EVALUATION OF ONTOLOGY BUILDING METHODOLOGIES
A Method based on Balanced Scorecards

Michela Chimienti, Michele Dassisti
Dipartimento di Ingegneria Meccanica e Gestionale, Politecnico di Bari, Viale Japigia 182, Bari, Italy

Antonio De Nicola, Michele Missikoff
IASI-CNR, Viale Manzoni 30, Rome, Italy

Keywords: Ontology Building Methodologies, Evaluation of Ontology Building Methodologies.

Abstract: Ontology building methodologies concern techniques and methods related to ontology creation that starts from capturing ontology users' requirements and concludes by releasing the final ontology. Despite the several ontology building methodologies (OBMs) developed, endowed with different characteristics, there is not yet a method to evaluate them. This paper describes an evaluation method of OBMs based on Balanced Scorecards (BSCs), a novel approach for strategic management of enterprises that we apply to the assessment of OBMs. Then, as a case study, the proposed evaluation method is applied to the UPON OBM. Finally, we show the major strengths of the BSCs' multi-disciplinary approach.

1 INTRODUCTION

An ontology is an explicit specification of a conceptualization (Gruber, 1993). Development of ontologies requires collaboration between a team of knowledge engineers (KEs) with a technical background, and domain experts (DEs) with adequate know-how in the domain to be modelled.

An ontology building methodology (OBM) is a set of techniques and methods, aimed at ontology creation, that starts from capturing ontology users' requirements and concludes by releasing the final ontology (Chimienti, 2006). Five approaches to ontology building are available (Holsapple, 2002). Using the *inspirational approach*, an ontology is built starting from its motivation. With the *inductive approach*, an ontology is built starting from observing, examining, and analyzing one or more specific cases in the domain of interest. With the *deductive approach*, an ontology is built starting from general principles and assumptions that are adapted and refined. Using the *synthetic approach*, an ontology is built starting from a base set of ontologies that are merged and synthetized. Finally, according to the *collaboration approach*, an ontology is built reflecting experiences and viewpoints of persons who cooperate and interact with each other. Existing OBMs usually adopt approaches that can be considered as hybrids of the five above mentioned. From a literature survey, among the most important OBMs, we cite: *SENSUS methodology* (Swartout, 1997), *On-To-Knowledge* (Sure, 2002), *Ontology Development 101* (Noy, 2001), *Methontology* (Corcho, 2003), *DILIGENT* (Tempich, 2006), and *UPON* (De Nicola, 2009).

Despite a growing literature on metrics aimed at assessing quality of ontologies (Burton-Jones, 2005), (Guarino, 2002), works related to evaluation of OBMs are still preliminary. In (Fernández-López, 1999), an approach to analyse OBMs inspired by the "IEEE 1074-1995: Standard for Developing Software Life Cycle Processes" (IEEE, 1996) is proposed. Since ontologies are part of software products, the author asserts the quality of an OBM is connected to the compliance with the processes for software development. The analysis criteria are established without defining how these should be measured and no additional perspectives, e.g., training facilities, development time, and involved human resources, are considered.

(Paslaru, 2006) proposes a framework to estimate costs of ontology engineering projects, consisting of a methodology to generate a cost model, an inventory of cost drivers, and the

Chimienti M., Dassisti M., De Nicola A. and Missikoff M. (2009).
EVALUATION OF ONTOLOGY BUILDING METHODOLOGIES - A Method based on Balanced Scorecards.
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development*, pages 141-146

ONTOCOM cost model. This methodology mainly focuses on the economical aspects of ontology engineering and does not provide a complete evaluation of OBMs, considering also ontology quality aspects (e.g., syntax and semantics).

Finally, (Hakkarainen, 2005) proposes a framework to evaluate ontology building (OB) guidelines according to five categories, mainly focusing on the usability of OB methods. The framework does not focus on ontology product and economic criteria, as development time and costs.

