the institutional ontology and the actual world, or putting it bluntly, they make sure that what happens inside the institution is in line with what is intended to happen in the real world.[2]
   In the lonja, grounding conventions that connect what happens inside the lonja with the physical world include, for instance, the way boxes are labeled, how buyers identify

---

[2] I use the term *grounding* with a meaning that is slightly different than the one used in Logic. To avoid confusion in this paper, instead of talking about *grounded* formulae —in classical logical parlance— I will say that a formula is *fully instantiated* and I use the term *grounding conventions* to stand for the mapping that *interprets* the terms of the object language as entities in the real world.

themselves and prove they are entitled to bid, how buyers open a credit line with the lonja and how it is managed by the lonja, how the lonja gets and transfers money from and to participants, what physical devices are used to convey information between buyers and auctioneer —e.g., a "bidding clock" and a remote control device or a button that stops it— or how is actual fish delivered to the buyer.

3. *Enabling conventions* are there to guarantee that what happens inside the institution *counts as* an action in the world.

   For the lonja —and *mutatis mutandi* in other electronic institutions— these constitutive conventions are necessary in order guarantee that the lonja, as an organization, performs effectively its duty as a mediator between buyers and sellers. Namely, that the lonja is active every day at the scheduled time, that sellers dutifully deliver their catch to the lonja and not sell it directly to buyers, that participant buyers are entitled to draw money from their credit line to pay for their purchases, that the property of fish is actually transferred from seller to buyer, or that the lonja as an organization pays taxes. Notice that this last type of conventions are usually reflected in contracts and other legally binding devices that may have a private character (say, between sellers, lonja and banks) or public (like those between the lonja and the fiscal authorities).

In devising a framework to build electronic institutions as *artifacts*, we would like to be able to capture and express those conventions, however there is a trade-off between how universal the framework may be and how fit is the framework to express and implement properly the conventions of a particular institution. In the next section I outline a framework that follows a pragmatic middle road. The framework is general enough to describe institutions whose domain of application is arbitrary but it is not designed to operationalize all types of conventions. I will discuss the limitations of our current framework in Section 4.4 but for the time being let me point out that it deals rather well with a large class of interaction conventions, proposes a simple way of addressing grounding conventions and is able to implement those enabling conventions that are inherent to a computational implementation of a traditional institution, which I think is what one would expect of such a framework.

3   An outline of electronic institutions

I will present very briefly the EI framework that has been under development for a decade in the IIIA [11, 15, 4].

We take EI to be a coordination artifact that facilitates the interactions of agents who pursue some explicit purpose. In H. Simon engineering design terms [18], we think of EI as the social interface layer between, on one side, the problem space the participating systems deal with, and on the other the internal decision or functional intricacies of the various participating systems. EI —as artifact— involve a *conceptual* framework to describe agent interactions as well as an *engineering* framework to specify and deploy actual interaction environments.

3.1 EI conceptual framework

Our work on EIs is based on the following assumptions:

1. *Agent neutrality.* Participating entities are agents. In the accepted sense of being persistent, identifiable, communication-capable software or humans. We do not assume anything about the rationality, capabilities or intentions but we do assume they are able to communicate with other agents.
2. *Dialogical Stance.* All interactions are construable as speech acts. We assume there are some enabling and grounding conventions that link speech acts uttered within the institution with the actual problem environment in such a way that actions in the institution count as actions in the world. We also assume that there is a shared language whose semantic and pragmatic content is somehow fixed by the institution and adopted by the participants.
3. *Repetitive Interactions.* We assume that it makes sense to institute interaction conventions when there are interactions that happen not once but many times following a regular pattern. Furthermore, we assume that such patterns of interaction apply not to specific individuals but to any agent that performs a given role during those interactions. We further assume that those repetitive interactions may be organized into some hierarchical system composed by sets of speech acts and relations among these sets.
4. *Agent-mediated commitment making.* We assume that when participants communicate with other agents they are able and entitled to establish and fulfill commitments, and eventually abide by their consequences.
5. *Institutional commitments.* Only illocutions uttered by participating agents have effect on the shared environment. The institution is the trustee of the intended conditions for illocution utterance and effects, hence of the (social) commitments established through agent interactions within the institution.

These assumptions allow us to think of institutions as a set of conventions that constrain the class of dialogues to a set of admissible ones and their intended effects.

In order to operationalise those assumptions we commit to a language for dialogical exchanges and a way to express the restrictions that define those that are admissible in a given institution.

We have an abstract operationalization of these assumptions, we call it $\mathcal{EI}$ and we have made precise and implemented that operationalization as EI_0.

