Peer-to-peer, agents, and knowledge management: a design framework

D. Bertolini, P. Busetta, A. Molani, M. Nori, and A. Perini

ITC-Irst, Via Sommarive, 18, I-38050 Trento-Povo, Italy
{bertolin,busetta,molani,nori,perini}@irst.itc.it

Abstract. This paper presents preliminary work on a framework for developing knowledge management (KM) applications. The objective of this framework is to support a particular KM paradigm that emphasizes aspects such as *autonomy* and *distribution* of knowledge sources. The framework, once completed, will include three main elements: a methodology; a set of architectural patterns; and, a technological infrastructure. The first element is based on *Tropos*, an agent oriented software engineering methodology presented in earlier papers. Patterns are derived from the analysis of common peer-to-peer systems; a specific technological framework, JXTA, is considered. Finally, the infrastructure aims at integrating multi-agent and peer-to-peer technologies. This paper focuses on a specific aspect: how the framework supports the architectural design process. In the discussion, we refer to an example taken from the Health Care domain.

1 Introduction

Common knowledge management systems support the collection and categorization of knowledge with respect to a single, shared perspective, and its redistribution to its users by a variety of means. In many instances, this leads to the construction of one or a few repositories of documents, organized around a single ontology or other meta-structures. Users are given tools to search, browse, or receive documents as soon as they become available, varying from simple Web interfaces to sophisticated agents running on the users' desktops.

Building the shared perspective at the core of a KM systems is an expensive task, sometimes impossible to achieve when users have substantially different goals and background. To tackle this issue, our research group[1] is investigating a novel approach called *distributed knowledge management* (DKM) [6]. In short, the idea is to support the integration of autonomously managed KM systems, without forcing the creation of neither centralized repositories nor shared ontologies and other categorization structures. This integration involves the deployment of a set of complex protocols and algorithms for information retrieval,

[1] This work is conducted in the framework of the EDAMOK project, funded by the Provincia Autonoma di Trento (P.A.T.) http://sra.itc.it/projects/edamok/index.html.

natural language processing, machine learning, deductive and inductive reasoning, and so on, sometimes requiring the direct participation of human users.

To support the DKM approach, we are developing a framework that gives a methodology and some technological support to analysts, designers, and developers. The framework integrates and extends concepts taken from an agent oriented methodology, Tropos, as well as from the peer-to-peer (P2P) and multi-agent (MAS) communities. The methodology focuses on the identification of the social actors – single users, departments, companies, or other forms of communities – that have an advantage in maintaining their own autonomous KM systems, and of the kind of knowledge exchanges they need to perform with other actors in order to achieve their goals. Some P2P concepts and technologies – such as *peer* and *peer group* – are well suited to capture and support certain aspects, such as the dynamism of the social organizations where DKM is applied. Typical MAS concepts – such as *agent* and *multi-agent systems* – are better suited to capture reasoning and collaboration among distributed components.

This paper focuses on a specific aspect of the framework: the architectural design processes. To this end, we analyze some well-known P2P systems, derive a general architectural pattern, and define a set of guidelines for a DKM system designer.

The paper is structured as follows. Section 2 introduces some basic Tropos concepts and notations, which are used to discuss some typical peer-to-peer architectures, to identify their common pattern, and to focus on JXTA, the P2P framework we are currently using. Section 3 presents the guidelines for a DKM system architectural design. Section 4 describes a scenario taken from the Health Care domain, and shows how to apply the guidelines. Finally, some related works and conclusions are presented in Section 5.

2 Logical architectures of peer-to-peer systems

In this section, we consider functional and non-functional requirements that are successfully satisfied by peer-to-peer systems [7]. The analysis is conducted using the *Tropos* methodology, briefly introduced below.