The aim of this paper is to present a method to evaluate an OBM based on Balanced Scorecards (BSCs) (Kaplan, 1996). This method takes into account different aspects of OB (e.g., financial, modelling, and ontology quality). The paper is organized as follows. Section 2 presents basic notions on BSCs and their application in the OB domain. Section 3 describes the proposed evaluation method and Section 4 demonstrates its feasibility by describing the evaluation of UPON OBM. Finally, in Section 4, conclusions and future work are discussed.

2 BSCS: BASIC NOTIONS

BSCs' approach is defined as "*a multi-dimensional framework for describing, implementing and managing strategies at all levels of an enterprise and linking objectives, initiatives and measures to an organization's strategy*" (Kaplan, 1996). It allows assessing business and enterprises according to four perspectives or scorecards: *financial/stakeholder*, *internal business process*, *innovation and learning*, and *customer*. Each perspective is analysed according to four components: *objectives*, *metrics*, *targets* and *initiatives*. BSCs' approach has been applied in several contexts and, among them, in the ICT domain (Buglione, 2001), (Ibáñez, 1998). We propose to apply BSCs to a particular ICT scenario: ontology building that, together with ontology maintenance and ontology reuse, constitutes the three areas of the ontology engineering process. The basic idea is to assimilate OBMs to organizations, ontologies to products, and DEs and KEs to employees of an enterprise. The perspectives already listed in the organization context, respectively, correspond to *methodology engineer*, *processes for ontology building*, *innovation and learning*, and *ontology user* in the OB context.

3 THE EVALUATION METHOD

In this section, BSCs in the OB context are analysed. Based on the assertion that you cannot control what you cannot measure, authors tried to refer to metrics as more objective as possible to support ontology modeller in evaluating an OBM. Since BSCs should consist of a linked series of objectives and measures that are both consistent and mutually reinforcing, in few cases, the same metric has been used to assess different but tightly coupled objectives.

3.1 The "Methodology Engineer" Perspective

This perspective addresses the problem of assessing whether an OBM adds value to company adopting it and, consequently, to whom have designed it. This evaluation is left to ontology engineers (OEs), a team of KEs and DEs, executing OBM tasks in order to build an ontology.

The first objective, ontology engineers' satisfaction, is measured using a multi items ordinal scale, as Likert's one (Likert, 1932), widely used to measure attitudes, opinions, and preferences. The adopted scale is constituted by a set of statements, with specific format features, related to the explanation of the methodology process steps, the provided knowledge resources (e.g., manuals, training material, procedures, etc.), and the functionalities supporting OBM development. The agreement of the individual to the value statement is assessed by grades anchored with consecutive integers. Ontology engineer satisfaction is thus measured by the *ontology engineer satisfaction overall score* (*OESOS*) based on the arithmetic mean of the response levels for the statements of the scale.

Resources optimization objective concerns time, financial, and human resources (KEs and DEs) involved in OB process. These resources are tightly coupled with the specific ontology to be realized and with the selected OBM. Metrics to be considered are *knowledge engineer-month effort* and *domain expert-month effort (KEE* and *DEE)*, representing the amount of time that KEs and DEs spends in OBM implementation (Paslaru, 2006).

3.2 The "Process" Perspective

The process perspective addresses the simplicity and efficiency of ontology building processes.

The first objective to reach is the degree of simplicity of methodology implementation. The related metrics are: *ontology building needed time*

(T_{OB}), measuring time spent, in man-month, by modellers in ontology development and maintenance activities; *methodology granularity (MG)* quantified by the number of methodology process steps, and *degree of details of methodology process steps (DoD). DoD* is "*an important aspect in evaluating whether a methodology covers a particular process step*" (Dam, 2004). It will range in a scale from very high to very low judging how well the process steps are identified, explained (e.g., whether examples are provided), and elaborated.