$\mathcal{EI}$ is based on three simple concepts : the elements needed to make sense of illocutions, *dialogical framework*; the structuring of interactions *performative structure*; and *rules of behavior*, the conventions that regulate commitments. More specifically:

1. The *Dialogical Framework* is the way of expressing the communication conventions that will prevail within the institution, including the domain ontology (which is then part of the institution) and the social structure of the intervening roles. Because of our dialogical stance, we hold a radical nominalistic view: the ontology of the institution is the vocabulary of the illocutionary formulae that will constitute the admissible dialogues.

Notice that by establishing a dialogical framework that is associated with a particular institution we are making explicit all the terms whose meaning needs to be *grounded*. In other terms, the grounding conventions of an electronic institution map entities of the institutional language to entities in the domain of application of the electronic institution.

2. *Performative Structure.* It establishes the intended constraints on conversation or interaction flow within the institution.
3. *Rules of Behavior.* The deontological conventions of an EI are partly expressed by the Performative Structure, but how social commitments are established by individual agents is expressible through what we have been calling rules of behavior. These rules state the pragmatics of institutional interactions; i.e., when is an agent entitled to utter an institutionally valid illocution and what are the social obligations entailed by uttering and receiving illocutions.

While this is a rather abstract operationalization of our assumptions, notice that the assumptions are consistent with the type of features that the lonja example exhibited and that the three components of $\mathcal{EI}$ were foreshadowed in my discussion in the previous section.

Since we are earnest in our purpose of constructing actual electronic institutions we need to be more precise in our operationalization. EI_0 is a precise instance of $\mathcal{EI}$ that we have been able to implement. Here is how we define it.

Dialogical Framework A traditional institution, say an auction house, restricts and gives meaning to interactions participants may take, and sees to it that the consequences of any interaction that takes place within the institution actually happen: if a good is being offered, the only action buyers can take is to rise their hand, indicating they take the bid, any other action is meaningless or inadmissible (and interpreted as a silent "no" to the bid). If a buyer wins a bid, the auctioneer will adjudicate the good to the buyer, charge the buyer and pay the seller for it; thus making the interactions involved relevant and meaningful to all participants. If we want to be able to build an electronic auction house we should be able to express that kind of conventions in a way that can be implemented, ascribed to by independent agents, and –basic to our purpose– in such a way that actual transactions can properly be carried out in the implemented institution.

We need to settle on a common illocutory language that serves to tag all pertinent interactions, or more properly, the valid speech acts. Formally, we take interactions to be illocutory formulae:

$$\lambda(speaker, hearer; \phi; t)$$

Speech acts, hence, start with an illocutory particle (declare, request, promise) that a *speaker* addresses to a *hearer* (who may be an individual or a group), at a time t, and the

content ϕ of the illocution is expressed in some object language whose vocabulary is the EI's ontology.

To fill in these formulae we need a vocabulary and grammar, but we also need to refer to speakers and listeners, actions, to time,.... We call all that the *Dialogical Framework* because it includes all what is needed for agents to participate in admissible dialogues in a given EI. We define it as a t-uple:

$$DF = \langle O, L, I, R_I, R_E, R_S \rangle$$

where

1. O stands for the EI vocabulary (ontology) ;
2. L stands for a content language to express the information exchanged between agents;
3. I is the set of illocutionary particles;
4. R_I is the set of internal roles (*i.e.*, roles that are performed by staff members of the institution);
5. R_E is the set of external roles;
6. R_S is the set of relationships over roles.

Performative Structure: Scenes and Transitions We assumed repetitive interactions are typical of institutions. As in our example of a traditional lonja, activities in an electronic institution are the composition of multiple, distinct and possibly concurrent, dialogical activities, each one involving different groups of agents playing different roles. Interactions between agents are articulated through recurrent dialogues which we call *scenes*, that follow some type of conversation protocol. The protocol of each scene restricts the possible dialogical interactions between roles, which can be multiply instantiated by different group of agents.

In EI_0, a scene protocol is specified by a directed graph whose nodes represent the different states of a dialogical interaction between roles. Its arcs are labeled with illocution schemata or timeouts.

More formally, a *Scene* is a t-uple:

$$s = \langle sR, sL, W, w_0, W_f, (WA_r)_{r \in R}, (WE_r)_{r \in R}, \Theta, \lambda, min, Max \rangle,$$

where

− sR is the set of scene roles involved in that scene;
− sL is the restriction to the scene of the EI dialogical framework as defined above;
− W is the set of scene states;
− $w_0 \in W$ is the initial state;
− $W_f \subseteq W$ is the set of final states;

- $(WA_r)_{r \in R} \subseteq W$ is a family of sets such that WA_r stands for the set of access states for role $r \in R$;
- $(WE_r)_{r \in R} \subseteq W$ is a family of non-empty sets such that WE_r stands for the set of exit states for role $r \in R$;
- $\Theta \subseteq W \times W$ is a set of directed edges;
- $\lambda : \Theta \longrightarrow L$ is a labeling function, where L can be a timeout, or an illocution schemata and a list of constraints;
- $min, Max : R \longrightarrow \mathbb{N}$ $min(r)$ and $Max(r)$ return the minimum and maximum number of agents that must and can play role $r \in R$.

