IMPROVEMENT OF THE DOCUMENT MANAGEMENT SYSTEMS BY THE DEVELOPMENT OF AN ONTOLOGY

Alba Fuertes [1], Núria Forcada [2] and Miquel Casals [3]

ABSTRACT

Electronic Document Management Systems (EDMS) is an Information Technology (IT) application that has started to be used in the construction industry as a tool to reduce some of the problems generated by fragmentation. However, EDMS have also some limitations, most of them related to the interoperability and information exchange between systems. In order to solve these problems, different projects, standards and initiatives based on information classification systems and ontologies are being developed, such as ISO 12006 series, Industry Foundation Classes (IFC), Lexicon and e-Construct, e-Cognos, among others.

This paper describes the development of an ontology for the AEC/FM projects' documentation management aimed at establishing a hierarchical structure of the different areas that conform the lifecycle of AEC/FM projects and an interrelationship system between them, where all the documentation created along a project is classified. Therefore, this ontology provides a context-related metadata, that locate the documents along the project lifecycle, and moreover, a content-properties metadata such as identifier, version,..., that relates information to the document to facilitate its research and understanding.

The improvement of the Document Management Systems in AEC/FM projects and their interoperability limitations can be solved by the establishment of this ontology that allows a better research, storage, understanding and classification of the documentation created along a construction project lifecycle.

KEY WORDS

ontology, document management, interoperability, construction project.

INTRODUCTION

The architecture, engineering, construction and facility management industry (AEC/FM) processes are subjected to the influence of highly variable and sometimes unpredictable factors, the team, which includes architects, engineers, subcontractors, and others, changes from one job to the next and, on the other hand, each construction site has its particular complexities. Consequently, AEC/FM projects are typified by their complexity and diversity and by the non standardized nature of their production. From this situation, a huge amount of organizational information is formalized in unstructured documents presenting critical issues such as difficult information search and retrieval, poor interoperability among information systems, poor reuse of content, as well as of business information, related to the context of use of documents in organizations (i.e. business processes and organizational schema).

In order to solve this problematic situation different Information Technology (IT) applications have been developed to be used in the AEC/FM industry. An example are the Electronic Document Management Systems (EDMS) which create an environment where

[1] Research Engineer. Technical University of Catalonia. Department of Construction Engineering, C/Colom 11, Edifici TR-5, 08222 Terrassa (Barcelona) Spain, alba.fuertes@upc.edu

[2] PhD. Technical University of Catalonia. Department of Construction Engineering, C/Colom 11, Edifici TR-5, 08222 Terrassa (Barcelona) Spain, nuria.forcada@upc.edu

[3] PhD. Technical University of Catalonia. Department of Construction Engineering, C/Colom 11, Edifici TR-5, 08222 Terrassa (Barcelona) Spain, miquel.casals@upc.edu

disparate forms of information can be linked together, within the context of a project or organization, to achieve easy access, interaction and control. EDMS use Information Classification Systems (ICS) to classify the internal information and to reduce their interoperability between systems and information exchange limitations.

From this, different standards, aimed at establishing internationally recognized information classification structures, projects based on them and ontologies, another kind of system of representation of the concepts that a domain contains and the relations that exist between them, are being developed (e.g., Lima, Zarli and Storer 2007). Such example, ISO 12006 (2001) and Industry Foundation Classes (IFC) (e.g., IAI 2004), standards that establish a structure for the classification of objects in AEC/FM sector, Lexicon (e.g., Woestenek 2002), a project based on ISO 12006 that provides general information such as building regulations, product information, cost data and quality assessments in a common and standardized language, and eCognos (e.g., Ei-Diraby, Lima and Feis 2005), a project that establishes and deploys a domain ontology for knowledge management, between others. Although all of pretends to classify objects and aspects of the AEC/FM sector, none of them is based on the document management along the lifecycle of a construction project. According to this lack and the existing need to improve the control, reuse and exchange of information, also in terms of documentation, an ontology for EDMS was developed pretending to classify the documents created along a construction project in all its lifecycle.