2.1 An overview of Tropos

The *Tropos* methodology [3] adopts an agent oriented approach to software development starting from the very early stage of requirement specifications, when the environment and the system-to-be are analyzed, down to system design and implementation. The methodology identifies a number of phases (*early requirements*, *late requirements*, *architectural design*, *detailed design*, and *implementation*) and has been applied to several case studies [4]. The core process of the methodology consists in performing conceptual modeling using a visual language that provides intentional and social concepts such as actor, goal, belief, plan and dependency. An *actor* models an entity that has strategic goals and intentionality, such as a physical *agent*, a *role* or a *position*. A *role* is an abstract

characterization of the behavior of an actor within some specialized context, while a *position* represents a set of roles, typically covered by one agent. The notion of actor in Tropos is a generalization of the classical AI notion of software agent. *Goals* represent the strategic interests of actors. A *dependency* between two actors indicates that an actor depends on another in order to achieve a goal, execute a plan, or deliver a resource. Tropos distinguishes between hard and soft goals, the latter having no clear-cut definition and/or criteria as to whether they are satisfied. Softgoals are useful for modeling software qualities [2], such as security, performance and maintainability. A Tropos model is represented as a set of diagrams. In particular, actor and dependency models are graphically represented as *actor diagrams* in which actors are depicted as circles, their goals as ovals. The network of dependency relationships among actors are depicted as two arrowed lines connected by a graphical symbol varying according to the dependum: a goal, a plan or a resource. Actor goals and plans can be analyzed from the point of view of the individual actor using three basic reasoning techniques: *means-end analysis*, *contribution analysis*, and *AND/OR decomposition*. From this analysis, new actor dependencies can be identified.

Tropos is currently being extended with concepts suitable to model some specific notions – such as distributed knowledge, autonomy and coordination – that are peculiar to DKM, in order to support an early requirement model of a DKM domain.

In this paper, we focus on a later phase in software development: given a requirement analysis of a DKM problem, we provide guidelines for defining one or more suitable system architectures. According to *Tropos*, this type of issues are typically faced during the *architectural design* phase.

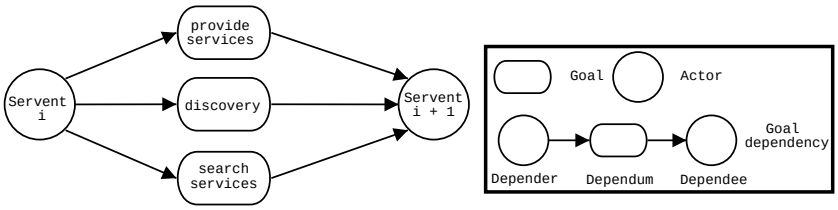

Fig. 1. Gnutella: architectural design, according to the Tropos methodology

2.2 The Peer-to-peer virtual community pattern

In order to understand what kind of requirements are successfully satisfied by a peer-to-peer approach, we analyze two well known systems, Gnutella and Napster, both supporting MP3 files sharing.

The logical architecture of *Gnutella*[2] is described in the *Tropos* actor diagram depicted in Figure 1. Both actors model a basic role, the Servent (Gnutella's ter-

[2] `http://www.gnutella.com`

minology for peer). Each servent has the goals of discovering other servents, looking for MP3 files, and providing its own MP3s, and depends on all other servents for achieving them. In other words, the goals of a servent are achieved only by means of a virtual community of peers, each one playing the role of servent. In the diagram, this is represented as goal dependencies between two generic servents. The dependencies are symmetric for every servent; for simplicity, we have shown them going only in one sense. The discovery goal enables the community to be dynamic, since peers can join and leave the community at any time without having to register anywhere their presence. There is no service centralized in any one peer; conversely, all the peers provide (and take advantage of) the same set of services.

In *Napster*[3], each client, represented by Peer actors in Figure 2, has similar goals of Servents in Gnutella (search and download files). However, in order to obtain a search result, the client must contact a central server (Napster Server). This owns the list of all the active clients and the list of all the shared MP3 files, and depends on them for building these lists. This is represented by a set of goal dependencies in Figure 2: the generic Peer depends on Napster Server for getting a search result (goal dependency GD3) and Napster Server depends on Peer to get lists of available MP3 (goal dependency GD2). Analogously to Gnutella, two generic peers depend on each other for downloading MP3 files (goal dependency GD1). So, the individual goal is obtained by means of a community of peers (Napster clients) that coordinates with an actor playing a distinct role in the community, i.e. the server.