In EI_0 we allow agents to enter a scene only if they are performing a role whose access is permitted at some given scene state. Likewise, agents may leave a scene only at those states where their role is allowed to exit. These restrictions or role-flows are needed in order to express conventions like allowing buyers to enter an auction room only when the auctioneer is ready and goods are available to be sold, or letting a buyer leave the auction house only when he or she has paid all dues.

Figure 4 is an example of a scene that defines the auctioning protocol for the Blanes Llotja. It is an actual specification in the ISLANDER graphic specification languag. At execution time agents interact by uttering instantiated illocutions matching the specified illocution schemata, and thus binding their variables to values.
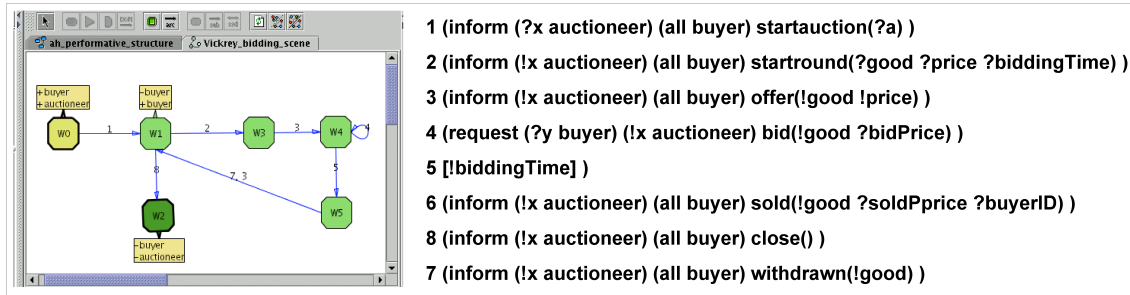

1 (inform (?x auctioneer) (all buyer) startauction(?a))

2 (inform (!x auctioneer) (all buyer) startround(?good ?price ?biddingTime))

3 (inform (!x auctioneer) (all buyer) offer(!good !price))

4 (request (?y buyer) (!x auctioneer) bid(!good ?bidPrice))

5 [!biddingTime])

6 (inform (!x auctioneer) (all buyer) sold(!good ?soldPprice ?buyerID))

8 (inform (!x auctioneer) (all buyer) close())

7 (inform (!x auctioneer) (all buyer) withdrawn(!good))

Fig. 4. The *bidding scene* of an electronic auction house. This ISLANDER specification shows the downward bidding conventions used in the Blanes auctions. Arcs are labeled by the messages participants may utter and nodes represent the state of the bidding.

While a scene defines a particular dialogical environment, the causal, temporal and other content relationships among scenes are expressed through a special type of scene we call *transitions*. Transitions can be thought of as gateways between scenes or as a change of conversation. Transitions are depicted by directed (annotated and labeled) arcs joining two scenes (through acceptable enter or exit states). Labels on the directed arcs determine which agents, depending on their roles, can progress from scenes to transitions, or from

transitions to scenes. In EI_0 we assume that there is always an initial and a final scene, which are the entry and exit points of the institution. We also allow, at run time, to have multiple instances of the same scene active.

The interlacing of scenes and transitions is captured through the *Performative Structure* of the EI.

Technically, we may now define a *Performative Structure* as follows:

$$PS = \langle S, T, s_0, s_\Omega, E, f_L, f_T, f_E^O, \mu \rangle, \text{ where}$$

1. S is a set of scenes;
2. T is a set of transitions;
3. $s_0 \in S$ is the *initial* scene;
4. $s_\Omega \in S$ is the *final* scene;
5. $E = E^I \bigcup E^O$ is a set of arc identifiers where $E^I \subseteq S \times T$ is a set of edges from scenes to transitions and $E^O \subseteq T \times S$ is a set of edges from transitions to scenes;
6. $f_L : E \longrightarrow FND_{2^{V_A \times R}}$ maps each arc to a disjunctive normal form of pairs of agent variable and role identifier representing the arc label;
7. $f_T : T \longrightarrow \mathcal{T}$ maps each transition to its type;
8. $f_E^O : E^O \longrightarrow \mathcal{E}$ maps each arc to its type (one, some, all or new);
9. $\mu : S \longrightarrow \{0, 1\}$ states whether a scene can be multiply instantiated at run time or not.

