ISSN 1392 – 124X, ISSN 2335 – 884X (online) INFORMATION TECHNOLOGY AND CONTROL, 2013, Vol. 42, No. 2

An Approach for Extracting Business Vocabularies from Business Process Models

Tomas Skersys[1,2], Rimantas Butleris[1,2], Kestutis Kapocius[1,2], Tomas Vileiniskis[2]

[1] *Department of Information Systems, Kaunas University of Technology, Studentu str. 50-309, Kaunas, Lithuania*

[2] *Center of Information Systems Design Technologies, Kaunas University of Technology, Studentu str. 50-313a, Kaunas, Lithuania*
e-mail: {tomas.skersys, rimantas.butleris, kestutis.kapocius, tomas.vileiniskis}@ktu.lt

crossref http://dx.doi.org/10.5755/j01.itc.42.2.2310

Abstract. Being a part of business process management (BPM) life cycle, business process modeling has found its place in information systems development (ISD) practices as well. At the same time, concepts of business vocabularies and rules are also the hot topics among BPM and ISD practitioners and academics. Nevertheless, in ISD, the integration of business process models with business vocabularies and rules is still not standardized and remains quite empiric. In this paper, basic aspects of the approach for business vocabularies' extraction from business process models are presented. The approach is based on novel business level OMG standards "Business Process Model and Notation" (BPMN) and "Semantics for Business Vocabularies and Business Rules" (SBVR), thus contributing to OMG's vision about Model-Driven Architecture (MDA) and to model-driven development in general.

Keywords: business vocabulary, business vocabulary extraction, business process diagram, SBVR, BPMN, VeTIS tool.

1. Introduction

From business process management (BPM) perspective, management of business vocabularies (BV) and business rules (BR) as well as business process (BP) modeling are interpreted as an integral part of BPM life cycle. Lately, information systems development (ISD) became yet another discipline, in which BP modeling and some aspects of business rules approach found their practical application.

Today, many come to an agreement that BPM and BR management (BRM) should not be treated as competing but rather as complementary technologies. Due to the need of ensuring the maximum flexibility and configurability of computerized information systems and BP execution/monitoring solutions, business rules are now acknowledged as a cr itical component in BPM and ISD activities. They also greatly contribute to organizational communication and other business knowledge exchange and transformations-oriented activities [14], [15], [24], [25], [29].

In support to its vision about Model-Driven Architecture (MDA [11]), Object Management Group (OMG) has also contributed to the standardization of business modeling discipline by providing such business modeling-level standards as "Business Process Model and Notation" (BPMN [1]) and "Semantics for Business Vocabularies and Rules" (SBVR [22]). Despite these facts, integration among different interrelated concepts (aspects) of the whole Business model itself still remains formally undefined, and the process of developing such model holds empiric manner. At the very core, all MDA-compliant OMG standards are based on the same meta-meta-model, called Meta Object Facility (MOF), however, none of the current OMG developments define how these business-level standards interrelate and link to each other. Any process-related concepts are out of scope of SBVR specification, and from the other side, BPMN has very poor support for the concepts related to business vocabulary and rules. Indeed, today, one of the main concerns of the professionals working in the area of ISD and using contemporary CASE tools is the lack of efficient, standards-based approaches that allow the development of business process models synchronized with formalized, well-structured business vocabularies and rules (BV&R) specifications.

Well-defined and practical "Business process model ↔ Business vocabulary & rules" integration

methods (BP↔BV&R; here, "↔" denotes bi-directional integration) as well as their implementations could be of great use to both ISD and BPM communities. The main driving factors for the full-scale of BP↔BV&R integration are: to let business experts define business vocabularies and rules by themselves and at the same time to ensure that such business knowledge is formal and integrated enough to pass them to the next stages of BPM (e.g., BP execution, simulation, monitoring); to accelerate the development of high quality business models in ISD projects.

In 2013, we started the European Social Fund (ESF) supported R&D project „Integration of Business Processes and Business Rules on the Basis of Business Semantics" (VEPSEM) at Kaunas University of Technology. At the end of this project, one of the work-packages will result in a certain method and tool to facilitate the BP↔BV&R integration process in the area of ISD, and also BPM. The scope of this paper is limited to just one aspect of such BP↔BV&R integration, i.e. semi-automated one-way extraction of SBVR business vocabularies from BPMN business process diagrams (BP →BV). Along with other VEPSEM research results (e.g. [24]), the results presented in this paper will contribute to the overall BP↔BV&R integration method and its implementation in a CASE tool *MagicDraw UML*.

Further in this paper, an overview of the related work is presented in Section 2. Section 3 deals with the basics of BPMN and SBVR standards. Arguments on why to use those standards for the purposes of this research are also presented. In Section 4, a BP→BV approach is presented by showing the mappings between the elements of BPMN and SBVR; some prototype implementation details are also given followed by the description of the extraction algorithm. Results of the experiment and the evaluation of the approach are presented in Section 5, and conclusions – in Section 6..

2. Related Work

Most of the research carried out in the area of BP↔BV&R integration deals with the development of business process models from BV&R specifications, which is the opposite direction of transformation (or extraction) to what the scope of this paper is. Nevertheless, these works are also relevant to the subject as they contribute to general understanding of this complex research area and also give certain insights and ideas towards our contribution. In 2008, a multi-disciplinary network of research excellence OPAALS ("Open Philosophies for Associative Autopoietic Digital Ecosystems") provided a D2.2 deliverable "Automatic code structure and workflow generation from natural language models" [3]. The deliverable presents six scenarios of model-to-model transformation for the workflow and code structure generation out of SBVR specifications. Among those scenarios, there is one about the development of BPMN/XPDL models using SBVR business vocabularies and rules, however, only the very basic transformations are provided in the document. Raj et al. [18] and Steen et al. [26] also investigate the automated support for transformation of SBVR-based BV&R specifications to BPMN-based BP models. These approaches focus on modeling correct sequences of activities in a business process model. Business rules are constructed using "IF <condition> THEN <action>" template, which is useful for modeling sequences of activities. However, such a solution restricts the expressiveness of SBVR itself. Steen et al. [26] also address business process model optimization issues.

An extensive research on business processes' specification using business vocabularies and rules was carried out by the EM-BrA2CE (Enterprise Modeling using Business Rules, Agents, Activities, Concepts and Events) project [5]. EM-BrA2CE is a framework unifying vocabulary and execution model for declarative process modeling. The vocabulary is specified using SBVR standard and the execution model is presented as a Colored Petri Net (CP-Net). Authors of this research state that the framework is intended to be a foundation for integrating and developing all forms of declarative business process modeling. However, being dedicated to a declarative BP modeling, the framework does not support functional and operational perspectives; also, it does not use any graphical business process modeling language (i.e. has no graphical interpreter) to visualize BP models in an appropriate graphical form.