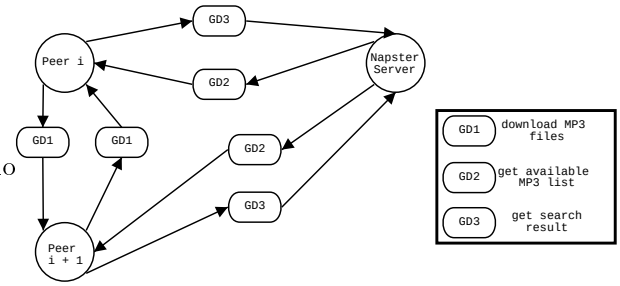

Fig. 2. Napster: architectural design

We abstract these two architectures, and others not discussed here, in the basic model depicted in Figure 3. The Individual actor models somebody who has at least one of two goals: accessing a resource (or, equivalently, using a service); and, letting others accessing her own resources (or using her own services). An individual has a set of constraints and preferences (shown as softgoals in the diagram) which drive her behavior when achieving these goals. In particular, an important requirement is being autonomous, i.e. being able to control what to access or to make available to the external world, and when.

The actor Virtual P2P community has the main goal of letting its members cooperate for sharing a specific type of resource or service. Three additional requirements (shown as softgoals) have been identified: being highly dynamic – available resources and services provided by members can be added or removed at any time –; being decentralized, i.e. the community is able to achieve its main goal independently of any specific member or component; and finally, being

[3] `http://opennap.sourceforge.net`

composed of peers on an equal basis, that is, every member is compelled to provide (at least potentially) resources or services, as well as having a right to access the others'.

The dependency be member of captures the fact that the individual's goals can be satisfied by joining the community. The act of joining implies the acceptance of the community's main goal and rules, highlighted above. Conversely, a community can achieve its goal only if it is active, and this can be accomplished only by having individuals as members. Finally, the dependency provide all services models the rule that in a peer-to-peer community all members are equal, i.e. provide the same set of services.

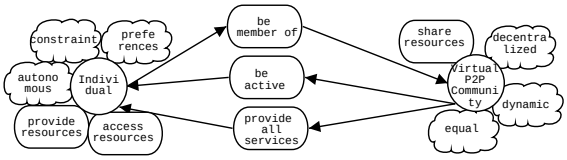

Fig. 3. The peer-to-peer virtual community pattern

Going back to the systems discussed previously, Gnutella is a "perfect" P2P system, since it satisfies all non-functional requirements of the community highlighted in Figure 3. Similarly, it may be argued that Napster is not a real P2P system, since the community depends on a centralized service.

2.3 JXTA P2P: logical architecture elements

We focus now on *JXTA*[4], a set of open, generalized peer-to-peer protocols that allow devices to communicate and collaborate through a connecting network.

From our perspective, the fundamental concepts of JXTA are three: *peer*, *peer group* and *service*.

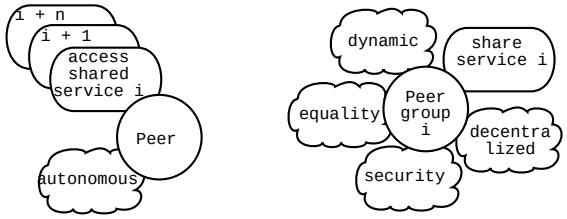

Fig. 4. JXTA: architectural design overview

A *peer* is "any device that run some or all the Project JXTA protocols". The complex layering of the JXTA protocols and the ability for a peer to simultaneously participate to more than one peer group (described below) imply that a peer is – at least conceptually – a multi-threaded program.

A *peer group* is a collection of peers that have agreed upon a common set of rules to publish, share and access "codats" (shorthand for code, data, applications), and communicate among themselves. Each peer group can establish its own membership policy, from open (anybody can join) to highly secure and protected (join only if you have sufficient credential). Thus, a peer group is used to support structuring (based on social organizations or other criteria) on top of the basic peer-to-peer network. As mentioned above, a peer may be part of many groups simultaneously.

[4] A P2P open source effort started in April 2001. `http://www.jxta.org/` and `http://spec.jxta.org/v1.0/docbook/JXTAProtocol`

A peer provides to each of its group a set of *services*, which are advertised to the other members of the group. In other words, groups in JXTA support a model for service publishing which is alternative to the traditional, centralized directory services model. Each group specifies a set of services that have to be provided by its members; they may include supporting basic JXTA protocols (e.g. discovery) as well as user-written applications, such as file sharing.

From a logical perspective, these concepts can be modeled in terms of roles in architectural design diagram depicted in the Figure 4, which extends the pattern depicted in Figure 3. A Peer has n goals access shared service i, one for each type of service it needs, and the requirement (specified as a softgoal) of being autonomous. A generic actor PeerGroup i has the goal of sharing a service of a specific type i and a set of requirements: supporting a decentralized distribution of services, allowing for dynamic membership to the group, supporting security management, and requiring all members to provide equal services.