In Figure 6 we can see the performative structure of the Blanes Llotja. Buyers move from the initial scene to the buyer admission scene (BA at the top left of the diagram), from there they proceed to the auction room (AR, in the lower left) and from there to one or both of the buyer settlement scene (BS) or the last scene (OUTPUT). Similarly for sellers and staff members like the seller-admitter or the auctioneer, whose roles label the corresponding arcs. In this example, an auctioneer opens an auction and one may have many auction rooms active at any time, hence the label "new" in the arc going from ROOT to AR.

Rules of behavior and obligations So far we have dealt with the way interactions within an EI can be defined and organized, we now need to say how they are going to have the intended effect.

Recall that the main purpose of the institution is to make sure that what happens inside has the effects participants have agreed to while agreements, misconducts, or whatever else that happens outside is in principle disregarded or impertinent to the institution . Actions within an institution, we said, are speech acts. Consequently, those speech acts that are made as prescribed by the performative structure of an institution (during an enactment) create obligations on participants: socially binding commitments whose fulfillment is warranted by the institution. We make such intended effects of commitments explicit through what we have been calling *rules of behavior*.

We need two predicates to express the connection between illocutions and rules:

- *uttered*$(s, w, \mathbf{i})$ denoting that an instantiated illocution unifying with the illocution scheme $\mathbf{i}$ has been uttered at state w of scene s.
- *uttered*$(s, \mathbf{i})$ denoting that an instantiated illocution unifying with the illocution scheme $\mathbf{i}$ has been uttered in a state of scene s.

This way we can refer to the utterance of an illocution within a scene or when a scene execution is at a specific state.

We now define behavioral rules as first-order formulae of the form,

$$\left(\bigwedge_{j=1}^{n} uttered(s_j, w_{k_j}, \mathbf{i}_{l_j}) \wedge \bigwedge_{k=0}^{m} e_k \right) \rightarrow \left(\bigwedge_{j=1}^{n'} uttered(s'_j, w'_{k_j}, \mathbf{i}'_{l_j}) \wedge \bigwedge_{k=0}^{m'} e'_k \right)$$

where s_j, s'_j are scene identifiers, w_{k_j}, w'_{k_j} are states of s_j and s'_j respectively, $\mathbf{i}_{l_j}, \mathbf{i}'_{l_j}$ are illocution schemata l_j of scenes s_j and s'_j respectively, and e_k, e'_k are boolean expressions over variables from the illocution schemata $\mathbf{i}_{l_j}$ and $\mathbf{i}'_{l_j}$, respectively.

The intuitive meaning of behavioral rules is that if fully instantiated illocutions matching $\mathbf{i}_{l_1}, \ldots, \mathbf{i}_{l_n}$ are uttered in the corresponding scene states and the expressions $e_1, \ldots, e_m$ are satisfied, then fully instantiated illocutions matching $\mathbf{i}'_{l_1}, \ldots, \mathbf{i}'_{l'_n}$ satisfying the expressions $e'_1, \ldots, e'_{m'}$ must be uttered in the corresponding scene states.

Notice that $\mathbf{i}'_{l_1}, \ldots, \mathbf{i}'_{l'_n}$ can be regarded as *obligations* that agents acquire where the antecedent of the rule of behavior is satisfied. Therefore, agents must utter fully instantiated illocutions matching these schemata and satisfying $e'_1, \ldots, e'_{m'}$, in the corresponding scenes in order to fulfill the commitment.

3.2 EI building tools

In the IIIA, we have developed a suit of programs for engineering EI_0 electronic institutions. It is composed of:

ISLANDER A graphical tool that supports the specification and verification of dialogical frameworks, performative structures and norms [3].

SIMDEI Simulation tool to animate and analyze ISLANDER specifications prior to the deployment stage.

aBUILDER Agent development tool which given an ISLANDER specification gives support to the generation of agent skeletons for that institution. The generated skeletons can be used on simulations supported by SIMDEI, or in the actual execution of the institution supported by AMELI.

AMELI Software platform to run EI_0 institutions. The platform facilitates agent participation in the institution while enforcing the institutional rules [5]. Electronic institutions specified with ISLANDER are run by AMELI.

Monitoring tool A tool which permits the monitoring of EI executions run by AMELI. It graphically depicts all the events occurring during an EI execution.