There is a number of research works being carried out, which use SBVR specifications to develop other information technology artifacts, e.g. web services [9], SPARQL and SQL queries ([27] and [12], respectively). Also, some approaches for the extraction of SBVR rules from legacy systems (program code) [16] and natural language text [6] [10] are present. All these works once again confirm the relevance of our research to the broad area of SBVR applications; however, they do not have direct impact on the goals of this paper.

Based on the above analysis, it is quite safe to presume that despite the growing number or SBVR research works and applications, the extraction of SBVR specifications from the existing BPMN business process models is researched very poorly. Indeed, some argue that business vocabularies, let it be business object models, facts models or textual specifications, are defined before the development of business process models – this is a common rule, however, not a universal one. In the real world, there are cases when BP models exist without any formal business vocabularies – the need for such vocabularies arises only when larger BPM or ISD projects are initiated. Also, such business vocabularies extraction approach contributes to the overall BP↔BV&R integration method. Currently, only Malik et al. [8]

proposed some guidelines of the approach to business rules generation from BPMN business process models. The authors of this paper also indicated that there were no other published research works contributing to this exact area of research.

3. BPMN and SBVR: the Basics and Argumentation

3.1. Business Process Model and Notation (BPMN)

Business process modeling is a common technique, which is widely used within organizations to identify, accumulate, structure, visualize and perceive the knowledge behind business process and to use that knowledge for very different purposes within the life cycle of BPM as well as ISD. Business process models are not only the basis for process simulation and optimization, but they also increase the understanding of business itself, its goals, allow to prepare technical requirements if necessary, experiment with the integration of new concepts in BPM, etc. [7].

Definition of BPMN. The primary goal of Business Process Model and Notation (BPMN) [1] is to provide a notation that is understandable by various people with very different expertise: business analysts who model real world business processes, IT experts who are responsible for the implementation of processes, and finally business people who manage and monitor these processes or work according to the given instructions (expressed via BP models).

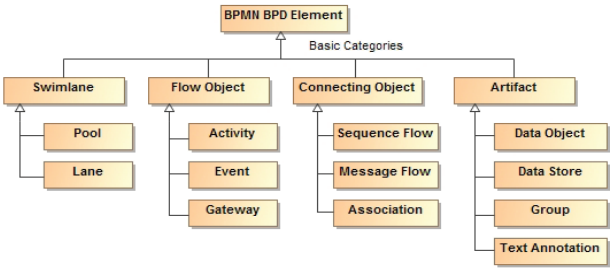

Figure 1. Classification of the core elements of BPMN business process diagram

The main diagram of BPMN is a Business Process Diagram (BPD), which allows the definition of process, its execution circumstances, some simple constraints, responsible actors and also to simulate the modeled process [1]. There are four categories of core elements in BPD: *Swimlanes*, *Flow Objects*, *Connecting Objects* and *Artifacts* (Figure 1). It should be mentioned that all of the core elements are the primary or secondary sources of knowledge in the process of business vocabulary formation (Section 4).

Why BPMN – argumentation. Analysis of Bunge-Wand-Weber ontology [19] showed that compared to other BP modeling languages BPMN covers the largest amount of real world concepts and is well-understood and accepted by business experts [4], [13]. Judging from these results one can conclude that BPMN models provide the most of the formalized business knowledge, from which business vocabularies (and later, business rules) are to be extracted.

The survey made by Wahl and Sindre [31] also indicated that BPMN is clearly understandable and well-suited for BP modeling. According to the survey of BPMN application in companies [20], the main reasons of using this standard are: documentation of organization; redesign of the organization; knowledge management; supporting continuous process management; requirements specification for software development. Compared to BPMN, software systems modeling-oriented standards, like UML and IDEF, also lack sufficient expressiveness compared to BPMN; moreover, such languages hold a number of concepts, constructions and rules that can be treated as excessive for BP modeling needs [17].

Another argumentation to use BPMN is the fact that this standard fluently integrates into Model Driven Architecture (MDA [11]), which is a widely accepted standard by CASE tools developers and other R&D communities. Integration of BPMN into MDA is achieved via the meta-meta-model (Meta Object Facility – MOF), which is common with other OMG modeling standards, and is supported by OMG's constant efforts to integrate their standards on all levels of ISD. From MDA perspective, BPMN model is a part of Business model.

3.2. Semantics of Business Vocabulary and Business Rules (SBVR)

Business rules "mantra"[1]. In general, business vocabularies (BV) fall under the discipline of business rules management (BRM). Depending on the area of application, the concept "Business rule" has quite a number of definitions. In the context of this paper, a business rule (BR) is defined as a logical statement that defines or constrains some aspect of the business in the concrete situation [30]. A BV is defined to contain a set of "specialized terms and definitions of concepts that a given organization or community uses in their talking and writing in the course of doing business" [22].

According to so-called business rules "mantra" (followed from "Business Rules Manifesto" [28]), business rules are built on facts and facts are built on terms (Figure 2). Terms (concepts) and facts are the ones that form the basis of any business vocabulary. In its turn, one cannot properly specify and manage business rules without having a proper business vocabulary at the background. In other words, one cannot effectively manage business rules without

[1] One may argue over the use of the word "mantra" in this context; however, an expression "Business rules "mantra" is commonly used in Business Rules Community and we adopt it without further argumentation.

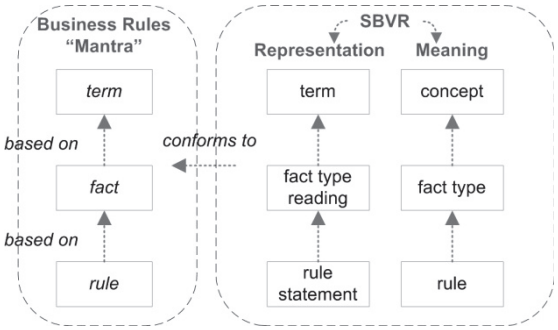

Figure 2. SBVR support for business rules "mantra" (adopted from [22])

taking a constant and proper care of business vocabularies.

Expressing business vocabularies and rules. Naturally, BR are being expressed in a variety of forms and languages depending on the area of application – procedural, executable rule languages that are used in BRM systems (such as FICO *Blaze Advisor* or IMB *Ilog*) is a good example of that. Aside from specialized areas of application, business people and even IS developers tend to express BR in natural language statements followed by some loose list of common terms and definitions. It is obvious that such unstructured business knowledge cannot be passed to the next stages of BPM/ISD life cycle or communicated unambiguously within or outside the organization. During the last decade, people from Business Rules community (http://www.brcommunity.com) such as R. G. Ross, B. von Halle and other researchers and practitioners proposed various approaches of specifying business rules in a declarative way using natural language constructions [21], [23], [30].

