Learning from Quality Issues of BPMN Models from Industry (Extended Abstract)

Henrik Leopold[1] Jan Mendling[2] Oliver Günther[3]

Abstract: Many organizations use business process models for documenting their business operations. In recent years, the Business Process Model and Notation (BPMN) evolved into the leading standard for process modeling. However, BPMN is complex: the specification offers a huge variety of different elements and often several representational choices for the same semantics. This raises the question of how well modelers can deal with these choices. Empirical insights into BPMN usage from the perspective of practitioners are still missing. We close this gap by analyzing a large set of BPMN 2.0 process models from practice. We found that particularly representational choices for splits and joins, the correct use of message flow, the proper decomposition of models, and the consistent labeling appear to be connected with quality issues. Based on our findings we give five recommendations how these issues can be avoided in the future. The work summarized in this extended abstract has been published in [LMG16].

Keywords: Process Model Quality, BPMN Modeling Guidelines, Modeling Recommendations

1 Introduction

Business process models play an important role for documenting business operations and for formalizing business requirements in software engineering. In recent years, the Business Process Model and Notation (BPMN) has become de-facto standard for process modeling. A major challenge of BPMN in practice concerns its complexity resulting from the considerable number of elements it offers, its sophisticated semantics, and its representational choices.

While the uptake of BPMN in practice has triggered the research community to study its usage, empirical studies on how BPMN is actually used are scarce [zMR08, MH08, Re10]. The scope of most contributions is restricted to language properties, e.g. [Re11, FMS13], instead of actual usage.

With this paper, we aim to shed light on the actual usage of BPMN, and conducted a study with six companies from industry. This way, we wanted to understand if quality issues arise and how they can be prevented. The participating companies provided us access to a total of 585 BPMN process models. We implemented an automatic guideline checker that covers rules described in BPMN textbooks [Al09, Si11, WM08]. The results helped us to learn about the frequency of different classes of modeling problems and to suggest a set of measures to overcome them.

[1] VU Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands, h.leopold@vu.nl
[2] WU Vienna, Welthandelsplatz 1, 1020 Vienna, Austria , jan.mendling@wu.ac.at
[3] Universität Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

2 A Study on BPMN Use in Industry

For our study, we collected a total of 585 BPMN 2.0 process model from six companies. The companies came from different industries and varied in size as well as in their degree of modeling experience. Using a variety of automated techniques, we developed a tool for checking a set of 35 well-known BPMN guidelines and correctness rules. This set covers in particular the guidelines proposed by Silver [Si11] and Allweyer [Al09] as well as the recommendations by White and Miers [WM08].

Figure 1 gives an overview of the 15 most frequent quality issues[3]. As indicated by the different bar colors, the quality issues can be subdivided into the three categories *structure*, *layout*, and *labeling*.

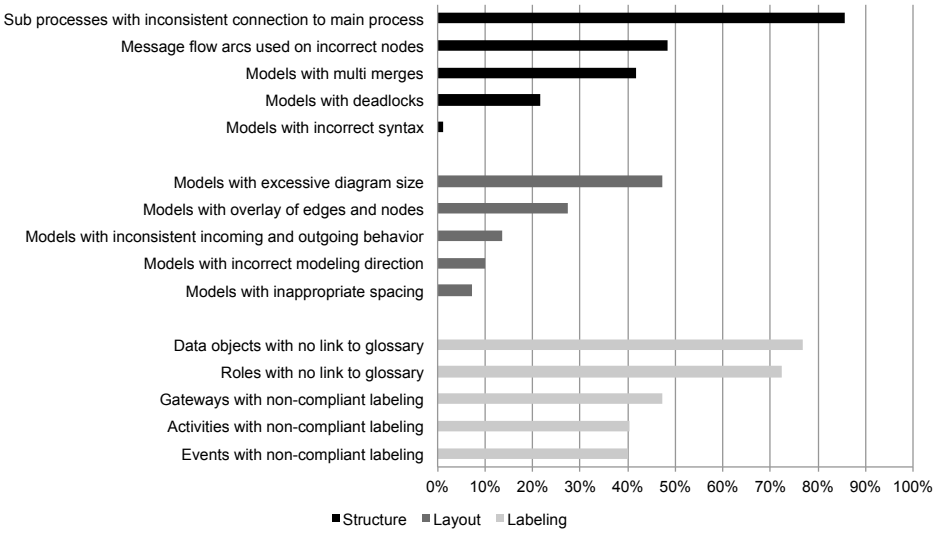

Fig. 1: Violation of modeling guidelines in practice

The *structure* category refers to the consistent and correct use of modeling elements such as activities, gateways, events, pools, and flow connectors. For this category, our study yielded mixed results. On the positive side, we found that about 99% of the investigated models are free from any syntactical errors. On the negative side, we observe that 22% of the models contain deadlocks and 42% contain multi merges. The biggest problem, however, is caused by the inconsistent association of main and sub processes. In 86% of all models containing sub processes, the roles of the sub process do not match the corresponding role of the main process.