These tools constitute an integrated development environment for engineering electronic institutions (*c.f.* [1]).[3] The suit allows the specification and deployment of any EI_0 institution, regardless of the purpose or content of the interactions. The ISLANDER tool facilitates the specification of the performative structure of an EI through a graphic editor and produces an XML document with the specification. That XML specification may be used by external users of the institution to build agents that comply with the specified conventions. In addition, aBUILDER generates agent templates that are capable of interacting with the specified institution and may be used to debug the specification or to build external agents on top of them. SIMDEI and the monitoring tool are convenient resources for debugging, texting and auditing an ISLANDER specified institution. AMELI produces run-time institutions from ISLANDER specifications. These institutions run on top of generic agent handling platforms like FIPA-compatible JADE [http://jade.tilab.com]. The use of these tools in the development cylce of an EI_0 institution is explained in Figure 5.
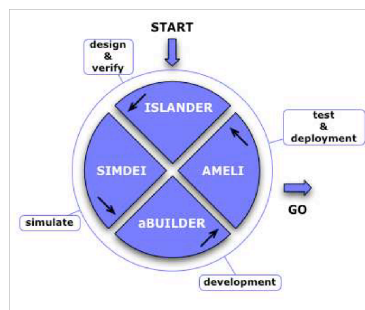

Fig. 5. The IIIA team has deployed an integrated development environment that can be used to specify and test electronic institutions following the cycle depicted here

Figure 6 shows a screen-shot of an ISLANDER specification of the Blanes Llotja, labeled *fm_institution*. The central diagram shows the performative structure. Panels on the left show synthetic information about the *fm_institution*, the list of the scenes (top) and the components of the performative structure: scene labels, transition data (bottom). The bottom panel shows the results of the automated debugging facilities of ISLANDER.

4 Revisiting EIs with legal eyes

4.1 Institutions as deontic theories

In the previous sections I have been using a theatrical metaphor to describe the way interactions are structured in an institution: agents exchange illocutions according to some

---

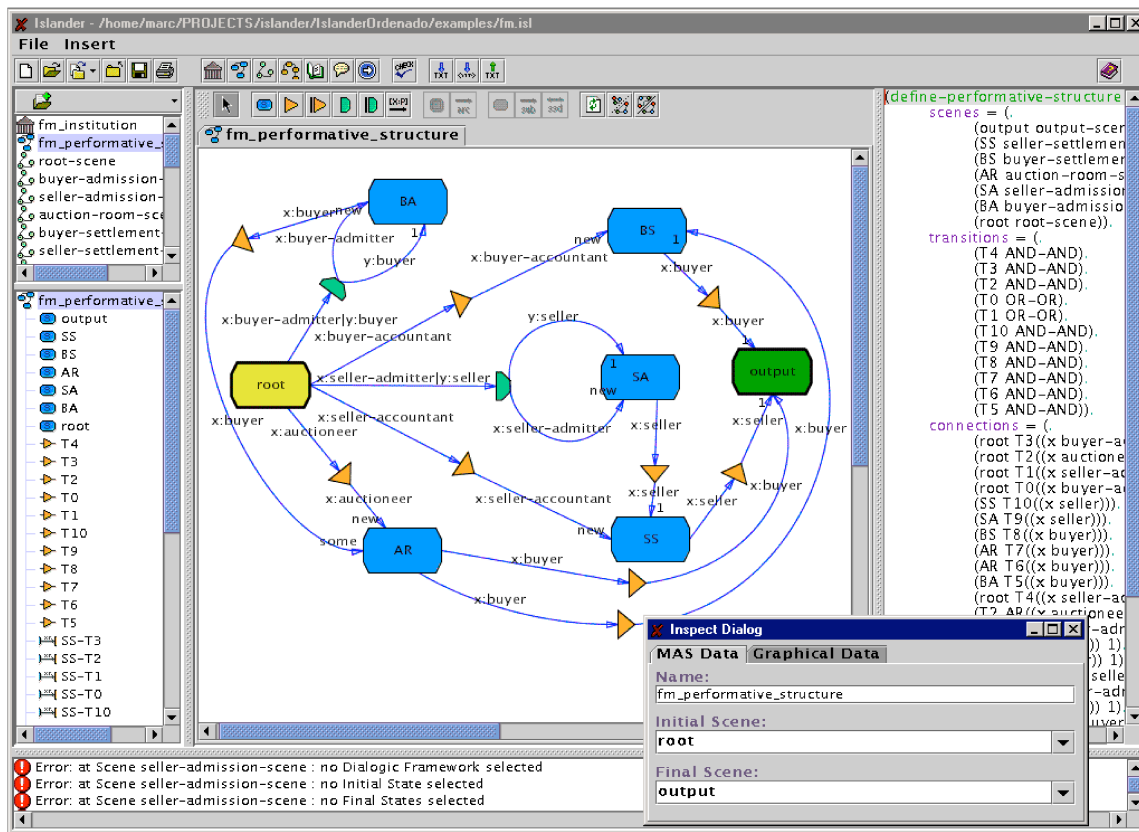[3] These tools are available on-line at http://e-institutions.iiia.csic.es

Fig. 6. A screen shot of the ISLANDER specification of the *Performative Structure* of an electronic auction house (fm in this case). It defines the interrelation between institutional *scenes* (boxes) –buyer admission, auction room, buyer settlements, etc.— by expressing the conditions for participants to move from one scene to another (labeled arcs).

sort of a script within the scene, and scenes are linked —by transitions— so that a more or less elaborate play is repeatedly performed. That metaphor is useful to explain institutional conventions as constraints that are imposed on dialogues: a scene is a dialogue that follows an explicit protocol and deals with some particular set of issues, a transition is the way agents may get into different conversation.

