ENTERPRISE ONTOLOGY MANAGEMENT
An Approach based on Information Architecture

Leonardo Azevedo, Sean Siqueira, Fernanda Araujo Baião
Jairo Souza, Mauro Lopes, Claudia Cappelli and Flavia Maria Santoro
NP2Tec – Research and Practice Group in Information Technology
Federal University of the State of Rio de Janeiro (UNIRIO), Rio de Janeiro, Brazil

Keywords: Ontology management, Knowledge management, Information architecture.

Abstract: Ontologies have gained popularity, but its promises of being a key point to the solution of real-world problems and mitigating interoperability problems at a large scale have not yet been accomplished. Ontology management is at the kernel of this evolution, and there is a lack of adequate strategies and mechanisms for handling it in such a way to contribute to a better alignment between business and IT. This work proposes an approach for enterprise ontology management as part of an Information Architecture initiative. This approach provides a more complete foundation of the ontology lifecycle while guiding the enterprise in this management, by defining a set of processes, roles and competencies required for ontology management. It was applied to a big enterprise in Brazil at the Data Integration department.

1 INTRODUCTION

Gruber (2008) defined ontology as "a set of representational primitives with which to model a domain of knowledge or discourse". There are many other definitions in the literature, but all these definitions converge to consider an ontology as a common representation of a domain of discourse. Once built, ontologies can be used in several areas such as (Damjanovic *et al.*, 2004; McGuinness, 2005): data integration; business process modeling; database design; information retrieval and extraction; knowledge management; qualitative modeling; language engineering; e-commerce; and configuration support. Many of these scenarios pose challenges and opportunities for any enterprise.

However, building an ontology is only a part of the way when pursuing to effectively use it within a corporate environment. There is a need to envision ontology construction and maintenance as part of an initiative towards information governance, which includes not only technical but also management activities. Ontology management is the task of producing and maintaining consistency between formal and real-world semantics (Fensel, 2008). This work extends this definition including a set of processes to achieve this consistency.

The proposed set of processes is based on an Information Architecture viewpoint. Enterprise architecture is a set of principles, methods and models that help bridge the communication gap between IT architects and stakeholders (Lankhorst, 2005). In this context, ontologies may be considered as conceptual data models that compose the set of artifacts to build the enterprise information view.

This paper is divided as follows. Section 2 presents an overview of ontology management. Section 3 presents in details our proposal, and finally section 4 concludes this work.

2 OVERVIEW OF ONTOLOGY MANAGEMENT

Corcho *et al.* (2003) explain that in order to build a new ontology, several basic questions are related to the methodologies, tools and languages to be used in the development process. In the survey conducted to base our approach, a set of disciplines that are important for ontology management were studied, such as: methodologies for constructing ontologies, representation languages, query languages, ontology reuse/compatibility, assessment of ontology quality, ontology evolution, and supporting tools. The studied authors do not, however, present a set of activities, roles and competencies in order to apply these concepts in an enterprise environment. This

Azevedo L., Siqueira S., Araujo Baião F., Souza J., Lopes M., Cappelli C. and Maria Santoro F. (2009).
ENTERPRISE ONTOLOGY MANAGEMENT - An Approach based on Information Architecture.
In *Proceedings of the 11th International Conference on Enterprise Information Systems - Information Systems Analysis and Specification*, pages 243-248
DOI: 10.5220/0002009802430248
Copyright © SciTePress

section complements the concepts presented in (Corcho *et al.*, 2003).

There are several methodologies for ontology construction in the literature and they can be compared regarding several different characteristics, such as creation strategy, application dependence, and existence of validation and evolution phases.

In order to build the ontology it is necessary to abstract reality according to some conceptualization. When represented as a concrete artifact, a model can support communication, learning and analysis of relevant aspects of the domain. To make it possible, an ontology must be expressed in a language. Diverse representation languages of real world conceptualizations exist, with distinct purposes (e.g., LINGO, ER, OWL, OntoUML). The OWL is by far the most adopted language for ontology representation, since it is a W3C standard.

In addition to representation languages it is also required to consider the query/retrieval mechanisms. Query languages in ontologies are fairly new and different proposals can be found in the literature. They can be based on database query languages (i.e., SQL, OQL), based on rule languages (i.e., Prolog, Lisp) and based on path expressions (specially those languages for querying XML documents such as XPath and XQuery). Examples of query languages are: nRQL, OWL-QL and SPARQL.

