Chapter 1
Modeling Agent Institutions

Nicoletta Fornara and Henrique Lopes Cardoso and Pablo Noriega and Eugenio Oliveira and Charalampos Tampitsikas

Abstract Everyday uses of the notion of institution and some typical institutions have been studied and formalized by economists and philosophers. Borrowing from these everyday understandings, and influenced by their formalizations, the notion of institution has been used within the agents community to model and implement a variety of socio-technical systems. Their main purpose is to *enable* and *regulate* the interaction among autonomous agents in order to achieve some collective endeavour. In this chapter we present and compare three frameworks for agent-based institutions (i) ANTE, a model that considers electronic institutions as computational realizations of adaptive artificial environments for governing multi-agent interactions; (ii) OCeAN, extended in MANET, a model for specifying Artificial Institutions (AIs), situated in agent environments, which can be used in the design and

Nicoletta Fornara
Università della Svizzera italiana, via G. Buffi 13, 6900 Lugano, Switzerland e-mail: nicoletta.fornara@usi.ch. Supported by the Hasler Foundation project nr. 11115-KG.

Henrique Lopes Cardoso
LIACC / Dep. Eng. Informática, Faculdade de Engenharia, Universidade do Porto Rua Dr. Roberto Frias, 4200-465 Porto, Portugal e-mail: hlc@fe.up.pt

Pablo Noriega
IIIA, Artificial Intelligence Research Institute, CSIC, Spanish Scientific Research Council e-mail: pablo@iiia.csic.es. Partially supported by the Consolider AT project CSD2007-0022 INGENIO 2010 of the Spanish Ministry of Science and Innovation and Generalitat de Catalunya 2009-SGR-1434

Eugenio Oliveira
LIACC / Dep. Eng. Informática, Faculdade de Engenharia, Universidade do Porto Rua Dr. Roberto Frias, 4200-465 Porto, Portugal e-mail: eco@fe.up.pt

Charalampos Tampitsikas
Università della Svizzera italiana, via G. Buffi 13, 6900 Lugano, Switzerland, University of Applied Sciences Western Switzerland, 3960 Sierre, Switzerland e-mail: charalampos.tampitsikas@usi.ch. Supported by the Swiss State Secretariat for Education and Research, COST Action IC0801 “Agreement Technologies”, project title “Open Interaction Frameworks, towards a Governing Environment”.

implementation of different open interaction systems; and (iii) a conceptual core model for Electronic Institutions (EIs), extended with EIDE, based on open, social, decomposable and dialogical interactions. Open challenges in the specifications and use of institutions for the realization of real open interaction systems are discussed.

1.1 Introduction

In everyday language, the notion of “institution” is used in different contexts, for example when one talks about the “institution of marriage”, when we say that a given university is an “institution of higher education”, or when we say that a politician does not behave “institutionally”. Those everyday uses and some typical institutions have been studied and formalized by economists, political scientists, legal theorists and philosophers (see [2, 49]). There are three features that these conventional understandings have. The first is the distinction between “institutional” and “brute” (or actual, physical or real) facts [51, 34], and the correspondence between the two. Another key conceptual element is the separation between the institution itself and the agents that participate in the collective endeavor that is the purpose of the institution. Finally, the assumption that institutions involve regulations, norms, conventions and therefore some mechanism of governance that make those components effective. In fact, most theoretical approaches to conventional institutions may be distinguished by the way this last assumption is made operational. In particular, while some approaches (for instance North [44] and Ostrom [46]) take institutions to be the conventions themselves–and consequently draw a clear distinction between institutions (conventions) and organizations (the entities that put the conventions in practice)— others (like Simon [52]) take institutions to be organizations (with rules or norms, institutional objects and due processes or procedures) but still keep individuals out of the institution.

Borrowing from these everyday understandings, and influenced by their formalizations, the notion of institution has been used within the agents community to model and implement a variety of socio-technical systems that serve the same purposes that conventional institutions serve. Artificial, electronic, agent-mediated, agent-based or, simply, agent institutions are some of the terms that have been used to name such computational incarnations of conventional institutions in the agents community, and for the sake of economy we take them as synonymous in this introduction. Their main purpose is to *enable* and *regulate* the interaction among autonomous agents in order to achieve some collective endeavour.

These agent institutions, as agent-based organizations do, play a crucial role as *agreement technologies* because they allow to specify, implement and enact the conventions and the services that enable the establishment, execution, monitoring and enforcement of agreements among interacting agents.

Agent institutions have been implemented as multiagent systems using different “frameworks” (conceptual models that have associated tools and a software architecture that allow implementation of particular institutions). However, these artificial institutions all hold three assumptions that mirror the three features of conventional institutions mentioned above:

1. Institution, on one hand, and agents, on the other, are taken as first-class entities. A particular institution is specified through a conceptual model, based on a metamodel, that may be more or less formalized, then it may be implemented on some type of institutional environment and enacted through interactions of some participating entities.
2. Institutions are open MAS, in the sense that: (i) it is not known in advance what agents may participate in an enactment, now when these agents may decide to enter or leave an enactment; (ii) the institution does not know what the particular goals of individual agents are; (iii) the institution has no control over the internal decision-making of agents (iv) agents may not necessarily comply with institutional conventions.
3. Institutions are regulated systems. Interactions in the agent institution must comply with some conventions, rules, and norms that apply to every participant agent and are somehow enforced. Regulations control interactions and are applicable to individual agents in virtue of the activities they perform and not because of who they are.

There are several ways that these assumptions lead to more precise notions of what constitutes an institution and how these may be implemented. This chapter discusses three frameworks that actually achieve that objective but before discussing those frameworks we would like to provide some background.

Institutions are Normative MAS. Institutions are a class of "normative multiagent systems" (norMAS) [7, 6]:

> A *normative multiagent system* is a multiagent system organized by means of mechanisms to represent, communicate, distribute, detect, create, modify, and enforce norms, and mechanisms to deliberate about norms and detect norm violation and fullfilment.

The ground assumption in normative MAS is that norms are used to constrain undesired behaviour, on one hand, but they also create a space of interaction where successful social interactions take, which as we mentioned before is what agent institutions do by setting and enforcing the rules of the game, creating an institutional reality where these rules apply and are enforced. Not surprisingly, agent institutions do have mechanisms that are similar to the ones listed in the description above because institutions (by definition) create the space of opportunity and constrain interactions to better articulate towards the common endeavour. The class of normative MAS and agent institutions are not the same because the mapping between the ideal mechanisms and the way an agent institution framework captures the mechanism is not obvious and is seldom fully established. The following sections will give substance to this last claim but some three prior qualifications are due.

- It is usually assumed that norms ought to be expressed as deontic formulas with a standard proof-theoretic notion of consequence associated to them. This is useful for a declarative description of conventions that is easy to communicate, promulgate and perhaps reason about (at design time as well as at run time). However it is not absolutely necessary, this because there may be other convenient ways of expressing different types of norms. For example, an artificial institution may express conventions that constrain agent actions in procedural (non-declarative) form, for instance using commitment-based protocols and dialogical games, and still use, say, model-checking devices to prove normative properties of the protocol. Likewise, an electronic institution describes permissions, obligations and prohibitions through finite state machines whose transitions are in fact conditional statements in a first order language and paths and propagation take the function of the modal operator; and in these networks, colored Petri nets may provide appropriate semantics for on-line and off-line normative conflict detection, for example.

- It is usually understood that such deontic formulas are enough to fully specify and govern a multiagent system. Not really. In addition to a collection of norms, a normative MAS requires several institutional constructs in order to legislate, apply, enforce and modify norms. Constitutive conventions for example may need extra-normative devices like bonds and identity certificates to provide entitlements to participating agents. Govenance mechanisms may require the existence of institutional agents that perform norm-enforcement functions, etc.
- Normative notions are pertinent only if norms may be violated. The actual situation is richer. There are application contexts where governance may need to be fully regimented (in electronic markets, for instance) and others that may not (conflict resolution, for example). Hence, enforcement mechanisms in an agent institution may involve a variety of components dealing with observability of actions, institutional power, law enforcement roles, reparatory actions, etc.

Institutions vs organizations The notions of *institution* and *organization* are closely related. The essential distinction, bluntly speaking, is that the institution is focused on what can be done, while organizations on who does it. Institutions, thus deal mainly with norms and governance, while organizations involve individuals, resources, goals. An institution creates a virtual environment, an organization is an entity in the world (a crude physical reality). An organization has boundaries that establish a clear differentiation: some rules apply inside, others apply outside; there is organizational staff, and there are customers and suppliers; there is a macroeconomic environment and there are objectives of the firm. On the other hand the organization also has several institutional components: best practices, social structure and roles, decomposable activities, internal governance. Although the distinction exists and may be formally stated in a crisp way, when we treat agent institutions, we tend to bundle together the specification of the institution with the implementation of that specification and what really blurs the distinction, we tend to identify the electronic institution (the virtual environment) with the running system that deals with actual transactions: that is, with the computational system *and* the firm that runs it.