Definition of SBVR. By that time, OMG joined forces with many from BPM and BR communities and, in 2008, released "Semantics of Business Vocabulary and Business Rules" (SBVR) standard [22], which was welcomed by both business and IT sectors. The vision of SBVR is to express business knowledge in a controlled natural language, which would be unambiguous and understandable to business and IT people as well as computer systems.

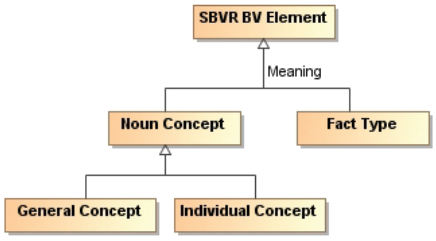

Figure 3. Classification of the core elements of SBVR business vocabulary

The complete SBVR business vocabulary meta-model contains over one hundred concepts defining various aspects of a business vocabulary. Due to the scope of this paper, this large structure can be scaled down to the very few core elements (Figure 3):

- *Noun Concept* can be specialized to *General Concept* and *Individual Concept* (and also, *Role*, which is out of scope in this paper). *General Concept* is a noun concept that classifies things on the basis of their common properties. *Individual concept* is a noun concept that corresponds to only one object (*thing*).
- *Fact Type* is a concept that denotes some type of relationship between two or more noun concepts or a characteristic of the noun concept. Following the definition, fact types are defined using the existing noun concepts, which have been already defined in BV. A *Fact Type* has a final set of specializations.

Four types of font styles with concrete formal meaning are used to *represent* noun concepts, fact types and business rules in SBVR-based Structured English or any other chosen language (e.g., Lithuanian):

- 'term' font is used to represent general concepts (object types) and **roles, e.g.** 'customer'. Terms are defined in singular form using lower case letters.
- 'Name' designates individual concepts that usually are proper nouns, e.g. 'Lithuania', 'IBM'. The first letter of a name is capitalized. One of the exceptions to the latter is the presentation of numerical values that are also shown in this style, e.g. '25'.
- '*verb*' font style represents a verb, a preposition, or a combination of these two. Verbs are used in singular active or passive forms – those are synonymous forms. For example, for the active form of an associative fact type 'customer *provides* order' there is a synonymous passive form 'order *is_provided_by* customer'. Characteristic, which is one of the subtypes of a *Fact Type*, is always used in passive form, e.g. 'customer *is_reliable*'.
- 'keyword' font represents linguistic symbols that are used to construct statements and definitions, e.g. 'each', 'It is obligatory that', 'greater then'.

SBVR business vocabulary has glossary-like *entries*, which specify concepts having representations in the vocabulary. Each entry is for a single concept. Each vocabulary entry starts with a *primary presentation* denoting a name of the concept. Additionally, a concept can be defined by other optional fields – SBVR presents more than ten of such fields, e.g. *Definition*, *General Concept*, *Concept Type*, *Synonym*. Due to the paper page limitations, we will not go into more descriptive details about the composition of an entry in this paper.

Why SBVR – argumentation. As of yet, SBVR is probably the first initiative to formalize and standardize the definition of business vocabularies suitable for wide-range practical application. Being an OMG standard, SBVR initially has a strong support from the world's BPM community, CASE tools

developers and other R&D parties. From the MDA perspective, business vocabularies and rules have to be defined at CIM level of MDA, i.e. in parallel with BP modeling. SBVR is fully integrated into the OMG's Model-Driven Architecture via MOF or Eclipse Metamodeling Framework (EMF).

A few years ago, our research group finished a state-funded R&D project *VeTIS* [2]. The main tangible result of the project was *VeTIS* tool [14], which realized automated transformations of SBVR specifications written in SBVR Structured English into SBVR 1.0 XMI format and subsequently – into the UML class models enriched with OCL constraints (in EMF UML 2.1.2 XMI). The project showed the possibilities to integrate SBVR with other OMG standards and perform business knowledge transformations in practice.

4. Extracting SBVR Business Vocabularies from BPMN Business Process Diagrams

In this section, basic principles of the BP→BV approach are presented by providing meta-model mappings of BPMN and SBVR elements, basic implementation aspects and the extraction algorithm.

4.1. Mapping the Elements of BPMN and SBVR

Theoretical research and practical experimentation with BPMN and SBVR models allowed us to identify the mapping sets of elements on meta-models' level (Table 1).

In BP→BV approach, at the initial phase, certain business rumblings are being extracted from BP diagram, and candidate business objects and expressions (O&Es) are formed (upper right GUI form in Figure 5). After that, SBVR vocabulary entries for noun concepts and fact types are being formed from these business O&Es. During the extraction of business rumblings, markings "A", "M" and "Alt_M" (Table 1) have the same meaning – they all mean that the marked BPD elements are used as sources of knowledge for automatic extraction of certain business rumblings. When all rumblings are being extracted, the meaning of the markings comes into play: "A" implies that a certain candidate business O&E can be *automatically* identified and extracted from the particular business rumbling; "M" implies *manual* (semi-automatic) identification and extraction of certain business O&E, and "Alt_M" designates the secondary (alternate) source for manual (semi-automatic) identification and extraction of one or more candidate business O&Es. For example, if the intersection of a certain BPMN element, e.g. *Lane*, with some SBVR element, e.g. *Noun Concept*, is marked with "A" then the actual business objects of the type *Lane* in any BPMN BP diagram will be the subject for *automatic* identification and extraction of certain business objects in order to form *Noun Concept* entries in SBVR business vocabulary ('customer' and 'supplier', from the BPD in Figure 5). Another example is that from any BPMN *Event*-type flow object the system will automatically extract certain business rumbling, which will be the subject for *manual* identification and extraction of certain candidate business expression in order to form a corresponding *Fact Type* (e.g. 'order *is_received*'). In order to extract the most of the business vocabulary-related knowledge from a BPD, one can also work with other diagram objects that are interactively presented by the system as alternative sources of knowledge for manual extraction and formation of corresponding BV entries – such cases are marked with "Alt_M" in Table 1.

Table 1. Mapping pairs of elements of SBVR and BPMN meta-models for BV extraction

BPMN BPD		SBVR BV	
Group	**Group Element**	**Noun Concept**	**Fact Type**
Swimlanes	Lane	A	A
	Pool	A	-
Flow Objects	Event	Alt_M	M
	Activity	Alt_M	M
	Gateway	Alt_M	M
Connecting Objects	Sequence Flow	-	-
	Sequence Flow (with Condition)	Alt_M	M
	Message Flow	-	-
	Message Flow (with ref. Message)	A	A
	Association	-	-
Artifacts	Data Object	A	-
	Data Object (with State)	A	A
	Data Store	A	-
	Group	A	-
	Text Annotation	Alt_M	Alt_M

4.2. Prototype Implementation of the Approach

On the implementation level, the proposed BP→BV approach may be viewed as an extension to one of our recent developments – *VeTIS* tool [14], which may be used as a stand-alone tool or as a plug-in of the CASE tool *MagicDraw UML* (Figure 4).