RESEARCH OBJECTIVE AND METHODOLOGY

The objective of this paper is to present the progresses carried out in the study based on the development of an ontology for EDMS for AEC/FM sector. Therefore, this research pretends to analyze and improve the first and already existing prototype of this ontology by adding the interaction of the actors participating in the project as well as the classification of the documents in accordance with different aspects such as the type of contractual arrangement or the group of related documents. Thus, it is claimed:

- To analyze the first prototype of the Ontology for the Document Management Systems in AEC/FM sector.
- To identify its lacks according to the needs of the future users in terms of interaction with the documentation and its efficient classification.
- To develop a new Concept Model of the Documentation Flow based on the already existing one in the Ontology for the Document Management Systems in AEC/FM.
- To define the elements (classes, subclasses, properties and individuals) that will constitute the extended Ontology according to the previous defined Concept Model.

EXISTING ONTOLOGY FOR EDMS FOR CONSTRUCTION

INTRODUCTION

The first prototype of the Ontology for the Document Management Systems was developed to be used in AEC/FM industry with the aim of reducing the interoperability and information exchange problems from the establishment of a hierarchical structure of the documentation created along the different areas that conform the lifecycle of AEC/FM projects and an interrelationship system between them.

The development of this ontology implied the previous definition of the domain, the important terms that it may contain and their classification system. Consequently, a Concept Model of the Documentation Flows was defined in order to identify which terms should be included in the ontology and how they should be classified and related.

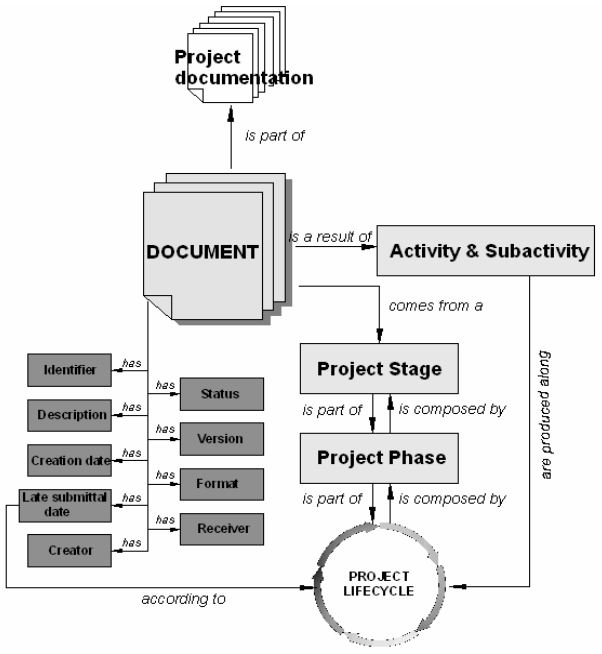
CONCEPT MODEL OF THE DOCUMENTATION FLOWS

The Concept Model of this ontology identifies as important terms in the classification of the documents the lifecycle of a construction project, the different areas of work and all the documentation created along a construction project.

The construction project lifecycle is divided in Phases, the period in the duration of a construction project identified by the overall character of the processes which occur within it, and Stages, defined as sub-processes of the project Phases. On the other hand, the documentation is considered as result of an Activity, defined as a working area of the project, and a Subactivity, understood as the type of information of special importance. An finally, in the definition of the Documents that should be included in the ontology, the actors and the type of contractual arrangement are considered as basic specifications, bearing in mind that the documentation flows created along a construction project depend on the actors involved in it as well as the roles they play in the organization chart defined by the kind of contractual arrangement. In reference to this, the Traditional procurement arrangement is the one selected considering that is the most used for ordinary projects of moderate size and complexity. On the other hand, three different actors, Client, Designer and Contractor, are identified pretending to include all the possible actors involved in a construction project. Up to this point, all the documents generated in a construction project are identified and implemented in the ontology.

From the definition of the elements to include in the ontology, two kinds of metadata are created and a Concept Model is developed. On one hand, Content-related metadata is defined as the metadata that relates the documents with the Phases, Stages, Activities and Subactivities in order to situate them along the project lifecycle. This metadata is inherent in the document. On the other hand, Content-properties metadata is also defined and is the one related to what the document contains or is about, thus providing to users and applications useful hints to help document search and retrieval and to improve the reuse of documented information. This metadata is not compulsory and depends on the author needs. As example the name of the creator or the receiver, the type of format, the creation date, the version,.... Therefore, these two kinds of metadata constitute the Concept Model of the documentation flow. See Figure 1.