However, it is also possible to think of performative structures as sets of formulae and of institutional enactments as a grounded version of that set. I explain: valid illocutions are formulae, as we have seen and there are formulae that correspond to preconditions or postconditions of illocutions, there are formulae that correspond to the conditions that belong to transitions and, finally, there are formulae that are the obligations generated during the enactment. All four types of formulae are instantiated during run-time through the actual illocutions that are uttered by participants. In other words, an enactment of an institution produces a particular collection of formulae: the set of formulae whose variables have been instantiated during the enactment of the institution by the utterance of some illocutions and the obligations that are the consequences of those utterances.

In that sense, the institutional conventions restrict the collections of formulae that may be expressed to the subset of those that are consistent with the performative structure and the obligations generated through the enactment. That is, when a legitimate illocution is uttered, the institution passes from one valid state of the institution to another but, according to the conventions of the institution, from that new state there will be subsequent states that are valid and some that are not. Agents that utter, or hear, a given legitimate illocution have *obligations* that restrict their future utterances. Obligations mean that because certain illocution is expressed by one agent at some moment in time, then that agent or the agents that receive that illocution *cannot say* something else at some time in the future, or that that those agents *has or have to say* something at some time in the future.

Without going into the details it is readily seen that interaction protocols and obligations may be stated as modal formulae that state prohibitions, permissions and commitments with respect to agents institutional actions. The illocutionary status of each agent's own theory depends on the illocutions it has uttered or heard and how these have taken it from one institutional state to another.

More technically speaking, we may say that institutions are deontic theories, and each step of each enactment is a possible world that is consistent with that theory whenever legitimate illocutions are uttered.

Although it is easy to show that for any given ISLANDER specification of an institution, scenes can be translated into first order deontic formulae, the full complexity of the dialogical dynamics of enactments in arbitrary ISLANDER-specifiable performative structures, is considerable. The formal pragmatics of EI_0 institutions is still to be developed. Notwithstanding that daring purpose, we have advanced in the complementary direction and proposed deontic languages that, being expressive enough to capture typical institutional conventions, are also executable as EI_0 institutions. We have proposed [6] a machine

executable language that is able to capture some of the "normative languages" common in agent-based systems (*e.g.* [17, 9, 10]). A collection of expressions in those languages is "interpreted" into a set of production rules whose notion of consequence is interpreted with a typical (JESS) deduction engine. The advantage of this simple approach is that normative expressions that are interesting enough are expressible in this language, that their consequences can be properly managed and that it is compact enough to attach the whole interpreter to the software device —*governor* (see below)— that we attach to each agent during the enactment of an EI_0 institution. Further work is being done in our group in that direction, so that there is a machine-executable language that satisfies a well defined set of desiderata to deal properly with agents' normative positions (citeAndresAAMAS06.

4.2 EI_0 as normative systems

When seen as deontic theories, EI_0 institutions —as specified in ISLANDER— happen to posses the following distinguishing features:

1. The conventions are specified at design-time and once set cannot be changed unless a new specification is made.
2. All conventions are explicit as part of the performative structure or as rules of behavior. As such, all conventions may become known by external agents in an objective and precise way that is not open to interpretation.
3. Convention compliance is obligatory for all agents.
4. Conventions are strictly enforced by the electronic institution.

Such strict and rigid conception of electronic institutions is appropriate for many applications in which the interaction is straightforward and specially when liability of participants is relevant. In those cases risk may be properly allocated at design time through appropriate institutional conventions and no misbehavior is tolerated. The typical case is electronic commerce: it is convenient to have stable conventions for transactions, it is necessary to guarantee that transactions are carried out in exactly the way they are supposed to be accomplished and that all possible wrongdoing is properly dealt with [12].

EI_0 institutions are designed to behave that way and they can in practice achieve that degree of compliance with conventions thanks to a software device we call "governor" that is attached to every participating agent. Governors are fully controlled by the institution since they form part of its infrastructure. In fact, each EI_0 governor is a local version of the EI that governs its agent according to the EI conventions, In practice the governor controls illocution utterances in time and form, it controls scene transitions and it also keeps track of all commitments that involve its agent.