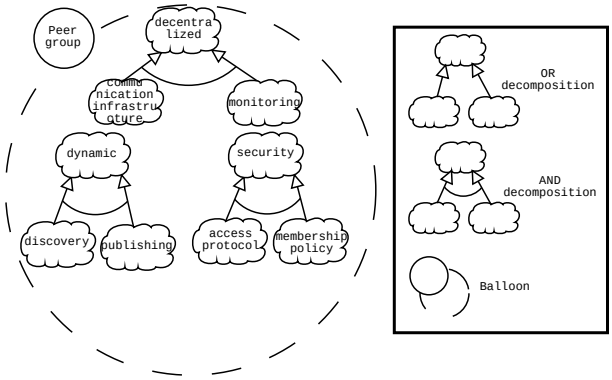

Fig. 5. JXTA: PeerGroup architectural design

JXTA tackles these requirements with a number of mechanisms, represented as goals decomposition in Figure 5. The publishing and discovery mechanisms, together with a message-based communication infrastructure and peer monitoring services, support decentralization and dynamism. security is supported by a membership policy (which authenticates any peer applying to a peer group) and an access protocol (for authorization control).

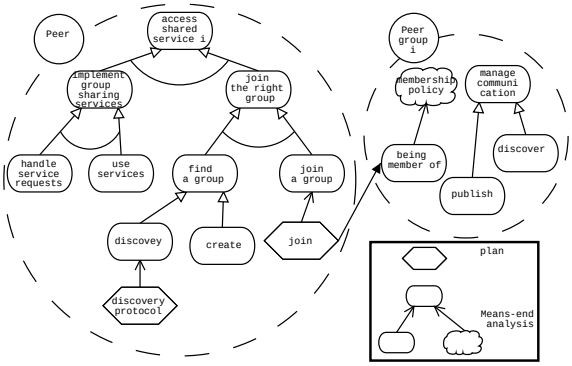

Fig. 6. JXTA: Peer architectural design

In summary, it can be said that the creation of a PeerGroup derives from the need of defining a set of peers that are able to communicate and interact with a common set of capabilities and protocols, and that are aggregated for security reason (i.e., no extraneous peer can interfere with them).

Figure 6 shows that a Peer's main goal (access shared service i) can be decomposed into two subgoals: join the right group and implement group sharing services. The first means that the peer must find a group that it is authorized to join and that provides the services it needs. This goal can be further decomposed into join a group and find a group. The latter can be satisfied by using the JXTA discovery service; if no adequate group is found, a new one can be created. Joining a group that has been discovered or created depends on its membership policy. Once a group has been joined, a Peer must implement all the services required by the group, implementing both the client side (use services) and the server side (handle service requests). It is important to stress that a Peer is autonomous in deciding how to provide a service (only protocols are predefined), and that a Peer can join different PeerGroups in order to achieve different goals.

3 Architectural design of communities in a DKM systems

In this section, we sketch the guidelines for designing the architecture of a DKM system, focusing on the support of virtual communities. Input to the process described here is, in Tropos terms, the output of a late requirements phase (Section 2.1); that is, the output of a domain analysis which included the system-to-be. The late requirements phase led to the identification of the stakeholders, including knowledge and services sources and of their users; how this analysis is performed is outside the scope of this paper. The guidelines consists of the following steps:

1. apply the P2P virtual community pattern;
2. design the JXTA peer group implementing the virtual community;
3. design the agents that implement a peer service.

Each steps is briefly described below, while next section presents their practical application.

Applying the P2P Virtual Community Pattern At this step, the designer has to answer to a basic question: can a Tropos late requirement model of a DKM problem be refined as a virtual P2P community model?

By definition, actors in a DKM scenario, as well as individuals in the community pattern, want to keep their autonomy. A designer, then, should look for the following conditions to be realized (they are illustrated in Figure 7 where two P2P virtual communities can be identified):

- there is a subset C of the actors that have pair-wise dependencies (e.g., actor A1 depends on A7 for goal α and A7 depends on A1 for β) in such a way that any actor is dependent on many, if not most, other actors in C. In other words, each actor is at the center of a sort of "star" network of reciprocal dependencies with many others;
- these dependencies can be generalized to one or a few types only. In particular, they are of type "access" and "provide" a service or a resource.