Institutional Frameworks In this chapter from Section 1.2 to Section 1.5 we will present three frameworks for agent-based institutions that illustrate how the previously mentioned ideas about institutions are made precise enough to model actual institutions and implement them as multiagent systems. Those frameworks are: (i) ANTE, a model that considers electronic institutions as computational realizations of adaptive artificial environments for governing multi-agent interactions; (ii) OCeAN extended in MANET, a model for specifying Artificial Institutions (AIs), situated in agent environments, which can be used in the design and implementation of different open interaction systems; and (iii) a framework for Electronic Institutions (EIs), extended with the EIDE development environment, based on open, social, decomposable and dialogical interactions. In Section 1.6 we discuss and compare those three frameworks for agent-based institutions. Finally in Section 1.7 some open challenges in the field of specifications and use of institutions for the realization of real open multiagent systems are discussed.

We should mention that in addition to these three frameworks, there are at least three other proposals that share the above principles. The first is the *OMNI* model [18], which derives from the *OperA* and *HARMONIA* frameworks introduced in the dissertations of Virginia Dignum [17] and of Javier Vázquez-Salceda [56] respectively. The *OMNI* model allows the description of MAS-based organizations where agent activities are organized as agent scripts (scenes) that are built around a collective goal. The admissible actions of each scene are regulated by a set of norms. The *OMNI* model contains three types of institutional component: normative, contextual and organizational; whose contents are specifiable in three levels of abstraction: descriptive, operational, implementation. Lately, they have developed the *OperettA* framework [1], to support the implementation of real MAS. The second one is the *instAL* framework that puts together the research developed over may years in the University of Bath [15, 13]. *InstAL* is a normative framework architecture and a formal mathematical model to specify, verify and reason about norms that are used to regulate an open MAS. Finally, the third one is the recent proposal by J. Pitt et al. [48] that stems from [5] and draws on institutional notions proposed by E. Olstrom [47].

1.2 The ANTE framework: Electronic Institutions as Dynamic Normative Environments

In this section we will consider electronic institutions as computational realizations of adaptive artificial environments for governing multi-agent interactions.

The use of an *Electronic Institution* as an infrastructure that enables regulation in multi-agent systems presupposes the existence of a common environment where norms (see Part **??**) guide the way agents should behave. The role of an *institutional normative environment* [37], besides providing a set of regulations under which agents' collective work is made possible, is twofold: to check whether agents are willing to follow the norms they commit to (through monitoring), and further to employ correction measures as a means of coercing agents to comply (through enforcement) (see also Chapter **??** on this).

Furthermore, when addressing open systems, the normative environment should enable the run-time establishment of new normative relationships, which are to be appropriately monitored and enforced. Hence, instead of having a predefined normative structure, the shape of the environment will evolve and adapt to the actual normative relationships that are established.

In order to make this feasible, we believe it is important to provide some infrastructure that facilitates the establishment of norm governed relationships. For that, we propose the provision, in an electronic institution platform, of a supportive and extensible *normative framework* [38]. Its main aim is to assist software agents in the task of negotiating and establishing electronic contracts.

Having in mind real-world domains such as agreements guided by electronic contracting, the normative environment will, while monitoring the compliance to norms that apply to specific contracts, record a mapping from the relevant interactions that take place (which concern electronic contracting exchanges). The connection between real-world interactions and the institutional environment is made through illocutions (speech acts) that empowered agents [34] perform with the intent of informing the institution that certain contract-related events have occurred. With an appropriate interface between the normative environment and the statements that agents make, we incrementally build a state of *institutional reality* [51], which is an image of relevant real-world transactions that are, through this means, institutionally recognized (i.e., transactions are turned into *institutional facts* inside the normative environment).

Hierarchical normative framework. In order to facilitate the establishment of electronic contracts, the normative environment should provide a supportive and extensible normative framework. This framework may be inspired by notions coming from contract law theory, namely the use of "default rules" [16] – background norms to be applied in the absence of any explicit agreement to the contrary. We therefore propose that this normative structure is composed of a hierarchy of *contexts* [39], within which norms are created that may apply to sub-contexts. The context hierarchy tries to mimic the fact that in business it is often the case that a B2B contractual agreement forms the business context for more specific contracts that may be created. Each contract establishes a new context for norm applicability.

A *norm defeasibility* approach [38] is also proposed in order to determine whether a norm should be inherited, for a specific situation, from an upper context. This feature allows the normative framework to be adapted (to better fit a particular contract case) and extended (allowing new contract types to be defined). Furthermore, the rationale behind the possibility of overriding any norm is based on the assumption that "default rules" should be seen as facilitating rather than constraining contractual activity [35] (see also Chapter **??** on defeasibility of rules in law).

Adaptive norm enforcement. Adaptive enforcement mechanisms are important in open environments, where the behavior of an agent population cannot be directly controlled. When the normative specification of contracts includes flaws, namely by omitting normative consequences for some contract enactment outcomes, self-interested agents may try to exploit their potential advantage and intentionally violate contract clauses.

In general, an institution may employ two basic kinds of sanctions in order to incentive norm compliance. Direct *material sanctions* inflict immediate penalties, whereas indirect *social sanctions* have a more lasting effect, e.g. by affecting an agent's reputation. The effectiveness of these alternatives may differ according to the agents that interact within the institutional environment. If agents are not able to take advantage of reputation information, the use of material sanctions is probably a better alternative. Having in mind the deterrence effect of sanctions (i.e., their role in discouraging violations), an institution may use an adaptive sanction model to maintain order (by motivating agents to comply) and consequently trust in the system.

Economic approaches to law enforcement suggest analyzing sanctions by taking into account their effects on parties' activities. Based on this understanding, we have designed and experimentally evaluated a model for *adaptive deterrence sanctions* [40] that tries to enforce norm compliance without excessively compromising agents' willingness to establish contracts. Raising deterrence sanctions has a side effect of increasing the risk associated with contracting activities.

We believe that our approach, which has been implemented as part of the ANTE framework [41], has the distinctive features of being both an open and a computationally feasible approach to the notion of artificial institution. In fact, an *institution* is grounded on some notion of regulation, which is materialized through rules and norms. While some researchers, mostly from fields other than computer science, take an abstract and immaterial perspective to institutions, we find it natural, when addressing electronic institutions, to follow a more proactive stance and ascribe to an electronic institution the role of putting its regulations into practice. These regulations are seen as evolving according to the commitments that agents, when interacting in an open environment, are willing to establish amongst themselves, relying on the institutional environment for monitoring and enforcement purposes. The guiding line for our approach has been the field of electronic contracting.

1.3 The OCeAN metamodel for the specification of Artificial Institution

OCeAN (Ontology CommitmEnts Authorizations Norms)[30, 27] is a metamodel that can be used for specification of Artificial Institutions (AIs). Those institutions thanks to a process of contextualization in a specific application domain can be used and re-used in the design of different *open systems* thought for enabling the interaction of autonomous agents. The fundamental concepts that need to be specified in the design of artificial institutions are:

- an *ontology* for the definition of the concepts used in the communication and in the regulation of the interaction. With an application independent component with concepts and properties that are general enough (like the notion of time, action, event, obligation, and so on) and an application dependent part;
- the possible *events*, *actions*, *institutional actions and events* that may happen or can be used in the interaction among agents, this mainly in terms of preconditions that need to be satisfied for their successful performance and effects of their performance;
- the *roles* that the agents may play during an interaction and the rules for playing such roles;
- an *agent communication language (ACL)* for enabling a communication among agents, for example for promising, informing, requesting, agreeing and so on;
- the set of *institutional powers* for the actual performance of institutional actions;
- the set of *norms* for the definition of *obligations*, *prohibitions*, and *permissions*.

In our past works we have proposed a commitment-based semantics of an agent communication language [26] that is regulated by the basic institution of language [30]. We have formalized the concepts for the specification of AIs using different formalisms, and we have used them for specifying the institutions necessary for the design of different types of electronic auctions. In particular initially we specified our metamodel with a notation inspired by the UML metamodel and we used the Object Constraint Language [45] as notation for expressing constraints [31]. Subsequently, due to difficulties of efficiently matching the norms that regulate agents interaction with the actions performed by the agents and the need to perform automatic reasoning on the content of messages and norms, we decided to formally specify the basic concepts of our metamodel by using the Discrete Event Calculus (DEC), which is a version of the Event Calculus. The Event Calculus is a formalism that fits well for the purpose of reasoning about action and change in time, it has has been introduced by Kowalski and Sergot in 1986 [36]. DEC has been introduced by Mueller [42] to improve the efficiency of automated reasoning by limiting time to the integers. This formalism has the advantage of making easier the simulation of the dynamic evolution of the state of the interaction and making possible to perform automated reasoning on the knowledge about the state of the interaction. The main limits of this approach are that the DEC formalism is not widely known among software engineers and the performances of the prototype that we implemented for simulating a run of the English Auction did not scale well with the size of the concepts represented and the number of participating agents.