Figure 1: Concept Model of the Documentation flow. Document metadata



From the definition of the Concept Model of the documentation flow, a document, that *is part of* the Project documentation, is considered *as the result of* the intersection of an Activity and a Subactivity and is classified along the project lifecycle by the Stages, that *are part of* a Phase, from where it *comes from*. On the other hand, this document *has* related a group of non compulsory information such as the identifier, name of the creator or the receiver, the type of format, the creation date, the version,....

Therefore documents are classified as the result of the intersection of an Activity and a Subactivity that take place along one or more Stages that are part of a Phase. By this way documents are located along the lifecycle of the construction project basing on a three dimensional model.

IMPLEMENTATION

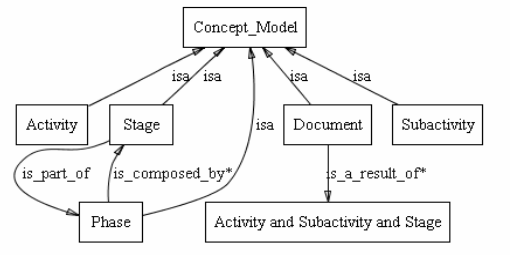
From the concepts identified in the Concept Model, classes, subclasses and properties are defined using Protégé editor, because of its free access, its opened code and its simplicity. Different steps are followed in the development of an ontology. First of all it is necessary to describe the domain of the ontology in order to identify the elements that it will contain. These elements become classes, subclasses, properties and individuals when the ontology is implemented in the editor Protégé.

Classes describe concepts in the domain and subclasses kinds of the already defined classes. In this proposed ontology the classes and their subclasses are: the class Phase composed by the subclasses: Conception, Technical design, Tender and contracting, Execution and Facility management; the class Stage and their subclasses: Conceptual definition, Feasibility study, Preliminary design, Detailed design, Tender action, Contracting, Construction planning, Construction, Construction delivery and Maintenance; the class Activity composed by: Advance, Changes, Contractings, Costs, Environment, Project, Quality and Safety & Health; the class Subactivity that includes: Communication, Documentation, Control and Planning; and finally the class Document that is composed by all the documents identified along a construction project carried out in Spain.

To classify the documents in the hierarchical structure created, properties of classes and subclasses are defined. Three kinds of Protégé properties have been used (See Figure 3):

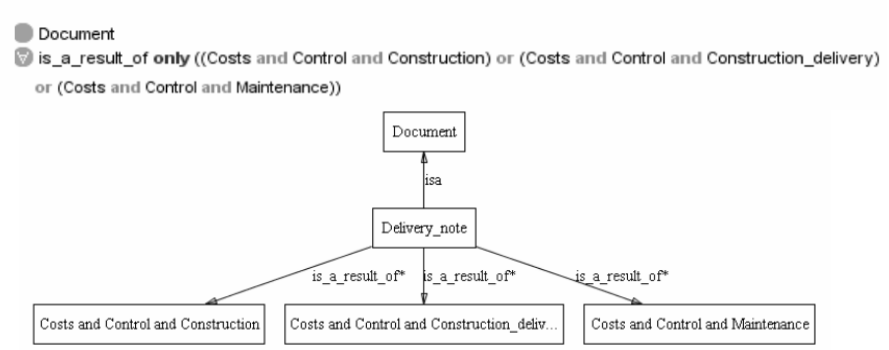
- Object-properties relate the class Document and their subclasses to the classes Stage, Activity and Subactivity. *is a result of* relates each document with the intersection of an Activity, a Subactivity and a Stage, *is part of* relates each Stage with its particular Phase, and finally, *is composed by* relates each Phase with all its Stages;
- Datatype-properties add the content-properties metadata already defined such as the *creator*, the *receiver*, the *format*,... to each document;
- Annotation-properties that, in this case, have been used to provide multi-lingual names for ontology elements (Spanish, Catalan and English) and some comments to make easier their understanding.

Figure 3: Object-Properties in Protégé



Documents are related to the Stages, Activites and Subactivites from which they come from by logic expressions such as *and* and *or*, that express intersection and union, respectively. Therefore, and expression is used to state that a document is a result of the intersection between an Activity, a Subactivity and a Stage (the Phase is already related with the Stage), and or expression is used to state that this document can be located in different locations (intersections) along the project lifecycle. See Figure 4.