One application of ontologies is for semantics interoperability between different information sources. Ontologies compatibilization allows reuse of knowledge structure in different applications. The main mechanisms are: ontology matching (Noy and Musen, 2001); ontology merging and alignment (Noy and Musen, 1999); ontology integration (Pinto *et al.*, 1999); ontology mapping (Noy and Musen, 2003); and others can be found in (Klein, 2001).

Brank *et al.* (2005) highlights that ontology analysis is a process of value an ontology according to some criteria, in order to define which one of many ontologies are most appropriate for a purpose.

There are many proposals for ontology validation, and they can be divided in approaches for lexical validation (Maedche and Staab, 2002), objective validation (Gangemi *et al.*, 2006), and for semantic validation (Guarino and Welty, 2002).

Ontologies changes according to updates, insertions, deletions, and structure reviews, since a better understanding of existing concepts or an ontology scenario change. Besides it is possible to change ontology formalism or the domain of representation (Klein and Fesel, 2001). One change in an ontology item can produce inconsistencies in other parts of it. The bigger is the ontology, the more complex is to understand completely the extension and meaning of changes (Stojanovic *et al.*, 2002). The implementation of changes must be propagated to other dependent ontologies and applications. Ontologies' instances must be changed to preserve consistency.

Gómes-Pérez *et al.* (2002) show that many tools appeared in this decade for ontology development (construction, annotation, integration etc) and for ontology use in applications. A detailed list of tools is presented by Gómes-Pérez *et al.* (2002), Welty (2004) and Cardoso (2007).

3 THE PROPOSED APPROACH

Spewak and Hill (1992) proposed a set of processes of an information technology architecture (ITA): Building the Current and the Future Architectures, Maintaining Current and Future Architectures, Defining Policies and Standards, Prospecting Technology, Participating on Committees, Evaluating the IT Quality, and Monitoring and Measuring Activities.

Considering ontologies as conceptual models, thus playing an important role within ITA, we claim that those set of processes may be specialized in order to drive enterprise ontology management (Figure 1). All these processes were detailed in macro-processes, Event-Driven Process Chains (EPC) were modelled as well as function trees (Sharp and McDermott, 2001).

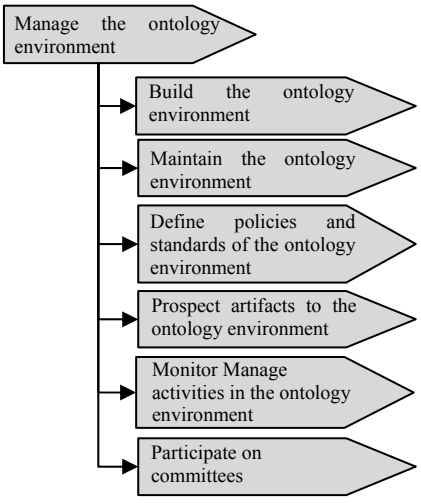

Figure 1: Manage the ontology environment macro-processes.

The first were used to detail critical event flows to the ontology management in those processes in which the activity sequence is known. The function

trees were used to present a set of activities that are necessary to the processes to which the execution sequence is not yet known, at least not exactly. In some cases, the execution sequence of activities is not critical to the purpose of providing an understanding of the ontology management proposal while in other cases it is extremely dependent of the business practices of the organization.

3.1 Build the Ontology Environment

The process "Build the ontology environment" has a sub-process to define roles, responsibilities and competences related to the ontology management in the organization as well as to attribute them to specific professionals.

Another sub-process is to define the infrastructure of ontology management environment. This process is related to the definition of appropriate tools for manipulation, usage and storage of global ontology as well as the application ontologies. Several categories of tools must be provided: editors, storage and retrieval servers for the knowledge structures, backup server, ontology visualization, etc. This environment has also to provide information security and allow the definition of different access levels to the users, be scalable and integrated to other tools of the enterprise.

After the environment infrastructure is operational, the global (or domain) ontology is created. Competence questions are formulated in order to understand and determine the ontology scope and objective. It is then validated according to syntax and semantics. If necessary, adjustments are applied. After the validation activities, ontology concepts are mapped to the integrated data model. Therefore, ontology items are associated to tables and attributes. If database concepts are not fully covered by the ontology, new items are added to it.

Other sub-process is related to the use of the ontology. The ontology must be published, highlighting benefits and possible uses. Access mechanisms and strategic uses must be defined. Training, specially the ontology maintenance and use is also very important to guarantee the success.

3.2 Maintain the Ontology Environment

It is important to maintain the ontology environment so that ontologies have the concepts that are used in the enterprise, especially those presented in the integrated database.