Consequently in 2009 we started to investigate the possibility to specify our model using Semantic Web Technologies [28, 25] (see also Part **??**). We proposed to specify the concepts (classes, properties, and axioms) of the OCeAN metamodel using OWL 2 DL: the Web Ontology Language recommended by W3C, which is a practical realization of a Description Logic system known as *SROIQ(D)*. We proposed an *upper level ontology* for the definition of the abstract concepts used in the specification of every type of artificial institution, like the concept of *event*, *action*, *time event, change event*, *temporal entity, instant of time* and so on. In particular for modeling time we used the standard OWL Time Ontology[1] enriched with some axioms useful for deducing information about instant of time and intervals. We specified the *OWL Obligation Ontology* [25] that can be used for the specification of the obligations that one agent has with respect to another agent to perform one action that belongs to a class of possible actions, within a given deadline, if certain activation conditions hold, and certain terminating conditions do not hold. Those obligations can be used to specify constrains on the behavior of the interacting agents and to express the semantics of conditional promises communicative acts [29]. The *OWL Obligation Ontology* together with some functionalities realized for performing closed world reasoning a certain classes can be used for *monitoring* the evolution in time of the state of the obligations on the basis of the events and actions that happens during the interaction. In fact reasoning in OWL is based on an *open world assumption* but in our model, in order to be able to deduce that an obligation to perform an action, when the deadline is elapsed, is violated, we need to implement closed-world reasoning and assuming that in the interaction contexts where this model will be used, not being able to infer that action has been performed in the past is sufficient evidence that the action has not been performed. Regarding monitoring it is also important to solve the problem of finding an efficient and effective mechanism for mapping real agents' actions in element of the OWL ontology for being able to perform automated reasoning on them and deducing that an obligation to perform a given action is fulfilled or violates. Currently the OCeAN meta-model has not been completely specified using Semantic Web Technologies, we plan to do it in our future works.

The main advantage of the choice of using Semantic Web technologies is that they are increasingly becoming a standard for Internet applications, and given that the OWL logic language is decidable, it is supported by many reasoners (like Pellet and HermiT), tools for ontology editing (like Protégé) and library for automatic ontology management (like OWL-API and JENA). Moreover the specification of artificial institutions in OWL makes them easily reusable as data construct in many different applications in different domains.

1.4 Artificial Institutions Situated in Environment: the MANET model

Thanks to the Agreement Technology COST Action in 2009 we started to investigate how to integrate the studies on the model of agent environments [57], in particular the model presented in the GOLEM framework [10], with the OCeAN meta-model of AI. As first result of this work we proposed the MANET (Multi-Agent Normative EnvironmenTs) model where AI are situated in agent environments [54].

One of the most important tasks of an *environment* is to mediate the actions and events that happen, where *mediate* means that an environment is in charge of regis-

[1] http://www.w3.org/TR/owl-time/

tering that an event has happened and of notifying this event to all agents registered to the template of this event (the agents that have a sensor for this type of events) [10]. An environment is composed of *objects* and *physical spaces*, and is the place where *agents* interact. A physical space describes the infrastructure of the system and its infrastructural limitations to the agents behavior in terms of physical rules.

Given that AIs are abstract description specified at design time, it is crucial to specify how certain AI can be concretely used at run-time for the definition and realization of open systems. Therefore we proposed to introduce in the model of environments the notion of *institutional space* that is used for having a first-class representation of AIs. In particular institutional spaces represent the boundaries of the effects of institutional events and actions performed by the agents, they may contain sub-spaces, and they enforce the norms of the system in response to the produced events.

Given that institutional spaces may contain sub-spaces, it is possible that the different AIs, used for the specification of different institutional spaces, may present some interdependencies. For example in a marketplace we can have many different auctions represented with sub-spaces created using different AIs. Given that agents may contemporarily participate in more than one space, it may happen that the norms of one space, for example the marketplace, regulate also some events of its sub-spaces, for example by prohibiting to an agent to do bid in an auction represented in a sub-space if it has a specific role in the market-place. For soling this problem it is necessary to give to the designer of the system the possibility to define events that may be *observed* outside the boundaries of the space. Another problem may arise when the rules a space (for example an auction) regulate for instance the participation of an agent to another space (another auction or a contract). In this case we need to introduce in the model the possibility for one space to *notify* another space about the fact that a specific event is happened.

The MANET model of artificial institutions situated in environment has been implemented in Prolog on top of GOLEM platform [10] and it was used for formalizing and running an e-energy marketplace [54] where agents representing different types of energy producers try to sell energy to potential consumers.

1.5 Electronic Institutions

The work we have been doing in the IIIA on electronic institutions (EIs, for short) may be observed from four complementary perspectives:

1. *The mimetic perspective:* EIs can be seen as computational environments that mimic the coordination support that conventional human institutions provide.

2. *The regulated MAS perspective* understands EIs as open multiagent systems, that organise collective activities by establishing a restricted virtual environment where all interactions take place according to some established conventions.

3. *EIs as "artifacts" perspective* takes EIs to be the operational interface between the subjective decision-making processes of participants and the social task that is achieved through their interactions.

4. *The coordination support perspective:* EIs are a way of providing structure and governance to open multiagent systems.

These four characterizations are supported by one single abstract model whose assumptions and core components we briefly discuss below. In turn, as we'll also see below, this abstract model is made operational through a set of software components that follow one particular computational architecture.

Over the past few years we have had the chance to build numerous examples of electronic institutions in a rather large variety of applications with those tools [19][2].

A conceptual core model for Electronic Institutions. Electronic institutions are grounded on the following basic assumptions about interactions:

- *Open.* Agents are black-boxes, heterogenous, self-motivated and may enter and leave the institutional space on their own will.
- *Social.* Agents come together in pursuit of an endeavour that requires a collective participation; thus agents need to be aware of other agents and their roles and of the capabilities needed to achieve a particular goal in a collective activity.
- *Decomposable.* To contend with the possibility (due to openness) of large number of agents being involved in the social interaction we allow the collective endeavour to be decomposed into atomic activities (*scenes*) that achieve particular goals with the participation of fewer individuals. The decomposition requires that scenes be connected in a network in which the achievement of individual and collective goals correspond to paths in that network.
- *Replicable.* Simple activities may be either re-enacted by different groups of agents or enacted concurrently with different groups.
- *Co-incident.* An agent may be active, simultaneously, in more than a single activity[3].
- *Contextual.* Openness and decomposability limit the knowledge agents have of each other, thus interactions are naturally *local* within subgroups of agents that share a common "scene context", while as a dynamic virtual entity, the collectivity of agents is itself immersed in a larger "institutional context".
- *Dialogical.* Activities are achieved through interactions among agents composed of non-divisible units that happen at discrete points in time. Thus construable as point-to-point messages in a communication language, so that even physical actions may be thus wrapped[4].

[2] The IIIA model of Electronic Institutions is the result, mainly, of three dissertations [43, 50, 20]

[3] We will deal with to this ubiquity of a given agent as *agent processes* that stem from it, so that we have an objective ground for concurrency and control issues when implementing the institutional infrastructure.

[4] Messages make reference to an application domain and should be properly "anchored" (their meaning and pragmatics should be established and shared by participants), e.g. the term "pay" entails the real action of transferring funds in some agreed upon way; in a trial, the constant "exhibit A" corresponds to some object that is so labeled and available at the trial.

These assumptions allow us to represent the conventions that will regulate agent interactions with the few constructs depicted in Figure 1.1. The full detail of these constructs is presented in [3] but, broadly speaking, to specify an EI we need:

1. *A dialogical framework* that consists essentially of (i) a social model of roles and their relationships; (ii) a domain and a communication languages that will be used to express the institutional messages, plus a few other languages for expressing institutional constraints, and (iii) an information model to keep the *institutional state*, that is, the updated values of institutional variables.
2. *A performative structure* that captures the high level structure of the institutional interactions as a network of scenes connected by transitions.
3. *Procedural and behavioural constraints* that affect the contents of the performative structure; namely, (i) preconditions and postconditions of messages within scenes, (ii) constraints on the movement of roles between scenes and (iii) propagation of the effects of actions among scenes; for expressing all these constraints we make use of the tower of languages of the dialogical framework.

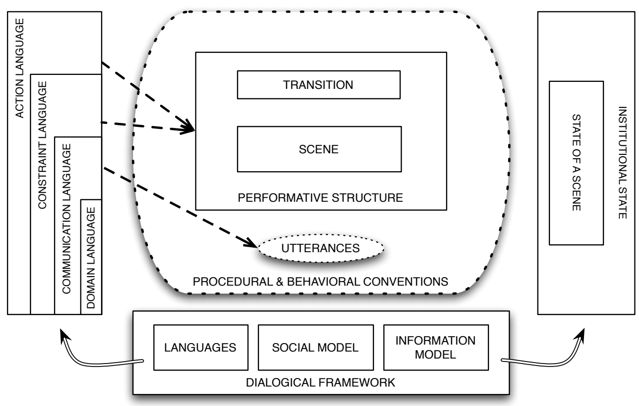

Fig. 1.1 Sketch of the Electronic Institutions Conceptual Model.

Our model has a straightforward operational semantics: institutional reality is changed through agent actions, but only those agent actions that comply with the institutional constraints have any institutional effect. More precisely, the institutional state is only altered through actions that comply with the procedural and behavioural constraints and in our model the only possible actions an agent can take are: to utter a message, to enter and leave the institution, and to move between scenes. Figure 1.1, hides the fact that an *electronic* institution also constitutes the infrastructure that *enables* actual interactions. Thus, we need that our conceptual model includes all those operations that need to be supported by the infrastructure; namely, those operations triggered by the actions of an agent that we just mentioned, plus those operations that the infrastructure itself needs to accomplish so that the first ones are feasible. Table 1.1 summarizes all those operations, the last column indicates the constructs that the operation updates.