Figure 4: *and* and *or* expressions



Up to this point, the hierarchical structure that allows the documentation classification of a general construction project is defined. The following and the last step depends on the particular project and users that work with the DMS where the ontology is applied. This step consists on the definition of individuals, that are considered as particular classes. These individuals are created by the users, who add the information to the already defined properties of each document to facilitate its research and understanding

REUSE AND IMPROVEMENT OF THE ONTOLOGY FOR EDMS FOR CONSTRUCTION

It is almost always worth considering what has already been done in order to check if it is possible to be refined and extended. One of the advantages and characteristics of ontologies is the facility to be reused in order to extend and improve the particular knowledge of a domain. For this reason, what is pretended in this paper is to reuse and improve the basis of the prototype of the ontology for EDMS by the identification and resolution of the limitations of its Concept Model and the development of a new and more complete one.

ANALYSIS OF THE CONCEPT MODEL OF THE DOCUMENTATION FLOWS

By the development of the Concept Model of the already existing ontology, documents are perfectly classified along the lifecycle of the project, bearing in mind the Activity and Subactivity from which they come from. Therefore, the user can know in each Phase, Stage, Activity or Subactivity of the project which documents should be created, and by the same way, along which Phases, Stages, Activity or Subactivity a particular document should be created. On the other hand, these documents have related some extra metadata such as the creator and the receiver that helps each person to identify which documents should receive/create. However, this application is not easy at all bearing in mind that the information *creator* and *receiver* is considered in this Concept Model as Content-properties metadata, and consequently, depend on the users and must be introduced by them. On the other hand, the creator and the receiver of a document depend on the role developed by each actor, and at the same time, this role is a consequence of the type of contractual arrangement applied in the project. As a result, this application becomes complex, heavy to carry out and, consequently, a waste of time.

Another limitation that seems to appear in this Concept Model is the idea to understand the document as an independent entity, with any relation to other documents or elements of the project. Most of the times each activity carried out along the lifecycle of the project creates a group of different documents that contain all the information related to this activity. For this reason the information of some documents can not be understood without the previous analysis of other related documents.

In conclusion, by the analysis of the developed Concept Model it can be identified two main limitations: the classification of the documents by the responsibility that each actor has on them depending on the role he/she develops in the project and the aptitude to relate these documents to each other to improve their understanding.

As a result, a development of a new Concept Model is proposed trying to solve these limitations and, consequently, improving the classification of the construction project documentation and, at the same time, the users' understanding.

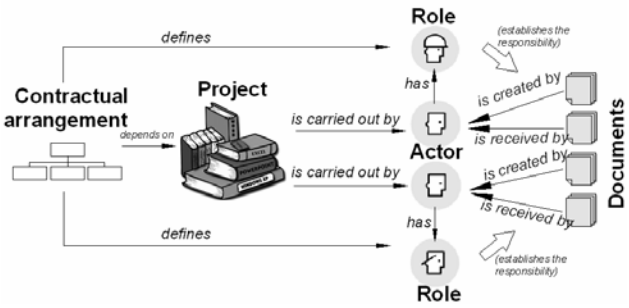
IMPROVEMENT OF THE CONCEPT MODEL OF THE DOCUMENTATION FLOWS

As previously mentioned, the existing Concept Model presents some limitations that affect to the classification of the documents and, consequently to the performance of the ontology and the future EDMS where it will be implemented.

The first question to be solved is the classification of the documents by the responsibility that the actors have on them depending on the role they play in the project that, moreover, is a consequence of the type of contractual arrangement applied.