Infrastructure Maintenance. In order to maintain the ontology management environment it is necessary to update tools, manage users for each tool and solve problems in the computational environment. Some required functions are: manage tools licenses, manage access permissions, create and publish tools' manuals, guarantee tools integration, update the environment whenever necessary, make contracts with the suppliers and partners, provide support to the environment etc.

Global Ontology Maintenance. Ontology changes can be required when a new business process is modelled, when domain concepts changes or when the enterprise knowledge changes.

Whenever a change is required, it is semantically evaluated and if applicable, the change is represented. After that, the change is implemented and the ontology is validated, capturing the changing meaning (conceptualisation) and verifying possible inconsistencies in other items. Finally, the new version of the ontology is propagated to the ontologies and dependent systems.

Global Ontology Extension. Global ontology can be extended to cover other knowledge areas or domains. The necessary information can come from different sources such as business process models, data models, domain specialists and documents.

Processes can provide good information about ontology items that have not been considered in the ontology or even is not really correct or complete. In addition, mapping process models and ontologies provides a better understanding of the processes and activities (Cappeli *et al.*, 2007).

The data model is also a good source of information because some ontology items are captured from tables (or entities), relationships, attributes and restrictions. A document detailing the process of building ontology from data models was also developed: "Good practices to build an ontology from logical or physical data models".

Domain specialists can be interviewed and help at capturing ontology items and are good sources of information as well as documents about the domain. Some examples of these documents are manuals, data dictionaries, training documents, conceptualization documents etc. It is also possible to use natural language processing to automatically capture ontology items, but manual processes were defined to the enterprise at this moment.

For the ontology items that are to be added or modified to the global ontology it is necessary to follow the changing processes.

Application Ontology Creation and Maintenance. In certain cases it is important to have application ontologies that extend the global ontology concepts and have concepts that are specific to a certain application. If a need for such kind of ontology is identified, then the application scope is observed and competence questions must be formulated. Existing ontologies in the same domain are analyzed to search for reusable items.

Then, the application ontology is planned; risks are identified and documented as well as the effort for the development. An execution chronogram is created. After that, the ontology is modeled. The necessary concepts that are in the global ontology are identified as well as those present in other existing ontologies. These concepts that have already been defined in other ontology are mapped while new concepts and items are codified.

Selecting concepts from the global ontology allows reusing these concepts that inherit characteristics from a class of the global ontology, for instance. Therefore it is possible to reduce development time and to allow disseminating concepts through the enterprise. The reuse is also applicable when identifying existing ontologies about the same domain. In some cases, the existing concepts must be adapted in order to keep the ontology consistence.

Representing concepts that do not exist in the global ontology can be through a graphical tool for representing concepts or in a textual way, indicating its name and other characteristics. For correctly identifying concepts, supporting documents can be used such as “Good practices to build an ontology from logical or physical data models” and “Questionnaire for gathering ontological items”, which were developed in this project. The relationships between concepts are also identified, presenting the concepts that have to be modelled and the dependence of concepts from the global ontology. The codification of the concepts and their relationships in OWL must follow description standards. Then the ontology is validated and the adaptations are executed.

3.3 Define Policies and Standards

This process is responsible for creation, maintenance, publishing and auditing policies and standards of the ontology environment.

Creating the Policies and Standards. This process starts with a need for standardizing some task or item of the ontology environment. Then the item or task is analyzed and the policies and standards are defined and approved.

Raising the item characteristics looks for improving the ontology quality, getting rules such as coverage (Which areas do the ontology items cover? the whole department or a part of the department, the whole enterprise or a department? How can this concept be generic to cover all the knowledge areas of the enterprise?), correctness (How correct is this item? Is it applicable to all cases? Is there an exception? How to validate all the aspects?), richness (Is it complete? Can it be more detailed? Is it worthwhile to detail? Are there other characteristics, properties or relationships to be considered?), commitment (Did all the possible areas that are going to use this item validate it adequately? Who was the responsible for the validation?), organization and modularity, reality correspondence, clarity, goal consistence, capture of the domain invariant structure, description form, syntax correctness, etc.

Then the standards are prospected in the market and if a proposal is found, it is discussed, adaptations are investigated and executed. After that, the standard is validated and approved.

Maintaining the Policies and Standards. This process starts with a need for maintenance, which can be a new standard in the market or new aspects to be considered in the organization standard. A group is created for standard revision and if changes are identified, they are analyzed as well as the impact of the change. If approved, the change is implemented. The changes are then validated.