Operation	**Called by**	**Effect on**
Speak	Agent	scene
RequestAccess	Agent	electronic institution
JoinInstitution	Agent	electronic institution, scene
LeaveInstitution	Agent	electronic institution, scene
SelectNewTargets	Agent	transition
RemoveOldTargets	Agent	transition
StartElectronicInstitution	Infrastructure	electronic institution
CreateSceneInstance	Infrastructure	scene institution
CloseSceneInstance	Infrastructure	scene
EnableAgentsToLeaveOrTransition	Infrastructure	transition
EnableAgentsToLeaveAndTransition	Infrastructure	transition
MovingFromSceneInstanceToTransitionInstance	Infrastructure	scene, transition
MoveAgentFromTransitionToScene	Infrastructure	scene, transition
RemoveClosedInstances	Infrastructure	electronic institution
Timeout	Infrastructure	scene

Table 1.1 Electronic institution operations

One computational architecture for Electronic Institutions. The model just presented may be implemented in different ways. We have chosen one particular architecture (see [23]) where we build a centralized institutional infrastructure that is implemented as a separate "social milieu" that mediates all the agent interactions, as Fig. 1.2 shows.

- *Governor* All communications between a given agent and the institution are mediated by a corresponding infrastructure agent that is part of the institutional infrastructure called the *governor* (indicated as G in Figure 1.2).[5] The governor keeps a specification of the institution plus an updated copy of the institutional state, thus when its agent produces an utterance, that utterance is admitted by the governor if and only if it complies with the institutional conventions as they are instantiated at that particular state; only then, the utterance becomes an institutional action that changes the state. Likewise, the governor communicates to the agent those institutional facts that the agent is entitled to know, the moment they happen. Additionally, the governor controls navigation of its agent between scenes, and the production of new instances of the agent itself (*agent processes*). It also keeps track of time for synchronization (time-outs) purposes. Note that in order to provide these services, a governor must coordinate with scene managers, transition managers, and the institution manager. In this realisation of the

[5] Agents cannot interact directly with one another, they use an agent communication language (like JADE) to interact with their governors who mediate their interactions *inside* the electronic institution.

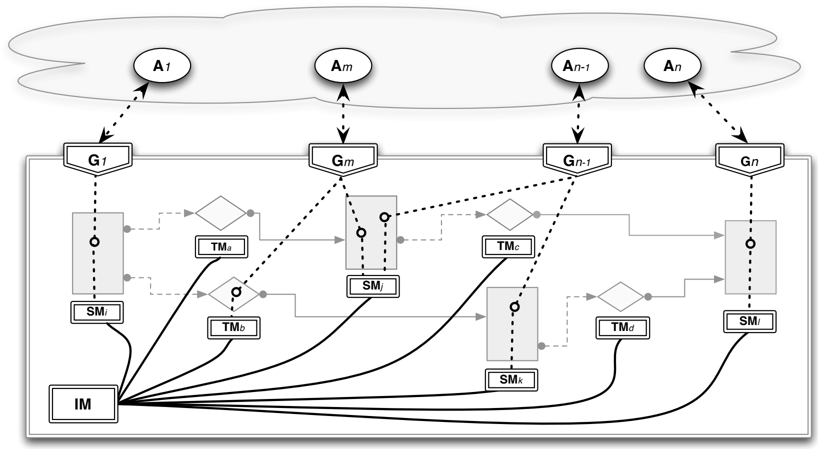

Fig. 1.2 An architecture for electronic institutions. Participating agents (A), communicate with (infrastructure) governor agents (G), which in turn coordinate with other infrastructure manager agents for each scene (SM) and each transition (TM) and with the institution manager agent (IM).

EI framework, therefore, governors are involved in the implementation of most of the operations in Table1.1.

- *Institution Management* Each institution has one *institution manager* agent (IM), which activates (*StartElectronicInstituion* operation) and terminates the institution. It also controls the entry (*RequestAccess*, *JoinInstitution*) and exit (*LeaveInstitution*) of agents, together with the creation the closing of scenes (*CloseSceneInstance*, *RemoveClosedInstances*). Finally, it keeps track of the electronic institution state.
- *Transition management* Each transition has a *transition manager* (TM) that controls the transit of agents between scenes by checking that requested moves are allowed (*EnableAgentsToLeaveOrTransition, EnableAgentsToLeaveAndTransition*) and, if so, allowing agents to move (*MovingFromSceneInstanceToTransitionInstance, MoveAgentFromTransitionToScene*).
- *Scene management* Each scene has an associated infrastructure agent, the *scene manager* (SM), who is in charge of: starting and closing the scene (in coordination with the institution manager); keeping track of agents that enter and leave the scene; updating the state of the scene by processing utterances (*Speak*) and time-outs (*Timeout*); and coordinating with transition managers to let agents in or out of a scene (*MovingFromSceneInstanceToTransitionInstance, MoveAgentFromTransitionToScene*).

Other architectures are feasible and we have, for instance, suggested a peer-to-peer variant of these ideas in [22].

A development environment based on that architecture. The computational model we just described, does not commit to any specific convention about the languages used in the specification of transitions and scenes, nor on the syntax and pragmatics of illocutions, nor on specific governance mechanisms. Those commitments come later when software tools to build actual electronic institutions become implemented. One way of implementing the computational model is the Electronic Institutions Development Environment (EIDE) [24] which includes the following tools:

ISLANDER: a graphical specification language, with a graphic interface [21]. It allows the specification of any EI that complies with the conceptual model and produces an XML file that the AMELI middleware runs[6].

AMELI: a software middleware that implements the functions of the social layer at run-time [23]. It runs an enactment, with actual agents, of any ISLANDER-specified institution. Thus it activates infrastructure agents as needed; controls activation of scenes and transitions, access of agents, messages between agents and institution, and in general guarantees —in coordination with infrastructure agents— the correct evolution of scenes and the correct transitions of agents between scenes. AMELI may be understood as a two-layered middleware. One public layer formed by governors, the other private layer —not accessible to external agents— formed by the rest of the infrastructure agents. External agents are only required to establish communication channels with their governors[7]. Infrastructure agents use the institutional state and the conventions encoded in the specification to validate agent actions and evaluate their consequences.

SIMDEI: is a simple simulation tool used for debugging and dynamic verification. It is coupled with a *monitoring tool* that may be used to display the progress of the enactment of an institution. It monitors every event that takes place and may display these events dynamically with views that correspond to events in scenes and transitions or events involving particular agents. Both tools may be used for dynamic verification.

aBUILDER: an agent development tool which, given an ISLANDER-specified institution, supports the generation of "dummy agents" that conform to the role

[6] ISLANDER allows static verification of a specification. It checks for *language integrity* (all roles and all terms used in illocutions, constraints and norms are properly specified in the dialogical framework), *liveness* (roles that participate in a given scene have entry and exit nodes that are connected and may be traversed), *protocol accessibility* (every state in the graph of a scene is accessible from the initial state and arcs are properly labeled), *norm compliance* (agents who establish "normative commitments" may reach the scenes where the commitments are due). ISLANDER may be extended to have a strictly declarative expression of scene conventions [33].

[7] The current implementation of the infrastructure can either use JADE or a publish-subscribe event model as communication layer. When employing JADE, the execution of AMELI can be readily distributed among different machines, permitting the *scalability* of the infrastructure. Notice that the model is communication-neutral since agents are not affected by changes in the communication layer.

specification and are able to navigate the performative structure, provided agent designers fill up their decision-making procedures[8].

Extensions of the framework: The Ei framework has been used in many application domains (see [19] p. 87 for an enumeration). This experience as well as more theoretically minded research has motivated adaptations and extensions to it. These are extensions are mainly due to(i) a normative understanding of electronic institutions [33, 55], (ii) the advantage of connecting the EI environment with other services [4], (ii) to achieve peer-to-peer architecture in order to address scalability [22] and institutions that evolve over time in order to adapt to changing conditions of the environment [11, 8]. A significant extension of the framework is that of the automatic generation of three dimensional immersive environments that represent the electronic institution. This work is described later in this book (Ch. **??**).

1.6 Conclusions: A comparison of the described institutional models

In this section we compare the three proposed models of institutions, ANTE, OCeAN/MANET, and EI, discussing their crucial differences and analogies on a set of relevant aspects.

- *Institutional reality*.
  All three models adhere to the representation of institutional reality proposed by John Searle in [51], in particular on the existence of an institutional reality that has a correspondence with the real or physical world, and on distinguishing between "institutional" facts and actions, on one side, and their possibly corresponding "brute" facts and actions, on the other.

- *Social model: roles and hierarchy of roles*
  - ANTE accommodates two types of roles within the institution. Agents providing core institutional services are seen as performing *institutional roles* that are under the control of the institution. Agents acting as delegates of external entities enact different roles that are normatively regulated by the institution, in the sense that they may be subject to norms and may further establish new normative relationships. Furthermore, some of these roles are empowered, through appropriate *constitutive rules*, by the institution to ascertain institutional reality (i.e. they act as trusted third parties from the institution's point of view).
  - OCeAN/MANET allows the definition of roles as labels defined by a given Artificial Institution (AI) and used in the AI to assign norms and institutional

[8] Based on the same ideas, there is an extension of aBUILDER [9] that instead of code skeletons produces a simple human interface that complies with the ISLANDER specification and is displayed dynamically via a web browser at run-time.

powers at design time to roles. This is necessary because at design time the name of the actual agents that will take part to the interaction is unknown. At run time AIs are realized in dynamically created institutional spaces, the agents in a space can start to play the roles defined in the space and coming from the AI. An agent can play more than one role contemporarily. During an interaction an agent can start to play a role and subsequently stop to play it.
  - EI allows for specification of role subsumption and the specification of two forms of compatibility among roles: "dynamic" (each agent may perform different roles in different activities) and "static" no agent may perform both roles in an enactment of the institution. It also distinguishes between *internal* roles (played by agents whose behavior is controlled by the institution), and *external* roles (the institution has no access to their decision-making capabilities) and this separation is static.