The main features of the *VeTIS* tool are:

1. Input and editing of SBVR BV&R with syntax validation;
2. Automatic transformation of SBVR specifications to UML class diagrams enriched with OCL constraints.

At this point, the main objective of the research is to extend the functionality of the *MagicDraw UML*

[2] Lithuanian State Science and Studies Foundation High Technology Development Program Project "Business Rule Solutions for Information Systems Development" (VeTIS).

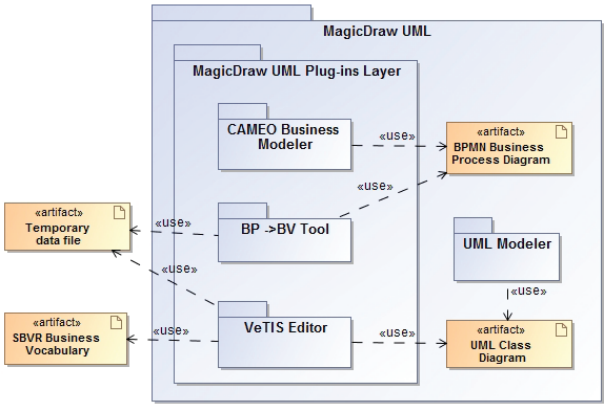

Figure 4. Top level implementation architecture of the approach in *MagicDraw UML*

tool by adding BP→BV extraction feature and thus allowing a user to work with business vocabularies in parallel to business modeling activities.

In *MagicDraw UML* environment, BP→BV extraction tool reads BPMN business process diagrams and stores the data in temporary data files (structured text format). During the formation of candidate business O&Es, temporary data files attain the format, which can then be used by *VeTIS* editor for further work with SBVR business vocabulary (Figures 4 and 5).

In Figure 5, Label 1 marks the automatic extraction of business rumblings from the modeled BP diagram; after that, candidate business O&Es are formed in the presented GUI window; Label 2 marks the formation of SBVR business vocabulary from the selected business O&Es.

MagicDraw UML fully supports BPMN-based BP modeling (via *CAMEO Business Modeler* plug-in); therefore, SBVR and BPMN modeling activities are fluently integrated under the same working environment (Figure 4). The approach uses standard MOF/EMF-based SBVR and BPMN meta-models supported by *VeTIS* and *CAMEO Business Modeler* accordingly. BP→BV extraction algorithm is to be implemented as a separate *MagicDraw UML* plug-in. In further developments, it is expected that BP→BV extraction tool will work with XMI files – this transition will be synchronized with the continuous improvement of *VeTIS* editor.

4.3. Basic Algorithm of the BP→BV Approach

On the highest level of abstraction, the algorithm of BP→BV approach is composed of three basic stages:

- *Stage 1*: Handling of business rumblings extracted from BPMN BP diagram;
- *Stage 2*: Formation of entries in SBVR business vocabulary;
- *Stage 3*: Validation of the developed overall SBVR business vocabulary with business domain expert.

Stage 1 deals with the extraction of business rumblings and formation of candidate business objects and expressions (O&Es), from which SBVR noun concepts and fact types will be formed. In *Stage 2*, the algorithm follows the basic principle of business rules

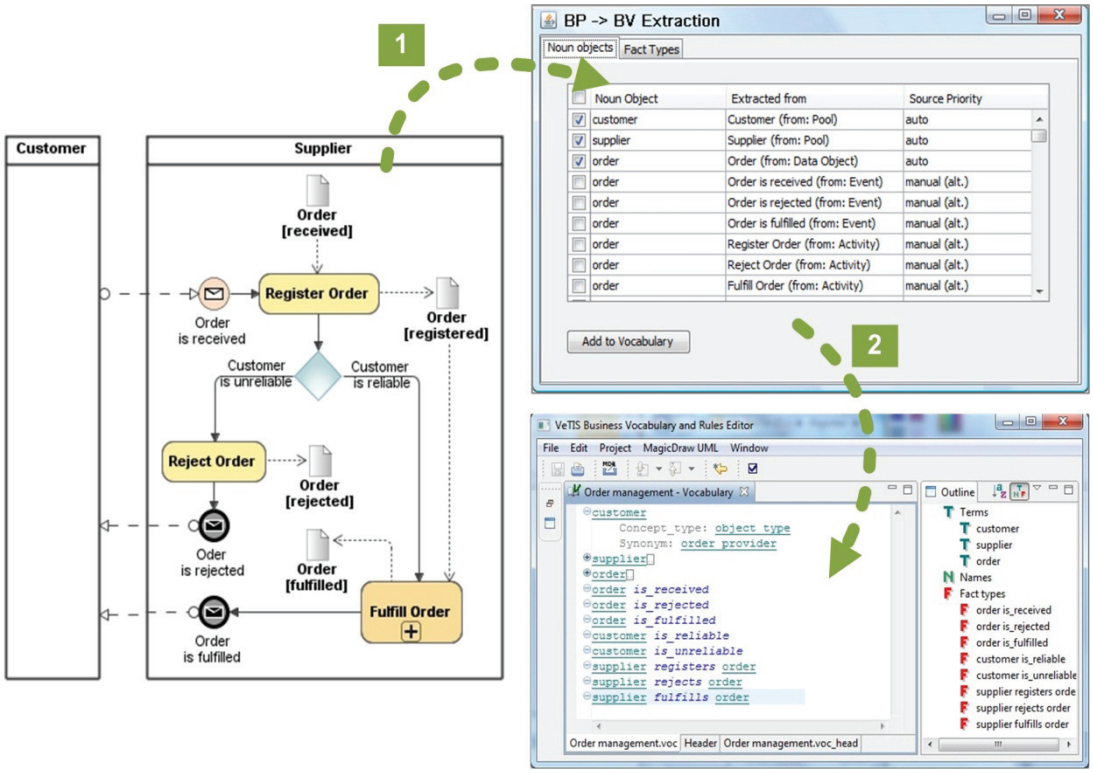

Figure 5. Three interacting components: BPMN modeler (*Cameo Business Modeler* plug-in for *MagicDraw UML*), BP→BV extraction tool (*MagicDraw UML* plug-in) and *VeTIS* editor (*Eclipse* or *MagicDraw UML* plug-in)

"mantra", which, once again, states that facts are built on terms – therefore, one must specify a set of noun concept entries before specifying fact types in SBVR business vocabulary. In this paper, *Stage 3* is assumed as a straightforward process of BV validation with business domain expert, therefore, will not be elaborated and discussed any further.

The algorithm covering all of the aforementioned stages is described in Table 2 and Figure 6.

