BPMN* - A Notation for Representation of Variability in Business Process Towards Supporting Business Process Line Modeling

Marcelo F. Terenciani*   Débora M. B. Paiva   Geraldo Landre   Maria Istela Cagnin†

Facom – Federal University of Mato Grosso do Sul (UFMS) – Campo Grande (MS) - Brazil

E-mail: terenciani@outlook.com, {debora, geraldo, istela}@facom.ufms.br

Abstract

This paper proposes an extension for the Business Process Model and Notation (BPMN), named BPMN, that based on the elements of the feature model (FM), commonly used to represent variability, intends to represent variability in Business Processes Line (BPL). This notation is evaluated by means of an empirical study whose main objective is to compare it with other notation named variant-rich BPMN (vrBPMN) regarding the productivity and correctness of the business process model template (BPMT), which is one of the artifacts that compose a BPL. From the results it was possible to observe that the proposed notation allows the elaboration of BPMT with less errors, although modeling time was kept almost the same.*

Keywords– *Business process line modeling, variability, notation, BPMN*

1. Introduction

Due to the strong competitiveness in the globalized world, it is necessary for organizations to establish a set of improvements that make their businesses evolve every day. However, for those improvements to be proposed, it is necessary to discover and document business processes from the organizations [7]. In this context, the Business Process Management (BPM) acts as an approach that aims to identify, document, model, operate, monitor and improve business process to achieve results that are aligned with the organizational objectives [2].

The activity of business process modeling supports the BPM, once it promotes better knowledge about organizations business in order to stay competitive in the market [8]; as well as it facilitates knowledge management, since it disseminates how the business functions for all the stakeholders. Regardless of its benefits, this activity is not always held due to the associated time and costs.

Nonetheless, according to Ladeira [6], business process models can be reutilized, which makes it possible to reduced the time and effort and improve quality in the elaboration of this type of artifact since they had been previously validated and improved. From this perspective, the use of software reuse techniques in the context of business processes, as in the case of BPL, has been utilized to enable the efficient reuse of business process models.

The term BPL emerged from adaptations of concepts and experiences from Software Product Line (SPL) [10] to the context of business processes. The BPL aims to manage a set of commonalities, which are the common parts of the business domain; a set of variabilities (composed by variation point[1], and variant[2]), that can be selected to accommodate the target process; a set of rules, which explicit the task of decision making to do the flexible composition of business process assets [1].

Basically, the modeling of a BPL is composed by a set of business processes from the same domain (BPL instances); a variability model that represents "what"; and "how" the business varies; a BPMT, which represents the variabilities of business processes in one domain; and a mapping between artifacts, utilized for the traceability between them [7, 3].

Two techniques are commonly utilized to model BPL in the literature [1, 3, 11, 13]: FM [4], utilized to represent variabilities in BPL, and BPMN [9], utilized to represent instances from the BPL. For the BPMT representation there are some notations, discussed in Section 2, but there is no consensus about the most appropriate notation.

The main objective of this work is to present an extension to the BPMN notation, called here BPMN*, which adds elements to the BPMN notation based on elements from the FM, with the intent to allow the creation of BPMT with less errors, once there is need to know another new notation be-

---

[1]Places where the variation can happen (vehicle color, for example) [10].

[2]Possible existent solutions for the variation point (for example, white, black, silver and red) [10].

sides the BPMN and the FM. This was observed by means of an empirical study presented in this paper.

2. Related work

Gröner et al. [3] propose the adoption of BPMN elements to model the BPMT and the adoption of the model FM to model the variabilities of a BPL. In that study, each variability is represented in the BPMT in a execution flow of the business process. In this case, it is not possible to explicitly distinguish the commonalities and the variabilities from the BPL in the BPMT, since the BPMN doesn't have specific elements for that purpose.

In another work, Schnieders and Puhlmann [13] propose an extension for the BPMN, referred as vrBPMN, whose objective is to explicitly represent the variabilities on business process models. However, since this extension has many stereotypes, its use implies the need for the business domain engineer to know another notation, in addition to the FM and the original BPMN.

Other notations to represent variabilities in business process were proposed in the literature, such as the C-EPC [12] and the C-YAWL [5], however they are not based on the BPMN and, thus, not discussed in this paper.

3. BPMN*

The BPMN* notation consists of a extension of the BPMN, since a set of stereotypes and tagged values based on the FM were added to the BPMN notation, as well as a new element was added to the BPMN metamodel to represent a variability association, that is, a relationship between a variation point and its respective variants. The BPMN was chosen to be extended in this work because it's the standard notation to represent business processes. The FM was chosen for being commonly used to represent variabilities [3].