Publishing the Policies and Standards. This process starts when a standard is created or changed. The enterprise communication mechanisms are used to publish the policies and standards. Training sessions are provided.

Controlling Policies and Standards. This process starts with a need for controlling the correct use of these policies and standards. A sample of projects is selected and it is verified the use of the policies and standards. The result of the auditing tasks is published as well as a list of needs for changes in the standard.

3.4 Prospect Artifacts

This process is responsible for prospecting technologies for the ontology management environment. It is a continuous process. First, information about tools for ontology management is

searched through the Internet, the participation on technical and scientific events and consulting.

Then, criteria for evaluating the tools are defined as well as the candidate tools. Criteria should cover mandatory and desirable characteristics for the type of tool in evaluation. For ontology management, they can be related to ontology edition (Is it possible to change concepts in a graphical mode? Is it possible to save the ontology in OWL?), ontology versioning (Is it possible to save more than one version of the ontology? Does it allow relating ontology versions?), etc. There are also essential criteria to the organization such as existence of technical support in the country, training facilities to the team, etc. Usually these business requirements are determinant to tool acquisition.

For each criterion it is given a weigh. Each tool is evaluated Then the result is analyzed and the recommendation is generated. Finally, the tool is chosen, directives for integrating technologies are defined and the evaluation result is published.

In addition to prospecting tools it is also desirable to prospect domain concepts. Prospecting can be through searches on the standardization organisms as well as institutions with similar domains, governmental organizations, etc. or through participation in inter-enterprise committees.

The concepts are then evaluated according to existing concepts of the global ontology. The evaluation result is published, presenting a comparative report of the concepts as well as the recommendation of the evaluation team. This recommendation can generate a change requirement for the global ontology.

3.5 Monitor Management Activities

This process is responsible for monitoring management activities in the ontology environment through quality indicators as well as communication of results. It is necessary to establish and monitor indicators to verify if (and how) internal activities of the area are executed and are according to what was expected. The results should be published.

3.6 Participate on Committees

This process is responsible for defining participation mechanisms of the management group of the ontology environment in institutional committees in the organization. Through these committees it is possible to establish relationships with other groups and areas in the organization.

Some possible functions of such participation can be: to analyze the contribution of using ontology inferences; support in systems modelling; support in data schema validation; share concepts in process models; create mappings between process models and ontologies; know strategies for making knowledge explicit; know business process meta-model; know infrastructure standards for ontology management; know information exchange standards; define systems that can use existing ontologies; discuss infrastructure requirements for ontology management; establish consensus in concept definition; identify data that can be used to refine concept definition; identify improvements in data quality; identify improvement in analysis tools; identify data integration opportunities; identify common processes with the data management; influence strategies for making knowledge explicit; influence standards for information interchange; identify improvements in the ETL process for the informational database; interact with data administration area; interact with informational area; interact with transactional area; interact with software development area; interact with knowledge management area; interact with support and infrastructure area; interact with concepts standardization groups; interact with process offices.

4 CONCLUSIONS

This work proposed a set of processes for ontology management in enterprises. The rationale for the processes definition was based on an Information Architecture viewpoint, where the ontologies play the role of conceptual data models. The proposed set of processes include: build the ontology environment (plan, organize, define roles and responsibilities, define and deploy infra-structure and construct global ontology); maintain the ontology environment (make changes on existing ontology and extend it by new concepts, update infra-structure, construct application ontologies, etc); define policies and standards (define new rules for guaranteeing ontology quality, correctness, commitment, ontology descriptions etc); prospect artifacts (prospect new tools and new domains); monitor and manage activities (define metrics to measure results and execute evaluation activity, and analyze if the expectations are accomplished); and participate in committees (establish relationships with other groups and areas in the organization in order to grow the contribution of ontology).

These proposed processes were presented and evaluated by consultants on Ontology Management and researchers, as well as enterprise professionals. They were a result of a project at the data

administration department at PETROBRAS, which is the largest and most important oil and gas company in Brazil. The data administration department at PETROBRAS is responsible for data integration in the domain of oil and gas exploration and production.

According to the three traditional abstraction levels of database design (Elmasri and Navathe, 2005), the administration department uses ontologies to represent the first level of abstraction (conceptual level). Ontology is also used to help data integration of concepts belonging to different areas. Besides, the use of some of the proposed processes has demonstrated good results. Different departments are developing ontologies, which have been integrated following the proposed activities, and they are used to support communication, to learn and to analyze relevant aspects of the company domains.

ACKNOWLEDGEMENTS

The authors would like to thank Petrobras, mainly TIC/TIC-E&P/GDIEP, for supporting this project.