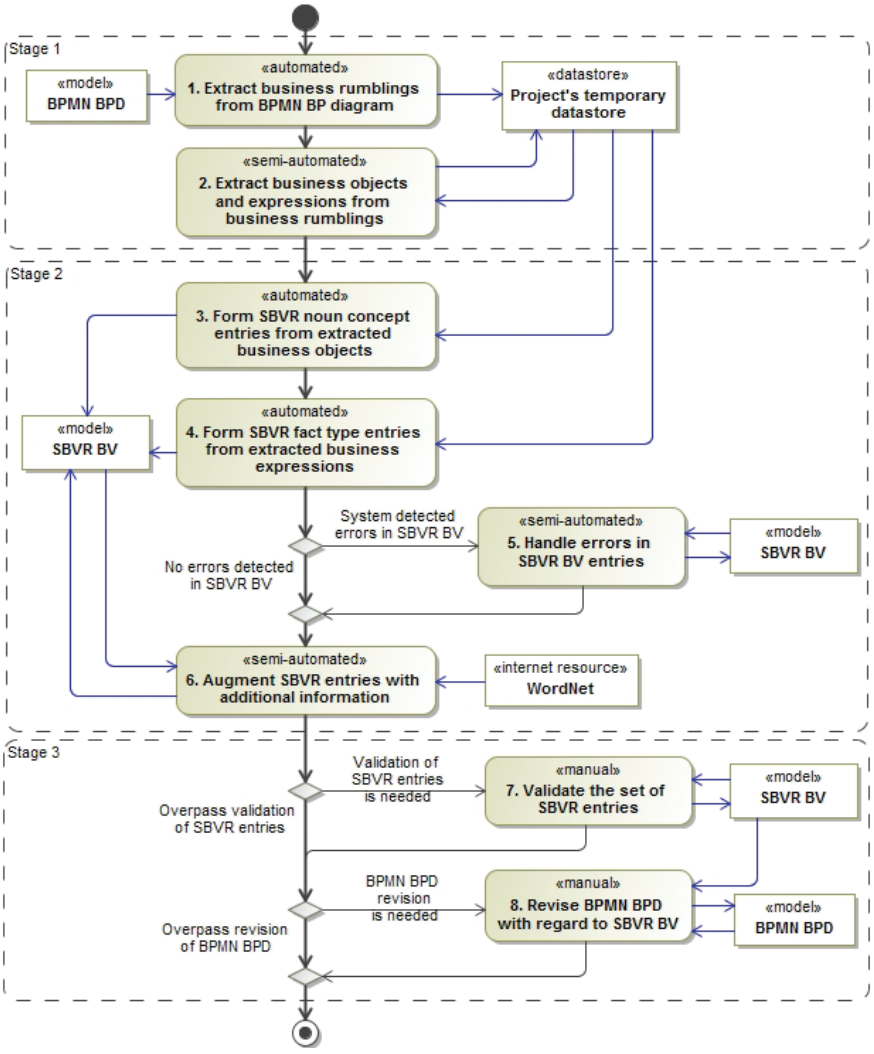

Figure 6. Algorithm of BP→BV approach

5. The Experiment and Evaluation of the Approach

To evaluate practical applicability of the approach, an experiment was run on a number of BPMN business process diagrams taken from OMG's non-normative document "BPMN by Example" [2]. The experiment and evaluation process consisted of the following steps: 1) Refactoring of the exemplary BPMN business process diagrams; 2) Application of the algorithm on the refactored diagrams; 3) Evaluation of the results. A more detailed review of this process is given below.

1) Refactoring of the exemplary BPMN business process diagrams. A common pre-condition to a successful application of any model-to-model transformation (extraction) algorithm is a sound input model. In our case, the quality of an input model directly influences the amount of manual work performed by a user during the business vocabulary extraction phase (i.e. good quality of the input model results in higher automation level in BP→BV extraction).

Arguably, business process modeling practice is not yet mature, especially, when talking about ISD projects. Using relatively new BPMN standard adds even more uncertainty to this discipline. Learning from the anti-patterns and best modeling practices is one way to improve one's process models; and there is a number of internet resources and books that provide such material. While researching on SBVR and BPMN models integration, we found out that the automated extraction of business vocabularies algorithm benefits from those best practices as well.

Table 2. Description of the algorithm of BP→BV approach (Figure 6)

Step	Description
1.	Business rumblings are extracted from BP diagram. The system performs this step automatically by taking full names of certain elements in BP diagram and presenting them as business rumblings. In *Step 1*, the extraction algorithm follows the predefined mappings as specified in Table 1 (though, notions A, M and Alt_M are ignored in this step). Examples of the extracted business rumblings from the BPD in Figure 5 are: "order is received", "order is rejected", "order is fulfilled", "customer", "supplier", "customer is reliable". Various free-text annotations and comments may also be a valuable source of knowledge, e.g.: "customer is unreliable, if he is on customer's black list".
2.	In general, a business rumbling may contain additional text, which is redundant in case of the formation of business vocabularies. Therefore, certain business O&E are extracted from the set of business rumblings. These business objects and expressions are then used for the automatic formation of SBVR BV entries (in *Step 3* and *Step 4*). In *Step 2*, the extraction of business O&Es is performed automatically or manually – depending on notions A, M and Alt_M from the mapping table (Table 1). Examples of automatic and manual extraction of business O&E from gathered business rumblings are: – *auto_extraction*(BPMN_BPD, "customer") => "customer"; – *auto_extraction*(BPMN_BPD, "order") => "order"; – *manual_extraction*(BPMN_BPD, "order is rejected") => "order"; – *manual_extraction*(BPMN_BPD, "order is rejected") = > "order is rejected". From the last two examples of manual extraction, one can notice that the same business rumbling was used for the extraction of business object "order" as well as business expression "order is rejected" – once again, manual extraction requires the involvement of some experienced expert (business analyst), who has certain knowledge in SBVR concept structures. In future developments, some steps of manual extraction might be also automated, e.g. automatic identification of business objects (nouns) and business expressions in free text business rumblings – we step into the area of NL processing here.
3.	*Step 3* is a straightforward automatic transformation of extracted business objects into SBVR noun concept entries. Each selected, non-duplicated business object is transformed into SBVR business vocabulary entry representing general or individual concept (Figure 5), e.g: customer supplier order
4.	*Step 4* is a straight-forward automatic transformation of extracted business expressions into SBVR fact type entries. Transformation of unary and binary fact types with a standard structures <term><verb> and <term><verb><term> is a straight-forward automatic process, e.g.: order is_received order is_rejected order is_fulfilled Expert assistance may be required in order to specify certain types of fact types if verbs in generated fact type expressions are not pre-associated with certain fact types, e.g. fact type phrases '*includes*' and '*is_included_in*' denote a partitive fact type. By default, any fact type, which does not fall into any other category of fact types, is declared as binary fact type (the most common type of fact types).
5.	In real life applications, the system with automatic extraction and transformation algorithms might sometimes give incomplete or even faulty results. *VeTIS* tool has an integrated SBVR syntax validation, which is used to detect syntactic errors in the generated SBVR business vocabulary. Detected errors are handled in *Step 5*.
6.	In *VeTIS* tool's environment, each SBVR business vocabulary entry may be manually augmented with additional information describing that particular concept, e.g.: customer Concept_type: object_type Synonym: order_provider *VeTIS* tool is integrated with *WordNet* internet resource (*http://wordnet.princeton.edu/*). If needed, this resource can augment BV entries with additional information, e.g. *Definitions*, *Synonyms*. In further developments, *WordNet* could be used to automatically identify synonyms in the existing business vocabularies.
7- 8.	*Step 7* and *Step 8* comprise *Stage 3* (Figure 6). These steps are processed with active participation of business analyst and business domain expert. In *Step 7*, SBVR business vocabulary is validated against the knowledge of business domain expert. To assure a feedback loop with the initial source of knowledge (i.e. BPMN BP diagram), *Step 8* is also considered in the algorithm. Depending on the situation, both steps may be highly iterative, involving modifications in both SBVR and BPMN models. The algorithm does not elaborate complex logic of these activities; however, *MagicDraw UML* and *VeTIS* editor may be successfully used here.