The second objective is the requirements capture excellence and the associated metric is *competency questions compliance (CQC)*. Competency questions (CQs) are questions, at a conceptual level, an ontology must be able to answer (Grüninger, 1995). They are essentially identified through interviews with DEs and ontology users brainstorming. *CQC* is measured by the ratio of number of answered CQs and total number of CQs.

The third objective is the methodology adaptability. The correspondent metric is the *domain applicability (DA)*, quantified by number of different domains in which the methodology can be applied. Pointedly, historical experience would be preferable to subjective judgements, unless collecting a significant number of expert's judgements.

The fourth objective is the reuse of existing knowledge bases and information, measured by the amount of *imported concepts (IC)*, *imported properties (IP)*, and *imported relations (IR)*. These values should be written in percentage terms with respect to existing ontology.

The fifth objective is methodology consistency. It is measured by the *contradictions count (CC)*, i.e., the number of contradictions detected in the methodology implementation.

The last objective is ontology quality. Here we consider both syntactic and semantic aspects of ontology quality. According to (Burton-Jones, 2005), the former aspect measures the quality of the ontology according to its formal style, the way it is written, while the latter aspect concerns the absence of contradictory concepts. Concerning syntactic ontology quality, the metrics to be used are *lawfulness (La)* and *richness (Ri)*. *La*, the degree of compliance with ontology language's rules, is assessed by the total number of syntax error reported in the ontology. *Ri*, referring to the proportion of modelling constructs (classes, subclasses, and axioms, or attributes) which have been used in the ontology, is assessed by the number of different modelling constructs. The higher is this number, the richer is the ontology. The semantic ontology quality metrics are *ontology consistency (Co)* and *ontology clarity (Cl)*. *Co* is checked by using a reasoner, such as Racer (Haarslev, 2001) or Pellet (http://www.mindswap.org/2003/pellet). This task is mainly performed by KEs, since the use of a reasoner requires technical skills. Besides the absence of contradictions, semantic quality also requires modelling constructs are correctly used (e.g., absence of cycles in the specialization hierarchy or the fact that classes and properties are disjointed) (Ide, 1993). Therefore consistency is assessed by the reasoner's result: true or false. *Cl* evaluates whether the context of terms is clear: an ontology should include words with precise meanings and should effectively communicate the intended meaning of defined terms (Gruber, 1993). The metric is assessed by the ratio of the total number of word senses and the total number of words in the ontology.

3.3 The "Innovation and Learning" Perspective

This perspective analyses whether people involved in OB activities have the adequate competencies and skills to perform the work and whether a certain degree of self-learning and capabilities improvement is allowed.

The first objective is personnel capabilities optimization (Boehm, 2000), representing both ability and efficiency required to each single actors involved. The capabilities are measured by *professional/technical interest (Q_{PTI})*, and by *teamworking and cooperation ability (Q_{TCA})*.

The second objective is personnel experience optimization. It is related to the required experience of KEs and/or DEs in conceptualizing a specific domain and using the selected OBM and its supporting tools. It can be measured by *communication skills (Q_{CS})*, *experience in using the OBM (Q_{EM})*, *experience in using supporting tools (Q_{EST})*, and *knowledge of domain (Q_{KD})*. Differently from Q_{TCA}, metric Q_{CS} considers the ability of DEs and KEs in interacting and interoperating among them.

The third objective is the OBM flexibility; the associated metrics are: *methodology customization (MC)*, *repair/cost ratio (RCR)*, and *self-learning capacity (SLC)*. *MC,* i.e., the capability of OBM in adapting to new, different, or changing requirements, is assessed by the percentage of customizable steps. *RCR* measures, ex post, the cost (in man-month) required to search and repair methodology defects detected and reported by ontology users. Finally, *SLC* metric addresses the methodology attitude in pushing OEs to implement self-learning functions and to improve methodology development process through a feedback process with methodology engineers.