- *Atomic interactions*

  - ANTE, concerning its institutional component, assumes an open setting in which there are two kinds of interactions going on in the system. On one hand, agents are free to interact with any other agents, without the institution even noticing that such interactions have taken place. On the other hand, illocutionary actions performed by agents towards the normative environment are seen as attempts to obtain *institutional facts* that are used by the latter to maintain the normative state of the system.
  - OCeAN/MANET defines *institutional actions* that in order to be successfully performed needs to satisfy certain conditions. One of these conditions is that the actor of the action needs to have the power to perform the institutional action, otherwise the action is void. The model defines also *instrumental actions*, for example the exchange of messages that should be used to perform institutional actions. Finally in the model it is possible to represent actions performed in the real world and that are relevant for the artificial interaction, for example the payment of an amount of money or the delivery of a product.
  - EI: There are essentially only two types of institutional actions: *speech* acts (represented as illocutions) and the *movement* actions which are accomplished in two steps exiting from a scene to a transition and entering from a transition to a scene (in some contexts an agent may *stay-and-go*, i.e remain active in the scene while at the same time becoming active in one or more different scenes)[9]. Consequently, on one side, an agent can act only by uttering an illocution or notifying the institutional environment its intention to move in or out of a transition (possibly changing role); on the other side, the perception of any given agent is restricted to those illocutions that are uttered by another agent and have the given agent as part of the intended listeners of that illocution, and the indication of the institutional infrastructure that a movement has been achieved

[9] In fact, as indicated in Table 1.1 these movements are implemented with five operations, which include the two key actions of entering and leaving the electronic institution.

- *Institutional state*
  - ANTE: The institutional normative state is composed of two sorts of so-called *institutional reality elements*. *Agent-originated events* are obtained as a consequence of agent actions, comprising essentially *institutional facts* that are obtained from the illocutions agents produce. These institutional facts map relevant real-world transactions that are through this means institutionally recognized. *Environment events*, on the other hand, occur as an outcome of the process of norm triggering and monitoring. Norms prescribe directed obligations with time windows, which when monitored may trigger different enactment states, namely temporal or actual violations, and fulfilments. All these elements are contextualized to the normative relationships that are established within the environment.
  - OCeAN/MANET: In the last version of the model the state of the interaction is represented using OWL 2 DL ontologies, one of the international standard language of the Semantic Web. Therefore the state of the interaction is represented using classes of concepts, individuals that belong to classes, object and data properties that connect two individuals or an individual to a literal (scalar values) respectively. The terminological box of the ontologies is also enriched with axioms, used to describe the knowledge on a given domain of application, and with SWRL rules, both are used by software reasoners to deduce new knowledge on the state of the interaction. Taking inspiration from the environment literature the state of objects, agents, events, and actions in a space are perceivable by the agents in that space.
  - EI: Only atomic interactions that comply with the institutional regimented conventions may be institutional actions and therefor change institutional facts. There is a data structure called the *institutional state* that contains all the institutional facts; that is, all the constants in the domain language plus the updated values of all those variables whose values may change through institutional actions. For each scene there exists a projection of that structure called the *state of the scene*. Additionally, there are some parameters whose default values are set by the institution and may be updated during an enactment. These are *institutional variables* (like the number of active scenes, the labels of active scenes and transitions), *scene variables* (like the number of participants, the list of items that remain to be auctioned, performance indicators such as the number of collisions or the rate of successful agreements) and *agent variables* (the list of external agents that have violated any discretional convention, the credit account of a trader). These parameters are not accessible to external agents although by design they may be accessible to some internal agents who may use the values of these variables in their individual decision-making.

- *Structure of the activities or compound interactions (contexts)*
  - ANTE: Interactions that need to be observed are executed through empowered agents, which will then inform the institutional environment of the actual

real-world activities that are taking place. Such activities are segmented into different normative contexts, that is, they pertain to specific normative relationships that are established at run-time. Within each such context different empowered agents may need to act as intermediaries, since different kinds of actions may need to be accomplished in order to successfully enact the contract subsumed in the context.

- OCeAN/MANET: The activities are realized into *institutional spaces* or *physical spaces* of interaction. Institutional spaces are used to realize AI at run-time, they may be entered and left by the agents starting from the root space. Physical spaces contains physical entities external with respect to the system, such as external resources, databases, external files, or web services, offering an abstraction that hide the low level details from the agents. Institutional spaces are in charge of representing and managing the social interaction of agents by realizing the concepts described in AIs and the services for norms monitoring and enforcement. Spaces are in charge of registering that an event has happened and represents the boundaries for the perception and of the effects of the events and actions.
- EI: Activities are decomposable into *scenes* that are connected by *transitions* into a network of scenes called a *performative structure*.
  - Scenes are state transition graphs where edges are labeled by *illocutionary formulas* and nodes correspond to a scene-state. A new scene-state may only be attained with the utterance of an admissible illocution. An utterance is valid if and only if it complies with the regimented conventions that apply under the current state of the scene. At some scene-states agents may enter or leave or *stay-and-go* the scene. Every performative structure contains one "start" and one "finish" scene that have the merely instrumental purpose to delimit the structure for syntactic (in specification) and implementation purposes (for enactment of the electronic institution).
  - A transition is a device that is used for two main purposes, to control role flow and to control causal and temporal interdependence among scenes. In particular, (a) when an agent exits a scene, it exits with the role it was playing in that scene but inside the transition the agent may change that role to enter a new scene (provided some institutional conventions are satisfied) (b) Moreover, when an agent enters a transition and depending on the type of transition it enters, that agent may join one, several or all the scenes that are connected to that transition. (c) Several agents, possibly performing different roles and coming possibly from different scenes, may enter the same transition and each has to decide on its own where to go from there and wether it changes role or not. (d) The transition coordinates flow by determining whether agents may proceed to their intended goal scene as soon as each agent arrives or wait until some condition holds in the state of the scene.

- *Hierarchical organization of the structure of activities*

- ANTE: Normative relationships established at run-time are organized as a hierarchy of contexts. Each context encompasses a group of agents in a specific regulated organization, within which further sub-contexts may be created, allowing for norm inheritance to take place. An overall institutional normative layer is assumed to exist, of which every subsequently created context is a sub-context. Furthermore, each context may add its own norms, which may be used to inhibit norm inheritance or to enlarge the normative framework that will govern the context.
- OCeAN/MANET: Spaces may contain other spaces generated dynamically at run-time, which become sub-spaces of the space where they are created. This hierarchy of spaces and the fact that one agent may be simultaneously in two spaces create interesting problems due to the interdependencies of spaces, this because the events of a space may be of interest to the father-space where this is contained or for a sibling space.
- EI: All agent interactions within an electronic institution are organized, as we mentioned above, by what we call a performative structure which is a network of scenes and transitions between those scenes. Two aspects are worth stating: First, a performative structure may be be embedded into another as if it were a scene, thus forming nested performative structures of arbitrary depth. Second, a performative structure becomes instantiated at run-time, thus although it is defined *a priori*, so to speak, the actual scenes do not come into existence until appropriate conditions take place (if ever) and they disappear likewise. In particular, it is possible to specify conditions that empower an internal agent to spawn a particular scene or performative (sub)structure.

- *Procedural and functional conventions*
  - ANTE: The effects of institutional facts are expressed through norms and rules. When triggered, norms prescribe directed obligations that are due to specific agents within a normative context. Such obligations have attached time-windows that are conventionally understood as ideal time periods for obtaining the obliged state of affairs. Outside this window temporal violations are monitored which may lead to different outcomes depending on the will of the obligation's counterpart. This semantics is captured by a set of monitoring rules that maintain the normative state of the system. The normative consequences of each obligation state is determined by the set of norms that shape the obligation's normative context, which may be established at run-time.
  - OCeAN/MANET: Both are expressed through pre and post-conditions of the actions defined by the institution. An important pre-condition for the performance of institutional actions (actions whose effects change institutional attributes that exist only thanks to the collective acceptance of the interacting agents) is the fact that the actor of the action should have the institutional power to perform the specific action.
  - EI: Both are expressed as pre and post-conditions of the illocutionary formulas of the scene transition graphs and through the labeling of transitions between scenes (this labeling expresses conditions for accessing a scene or a group

of scenes or a nested performative structure, synchronization, the change of roles, the creation of new scenes or activation of an existing scene). In the current EIDE implementation, there is also the possibility of explicitly expressing norms as production rules that are triggered whenever an illocution is uttered, thus allowing the specification and use of regimented and not-regimented conventions. Notice that although EI use illocutionary formulas to label actions, there are no social semantics of illocutionary particles involved. Thus scene protocols are not commitment-based protocols as is the case with [26] or more generally, [14, 12].