The *layout* category is concerned with the proper positioning of the process model elements in terms of cognitive effectiveness. Hence, it is the goal of the rules and guidelines of this category to guarantee that a model can be easily read and understood. Our study shows that only a few models suffer from issues such as inappropriate spacing, arcs flow-

[3] Note that there are no overlaps among the quality issues.

ing into the wrong direction, or inconsistent incoming and outgoing behavior. Nevertheless, not all aspects are respected and implemented to the same degree. The biggest layout issue concerns models of extensive size. About 47% of all models exceed the maximum diagram size, i.e., they do not fit on a DIN A3 page.

The *labeling* category refers to the proper use of natural language in the process model. Our empirical results show that between 40% and 47% of the labels follow syntactic patterns that are potentially ambiguous and hence may negatively affect the understanding of the model [MRR10]. Moreover, glossaries are used by a fraction of modelers only. About 72% of all roles and 77% of all data objects are not linked to a glossary. As a result, an inconsistent usage of roles and data objects can be expected.

3 Five Measures for Improved BPMN Modeling

The results show that many quality and correctness criteria are well respected in the investigated organizations. However, we also observed that many advanced structural concepts, such as consistency among process models, process model size as well as the labeling of process model elements, appear to be connected with quality issues. Apparently, the available modeling recommendations and guidelines are not sufficiently clear. In the following, we discuss the five major problem areas we identified and give specific recommendations on how to avoid them.

1. *Avoid implicit splits and joins*: Implicit splits and joins via multiple outgoing and incoming arcs are the major cause for deadlocks and multi merges. This problem is caused by BPMN offering several options to represent such semantics. We therefore recommend prohibiting the use of multiple arcs. The semantics of splits and joins can be clearly and unambiguously defined using gateways.
2. *Provide tool support for proper model decomposition*: Our empirical results show that modelers may struggle with the proper decomposition of their models. Either the models are too big, or they are not fully consistent. Since both problems can be effectively enforced by a modeling tool, we recommend implementing respective mechanisms.
3. *Omit the throwing message event*: Our study suggests that message flow arcs may cause several problems. Particularly the throwing message event appeared to cause confusion. We hence recommend removing the throwing message event from the symbol set. It is easy to use activities for throwing events instead.
4. *Establish centrally maintained glossary*: The consistent reuse of central concepts such as roles and data objects is an important requirement for a sound process architecture. Thus, we propose to introduce a centrally defined glossary that either automatically monitors and imports new terms or is regularly updated by a dedicated glossary manager.
5. *Provide tool support for applying linguistic checks during the modeling process*: Achieving consistency with respect to the structural use of natural language seems

to be difficult. The most effective measure seems to communicate such inconsistencies already during the modeling process. Modeling tools could use techniques such as refactoring to automatically suggest a correct version of a non-compliant label [LSM12].

A closer look at our recommendation list reveals that particularly recommendations 1 and 3 can be traced back to the representational choices of BPMN. The modelers from the investigated organizations struggled with correctly dealing with these choices and incorporated errors that should be avoided. Our recommendations have the advantage that they do not restrict the expressive power of BPMN. Instead, they help the modeler to select a preferable representation when a specific pattern of behavior needs to be expressed. Recommendations 2, 4, and 5 refer to quality issues that may also occur in other process notations such as Event-driven Process Chains or UML activity diagrams. Still, our study demonstrates that also BPMN models may suffer from these problems. Hence, also these recommendations contribute to a consistent process architecture.

References

[Al09] Allweyer, Thomas: BPMN 2.0 - Business Process Model and Notation. Books on Demand GMBH, Norderstedt, 2nd edition, 2009.

[FMS13] Figl, Kathrin; Mendling, Jan; Strembeck, Mark: The Influence of Notational Deficiencies on Process Model Comprehension. J. AIS, 14(6), 2013.

[LMG16] Leopold, Henrik; Mendling, Jan; Günther, Oliver: Learning from Quality Issues of BPMN Models from Industry. IEEE Software, 33(4):26 – 33, 2016.

[LSM12] Leopold, Henrik; Smirnov, Sergey; Mendling, Jan: On the refactoring of activity labels in business process models. Information Systems, 37(5):443–459, 2012.

[MH08] Muehlen, Michael zur; Ho, Danny T: Service process innovation: a case study of BPMN in practice. In: Hawaii international conference on system sciences, proceedings of the 41st annual. IEEE, pp. 372–372, 2008.

[MRR10] Mendling, Jan; Reijers, Hajo A.; Recker, Jan: Activity Labeling in Process Modeling: Empirical Insights and Recommendations. Information Systems, 35(4):467–482, 2010.

[Re10] Recker, Jan: Opportunities and constraints: the current struggle with BPMN. Business Proc. Manag. Journal, 16(1):181–201, 2010.

[Re11] Recker, Jan; Rosemann, Michael; Green, Peter F.; Indulska, Marta: Do Ontological Deficiencies in Modeling Grammars Matter? MIS Quarterly, 35(1):57–79, 2011.

[Si11] Silver, Bruce: BPMN Method and Style, with BPMN Implementer's Guide. Cody-Cassidy Press, 2nd edition, January 2011.

[WM08] White, Stephen A.; Miers, Derek: BPMN Modeling and Reference Guide: Understanding and Using BPMN. Future Strategies Incorporated, 2008.

[zMR08] zur Muehlen, Michael; Recker, Jan: How Much Language Is Enough? Theoretical and Practical Use of the Business Process Modeling Notation. In: Proceedings of CAiSE. Springer-Verlag, pp. 465–479, 2008.