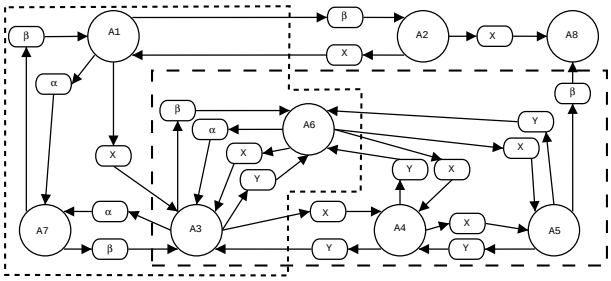

Fig. 7. Identifying P2P virtual communities. The first community (dashed line) is composed by actors involved in pair dependencies for goals X and Y, the second (dotted lines) by actors involved in pair dependencies for goals α and β.

If these conditions are satisfied, then the modeler can apply the P2P virtual community pattern to generate a model for a community supporting C. The main goals of this virtual community are suitable abstractions of the generalized dependencies linking the actors in C. The goals of an individual are also generated from each generalized dependency, and are two: achieving the objective of the dependency, and, conversely, satisfying it. A validation step is necessary, and consists of two main activities:

- verifying that the goals of the actors in C can be effectively satisfied by adopting the goals of the individual in the community model; that is, verifying that the P2P virtual community helps in achieving the actors' goals. This may be performed as a means-end analysis, which decomposes those hard goals into plans that, eventually, are satisfied by adopting the individual's goals;
- verifying that the general soft goals of a P2P virtual community – dynamism, equality, and decentralization – are at least partially achieved by the community being designed.

Designing a JXTA peer group Once that one or more P2P virtual communities have been identified, their supporting infrastructure can be designed. Assuming that JXTA is adopted as basic communication technology, our major – and natural – choice is then to associate a JXTA peer to each individual, and a JXTA peer group to each P2P virtual community. Thus, a social actor (no matter if a single person or a group) will have, as its supporting system, a peer participating to as many peer groups as communities that have been identified during the requirement analysis of its DKM problems.

Agents as JXTA services The final step is designing how individuals implement their goals with respect to each of the communities they are part of. From a JXTA perspective, it is necessary to specify which application services have to be provided by a peer, and how they are published on the network; however, JXTA leaves the designer total freedom concerning their communication protocols and

internal implementation. JXTA provides its own communication mechanisms (unreliable message queues called "pipes" and an XML based encoding format), but the application may decide to use something different, since JXTA supports publishing and discovery of communication end-points of any type (i.e., the address to which to send messages in order to obtain a service from a certain peer) as long as they can be represented in XML. Moreover, JXTA allows the publishing of additional service-specific information – also encoded as XML documents – along with their communication end-points; this gives a nice opportunity for targeted discoveries and filtering of potential peers.

The services to be provided by the peer associated to an individual are easily identified from the individual's goals with respect to the community. Once the peer protocols and publishing information have been defined, the design of the actors participating to the community can be performed in parallel. The Tropos methodology naturally leads to the design of multi-agent systems; thus, it is most likely that the implementation of services are agents themselves, interacting both with agents internal to their own actor and with other agents running for other actors of the same community.

4 A Case Study

We show a practical application of the architectural guidelines given above to a DKM problem, taken from Health Care (HC) domain. Our scenario consists of the hospital wards involved in a specific patient treatment and other related stakeholders. For instance, a patient with cancer may be admitted and pharmacologically treated by the oncological ward, operated by the surgical ward, and sent to the X-ray lab for analysis; all these stages may happen at different times, and possibly for different reasons. Each ward has its specific local knowledge and a specific way of organizing and managing activities, processes, and information, according to its own primary objective.

We assume that the output of a deep requirement analysis led to the model depicted in Figure 8, where the actors represent the knowledge management systems of the entities involved in the scenario. The main goal of each actor has been delegated to it by its local users (not shown in the diagram). For instance, the Oncological Ward has the goal define the therapy. In order to achieve this goal, the Oncological Ward has to collect information on the surgery that the patient was subjected to and on possible post operation events and treatments. This