- *Constitutive conventions*
  - ANTE: Obtaining institutional facts from brute facts (which are basically agent illocutions) is achieved through appropriate *constitutive rules*, which mainly describe empowerments of different trusted third parties. These constitutive rules, which can be easily extended and/or adapted, determine the ontology for brute and institutional facts that can be used in the institution. Furthermore, it is possible to define further constitutive rules within each context, in this case enriching the domain ontology by obtaining more refined institutional facts. As a basic implementation, three types of transactions are reportable to the normative environment, related with the flow of products, money and information.
  - OCeAN/MANET: In this model the content language used for communicative acts and norms is defined using domain ontologies written in OWL 2 DL or in RDF+RDF Schema. Those ontologies may be defined by the designer of the interaction system or may already exist as proposed standards on the Web, like the well known ontology FOAF[10] that may be used for describing agents. In many cases the link between the name of a resource (its URI) and the corresponding resource in the real world can be done using existing knowledge repositories[11].
  - EI: The EI framework does not include axioms or definition statements that establish basic institutional facts. Nevertheless, there is a *domain language* that is used for expressing illocutionary formulas and whose terms correspond with physical facts and actions (e.g a sculpture to be auctioned, pay 32 euros for the item that has just been adjudicated). The correspondence between language and real entities is established *ad-hoc* for the domain language. In practice, however, an electronic institution needs to have true constitutive conventions in order to establish the legal (actual) entitlements of intervening parties and the correspondence between institutional and brute facts and actions. Examples of constitutive conventions are the contracts that allow an old books dealer to offer a used book through Amazon.com and follow the process through from offer to book delivery.

[10] `http://www.foaf-project.org/`
[11] `http://linkeddata.org/`

- *Social Commitments*
  - ANTE: Social commitments, in a broad sense, are established as an outcome of a previous negotiation phase, the success of which obtains a new normative context within the institutional environment. Once a normative context is obtained, applicable norms dictate when (according to the normative state) and which commitment instantiations (directed obligations) are entailed.
  - OCeAN/MANET: A commitment-based Agent Communication Language (ACL) is used [26, 29]. In particular communicative acts exchanged among agents have a meaning that is a combination of the meaning of the content of the messages and a meaning of the illocutionary force of the communicative acts (for example promise, query, assert).
  - EI: although, in EI, illocutionary formulas label actions, there is no social semantics of the illocutionary particles involved. Thus scene protocols are not commitment-based protocols properly speaking. However, commitments are hard-wired in scene specifications , and their evolution is captured in the evolving state of the institution. It should be noted, though, that in EI some commitments are expressed crudely but explicitly when a given admissible action (say winning a bidding round) has a postcondition that entails preconditions for future actions in other scenes.
- *Governance*
  - ANTE: The approach adopted in ANTE is to bear with the autonomy of agents, by allowing them to behave as they wish. From the institution's perspective, we assume it is in the best interest of agents to publicize their abidance with any standing obligations, by using the necessary means to obtain the corresponding institutional facts. Normative consequences of (non)fulfilment are assured by triggering applicable norms. Permissions and prohibitions are not handled explicitly in the system, i.e., not permitted actions simply have no effect within the normative environment. Entitlements are handled by defining norms triggered upon the occurrence of specific institutional facts. Any obligation outcomes – (temporal) violations and fulfilments – may also have further effects within the ANTE framework by reporting such events to a computational trust engine, which provides a mechanism of indirect social sanctioning.
  - OCeAN/MANET: The openness of the interaction systems realized using this model requires a governance in order to create an expectations on the actions of the participants agents. Contemporarily the model has to take into account the autonomy and heterogeneity of the interacting agents and avoid to constrain their behavior in rigid protocols. The main concepts introduced in the model related to governance are: *institutional power* (if an agent has not the power to perform an action its effects are void), *permission* (if an agent has not the permission to perform an action its effect take place but the agent incur in a violation), *obligations* (the agent has to perform an action with-in a given

deadline) and *prohibitions* (the agent cannot perform an action, if it does it will incur in a violation).

– EI: There are three different approaches for the implemention of governance in the EI model.
    1. In the standard model, all regimented conventions may be encoded in the performative structure as part of the specification of scenes and transitions and are therefore enforced in a strict and automatic fashion by the runtime implementation. Non-regimented conventions are encodable in the decision-making capabilities of internal agents and it is a matter of design whether some regimented ones may also be embedded in internal agents code. One may thus establish different types of (internal) norm-enforcement agents. Notice that although an internal agent may fail or decide not to enforce a violation, every violation is observed (registered) by the institution nonetheless.
    2. In the current implementation of EIDE one may choose to specify a collection of normative statements that are not part of the performative structure. This collection is coupled with an inference engine that takes hold of every utterance before it may be validated by the performative structure (see [33]). The process is as follows (i) An illocution is first tested against the normative statements and if it is consistent, it is labeled as "admissible" or rejected otherwise. (ii) The admissible illocution is then added to the current collection and the engine is activated; (iii) If the illocution triggers a violation, the concomitant corrective actions are taken, otherwise control is given to the performative structure that deals with the illocution as in approach 1. This approach allows to deal with discretional enforcement with more flexibility than approach (1) because in addition to all the mechanisms available in that approach, this one allows for a declarative specification of norms, an explicit distinction between regimented and non-regimented norms, and a variety of contrary-to-duty devices encodable as corrective actions.
    3. There is a proposed extension of the EI model that deals explicitly with norms and normative conflicts through the use a a "normative structure" that deals exclusively with norms and propagation of normative consequences between scenes [32, 55].

- *Ubiquity and concurrent activities*

    – ANTE: Agents may freely establish new normative relationships, and hence many of them may be active at the same time. The institutional environment pro-actively monitors every active context. There is a strong distinction between the agent identity and the normative relationships in which it is engaged. There is no notion of "physical" displacement of the agents within the institution. Within the ANTE framework, several other activities may take place at the same time, such as negotiations and computational trust building, which is achieved by gathering relevant enactment data from the normative environment monitoring process.

  - OCeAN/MANET: An interaction system realized using one or more AIs consists of a root space that contains physical and institutional spaces. An agent situated in a given space can enter all its sub-spaces, therefore an agent can be in more than one space and it has a persistent identity.
  - EI: An electronic institution usually consists of multiple scenes that are active simultaneously. In many cases the number of active scenes changes during execution since new scenes are created, activated or closed as the enactment proceeds. A given agent may be simultaneously active in more than one scene but it has a persistent identity in the sense that the effects of its institutional actions are coherent (for example, in an electronic market where an agent may be closing deals in different negotiations, this agent has *one* variable that captures its credit so the value of that variable changes every time it commits to pay, in whatever scene they commit). The current EI framework does not include a "meta-environment" where multiple institutions co-exist, however the peer-to-peer architecture proposed in [22] would be suitable for the implementation of lightly-coupled (and uncoupled) institutions in a shared environment.

- *Performance Assessment*
  - ANTE: Agent performance is assessed and exploited from two different perspectives. The first one is based on computational trust: the enactment of contracts produces evidences that are fed into a computational trust engine, which then produces trustworthiness assessments of agents that can be used when entering into further negotiations. In the current prototype implementation, trust information may be used for pre-selection of negotiation peers or for proposal evaluation. Another assessment of performance is measured by the normative environment, which for the whole agent population is able to determine the average enactment outcome for instances of stereotyped normative relationships (types of contracts).
  - OCeAN/MANET: There are not yet available services for assessing system's or agent performance.
  - EI: This model does not capture system goals explicitly, however scene and institutional variables may be used to specify some assessment of the performance of the institution with respect to whatever goals are defined. Internal agents may be designed to use such information in order to improve performance.

- *Formal properties*
  - ANTE: No formal methods for analyzing normative relationships are employed – it is up to the system designer to ensure correctness. The normative environment does record on-line every possible event that is captured while monitoring norms, allowing for an off-line verification of correctness.
  - OCeAN/MANET: For the moment there is not the possibility to check formal properties of AI at design-time. At run-time one crucial service is the monitoring of the state of the interaction, the detection of violations, and the enforcement of norms. Moreover in every instant of time it is possible to de-

duce the list of the actions that an agent is obliged, prohibited, permitted and empowered to perform, from this list and from an ontological definition of the terminology used to describe the actions it is possible to single out possible contradiction in the prescribed behavior. At design time this check is harder because in this model all normative constrains are related to time.
- EI: There is off-line automatic syntactic checking of scene and transition behavior. For example, in every scene: all roles have entry and exit states and these are reachable; every role has at least one path that takes it from start to finish; every term used in an illocution needs to be part of the domain ontology. On-line monitoring of all the activities: every utterance and attempted move produce a trace that may be displayed and captured for further use. The extensions mentioned in [55] allow for some off-line and on-line formal and automated reasoning about an institution.

- *Institutional Dynamics*
  - ANTE: The normative environment is assumed to be open and dynamic, in the sense that it encompasses an evolving normative space whose norms apply if and when agents commit to a norm-governed relationship. While providing an institutional normative framework, this infrastructure enjoys the properties of adaptability and extensibility, by providing support for norm inheritance and defeasibility. Normative contexts can therefore be created that adapt or extend a predefined normative scenario according to agents' needs.
  - OCeAN/MANET: This model is based on the idea that a human designer specifies an AI and this AI may be used at run-time to dynamically create spaces of interaction. Similarly norms at design time are specified in terms of roles and have certain unspecified parameters, at run-time those norms will be instantiated more than one time having as debtor different agents and different values for their parameters. In general this model does not include meta-operations for changing the model of AIs.
  - EI: With the current model internal agents may be given the capability to create new scenes from repositories of available scenes and even graft nested performative structures into a running institution. In a similar fashion internal agents may create new internal agents when needed (say for a newly grafted performative structure) by invoking a service that spawns new agents that is outside of the electronic institution proper but is available to the internal agent. This mechanism is also used to embed the EI environment into a simulation environment [4]. The current model includes no primitive meta-operations that would allow agents to change the specification of an institution beyond what was just said, however here have been proposals for other forms of autonomic adaptation [8, 11].