The last objective addresses the supporting tools accessibility, i.e., the availability and usability of tools during the OBM development process. This objective can be measured by the *supporting tools coverage* on OBM *(STC)*, namely, the percentage of OBM development's steps covered by supporting tools, and by the *quality of supporting tools* (Q_{ST}), ranging from excellent to inadequate

3.4 The "Ontology User" Perspective

This perspective addresses end-users satisfaction with respect to the built ontology and its quality. The quality of ontology is a multidimensional feature and should be evaluated with respect to different characteristics (Burton-Jones, 2005). Besides the above discussed semantic and syntactic quality, the objectives to be also considered are: ontology user satisfaction, ontology social quality, ontology pragmatic quality, and ontology extendibility.

Ontology user satisfaction has been assessed by the *ontology user satisfaction overall score (OUSOS)* based on the response levels of a five-grade Likert's scale. The scale addresses the ontology completeness, its terminology consistency with general usage, and its ability to cover the domain it claims to cover.

The ontology social quality reflects the fact that ontologies exist in communities. It is measured by *authority (Au)*, i.e., the number of ontologies that link to it by defining their terms using its definitions, and *history (Hi)*, i.e., the total number of times the ontology is accessed (when public) from the internal or the external of the community managing it.

The ontology pragmatic quality refers to the ontology content and users' usefulness, regardless of its syntax and semantics. It is assessed by *fidelity (Fi)*, *relevance (Re)*, and *completeness (Com)*. *Fi* concerns whether claims an ontology makes are "true" in the target domain. It is measured by the ratio of number of terms due their description to existing trustable references and the total number of terms. *Re*, checked in conjunction with *Com*, assesses the correct implementation of ontology's requirements. This metric can be assessed by performing two tests (De Nicola, 2009). The first test concerns the *ontology coverage (Cov)* over the application domain. A DE is asked to semantically annotate the UML diagrams, modelling the considered scenario, with the ontology concepts. The second test concerns the CQs and the possibility to answer them by using the ontology content. The metric *competency questions compliance (CQC)* can be again used for this test.

In dynamic environments such as business one, ontology's usefulness highly depends on its extendibility (i.e., whenever new concepts can easily be accommodated without any changes to the ontological foundations) (Geerts, 2000). This objective can be assessed by *ontology extendibility score (OES)*, ranging from very high to very low.

4 CASE STUDY

In this section the application of the proposed BSC-based method to UPON OBM is illustrated. The built ontology represents the knowledge underlying the exchanged eBusiness documents in the Procurement domain. UPON is an incremental methodology for OB, developed along the line of the Unified Process, a widespread and accepted method in the software engineering community.

The application of the method is demonstrated by evaluating each metric of each perspective previously described. Most of the metrics were based on human judgments and thus were evaluated by means of interviews with the group of experts involved in the OB process (i.e., two KEs, two DEs, and two ontology users).

According to Table 1, UPON fits well in the methodology engineer perspective. Since both human and financial resources optimizations are reached, UPON adds value to its designers. Furthermore, both KEs and DEs are satisfied by the methodology development process.

In the process perspective (Table 2), the values of all the metrics respect the predefined targets. Although the methodology process steps are effectively and efficiently performed additional examples and explanations will increase the value of *DoD*. Note that *IR* metric is really far from target: the number of imported relations has to be increased.

The analysis of the innovation and learning perspective (Table 3) shows that personnel capabilities and experience do not completely accomplish the targets: an improvement of their capabilities and skills has to be pursued. Since the Athos ontology management system (http://leks-pub.iasi.cnr.it/Athos) covers only 70% of process steps, improvement of the coverage of supporting tools is also needed.

In the ontology user perspective (Table 4), the objective "ontology social quality" is not reached mainly because the developed ontology is not public and external actors can not access it.

Table 1: The methodology engineer perspective for the eProcurement application.