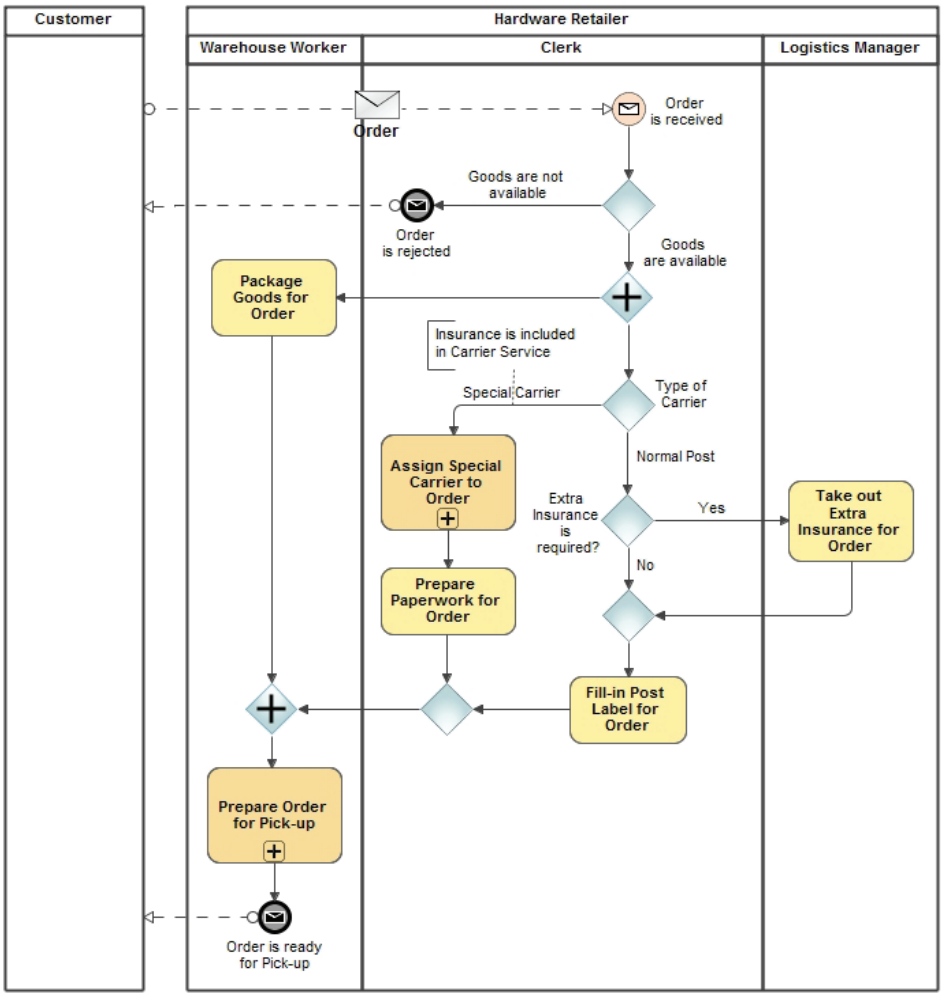

Figure 7. Refactored BPMN diagram of "Shipment Process of a Hardware Retailer" ([2], p. 3)

Below, we provide a list of main best practices for the development/refactoring of BPMN business process diagrams from the point of view of our approach:

– *Use consistent naming scheme for the elements in all diagrams*. Usually, this rule is applied on the names of *activities*; however, it is also important to formulate correct names for other named elements, which are used for the extraction of BV: *events*, *data objects*, *messages*, *gateways*, and conditioned *sequence flows*. Activity should have name pattern consisting of a strong verb and a domain specific noun; also activity name should be as short as possible (while retaining the required business information) and have no joins "and/or" in it. The occurrence of some *event* in a business process indicates that something has changed (usually it is a change of a state in some state machine); therefore, the name of the event should also indicate the fact that something has happened – a preferred name pattern for an event is a domain specific noun followed by a verb in passive form (e.g. "Client has arrived", "Order is sent"). The names for *data objects* and *messages* should be limited to domain specific nouns only. While naming *gateways*, only *exclusive gateways* should be considered. Name for an exclusive gateway can be formulated as a question (e.g. "Order is completed?") or as a noun representing a classification scheme of some business object (e.g. "Customer Status"). Outgoing *sequence flows* (branches) of exclusive and inclusive gateways should have names representing conditions, which may hold *false* or *true*; preferred name pattern for a conditional sequence flow is a domain specific noun or a noun and a verb in passive form. A noun name is used when an exclusive gateway has a name representing a classification scheme (e.g. gateway "Customer Status" has two outgoing branches: "VIP Customer" and "Ordinary Customer"); a noun with passive verb is used to name branches outgoing from unnamed inclusive and exclusive gateways.

– *Use pools and lanes in all diagrams.* This recommendation was formulated with regards to our approach. In common practice, pools/lanes may be omitted when modeling private business processes. However, from the point of view of our

approach, omitting pools/lanes from the diagrams results in reduction of automation level while formulating SBVR entries for noun concepts and fact types. E.g., when using pools, the BV extraction tool will automatically extract the candidate business expression "Order Manager registers Product Order", which then will be passed to *VeTIS* editor to form a fact type entry "order_manager registers product_order"; when pools are *not used*, the business vocabulary extraction tool would only extract "register Product Order" expression, which then would have to be augmented *manually* to a full candidate business expression "Order Manager registers Product Order".