- *Implementation architecture*
  - ANTE: The ANTE framework is realized as a Jade FIPA-compliant platform, where agents can make use of the available services (e.g. negotiation, contract monitoring, computational trust) through appropriate interaction proto-

cols, such as FIPA-request and FIPA-subscribe. Using subscription mechanisms agents are notified of the normative state of the system in which their normative relationships are concerned. The normative environment has been implemented using the Jess rule-based inference engine.
  - OCeAN/MANET: The model of AI has been fully formalized in Event Calculus and we are currently formalizing it using Semantic Web Technologies. An AI for realizing a Dutch Auction has been also specified in PROLOG and tested in a prototype realized above the GOLEM environment framework [54]. An implementation of a complete energy market-place based on Semantic Web Technologies and the GOLEM framework is under development.
  - EI: The model has been fully detailed [19] in the Z specification language [53] and deployed in the architecture sketched in Fig.1.2. This architecture creates a sort of "social layer" that is independent of the communication layer used to exchange messages between an agent and the electronic institution. The normative engine extension is also implemented in the same architecture. A peer-to-peer architecture has been proposed [22] and a prototype is now under construction.

- *Tools*

  - ANTE: The ANTE framework includes graphical user interfaces (GUI) that allow the user to inspect the outcomes of each provided service, including the evolution and outcome of a specific negotiation, the inspection of trustworthiness scores of the agents in the system, as well as the overall behavior of the agent population in terms of norm fulfillment. The framework includes also a complex API allowing for the specification of user agents, for which a set of predefined GUI are also available that enable the user to inspect the agent activity, namely its participation in negotiations and contracts. The API allows a programmer to easily encode agent behavior models in response to several framework activities, such as negotiation and contract enactment, which makes it straightforward to run different kinds of experiments (although Jade has not been designed for simulation purposes).
  - OCeAN/MANET: Thanks to the fact that we base our model of current standard semantic web technologies, it is possible to use the ontology editor Protègè for editing the ontologies used in the specification of the model of AI and spaces and to use one of the available reasoners (Pellet, HermiT, and so on) for checking their consistency. Our future goal is that once the model of a set of AIs is defined and a set of agents able to interact with a system getting its formal specification are developed, the interaction system can start to run and enable agents to interact using the available actions and constrained by the specified norms.
  - EI: As mentioned in the previous section, EIDE includes a graphical specification language (ISLANDER), an agent middleware for electronic institutions (AMELI) that generates a runtime version of any ISLANDER compatible specification. EIDE also includes an automated syntactic checker, a simple simulator for on-line testing and debugging, a monitoring tool, and a software

that generates agent skeletons that encode the navigational behavior that is compatible with an ISLANDER specification.

- *Agents*
  - ANTE: The framework is neutral in which user agents' internal architectures and implementation languages are concerned. It is assumed, however, that agents are able to communicate using FIPA ACL and the FIPA-based interaction protocols and ontologies interfacing each of the framework's services. It is also straightforward to admit human agents to participate, provided that appropriate user interfaces are developed.
  - OCeAN/MANET: The model of the interaction system realized using the AI is independent on the agents' internal structure. Nevertheless it is assumed that the participating agents are able to interact using the available communicative acts whose content should be expressed using shared ontologies.
  - EI: The model is agent-architecture independent. Agents are required only to comply with interface conventions that support institutional communication. Hence human agents may participate in an electronic institution enactment provided they have the appropriate interfaces. The tool HIHEREI [9] automatically generates such a human interface for any ISLANDER compatible specification of an electronic institution. In the current implementation, AMELI is communication-layer independent.

1.7 Challenges

There are many open challenges in the field of specification and use of institutions for the efficient realization of real open interaction systems in different fields of applications, going from e-commerce, e-government, supply-chain, management of virtual enterprise, and collaborative/socila resource sharing systems.

One interesting challenge goes into the direction of using those formal and declarative models of *hybrid* open interaction systems involving both software and human agents. In this perspective one possibly important use of these technologies is for designing flexible open collaborative/social systems able to exploit the flexibility, the intelligence, and the autonomy of the interacting parties. This in order to improve existing business process automation systems where the flow of execution is completely fixed at design time or groupware where the work of defining the context and the rules of the interaction is left to the human interacting parties and no automatic monitoring of the completion of tasks is provided.

When considering the automation of e-contracting systems through autonomous agents, another important challenge is to endow agents with reasoning abilities that enable them to establish more adequate normative relationships. Infrastructural components need to be developed that ease this task, e.g. through normative frameworks that agents can exploit by relying on default norms that may nevertheless need to be overridden. A complementary challenge is how to ensure reliable behaviours when agents act as human or enterprise delegates, that is, how to simultaneously cope with expressivity and configurability through human interfaces and agents' autonomy in institutional normative environments. Another interesting challenge is to look at the Environment as a structured medium not only to facilitate agents' interaction but also as an active representative of the "society" in which agent relationships take place.

References

1. H. Aldewereld and V. Dignum. Operetta: Organization-oriented development environment. In M. Dastani, A. E. Fallah-Seghrouchni, J. Hübner, and J. Leite, editors, *LADS*, volume 6822 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 2010.
2. M. Aoki. *Toward a Comparative Institutional Analysis*. MIT Press, 2001.
3. J. L. Arcos, M. Esteva, P. Noriega, J. A. Rodríguez-Aguilar, and C. Sierra. Engineering open environments with electronic institutions. *Eng. Appl. of AI*, 18(2):191–204, 2005.
4. J. L. Arcos, P. Noriega, J. A. Rodriguez-Aguilar, and C. Sierra. E4mas through electronic institutions. In D. Weyns, H. Parunak, and F. Michel, editors, *Environments for Multi-Agent Systems III.*, number 4389 in Lecture Notes in Computer Science, pages 184–202. Springer, Berlin / Heidelberg, 08/05/2006 2007.
5. A. Artikis, M. Sergot, and J. Pitt. Specifying norm-governed computational societies. *ACM Trans. Comput. Logic*, 10(1):1:1–1:42, Jan. 2009.
6. G. Boella, G. Pigozzi, and L. V. der Torre. Five guidelines for normative multiagent systems. In G. Governatore, editor, *Legal Knowledge and Information Systems. JURIX 2009*, pages 21–30, Amsterdam, October 22-24 2009. IOS Press.
7. G. Boella, L. van der Torre, and H. Verhagen. Introduction to the special issue on normative multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 17(1):1–10, 2008.
8. E. Bou, M. Lopez-Sanchez, and J. A. Rodriguez-Aguilar. Adaptation of autonomic electronic institutions through norms and institutional agents. In G. O'Hare, A. Ricci, M. O'Grady, and O. Dikenelli, editors, *Engineering Societies in the Agents World VII*, volume 4457 of *Lecture Notes in Computer Science*, pages 300–319. Springer Berlin / Heidelberg, 2007.
9. I. Brito, I. Pinyol, D. Villatoro, and J. Sabater-Mir. HIHEREI: Human Interaction within Hybrid Environments. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '09)*, pages 1417–1418, Budapest, Hungary, 2009.
10. S. Bromuri and K. Stathis. Distributed agent environments in the Ambient Event Calculus. In *Proceedings of the Third ACM International Conference on Distributed Event-Based Systems*, DEBS '09, pages 12:1–12:12, New York, NY, USA, 2009. ACM.
11. J. Campos, M. Lopez-Sanchez, and M. Esteva. A case-based reasoning approach for norm adaptation. In A. S. E. Corchado, M. Graña-Romay, editor, *5th International Conference on Hybrid Artificial Intelligence Systems (HAIS'10)*, volume 6077, pages 168–176, San Sebastian, Spain, 23/6/2010 2010. Springer, Springer.
12. A. Chopra and M. Singh. Nonmonotonic commitment machines. In F. Dignum, editor, *Advances in Agent Communication*, volume 2922 of *Lecture Notes in Computer Science*, pages 1959–1959. Springer Berlin / Heidelberg, 2004.
13. O. Cliffe, M. De Vos, and J. Padget. Specifying and reasoning about multiple institutions. In P. Noriega, J. Vzquez-Salceda, G. Boella, O. Boissier, V. Dignum, N. Fornara, and E. Matson, editors, *Coordination, Organization, Institutions and Norms in Agent Systems II - AAMAS 2006 and ECAI 2006 International Workshops, COIN 2006 Hakodate, Japan, May 9, 2006 Riva del Garda, Italy, August 28, 2006*, volume 4386 of *Lecture Notes in Computer Science*, pages 67–85. Springer Berlin / Heidelberg, 2007.