Objective	*Metric/Value*	*Target*	*Initiative*
OE satisfaction	*OESOS*=3,4	Close to 4	*Not needed*
Resources optimization	*KEE*=2	Smaller is better	*Not needed*
	DEE=2	Smaller is better	

Table 2: The process perspective for the eProcurement application.

Objective	*Metric/Value*	*Target*	*Initiative*
Degree of simplicity of methodology implementation	T_{OB}=2 man-month	Smaller is better	*Not needed*
	MG=16	Range [10,25]	
	DoD=medium	Very high	Provide more examples
Requirements capture excellence	*CQC*=0,9	Close to 1	*Not needed*
Methodology adaptability	*DA*=2	Bigger is better	*Not needed*
Reuse of existing internal and external KB and information	*IC*=0,8	Close to 1	*Not needed*
	IP=0,8	Close to 1	
	IR=0	Close to 1	Increase imported relations
Methodology consistency	*CC*=0	Smaller is better	*Not needed*
Syntactic ontology quality	*La*=0	Smaller is better	*Not needed*
	Ri=4	Bigger is better	
Semantic ontology quality	*Co*=True	True	*Not needed*
	Cl=1	Close to 1	

Table 3: The innovation and learning perspective for the eProcurement application.

Objective	*Metric/Value*	*Target*	*Initiative*
Personnel capabilities optimization	Q_{PT}=Medium	Very high	Personnel capabilities improvement
	Q_{TC}=Medium	Very high	
Personnel experience optimization	*CS*=Medium	Very high	Personnel experience improvement
	Q_{EM}=Very low	Very high	
	Q_{EST}=Very low	Very high	
	Q_{KD}=High	Very high	
Methodology flexibility	*MC*=0,8	Close to 1	*Not needed*
	RCR=Not Avail.	Smaller is better	
	SLC=High	Very high	
Supporting tools accessibility	*STC*=0,7	Close to 1	Supporting tools coverage improvement
	Q_{ST}=Good	Excellent	

Table 4: The ontology user perspective for the eProcurement application.

Objective	*Metric/Value*	*Target*	*Initiative*
Ontology user satisfaction	*OUSOS*=3,25	Close to four	*Not needed*
Ontology social quality	*Au*=0	Bigger is better	Ontology publication
	Hi=0	Bigger is better	
Ontology pragmatic quality	*Fi*=1	Close to 1	*Not needed*
	Cov=82%	Close to 100%	
	CQC=0,9	Close to 1	
	Com=139	Bigger is better	
Ontology extendibility	*OES*=High	Very high	*Not needed*

5 CONCLUSIONS

The positive features of this method are motivated by the following considerations.

The method focuses on different perspectives. In fact, there is not a unique way to correctly model a domain, but there are always several alternatives depending on several aspects (e.g., objectives of ontology users, skills of OEs, available economical resources, etc.).

The proposed method is supported by detailed usage procedures, relies on modellers' knowledge (by means of with DEs and KEs involved in the building process), and specifies quantitative and qualitative measurements. The former measures assure more objectivity whereas the latter involve matters of perception (i.e., human judgements based on the experience of OEs and ontology users).

The proposed evaluation method grounds on a benchmarking process, it is based on the quality of results (defect detection), and it also considers how to improve them (defect correction). This allows future improvements of the methodology.

The idea presented provides a ready-on-hand procedure for ontology developers to assess different methodologies. As future work, we intend to adopt the BSCs-based method to evaluate other OBMs and to compare them. The benchmarking results will support ontology engineers in selecting the most appropriate OBM for a particular application.

REFERENCES

Boehm, B.M., Abts, C., Chulani, S. (2000). Software development cost estimation approaches–A survey. In *Annals of Software Engineering*, 10, 177-205.

Buglione, L., Abran, A., Meli, R. (2001). How functional size measurement supports the balanced scorecard framework for ICT. In *Proceedings of European Conference on Software Measurement and ICT Control*, 259-272.