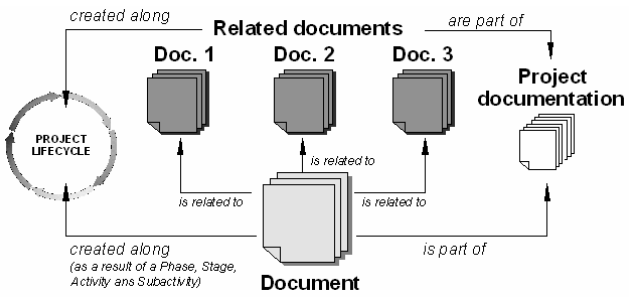
The analyzed ontology is based on the fact that documentation flows created along a construction project depend on the actors involved in it as well as the roles they play in the organization chart defined by the kind of contractual arrangement. Consequently, actors and contractual arrangement are considered basic specifications for the selection of the documents to study. However, with the aim of simplifying and reducing the extension of the ontology not all the types of each specification are used in the definition of the documentation flows. In reference to the type of contractual arrangement, only the Traditional procurement arrangement is selected. On the other hand, only three different actors, pretending to include all the possible actors involved in a construction project, are defined: Client, Designer and Contractor. Although the contractual arrangement and the actors are selected, they are only understood as predefined specifications for the identification and definition of the documentation flow tasks and they are not included as important terms in the Concept Model, and consequently, in the ontology. As a result, information related to the document such as the creator or the receiver are only considered as Content-properties metadata, which is neither compulsory nor pre-related to the document, that it means that must be introduced by the user. In order to solve this limitation in the Concept Model proposed terms such as Actor, Role and Contractual arrangement are included, as well as the relations existing between them. These terms should be considered as Content-related metadata and it is information inherent in the document's metadata, therefore, is previously incorporated. As a result, each type of contractual arrangement defines the roles that all the actors of the project will have and, as consequence, their responsibility on the document. By this way, a particular actor can know which documents should receive or create depending on the role he/she is playing and, as previously exposed, the phase, stage, activity and subactivity that it is been taking place. In reference to the already developed Concept Model, these modifications affect the content of the two kinds of metadata. The terms creator and receiver included in the Content-properties metadata disappear as terms to become relations (*is created by, is received by*) between the actor and the document terms in the new Concept Model. See Figure 5.

Figure 5: Contractual arrangement, actor and role elements in the Concept Model



On the other hand and previously analyzed, the Concept Model presents another limitation: the aptitude to relate the documents to each other to improve their understanding. Related Documents are those extra documents which are necessary for the entire understanding of the document. Some documents might be self understandable so they must not need extra information; others might not. To solve this problem and improve the comprehension of the information contained in the document by the user a term called *related documents* is added to the Concept Model. This term includes all the documents related to the specific studied document and is also considered as Content-related metadata. As a result, the user can know which documents have some relation to the document used and, bearing in mind that these ones are at the same time part of the project documentation and, consequently, considered documents as well, all the information referred to them is also available. See Figure 6.

Figure 6: Related documents in the Concept Model



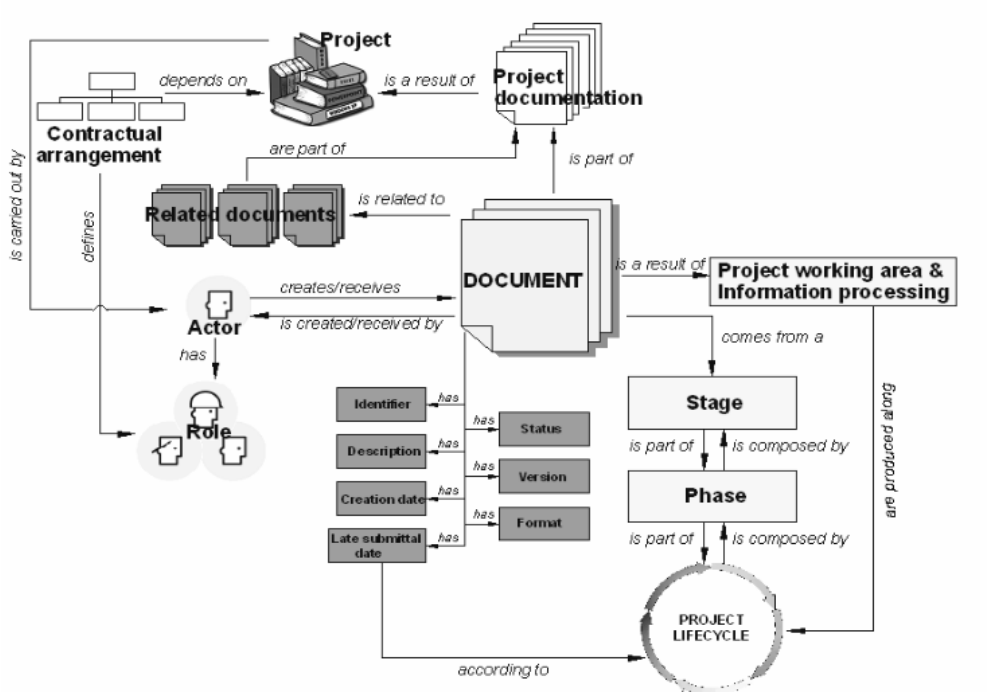
In reference to the other terms and relationships already included in the Concept Model that relate the document to the Phases, Stages, Activities and Subactivities, any modification has been produced. However and already mentioned, the terms *creator* and *receiver* have been taken out from the Content-properties metadata.