14. M. Colombetti. A commitment-based approach to agent speech acts and conversations. In *Proceedings of the Fourth International Conference on Autonomous Agents, Workshop on Agent Languages and Conversation Policies.*, pages 21–29, 2000.
15. D. Corapi, M. De Vos, J. Padget, A. Russo, and K. Satoh. Normative design using inductive learning. *Theory and Practice of Logic Programming*, 11:783–799, 2011.
16. R. Craswell. Contract Law: General Theories. In B. Bouckaert and G. De Geest, editors, *Encyclopedia of Law and Economics*, volume Volume III: The Regulation of Contracts, pages 1–24. Edward Elgar, Cheltenham, UK, 2000.
17. V. Dignum. *A model for organizational interaction: based on agents, founded in logic*. PhD thesis, University Utrecht, 2004.
18. V. Dignum, J. Vázquez-Salceda, and F. Dignum. OMNI: Introducing social structure, norms and ontologies into agent organizations. In *PROMAS*, volume 3346 of *Lecture Notes in Computer Science*, pages 181–198. Springer, 2004.
19. M. d'Inverno, M. Luck, P. Noriega, J. A. Rodriguez-Aguilar, and C. Sierra. Communicating open systems. *Artificial Intelligence*, 186(0):38 – 94, 2012.
20. M. Esteva. *Electronic Institutions: from specification to development. PhD thesis Universitat Politècnica de Catalunya (UPC), 2003*. Number 19 in IIIA Monograph Series. IIIA, 2003.
21. 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 Multiagent systems (AAMAS '02)*, pages 1045–1052. ACM Press, July 2002.
22. M. Esteva, J. A. Rodriguez-Aguilar, J. L. Arcos, and C. Sierra. Socially-aware lightweight coordination infrastructures. In *AAMAS'11 12th International Workshop on Agent-Oriented Software Engineering*, pages 117–128, 2011.
23. M. Esteva, J. A. Rodriguez-Aguilar, B. Rosell, and J. L. Arcos. AMELI: An agent-based middleware for electronic institutions. In C. Sierra and L. Sonenberg, editors, *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi-agent Systems (AAMAS '04)*, volume 1, pages 236–246, New York, USA, July 19-23 2004. IFAAMAS, ACM Press.
24. M. Esteva, J. A. Rodrguez-Aguilar, J. L. Arcos, C. Sierra, P. Noriega, and B. Rosell. Electronic Institutions Development Environment. In *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '08)*, pages 1657–1658, Estoril, Portugal, 12/05/2008 2008. International Foundation for Autonomous Agents and Multiagent Systems, ACM Press.
25. N. Fornara. Specifying and monitoring obligations in open multiagent systems using semantic web technology. In A. Elçi, M. T. Kone, and M. A. Orgun, editors, *Semantic Agent Systems: Foundations and Applications*, volume 344 of *Studies in Computational Intelligence*, chapter 2, pages 25–46. Springer-Verlag, 2011.
26. N. Fornara and M. Colombetti. Operational specification of a commitment-based agent communication language. In *Proceedings of The First International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2002, July 15-19, 2002, Bologna, Italy*, pages 536–542. ACM, 2002.
27. N. Fornara and M. Colombetti. Specifying Artificial Institutions in the Event Calculus. In V. Dignum, editor, *Handbook of Research on Multi-Agent Systems: Semantics and Dynamics of Organizational Models*, Information Science Reference, chapter XIV, pages 335–366. IGI Global, 2009.
28. N. Fornara and M. Colombetti. Representation and monitoring of commitments and norms using OWL. *AI Communications - European Workshop on Multi-Agent Systems (EUMAS) 2009*, 23(4):341–356, 2010.
29. N. Fornara, D. Okouya, and M. Colombetti. Using OWL 2 DL for expressing ACL Content and Semantics. In M. Cossentino, M. Kaisers, K. Tuyls, and G. Weiss, editors, *EUMAS 2011 Selected and Revised papers*, LNCS, page to appear, Berlin, Heidelberg, 2012. Springer-Verlag.
30. N. Fornara, F. Viganò, and M. Colombetti. Agent communication and artificial institutions. *Autonomous Agents and Multi-Agent Systems*, 14(2):121–142, April 2007.

31. N. Fornara, F. Viganò, M. Verdicchio, and M. Colombetti. Artificial institutions: A model of institutional reality for open multiagent systems. *Artificial Intelligence and Law*, 16(1):89–105, March 2008.
32. D. Gaertner, A. Garcia-Camino, P. Noriega, J. A. Rodriguez-Aguilar, and W. W. Vasconcelos. Distributed norm management in regulated multi-agent systems. In *Proceedings of the 6th international Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '07)*, pages 624–631, Honolulu, Hawaii, 14/05/07 2007. ACM Press.
33. A. Garca-Camino, P. Noriega, and J. A. Rodrguez-Aguilar. Implementing norms in electronic institutions. In S. T. Michal Pechoucek, Donald Steiner, editor, *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '05)*, pages 667–673, Utrecht, NL, 2005. ACM Press.
34. A. J. I. Jones and M. J. Sergot. A formal characterisation of institutionalised power. *Logic Journal of the IGPL*, 4(3):427–443, 1996.
35. L. Kaplow. General Characteristics of Rules. In B. Bouckaert and G. De Geest, editors, *Encyclopedia of Law and Economics*, volume Volume V: The Economics of Crime and Litigation, pages 502–528. Edward Elgar, Cheltenham, UK, 2000.
36. R. A. Kowalski and M. J. Sergot. A logic-based calculus of events. *New Generation Computing*, 4(1):67–95, 1986.
37. H. Lopes Cardoso. *Electronic Institutions with Normative Environments for Agent-based E-contracting*. PhD thesis, Universidade do Porto, 2010.
38. H. Lopes Cardoso and E. Oliveira. Norm defeasibility in an institutional normative framework. In M. Ghallab, C. Spyropoulos, N. Fakotakis, and N. Avouris, editors, *Proceedings of The 18th European Conference on Artificial Intelligence (ECAI 2008)*, pages 468–472, Patras, Greece, 2008. IOS Press.
39. H. Lopes Cardoso and E. Oliveira. A Context-based Institutional Normative Environment. In J. Hubner, E. Matson, O. Boissier, and V. Dignum, editors, *Coordination, Organizations, Institutions, and Norms in Agent Systems IV*, LNAI 5428, pages 140–155. Springer, 2009.
40. H. Lopes Cardoso and E. Oliveira. Social control in a normative framework: An adaptive deterrence approach. *Web Intelli. and Agent Sys.*, 9:363–375, December 2011.
41. H. Lopes Cardoso, J. Urbano, A. Rocha, A. Castro, and E. Oliveira. ANTE: Agreement Negotiation in Normative and Trust-enabled Environments. In *Handbook of Agreement Technologies (Chapter on Applications), in this volume*. Springer, 2012.
42. E. T. Mueller. *Commonsense Reasoning*. Morgan Kaufmann, San Francisco, 2006.
43. P. Noriega. *Agent-Mediated Auctions: The Fishmarket Metaphor. PhD thesis Universitat Autònoma de Barcelona, 1997.* Number 8 in IIIA Monograph Series. IIIA, 1999.
44. D. C. North. *Institutions, institutional change, and economic performance*. Cambridge University Press, Cambridge, 1990.
45. Object Management Group. UML 2.0 OCL Specification. http://www.omg.org/, 2005.
46. E. Ostrom. An agenda for the study of institutions. *Public Choice*, 48(1):3–25, 1986.
47. E. Ostrom. Institutional analysis and development: Elements of the framework in historical perspective. In C. Crothers, editor, *Historical Developments and Theoretical Approaches in Sociology in Encyclopedia of Life Support Systems(EOLSS), Developed under the Auspices of the UNESCO*. Eolss Publishers, Oxford, UK, 2010.
48. J. Pitt, J. Schaumeier, and A. Artikis. Coordination, conventions and the self-organisation of sustainable institutions. In *Proceedings of the 14th international conference on Agents in Principle, Agents in Practice*, PRIMA'11, pages 202–217, Berlin, Heidelberg, 2011. Springer-Verlag.
49. W. W. Powell and P. J. Dimaggio. *The New Intitutionalism in Organizational Analyisis*. University of Chicago Press, 1991.
50. J. A. Rodriguez-Aguilar. *On the Design and Construction of Agent-mediated Electronic Institutions, PhD thesis, Universitat Autònoma de Barcelona, 2001*. Number 14 in IIIA Monograph Series. IIIA, 2003.
51. J. R. Searle. *The construction of social reality*. Free Press, New York, 1995.
52. H. A. Simon. *The Sciences of the Artificial*. MIT Press, third edition, 1996.

53. J. M. Spivey. *Understanding Z: A Specification Language and its Formal Semantics*. Cambridge University Press, 1988.
54. C. Tampitsikas, S. Bromuri, N. Fornara, and M. I. Schumacher. Interdependent Artificial Institutions In Agent Environments. *Applied Artificial Intelligence*, 26(4):398–427, 2012.
55. W. W. Vasconcelos, A. García-Camino, D. Gaertner, J. A. Rodríguez-Aguilar, and P. Noriega. Distributed norm management for multi-agent systems. *Expert Systems with Applications*, 39:5990–5999, 04/2012 2012.
56. J. Vázquez-Salceda. *The role of Norms and Electronic Institutions in Multi-Agent Systems applied to complex domains. The HARMONIA framework*. PhD thesis, Universidad Politecnica de Catalunya, 2003.
57. D. Weyns, A. Omicini, and J. Odell. Environment as a first class abstraction in multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 14(1):5–30, 2007.