The intent of the BPMN* is to facilitate the modeling of variabilities in business processes of a BPL, that is, the elaboration of the BPMT. The main justification for the proposed extension is that business domain engineers don't need to have knowledge about a specific notation to represent variabilities in business processes, as in the case of the vrBPMN. Besides that, it is believed that the learning curve for the BPMN* is smaller when the business domain engineers already know the BPMN and the FM, which are commonly utilized for the modeling of BPL. That is observed in the results from the empirical study in the Section 4.

Table 1 presents the stereotypes, tagged values and the new element added to the BPMN to represent the variability in business processes of BPLs. A study to identify the BPMN elements where variability could happen was conducted based on [3, 13]. From this study, it was observed that variability can happen in the following BPMN elements: process, sub-process, activity, event, data object, pool and sequence flow. This way, the elements from the BPMN* can be used to represent variability in such elements during the modeling of variability in business processes.

Elements	Description
<<*varpoint*>>	Stereotype added to BPMN elements, in order to identify variation points, that is, where the variability happens.
<<*variant*>>	Stereotype added to BPMN elements, in order to identify variants, that is, the possible resolutions of a variation point. A variant is always associated to a variation point, through the element "Variability Association".
<<*mandatory*>>	Stereotype added to identify variation points and variants that must be resolved. This stereotype can be omitted.
<<*optional*>>	Stereotype added to identify variation points and variants that have an optional component.
<<*or*>>	Stereotype added to variability associations to identify the behaviour of the variation point. In this case, the stereotype indicates that one or more variants from the variation point should be selected.
<<*xor*>>	Stereotype added to variability associations to identify the behaviour of the variation point. In this case, the stereotype indicates that only one of the variants from the variant point must be selected.
feature	Tagged value added to variability elements to identify a correspondence between elements from the BPMT and the features from the FM.
·············▶	A variability association is an element utilized to represent an association between a variation point and its variants.

Table 1: Elements from the BPMN* notation

Figure 1a represents a model of variabilities in FM, composed by a variation point and its variants. The variation point "*Payment*" is associated with its variants "*Credit*", "*Debit*" and "*Cash*" through the relationship "*or*", which allows to select at least one of the variants of the variation point during the instantiation of the BPL.
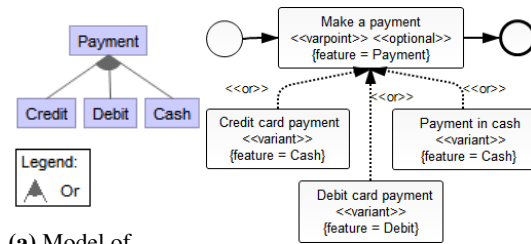
(a) Model of Variabilities in FM. **(b)** BPMT in BPMN*.

Figure 1: Representation of variation point "*Payment*".

Figure 1b is the representation of the variation point "*Payment*" in BPMN*. In this case, the stereotype <<*varpoint*>> added to the label of the activity "*Make a payment*" indicates that there is a variation point, the tagged value *feature* indicates the name of the corresponding feature, enabling the traceability between the FM (Figure 1a) and the BPMT (Figure 1b). The variants are identified with the stereotype <<*variant*>> and are associated with to their respective variation point through a *variability asso-*

ciation with the stereotype $<<or>>$ added, indicating the same behaviour from the FM.

4. Empirical Study

The objective of the empirical study is to analyze the BPMTs generated while using the BPMN* and vrBPMN notations, with the purpose of the evaluation in relation to the efficiency in terms of time spent for the elaboration of the BPMT (productivity) and the quantity of errors found in the resulting BPMT (correctness), by the point of view of business domain engineers. The hypotheses are described in Table 2.

Hypothesis	Description
$H0$	Time taken to model the BPMT utilizing the BPMN* notation is equal or greater than when using the vrBPMN notation.
H_a0	Time taken to model the BPMT utilizing the BPMN* notation is less than when using the vrBPMN notation.
$H1$	The quantity of errors made in the modeling of the BPMT utilizing the BPMN* notation is equal or greater than when using the vrBPMN notation.
H_a1	The quantity of errors made in the modeling of the BPMT utilizing the BPMN* notation is less than when using the vrBPMN notation.
	H: null hypothesis, H_a: alternative hypothesis

Table 2: Empirical study hypothesis

The participants of the empirical study are undergrad students from Computer Engineering, System Analysis, Computer Science, and Technologist in Analysis and Development of Systems courses from the Facom/UFMS.