– *Use "7+2" rule*. This rule states that no business process diagram should consist of more than 7+2 activities (others recommend to reduce this number to 5+2). Making a hierarchical business process model allows one to fragment and work with smaller (well-grained) pieces of the model. This is also good for a BV extraction as smaller fragments are easier to analyze. Using hierarchical model results in certain inconveniences while dealing with duplicates; however, *VeTIS* editor discovers duplicate entries that then are eliminated by a user.

Following the above mentioned recommendations, the selected exemplary BPMN diagrams (from [2]) were refactored and passed to the BP→BV extraction algorithm. For illustrative purposes, one of such diagrams, namely, "Shipment Process of a h ardware retailer" is presented in Figure 7.

2) Application of the algorithm on the refactored diagrams. Applied on the exemplary refactored BPMN BPD, the algorithm provided the results presented in Table 3. The results are additionally commented to add more details to a rather informal description of the extraction algorithm itself.

Table 3. Results of the application of the algorithm on the refactored exemplary BPMN BPD

SBVR Business Vocabulary	Comment
Noun Concepts	
customer	Auto-extraction from: *Pool/Lane*.
hardware_retailer	
warehouse_worker	
clerk	
logistics_manager	
order	Auto-extraction from: *Message Flow (with ref. Message)*. Alt_M extraction from: *Event*, *Activity*.
goods	Alt_M extraction from: *Sequence Flow (with Condition)*, *Activity*.
carrier	Alt_M extraction from: *Gateway*.
special_carrier General Concept: carrier	Alt_M extraction from: *Sequence Flow (with Condition)*. In *VeTIS* editor, we identified 'special_carrier' being a specialization of 'carrier'.
carrier_service	Alt_M extraction from: *Text Annotation*.
normal_post General Concept: carrier	Alt_M extraction from: *Sequence Flow (with Condition)*. 'normal_post' is also identified as a specialization of 'carrier'.
insurance	Alt_M extraction from: *Text Annotation*.
extra_insurance General concept: insurance	Alt_M extraction from: *Activity*, *Gateway*. In *VeTIS* editor, we identified 'extra_insurance' being a specialization of 'insurance' because both normal post and special carrier provide some kind of insurance.
paperwork	Alt_M extraction from: *Activity*.
post_label	Alt_M extraction from: *Activity*.
pick_up	Alt_M extraction from: *Activity, Event*.
Fact Types	
warehouse_worker *is_at* ardware_retailer	Auto- extraction from: *Lane*. Here, binary fact types are formed from the business expressions having a predefined pattern "<business object> is at <business object>". Reserved verb 'is at' is used to show the relationship between the business actors representing the pool and its lanes. In *VeTIS* editor, we added a synonymous noun form to one of the fact types for illustrative purposes. Such noun forms of fact types are usually used in business rules statements.
clerk *is_at* hardware_retailer Synonymous Form: clerk *at* hardware_retailer	
logistics_manager *is_at* hardware_retailer	
order *is_received*	M extraction from: *Event*.

SBVR Business Vocabulary	Comment
order *is_rejected*	Even though the extraction is marked as "manual", the extraction of business expressions from the business rumblings was a straight-forward task. Two unary fact types and one binary fact type was formed from those business expressions.
order *is_ready_for* pick_up	
warehouse_worker *packages* goods *for* order	M extraction from: *Activity*.
warehouse_worker *prepares* order *for* pick_up	
clerk *assigns* special_carrier *to* order	
clerk *prepares* paperwork *for* order	
clerk *fills_in* post_label *for* order	
logistics_manager *takes_out* extra_insurance *for* order	
customer *sends* order *to* clerk	Auto-extraction from: *Message Flow (with ref. Message)*. N-arry fact type is formed from the automatically extracted business expression 'customer sends order to clerk'. A predefined pattern "<business object> sends <business object> to <business object>" is used for the extraction of business expressions from message flows with referenced messages.
extra_insurance *is_required*	M extraction from: *Gateway*.
goods *are_available*	M extraction from: *Sequence Flow (with Condition)*. Only one unary fact type was formed from two business expressions: 'goods are available' and 'goods are not available', because one only shows the negation of another. When used in business rules statements, fact types are negated using predefined keyword 'not' and modal operations (e.g. 'It is impossible that').
insurance *is_included_in* carrier_service Concept type: partitive_fact_type Synonymous Form: carrier_service *includes* insurance	M extraction from: *Text Annotation*. In *VeTIS* editor, verb '*is_included_in*' denotes a partitive fact type.

3) Evaluation of the results. When applied to the refactored BPMN business process diagrams, the algorithm provided the expected results.

Automated extraction of candidate *business objects* from business rumblings was quite low because no data objects were used in the exemplary diagram. From Table 1, one can see that business rumblings extracted from swim lanes and artifacts are the primary sources for the automatic formation of candidate business objects and, consequently, SBVR noun concepts. However, one should keep in mind that the extraction of business rumblings, from which candidate business objects (and expressions) are formed, is a fully-automated task.

Currently, the extraction of *business expressions* involves more manual work by default (Table 1). However, the use of refactored diagrams resulted in relatively simple manual involvement of a user.

One could raise the level of automated formation of business objects and expressions by utilizing existing tools for lexical analysis. Currently, the main issue here is to allow the BP→BV tool to automatically identify nouns, verbs, adjectives and other parts of the language in the extracted business rumblings. The *WordNet* internet resource could be used for this purpose. We are also looking into *Stanford POS Tagge* (*http://nlp.stanford.edu/software/tagger.shtml*) as an alternative solution in our future plans.

6. Conclusions

An approach for semi-automated extraction of SBVR business vocabularies from BPMN business process models is presented in this paper.

It is safe to say that the introduction of SBVR standard has to some extent revived the business rules approach sparking interest from both researchers and practitioners. Despite that, five years on, tangible solutions on how to integrate SBVR into the IS development life cycle are still few and far apart. Based on the nature and features of SBVR, it is only natural to envision it as part of a Model-Driven Development paradigm. In this regard, it can be considered as a l ink between business analysis and information systems requirements specification and design.

Our earlier research has resulted in an industry accepted SBVR editing tool VeTIS, which bridges the gap between business vocabularies and IS design, namely, UML class diagrams. The approach presented in this paper is aimed at closing the loop by integrating business process models into the business vocabulary life cycle. Results of initial experiments with the proposed algorithm and future developments of the prototype for BP→BV extraction are promising.

Two directions for future work should be pointed out. First, the possibilities of business rules extraction from business process models must be thoroughly examined. In parallel, BP→BV&R two-way synchronization methodology must be elaborated. Certain research and development has already been done in this area [24].