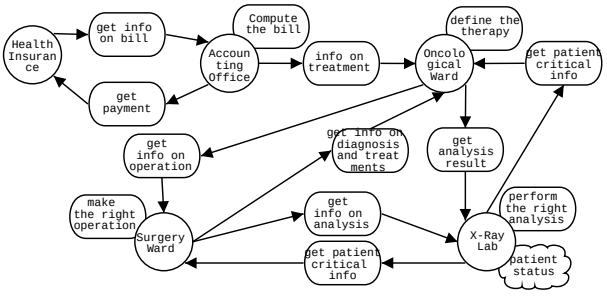

Fig. 8. The hospital wards scenario: an actor diagram

is represented by the goal dependency get info on operation between the actor Oncological Ward and the actor Surgery Ward. Analogously, the Oncological Ward needs information also on the MRI or x-ray analysis performed by the X-Ray Lab, represented by the goal dependency get analysis result. Conversely, the X-Ray Lab depends on the Oncological Ward to get critical information about patient. We represent additional requirements as softgoals; for example, X-Ray Lab's goal perform the right analysis is influenced by the patient status (e.g., a patient could be pregnant).

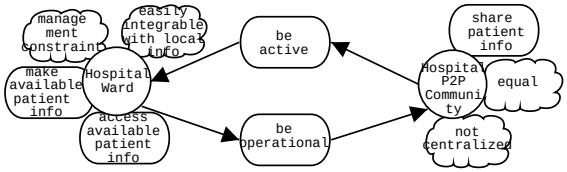

Fig. 9. Applying the P2P virtual community pattern.

To realize their mission, hospital wards need to cooperate, as highlighted by the goal dependencies among their KM systems. The result of this collaboration is a *distributed and virtual* clinical patient report. Privacy policies are a sensitive issues in this context. Following the first step of the guidelines presented in Section 3, we analyze the model depicted in Figure 8 and apply the virtual P2P community pattern (Figure 9). The result is a model where the community has the goal share patient info, and each individual has the generic goals of finding patient record informations (access available patient info) and making available patient record informations (make available patient info). It can be easily seen that these generic goals do actually contribute to achieving the wards' and labs' original goals.

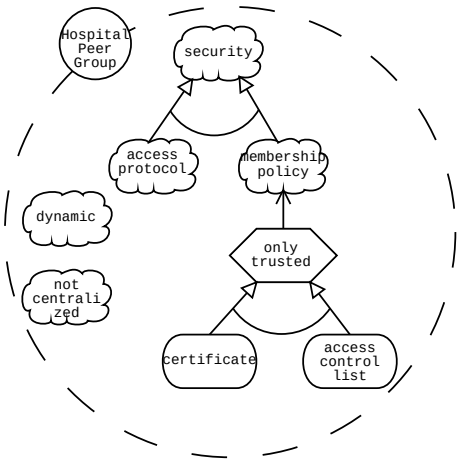

Fig. 10. The HC PeerGroup.

The community is not centralized, since all members keep their own information (as noted above, the members of this group have a strong interests in maintaining their autonomy), and all of them are on an equal basis because they need both to access and to provide information. The generic goal dependency, be member of, between an individual and a P2P community described in Figure 3 has been specialized as be operational. Dynamism of the community is probably minimal in this setting, since the number of wards and labs tend to be stable over time; however, new actors may join over time, for instance new wards, mobile emergency teams, out-of-hospital care for formerly hospitalized patients. One of the advantages of designing a virtual P2P community is that, *after* a system's deployment, new actors can fit in the community as long as their goals

can be reformulated to be the individual's goals. Figure 10 represents the output of the second step of our guidelines: a model for the peer group supporting the community identified above, according to the model of a JXTA peer group depicted in Figure 4. A custom security policy is specified, while all other other goals are adequately supported by the standard JXTA protocols and do not need to be analyzed any further. The membership policy allows only trusted peers to join the community; this control could be performed by using certificates and digital signatures, maintaining a list of trusted or revoked certificate, and granting particular credentials to every peer. Each time a peer use a service, the access protocol checks the credentials of the peer in order to verify if it is authorized to access that particular service (not shown in the diagram). The final result of this second step is the definition of an environment – the peer group – that supports dynamic joining or leaving and discovery of members, while providing a known level of reliability in terms of security and tolerance to failures of single components.

We finally apply the third step of our guidelines. The service provided by the peers of our community is an interface to their local databases containing patient information. We do not elaborate here on published information and protocols, which can be inspired by works on information integration systems, while we present a plausible architecture for a specific actor, the X-Ray Lab; Figure 11 contains only what is relevant to this discussion. The actor has a goal of its own (perform the right analysis), which, after a decomposition not shown here, includes the sub-goal get patient info.