And finally, as a recommendation to reduce confusion in the real meaning of the elements Activity and Subactivity, two new names for these concepts are proposed: the concept defined as a working area of the project, mentioned till now Activity, will be named *Project Working Area*. On the other hand, the concept understood as the type of information of special importance, considered till now as Subactivity, will be mentioned *Information Processing*.

As a result and to sum up, two types of metadata still exist, however they include different terms of the Concept Model. On one hand, the Content-related metadata contains all the terms that classify the document along the project lifecycle bearing also in mind the type of contractual arrangement and the role that the actors play. This information is inherent in the document and, therefore, is already related to it. The terms Project Working Area, Information Processing, Phase, Stage, Actor, Role, Contractual arrangement and Related documents are considered as Content-related metadata. On the other hand, the Content-

properties metadata are the already defined, bearing in mind that the terms *creator* and *receiver* have been taken out. See Figure 7.

Figure 7: Extended Concept Model of the Documentation flow



IMPLEMENTATION

As already mentioned, what is pretended in this paper is to reuse and improve the already existing ontology for EDMS. For this reason, all the terms and relationships that haven't been modified in the new Concept Model and are already included in the ontology developed in Protégé, are maintained and used as the basis of the improved and more extended ontology.

As a result, the classes *Phase*, *Stage, Activity* (now named *Project working area*), *Subactivity* (now named *Information processing*) and *Document*, as well as their subclasses, haven't been changed. The same happens with the Object-properties *is a result of, is part of* and *is composed by* that relate these classes and subclasses to each other, and the Annotation-properties that provide multi-lingual names to these elements. On the other hand, the Datatype-properties have been modified as a consequence of the disappearance of the terms *creator* and *receiver*.

From the new terms identified in the Concept Model new classes, subclasses and properties are defined. In reference to the classes, two new classes are created: *Role* and *Contractual arrangement*. As it can be observed, the elements *Related documents* and *Actor* are not considered as classes of the new ontology. Although both of them are important terms in the Concept Model and help to understand it, they are not directly used in the classification of the documents in the ontology. When we talk about the Related documents, we refer to all that documents that are already included in the class *Document* that have a relationship *is related to* (translated to a Object-property in the ontology) between them. Therefore, a task to carry out is to identify all relationships between the documents that appear along the project lifecycle. In reference to the element *Actor*, the actors that participate in a project can not be predefined; each project has a specific number of actors performing more or less roles. On the other hand, in all the projects a particular group of roles is performed, independently of the specific number of actors, understanding that different roles can be developed by only one

actor. For this reason, the class *Role* is created as the better way to classify documents by the responsibility that the actors, that at the same time play roles, have on them.

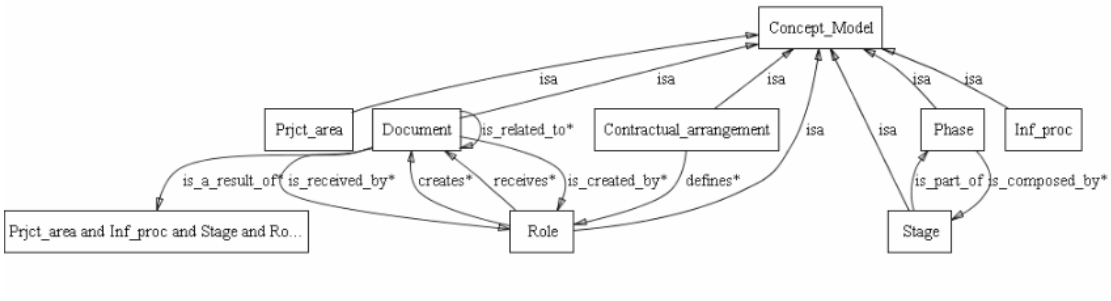
Therefore, bearing in mind different theories and working methods related to the definition of the actors/roles of a construction project (e.g., LOE 1999, IAI 2004, ISO 12006 2001), the following roles are identified as subclasses of the class *Role*: Client, Owner, Designer, Professional advisor, Main contractor, Project Manager, Supplier, Construction Manager, Quality Control Entity, Civil Engineer, Structural Engineer, Services Engineer, Services Suppliers, Maintenance Manager, H&S responsible man, Quality responsible man, Environment responsible man, Subcontractors.