4.3 Conventional compliance and enforcement on a wider perspective

As explained above, because of the way we currently specify and implement EI_0 institutions, only *legitimate* illocutions may be uttered. Interactions are regulated as strict proto-

cols, all whose possible outcomes are centrally defined *a priori* and cannot be changed at run-time, hence they are static in a rather strong way. Thus if we come to think of EI_0 institutions as normative theories, their norms are obligatory observance for all participating agents, and furthermore, their enforcement is strict, tthere is no way an agent may infringe a convention, its governor inhibits any misbehavior.

There are, however, many other types of on-line interactions that may not need to be that restrictive, or that could or should not be that strict and static. For those kinds of applications, we would like to extend the EI_0 class and be able to specify and implement other types of EIs. That is the reason behind our current efforts in more expressive deontic languages that may capture, declaratively, rules that, for instance, have sanctions and repair actions attached or rules whose application is time-dependent [7]. These new forms of expressing institutional conventions are being coupled with convention enforcement mechanisms that allow for less strict compliance. Example of such mechanisms is the inclusion of rule-enforcement roles that may be played by internal or external agents who in turn may be subject to strict or lenient compliance of the institutional conventions. Likewise, other more general operationalizations of $\mathcal{EI}$ are desirable because we might want to have flexible EIs that evolve with the changing environmental conditions –*e.g.*, competitive pressure in markets, unforeseen procedures in medical protocols. For that reason we have been looking into richer and less rigid performative structures [14].

Administrative procedures such as public procurement suggests these other types of needs EIs should in principle be able to satisfy. How can one decide the balance between the procedural aspects of the process and the declarative ones? The first type may be implemented in classical EI_0, but the second type would certainly need richer deontic expressions to be computed in real-time. In general, administrative procedures may follow some standard steps, but in many cases discretional behavior is admitted or unavoidable. If one sets up an EI for e-government procurement, for example, how strict should it bet? And if some inadmissible behavior is unavoidable, how should the EI deal with it? merely keeping track of wrongdoing? applying penalties? bringing human authorities into the ensuing process?

The case of medical protocols is also worth mentioning for other reasons: protocols in addition to having procedural as well as a declarative prescriptive content, exist within a hierarchical normative framework (hospital policies, medical practice guidelines, health regulations, human-rights law) that affects them. However, in contrast to administrative procedures that share these features, medical protocols need by force to have a certain flexibility in order to let physicians deal with unforeseen developments, atypical cases and unprecedented situations. How, then, should hospital management systems deal with these medical protocols? is an EI-like implementation of those environments feasible? safe lability-wise? legally admissible?

Other types of sophistication beyond the ones just mentioned need to be studied when applying the more general notion of electronic institution to other virtual interactions. A case at hand are contracts among firms that decide to constitute a virtual organization

[8]. A contract is made around some agreed upon conditions that are generally subject to a legal framework. The contract may change or be updated by the partners, and it amy also fail. In any case, contracts need to valid under a legal framework and if some violation of contract occurs, the plaintiff should be able to invoke the contract clauses or the legal framework to prove wrongdoing and probably to have compensation. In such cases, if an electronic institution is set up to facilitate virtual organizations, the system in addition to handling the dynamic nature of contracts should also be able to detect violations, assign blame and take corrective actions. Of course there are many alternatives to implementing these capabilities. it may be that we want the EI to perform all those tasks on its own, or possibly serve as a mechanism for due process to resolve conflicts derived from the execution of the contract.

4.4   A final remark

All examples I just mentioned satisfy the assumptions I made at the beginning of this paper: they are repetitive processes, where agents —whose internal motivations we do not or cannot control— establish commitments —through canonical expressions— that are to be fulfilled under the supervision of a third party: the electronic institution itself. For these examples for these types of electronic institutions —that are more sophisticated than the lonja—, we would like to be able to express their conventions properly and in such a way that from that expression we may build a computational counterpart where humans and their software counterparts may interact effectively. Consequently, for those conventions that constitute such types of institutions we need to capture subtle aspects like issuance, adoption, compliance, enforcement, blame assignment, punishment, rewarding, restoration, interlocking, evolution of conventions. Should I refrain from calling these aspects *normative*?

In order to deal with the design challenge of electronic institutions, we computer scientist have been guided by our habits of tool builders. However if we really want to design and build effective sophisticated electronic institutions we need the knowledge an experience of legal thinkers. I presume, legal thinkers may find that challenge worthy for them as well.

5   Acknowledgments

References

1. Josep Arcos, Marc Esteva, Pablo Noriega, Juan Rodriguez-Aguilar, and Carles Sierra. Environment engineering for multiagent systems. *Engineering Applications of Artificial Intelligence*, (18):191–204, 2005.