This is delegated to a specific agent, the X-Ray Assistant; after further decomposition, the goal is achieved by accessing the community (access remote patient info, which is a reformulation of the generic individual's access available patient info of Figure 9).

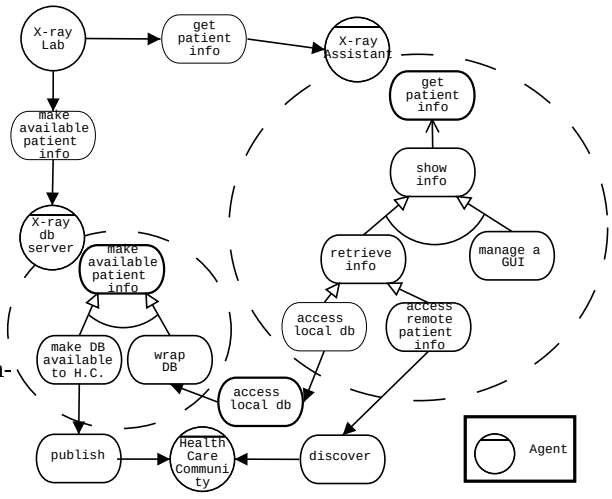

Fig. 11. The X-Ray Lab Peer and the PeerGroup

Participating to the community implies that the X-Ray Lab has to satisfy the individual's make available patient info goal. This is delegated to a second agent, the X-ray db server, which operates both for the community and for local agents (see the access local db goal delegated by the X-Ray Assistant).

5 Related works and conclusion

This paper sketches some preliminary results of a study aimed at defining a framework for designing DKM application covering all development phases. Its process is based on *Tropos*, an agent oriented software engineering methodology. We focused on the architectural design phase, when the application requirements are analyzed in the light of known architectural patterns, and defined some guidelines driving the development of DKM systems as a combination of a specific peer-to-peer infrastructure, JXTA, and multi-agent technologies. We illustrated this guidelines with reference to a scenario taken from the HC domain.

Different lines of research are relevant to our work, ranging from peer-to-peer and agents applications, to software engineering. Two recent attempts to blend peer-to-peer and agents into a single framework are worth to be mentioned: Anthill [1] and InfoQuilt [8]. Both see peer-to-peer as a paradigm for networks of autonomous systems that may join or leave at any time, while further structuring of these networks in peer groups is not yet contemplated. Relevant to our work are also studies devoted to the characterization of the peer-to-peer technology respect to critical issues, such as security [5] and studies in the area of agent-oriented software engineering [9].

Our long term objective is to complete the framework, working both on the conceptual aspects required to model a DKM application and on the technology.

References

1. O. Babaoglu, H. Meling, and A. Montresor. Anthill: A framework for the development of agent-based peer-to-peer systems. Tech. Rep. UBLCS-2001-09, University of Bologna, Italy, 2001.
2. L. K. Chung, B. A. Nixon, E. Yu, and J. Mylopoulos. *Non-Functional Requirements in Software Engineering*. Kluwer Publishing, 2000.
3. F. Giunchiglia, J. Mylopoulos, and A. Perini. The Tropos Software Development Methodology: Processes, Models and Diagrams. *Workshop AOSE02*, AAMAS Conference, Bologna, Italy, 2002.
4. F. Giunchiglia, A. Perini, and F. Sannicolò. Knowledge level software engineering. In Springer Verlag, editor, *In Proceedings of ATAL 2001*.
5. L. Liu, E. Yu, and J. Mylopoulos. Analyzing security requirements as relationships among strategic actors. Tech. Rep., University of Toronto, 2002.
6. P. Bouquet M. Bonifacio and P. Traverso. Enabling Distributed Knowledge Management: Managerial and technological implication. *Novatica and Informatik/Informatique*, 3(1), 2001.
7. Andy Oram, editor. *Peer-to-Peer Harnessing the Power of Disruptive Technologies*. O'Reilly Associates, 2001.
8. S. Patel and A. Sheth. Planning and optimizing semantic information requests using domain modeling and resource characteristics. *In Proceeding of CoopIS 2001*, Trento, Italy, September 2001.
9. M. Wooldridge, P. Ciancarini, and G. Weiss, eds. *Proc. of the 2nd Int. Workshop on Agent-Oriented Software Engineering 2001*, LNCS 2222.