On the other hand, taking into account the Spanish practice, the contractual arrangements are broadly classified under three headings (e.g., Heredia 1998). Thus, the subclasses of the class Contractual arrangement are: Traditional procurement arrangement, Turnkey project arrangement and Professional Construction Management arrangement.

To relate the documents with the new classes and subclasses added in the Concept Model and in the ontology, new properties of classes and subclasses are defined. According to the three kinds of Protégé properties already used in the existing ontology. See Figure 8:

- Object-properties; *is related to* relates all the documents included in the class Document according to the relationships that appear between them along the project lifecycle; *creates*, as well as *receives*, relates the Role class with all the documents that should create/receive according to the Phase, Stage, Project working area and Information processing that is taking place and to the contractual arrangement; *defines* relates each kind of contractual arrangement with the roles that participate in the project.
- Datatype-properties; due to the new classes defined, *creator* and *receiver* are deleted.
- Annotation-properties; multi-language is also used with the new classes, subclasses and properties defined.

Figure 8: Object-Properties in Protégé



In this new ontology, Documents are related to the Stages, Project area, Information processing and Role from which they come from by the same, and already exposed, logic expressions *and* and *or*.

Up to this point, the hierarchical structure that allows a better documentation classification of a general construction project is defined.

CONCLUSIONS

By the analysis the already developed Ontology applicable to a DMS for Construction a new and more extended Concept Model has been defined. On the one hand, new classes and properties have been added in order to improve the classification of the documents bearing in mind the contractual arrangement used in the project, the roles that it defines as well as the responsibility they have on the documents. On the other hand, documents are also related to

each other to improve their understanding. By this way, a document created during a construction project is classified along its lifecycle as a result of the intersection of a Project area, Information processing, a Stage and a Role, according to the contractual arrangement selected.

The future work of this research would be, on the one hand, the identification and definition of all the possible relations that exist between documents according to the Stage, Project area, Information processing and Role that is taking place. On the other hand, contractual arrangements must be analyzed in order to identify the roles they define as well as the responsibilities to create and receive a document these roles have. By the identification of all these relationships the extended and improved Ontology would be developed and used as internal information classification system of a tool such as BSC, a web page,... allowing users' interaction and the exchange of information by the web.

REFERENCES

Ei-Diraby T.E., Lima C., and Feis B. (2005). "Domain taxonomy for construction concepts: Toward a formal ontology for construction knowledge." *Journal of Computing in Civil Engineering,* 19 (4) 394-406.

Heredia R. (1998). *Dirección integrada de Proyecto -DIP- Project Management*. 3ª edición, Servicio de Publicaciones de la ETSII de la UPM, Madrid.

International Alliance for Interoperability (IAI) (2004). *IAI Web Site* (available at http://www.iai-international.org).

International Organization for Standardization ISO 12006-2 (2001). *"International Organization for Standardization. Building construction - Organization of information about construction works. Part 2: Framework for classification of information".*

Ley de Ordenación de la Edificación (1999). Building Act 38/1999. *Ley 38/1999, de 5 de noviembre, de Ordenación de la Edificación (BOE núm. 266, de 6 de noviembre de 1999).* Ministerio de Fomento. Madrid.

Lima C., Zarli A., and Storer G. (2007). *"Controlled Vocabularies in the European Construction Sector: Evolution, Current Developments, and Future Trends".* Complex Systems Concurrent Engineering. Collaboration, Technology Innovation and Sustainability, Sao Jose dos Campos, Brasil.

Woestenek K. (2002). "From lexicon to XTD." *European Conference on Product and Process Modelling in the Building and Related Industries. Proc. eWork and eBusiness in Architecture, Engineering and Construction*, Portoroz, Slovenia, The Netherlands: Balkema