Burton-Jones, A., Storey, V.C., Sugumaran, V., Ahluwalia, P. (2005). A semiotic metrics suite for assessing the quality of ontologies. In *Data & Knowledge Engineering*, 55(1), 84-102.

Chimienti, M., Dassisti, M., De Nicola, A., Missikoff, M. (2006) Benchmarking Criteria to Evaluate Ontology Building Methodologies. In *Proceedings of the EMOI-INTEROP'06* Luxembourg.

Corcho, O., Fernández-López, M., Gomez-Perez, A. (2003). Methodologies, tools and languages for building ontologies. Where is the meeting point? In *Data & Knowledge Engineering*, 46(1), 41–64.

Dam, K.H., Winikoff, M. (2004). Comparing Agent-Oriented Methodologies. In *Agent-Oriented Information Systems (AOIS 2003)*, Springer LNAI 3030, 78–93.

De Nicola, A., Navigli, R., Missikoff, M. (2009). A software engineering approach to ontology building. In *Information Systems*, 34 (2), 258-275.

Fernández-López, M. (1999). Overview of Methodologies for Building Ontologies. In *Proceedings of the IJCAI-99* Stockholm, Sweden.

Geerts, G.L., McCarthy, W.E. (2000). The Ontological Foundations of REA Enterprise. Information Systems. Annual Meeting of the American Accounting Association, Philadelphia, PA, USA.

Gruber, T.R. (1993). A translation approach to portable ontology specifications. In *Knowledge Acquisition, 5(2)*, 199–220.

Grüninger, M., Fox, M.S. (1995). Methodology for the design and evaluation of ontologies. In *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing.*

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

Haarslev, V., Möller R. (2001). Description of the RACER system and its applications. In *Proceedings of the International Workshop on Description Logics*.

Hakkarainen, S., Strasunskas, D., Hella, L., Tuxen, S. (2005). Choosing Appropriate Method Guidelines for Web-Ontology Building. In *ER 2005*. Lecture Notes in Computer Science, 3716, 270-287.

Holsapple, C.W., Joshi, K.D. (2002). A collaborative approach to ontology design. In *Communications of the ACM*, 45(2), 42-47.

Ibáñez, M. (1998). Balanced IT Scorecard Generic Model Version 1.0. In *Technical Report of European Software Institute*, ESI-1998-TR-009.

Ide, N., Véronis, J. (1993). Extracting knowledge bases from machine-readable dictionaries: have we wasted our time? In *Proceedings of KB&KS workshop*, Tokyo, 257–266.

IEEE (1996). IEEE Std. 1074-1995 Standard for Developing Software Life Cycle Processes. Standards Coordinating Committee of the IEEE Computer Society, April 26, 1996, New York, USA.

Kaplan, R., Norton, D. (1996). The balanced scorecard: translating strategy into action. Harvard Business School Press, Boston Massachusettes.

Likert, R. (1932). A technique for the measurement of attitudes. In *Archives of Psychology*, 140, 1–55.

Noy, N.F., McGuinness, D.L. (2001). Ontology Development 101: a guide to creating your first ontology, Knowledge Systems Laboratory, Stanford.

Paslaru Bontas Simperl, E., Tempich, C., Sure Y. (2006). ONTOCOM: a cost estimation model for ontology engineering. In *Proceedings of ISWC*.

Sure, Y., Studer, R. (2002). On-To-Knowledge Methodology. *On-To-Knowledge deliverable D-17*, Institute AIFB, University of Karlsruhe.

Swartout, B., Ramesh, P., Knight, K., Russ, T. (1997). Toward Distributed Use of Large-Scale Ontologies. In *Proceedings of the Symposium on Ontological Engineering of AAAI*, Stanford, California, 138-148.

Tempich, C., Pinto, H.S., Staab, S. (2006). Ontology engineering revisited: an iterative case study. In *Proceedings of the European Semantic Web Conference ESWC*, Springer LNCS 4011, 110–124.