REFERENCES

Brank, J., Grobelnik, M., Mladenic, D., 2005. A survey of ontology evaluation techniques. In *8th International mulit-conference Information Society IS*, pp. 166–169.

Cappelli, C., Baião, F., Santoro, F., *et al.*, 2007. An approach for domain ontology construction from business process models. In *Second Workshop on Ontologies and Metamodeling in Software and Data Engineering* (WOMSDE'07), João Pessoa, Brazil.

Cardoso, J., 2007. The Semantic Web Vision: Where are we?. IEEE Intelligent Systems, *22* (5), pp. 84-88.

Corcho, O., Gómez-Pérez, A., Guerrero-Rodríguez, D.J., *et al.*, 2003. Evaluation experiment of ontology tools' interoperability with the WebODE ontology engineering workbench. In *ISWC2003 Workshop on Evaluation of Ontology Tools*, USA.

Damjanovic, V., Devedžic, V., Djuric, D., *et al.*, 2004. Framework for Analyzing Ontology Development Tools. *AIS SIGSEMIS Bulletin*, *1* (3), pp. 43-47.

Elmasri, R., Navathe, S., 2005. *Fundamentals of Database Systems*, Addison-Wesley.

Fensel, D., 2008. Foreword. In: Hepp, M.; De Leenheer, P.; de Moor, A.; Sure, Y. (Eds.). Ontology Management: Semantic Web, Semantic Web Services, and Business Applications. p. 295.

Gangemi, A., Catenacci, C., Ciaramita, M., *et al.*, 2006. Modelling Ontology Evaluation and Validation, In *3rd European Semantic Web Conference* (ESWC2006).

Gómes-Pérez, A., Fernández-López, M., Fensel, D., 2002. Deliverable 1.3: A survey on ontology tools, Technical Report IST-2000-29243, OntoWeb - Ontology-based information exchange for knowledge management and electronic commerce.

Gruber, T.R., 2008. Ontology. In: Liu, L. and Özsu, M. (Eds.) Encyclopedia of Database Systems, Springer-Verlag.

Guarino, N., Welty, C., 2002. Evaluating ontological decisions with OntoClean. In *Communications of the ACM*, 45 (2), pp. 61-65.

Klein, M., 2001. Combining and Relating Ontologies: An Analysis of Problems and Solutions. In *Workshop on Ontologies and Information Sharing at the 17th International Joint Conference on Artificial Intelligence*, pp. 53-62, Seattle, USA.

Klein, M., Fensel, D., 2001. Ontology Versioning on the Semantic Web. In *Proceedings of the International Semantic Web Working Symposium* (SWWS), pp. 75-91, California, USA.

Lankhorst, M., 2005. Enterprise Architecture at Work: Modelling, Communication, and Analysis, Springer.

McGuinness, D.L., 2005. Ontologies Come of Age. In: Fensel, D., *et al.* (eds), Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, MIT Press.

Noy, N.F., Musen, M.A., 1999. SMART: Automated Support for Ontology Merging and Alignment. In: 12th Workshop on Knowledge acquisition, modeling and management, v. 4, pp. 1-20, Banff, Canada.

Noy, N., Musen, M.A., 2001, Anchor-PROMPT: Using Non-Local Context for Semantic Matching. In *17th Workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence*, *1*, pp. 63-70, Seattle, EUA.

Noy, N., Musen, M.A., 2003. The PROMPT Suite: Interactive Tools For Ontology Merging And Mapping, International Journal of Human-Computer Studies, 59 (6), pp. 983-1024.

Pinto, S.H., Gómez-Pérez, A., Martins, J.P., 1999. Some Issues on Ontology Integration. In *Workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends* (IJCAI99's), *18*, pp. 7-12, Stockholm, Sweden.

Sharp, A., McDermott, P. 2001. Workflow Modeling: Tools for Process Improvement and Application Development, Artech House.

Spewak, S. H., Hill, S. C, 1992. Enterprise Architecture Planning: Developing a Blueprint for Data, Applications, and Technology, John Wiley & Sons, Inc.

Stojanovic, L., Maedche, A., Motik, B., *et al.*, 2002. User-driven ontology evolution management. In *Proceedings of the 13th European Conference on Knowledge Engineering and Knowledge Management*, *Ontologies ad the Semantic Web*, pp. 285-300.

Welty, C., 2004. Ontology Maintenance Support: Text, Tools, and Theories. In *Presentation at the 7th International Protégé Conference*, Bethesda, USA.