The participants are divided in two groups balanced by background level and are composed by forty members each. The groups labeled as G-BPMN* and G-vrBPMN, used the notations BPMN* and vrBPMN respectively.

The training was held in two days. In the first day, with the two groups in the same place, an explanation was presented about the basic concepts of the BPMN and FM, with a length of three hours. For a better assimilation of the given concepts, exercises were applied and their solutions provided. In the second day of training, the groups G-BPMN* and G-vrBPMN were separated in distinct places for training about the BPMN* and vrBPMN to be ministered, together with a exercise for the fixation of the terms presented.

With the end of the training, each participant received a table containing the main elements from each notation, according to his or her group, and the correspondence from those elements to the FM; as well as supporting guidelines for the elaboration of the BPMT. They also received three instances of a BPL from the domain of rental services (Process of Video Rental, Process of Borrowing in the Library and Process of Renting Vehicle), the FM of that domain and the execution form. Twenty minutes were given for the participants so that they could understand the provided artifacts and ask any questions related to interpretation.

Even though eighty participants have answered the participant profile form, seventeen of them didn't attend in the day of the study or had given up participating, for not being a mandatory activity in the aforementioned classes. Two participants were removed for not reporting the end time of the elaboration of the BPMT in the execution form; and two were removed because the elaborated BPMT was illegible. Hence, 59 participants were considered during the analysis of the data.

First, an analysis of the data from the study was realized in order to identify outliers. To aid in the identification of outliers, graphs of the type box-plot were generated.

Based on data from the execution form, the time spent by each participant was calculated. This time was used to make a box-plot, where it was detected the absence of outliers related to the time.

A box-plot was constructed taking into consideration the quantity of errors found in the elaborated BPMTs. Hence, the participants that committed 11, 12 and 14 errors in the G-BPMN* group were classified as outliers and were removed from the sample. With the removal of outliers, the group G-BPMN* was left with 31 participants and the group G-vrBPMN with 25 participants, adding up to 56 participants.

Analysis of the data regarding the correctness of the BPMT: All of the BPMTs elaborated by the participants were analyzed with the intent of identifying the errors committed.

In Figure 2a, the participants were sorted in relation to the amount of errors found in the BPMT. It's possible to visualize that members from the group G-BPMN* committed less errors than the members of the group G-vrBPMN. Analyzing the graph it's given that from the eighteen (32,14%) participants that didn't commit errors, 77,78% of them are members of the group G-BPMN*. In relation to the highest number of errors committed, the member from the group G-BPMN* that had more errors, had 10 errors less than the member from the G-vrBPMN that committed more errors. This way, analyzing the two members that committed more errors in each group, it is observed an increase of 166,67% from the group G-vrBPMN in relation to G-BPMN*.
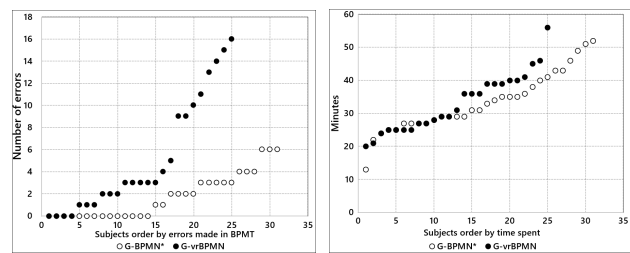
(a) Correctness of the BPMT.   **(b)** Elaboration time.

Figure 2: Analysis of the data dispersion.

Based on data from the graph presented in the Figure 2a, it's given that the average of errors committed by members of the groups G-BPMN* and G-vrBPMN is of 1.77 and 5.2 erros, respectively. With that, it is observed an average increase in errors of 193,78% for the G-vrBPMN group in relation to the G-BPMN* group.

Analysis of the data in relation to the elaboration time: In the Figure 2b is illustrated a dispersion graph of the time spent by the participants, sorted from the shorter time to the highest time. It's observed that the time spent to elaborate the BPMT with the BPMN* notation tends to be similiar to the time spent to elaborate the BPMT with the vrBPMN notation.