2. Guifré Cuní, Marc Esteva, Pere Garcia, Eloi Puertas, Carles Sierra, and Teresa Solchaga. MASFIT: Multi-Agent System for Fish Trading. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)*, Valencia, Spain, 2004.

3. M. Esteva, D. de la Cruz, and C. Sierra. Islander: an electronic institutions editor. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2002)*, pages 1045–1052, Bologna, Italy, July 15-19 2002.

4. Marc Esteva. *Electronic Institutions: from specification to development.* PhD thesis, Universitat Politècnica de Catalunya (UPC), Barcelona, Catalonia, Spain, 2003. Published by the Institut d'Investigació en Intelligència Artificial. Monografies de l'IIIA Vol. 19, 2003.

5. Marc Esteva, Juan A. Rodriguez-Aguilar, Bruno Rosell, and Josep Lluis Arcos. AMELI: An agent-based middleware for electronic institutions. In *Third International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS'04)*, pages 236–243, New York, USA, July 19-23 2004.

6. Andres Garcia-Camino, Pablo Noriega, and Juan Antonio Rodriguez-Aguilar. Implementing norms in electronic institutions. In *Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*, 2005.

7. Andres Garcia-Camino, Juan Antonio Rodriguez-Aguilar, Carles Sierra, and Wamberto Vasconcelos. Norm-Oriented Programming in Electronic Institutions: A Rule-Based Approach. In *Coordination, Organizations, institutions and Norms in Agent Systems (COIN'06) in Fifth International Joint Conference in Autonomous Agents and Multiagent Systems (AAMAS'06), Hakodate 2006*, 2006. (forthcoming).

8. Henrique Lopes-Cardoso and Eugénio Oliveira. Virtual enterprise normative framework within electronic institutions. In *5th International Workshop on Engineering Societies in the Agents World (ESAW 2004*, Toulouse, October 2004.

9. F. López y López, M. Luck, and M. d'Inverno. Constraining autonomy through norms. In *Proceedings of The First International Joint Conference on Autonomous Agents and Multi Agent Systems AAMAS'02*, pages 674–681, 2002.

10. Loizos Michael, David C. Parkes, and Avi Pfeffer. Specifying and monitoring market mechanisms using rights and obligations. In Peyman Faratin and Juan A. Rodrguez-Aguilar, editors, *Agent-Mediated Electronic Commerce VI: AAMAS 2004 Workshop, AMEC 2004, New York, NY, USA, July 19, 2004. Revised Selected Papers*, volume 3435 of *Lecture Notes in Artificial Intelligence*. Springer, 2006.

11. Pablo Noriega. *Agent Mediated Auctions: the Fishmarket Metaphor*. PhD thesis, Universitat Autònoma de Barcelona (UAB), Bellaterra, Catalonia, Spain, 1997. Published by the Institut d'Investigaci en Intelligncia Artificial. Monografies de l'IIIA Vol. 8, 1999.

12. Pablo Noriega, Juan Antonio Rodriguez-Aguilar, and Carles Sierra. *Aplicaciones de Tecnología de Agentes en Comercio Electrónico*, pages 231–241. Pearson, 2004.

13. Douglass C. North. *Institutions, Institutional change and economic performance*. Cambridge Universisy press, 40 west 20th Street, New York, NY 10011-4211, USA, 1990.

14. A Robles and Pablo Noriega. A framework for building ei-enabled intelligent organizations using mas technology. In M.P. Gleizes, G. Kaminka, A. Nowé, S. Ossowski, K. Tuyls, and K. Verbeeck, editors, *Proceedings of the Third European Conference in Multi Agent Systems (EUMAS 05)*, pages pp.344–354., Brussels, Belgium, December 2005.

15. Juan A. Rodríguez-Aguilar. *On the Design and Construction of Agent-mediated Electronic Institutions*. PhD thesis, Universitat Autònoma de Barcelona (UAB), Bellaterra, Catalonia, Spain, 2000. Published by the Institut d'Investigació en Intelligéncia Artificial. Monografies de l'IIIA Vol. 14, 2003.

16. Juan A. Rodríguez-Aguilar, Pablo Noriega, Carles Sierra, and Julian Padget. FM96.5 a java-based electronic auction house. In *Second International Conference on The Practical Application of Intelligent Agents and Multi-Agent Technology(PAAM'97)*, pages 207–224, 1997.

17. Marek Sergot. A computational theory of normative positions. *ACM Trans. Comput. Logic*, 2(4):581–622, 2001.

18. Herbert A. Simon. *The Sciences of the Artificial*. MIT Press, Cambridge, 1969.