Acknowledgements

The work is supported by the project VP1-3.1-ŠMM-10-V-02-008 „Integration of Business Processes and Business Rules on the Base of Business Semantics" (acronym: VEPSEM, duration: 2013-2015), which is funded by the European Social Fund (ESF).

References

[1] Business Process Model and Notation (BPMN). OMG, Version 2.0, Document Number: formal/2011-01-03, http://omg.org/spec/BPMN/2.0 (2011).

[2] BPMN 2.0 by Example. OMN, Version 1.0 (non-normative), Document Number: dtc/2010-06-02, http://omg.org/spec/BPMN/2.0/examples/PDF (2010).

[3] **R. Eder, T. Kurz, T.J. Heistracher, A. Filieri, M. Russo, et al.** D2.2 - Automatic code structure and workflow generation from natural language models, OPAALS Project (2008).

[4] **O. Glassey**. A case study on process modelling - Three questions and three techniques. *Decision Support Systems*, 2008, Vol. 44 No. 4, pp. 842-853. http://dx.doi.org/10.1016/j.dss.2007.10.004

[5] **S. Goedertier, R. Haesen, J. Vanthienen**. EM-BrA2CE v0.1: A vocabulary and execution model for declarative business process modeling. *FETEW Research Report KBI-0728, Katholieke Universiteit Leuven,* 2007.

[6] **T. H. Gomez, A. F. J. Franco.** Business Rules Extraction from Business Process Specifications Written in Natural Language. *Business Rules Journal*, 2010, Vol. 11, No. 7, http://www.BRCommunity.com/a2010/b543.html.

[7] **A. Groznik, A. Kovacic**. Business renovation: from business process modelling to information system modeling. In: *24th International Conference on Information Technology Interfaces*, 2002, pp. 405-409.

[8] **S. Malik, I. S. Bajwa**. Back to Origin: Transformation of Business Process Models to Business Rules. In: *BPM 2012 International Workshops, Revised papers. Estonia*, 2012, pp. 611-622.

[9] **A. Marinos, P. Krause.** An SBVR Framework for RESTful Web Applications. In: *International Symposium on Rule Representation, Interchange and Reasoning on the Web*, 2009, pp. 144-158.

[10] **J. L. Martinez-Fernandez, J. C. Gonzalez, J. Villena, P. Martinez.** A Preliminary Approach to the Automatic Extraction of Business Rules from Unrestricted Text in the Banking Industry. In: *13th International Conference on Natural Language and Information Systems,* 2008, pp. 299-310. http://dx.doi.org/10.1007/978-3-540-69858-6_29.

[11] Model-Driven Architecture (MDA): OMG spec., v. 2.0. http://omg.org/mda/ (2003).

[12] **S. Moschoyiannis, A. Marinos, P. Krause**. Generating SQL Queries from SBVR Rules. In: *International Symposium RuleML 2010, Lecture Notes in Computer Science*, Vol. 6403, 2010, pp. 128-143.

[13] **M. Muehlen, M. Indulska, G. Kamp**. Business process and business rule modeling languages for compliance management: a r epresentational analysis. In: *ACM International Conference*, 2007, Vol. 334, pp. 127-132.

[14] **L. Nemuraite, T. Skersys, A. Sukys, E. Sinkevičius, L. Ablonskis.** VETIS tool for editing and transforming SBVR business vocabularies and business rules into UML&OCL models. In: *16th Int. Conference on Information and Software Technologies*, 2010, pp. 377-384.

[15] **J. Nenortaite, R. Butleris**. Improving Business Rules Management through the Application of Adaptive Business Intelligence Technique. *Information Technology and Control*, 2009, Vol. 38, No. 1, pp. 21–28.

[16] **M. Osrof, R. Zaghal.** Ontology Based Business Rules Extraction Model and Algorithm (OBBREMA). In: *13th International Arab Conference on Information Technology*, 2012, pp. 66-73.

[17] **R. Paiano, A. L. Guido, A. Pandurino**. Designing Complex Web Information Systems: Integrating Evolutionary Process Engineering. In: *Information Science Reference*, 1st ed., 2009. http://dx.doi.org/10.4018/978-1-60566-300-5

[18] **A. Raj, T. V. Prabhakar, S. Hendryx**. Transformation of SBVR business design to UML models. In: *1st conference on India software engineering conference*, 2008, pp. 29-38.

[19] **J. C. Recker, M. Indulska, M. Rosemann, P. Green.** Do Process Modelling Techniques Get Better? A Comparative Ontological Analysis of BPMN. In: *16th Australasian Conference on Information Systems* 2005.

[20] **J. C. Recker.** Opportunities and constraints: the current struggle with BPMN. *Business Process Management Journal*, 2010, Vol. 16, No. 1, pp. 181–201. http://dx.doi.org/10.1108/14637151011018001

[21] **R. Ross.** Principles of the Business Rule Approach. Addison Wesley, 2003.

[22] Semantics of Business Vocabulary and Business Rules (SBVR). OMG, Version 1.0. December, http://omg.org/docs/formal/08-01- 02.pdf (2008).

[23] **T. Skersys, V. Peciulis, R. Simutis.** Business rules specification using natural language-based templates: approach and implementation. In: *14th Int. Conference on Information and Software Technologies, Kaunas, Lithuania*, ISSN 2029-0020, 2008, pp. 353-360.

[24] **T. Skersys, L. Tutkute, R. Butleris, R. Butkiene.** Extending BPMN business process model with SBVR business vocabulary and rules. *Information Technology and Control*, 2012, Vol. 41, No. 4, pp. 356-367. http://dx.doi.org/10.5755/j01.itc.41.4.2013

[25] **T. Skersys.** Business knowledge-based generation of the system class model. *Information Technology and Control*, 2008, Vol. 37, No. 2, pp. 145-153.

[26] **B. Steen, P. Ferreira Pires, M. E. Iacob.** Automatic Generation of Optimal Business Processes from Business Rules. In: *14th IEEE Int. Enterprise Distributed Object Computing Conference Workshops*, 2010, pp. 117-126. http://dx.doi.org/10.1109/EDOCW.2010.40

[27] **A. Sukys, L. Nemuraite, B. Paradauskas, E. Sinkevičius.** SBVR Based Representation of SPARQL

Queries and SWRL Rules for Analyzing Semantic Relations. In: *1st International Conference on Business Intelligence and Technology*, 2011, pp. 1-6.

[28] The Business Rules Manifesto, http://businessrulesgroup.org/brmanifesto.htm (2003).

[29] **L. Tutkute, R. Butleris, T. Skersys.** An approach for the formation of leverage coefficients-based recommendations in social network. *Information Technology and Control*, 2008, Vol. 37, No. 3, pp. 245-254.

[30] **B. von Halle.** Business Rules Applied – Business Better Systems Using the Business Rules Approach. John Wiley & Sons Inc. (2002).

Received August 2012.