Additionally, it is observed that, on average, participants from the group G-BPMN* took 33.35 minutes to elaborate the BPMT, wherein the participant that took less time took 13 minutes and the one that took more time spent 52 minutes in the elaboration. And for the group G-vrBPMN, participants took 33.16 minutes on average to finish the BPMT and, in this group, the participant that finished first took 20 minutes, and the last one took 56 minutes. With this data, it can be observed that the time spent with both notations for modeling the BPMT is approximately the same, even though the group G-vrBPMN is, on average, 0,57% faster than the group G-BPMN*. However, it is noticed an increase in time of 53,8% from G-vrBPMN in relation to the G-BPMN* in regard to the participant that took less time to elaborate the BPMT.

Hypothesis Analysis The hypothesis $H0$ was accepted, since the modeling time from the members of the group G-BPMN* was approximately equal (on average, 0,57% slower) to the group G-vrBPMN. Thus the hypothesis H_a0 was rejected, given that the modeling time of the BPMT making use of the BPMN* was greater than using the vrBPMN. Due to little difference of time spent between both notations, it is noted that other studies should be done to better analyze this hypothesis. The hypothesis $H1$ was refuted, since the amount of errors made in the modeling of the BPMT making use of the BPMN* was less (193,78% on average) than when using the vrBPMN. Therefore, the hypothesis H_a1 was accepted, that affirms that the quantity of errors when using the BPMN* was smaller than when using the vrBPMN. All the artifacts used during the training and execution of the empirical study conducted in this work are available at http://goo.gl/g4V2TU.

5. Conclusion and Future Work

This paper presented the BPMN* notation, which is an extension to the BPMN for the explicit representation of variabilities in business process models, useful to support the modeling of BPL. Once that the new elements incorporated into the proposed notation are based on the FM, the learning curve for its utilization is lower, propitiating the elaboration of BPMTs with a fewer amount of errors, as observed in the empirical study presented. The results from this study also allow to observe that, on average, the modeling time using both notations was about the same for the considered business domain.

As suggestions for future works, there are: *i)* leading of other empirical studies to better analyze the hypothesis $H0$, taking into consideration business processes models from real organizations; *ii)* development of a CASE tool to support the use of the BPMN* notation, aiming to encourage its use; and *iii)* incorporate the proposed CASE tool in an real environment of BPL management with the intent of observe its benefits.

References

[1] N. Boffoli, D. Caivano, D. Castelluccia, and G. Visaggio. Driving flexibility and consistency of business processes by means of product-line engineering and decision tables. In *3rd Int. Work. on Product Line Approaches in Soft. Eng.*, 2012.

[2] H. Eriksson and M. Penker. *Business Modeling With UML: Business Patterns at Work*. 2000.

[3] G. Gröner, M. Bošković, F. Silva Parreiras, and D. Gašević. Modeling and validation of business process families. *Inf. Syst.*, 38(5):709–726, July 2013.

[4] K. C. Kang, S. G. Cohen, J. A. Hess, W. E. Novak, and A. S. Peterson. Feature-oriented domain analysis (FODA): Feasibility study. Technical report, Software Eng. Inst., 1990.

[5] M. La Rosa, F. Gottschalk, M. Dumas, and W. van der Aalst. Linking domain models and process models for reference model configuration. In *Business Process Management Work.*, volume 4928, pages 417–430. Springer Berlin Heidelberg, 2008.

[6] S. Ladeira, R. Penteado, R. Braga, and M. Cagnin. Business modeling reuse based on views: a case study. In *22nd Brazilian Symp. on Soft. Eng.*, October 2008. in port.

[7] G. Landre, E. Palma, D. Paiva, E. Y. Nakagawa, and M. I. Cagnin. vrBPMN* and Feature Model: An approach to model business process line. In *5th Int. Work. on Process Model Collections: Management and Reuse*, 2014.

[8] K. Laudon and J. Laudon. *Essentials of Management Information Systems*. Prentice Hall, 2012.

[9] OMG. Business process model and notation (BPMN).

[10] K. Pohl, G. Böckle, and F. J. v. d. Linden. *Software Product Line Engineering: Foundations, Principles and Techniques*. Springer, 2005.

[11] C. Rolland and S. Nurcan. Business process lines to deal with the variability. In *43rd Hawaii Int. Conf. on System Sciences*, 2010.

[12] M. Rosemann and W. M. P. van der Aalst. A configurable reference modelling language. *Inf. Syst.*, 32(1):1–23, Mar. 2007.

[13] A. Schnieders and F. Puhlmann. Variability mechanisms in e-business process families. In *9th Int. Conf. on Business Information Systems*, 2006.