We still don't know how much BPMN is enough – but we are getting closer

Michael zur Muehlen[1] and Jan Recker[2]

[1] Howe School of Technology Management
Stevens Institute of Technology
Michael.zurMuehlen@stevens.edu

[2] Information Systems School, Science & Engineering Faculty
Queensland University of Technology
j.recker@qut.edu.au

Abstract. Process models expressed in BPMN typically rely on a small subset of all available symbols. In our 2008 study, we examined the composition of these subsets, and found that the distribution of BPMN symbols in practice closely resembles the frequency distribution of words in natural language. We offered some suggestions based on our findings, how to make the use of BPMN more manageable and also outlined ideas for further development of BPMN. Since this paper was published it has provoked spirited debate in the BPM practitioner community, prompted the definition of a modeling standard in US government, and helped shape the next generation of the BPMN standard.

Keywords: BPMN, process modeling, empirical research, research agenda

1 Motivation and Genesis of Paper

Process modeling is not a new phenomenon, but the notations for mapping out process diagrams seem to be in constant state of flux. Established notations are applied, refined, evolve, and are replaced with new notations. Like natural languages, process representations and their associated grammars seem to evolve. The Business Process Model & Notation (BPMN) has evolved since its inception in 2001 and has found its fair share of adopters in both modeling tool vendors, and industrial applications. Since the BPMN notation contains a large number of constructs (compared to older process notations such as Flow Charts, Petri Nets, or Event-driven Process Chains), we were interested in the question which subset of BPMN elements modelers would choose to represent models. Was the language mature? Was there a defined subset that modelers naturally gravitated toward? Both authors approached this topic from different starting points – one (Recker) from the empirical use of modeling notations, the other (zur Muehlen) from the evolution of standards over time. Both of

us expected modelers to use a problem-specific subset of the notation, but neither of us knew how large this subset would be, nor which symbols it would contain.

To address the research question, we began collecting BPMN diagrams. Both authors had worked in process modeling projects in industry, so the models generated as part of consulting engagements became a seed data set. This was complemented with models generated by students in process modeling courses, and models that were collected through Internet searches. Once we felt that we had obtained a sufficient number of models we began counting symbols, and tallied our results in a statistical software package. The results confirmed our hunch: Most models contained a small fraction of BPMN symbols, and the overall frequency distribution of symbols followed closely the exponential Zipf-curve that is indicative of the word distribution in natural languages. We concluded from this study that we can learn much about how we could and should use BPMN from our use of natural languages in different settings such as informal conversations, tech talks, essays and so forth.

2 Impact on Industry

2.1 Reception in practice: the feedback on the paper by practitioners

> *If you have even a passing interest in BPMN, you're probably aware of the great debate happening amongst a few of the BPM bloggers in the past week. […] It's worth taking the time to work your way through this debate, and keep and eye on Bruce and Michael's blogs for any further commentary.*
>
> Sandy Kemsley, Column 2, blog post 13 March 2008, http://www.column2.com/2008/03/the-great-bpmn-debate/.

One of the interesting phenomena that emerged around our paper was its reception by the BPM community of practitioners. To aid the transfer of research into practice, we decided to blog about what we believed to be main findings and implications (http://www.workflow-research.com/2008/03/03/how-much-bpmn-do-you-need/). This post led to some spirited comments and related blog posts. We were surprised by the number of commentaries and the critical feedback we received from the community, starting with Bruce Silver's challenge of the implications we laid out in our post (http://www.brsilver.com/wordpress/2008/03/09/on-how-much-bpmn-do-you-need/), and with the views of other participants that responded to this debate (e.g., http://processdevelopments.blogspot.com/2008/03/hottest-bpmn-process-modelling-debate.html).

Much of this debate was dedicated to interpreting the findings in a set of actionable implications for the community. While we had originally set out to study how BPMN *was* being used, it became clear that practitioners were interested in how BPMN *should* be used. We aspired to formulate recommendations especially for vendors and training providers; and obviously some of our arguments were deliberately challenging and provocative, in an effort to inspire certain changes around standard making, method and tool design and training development.

Some, but not all, of our recommendations and interpretations were lauded by respondents to the blogs; some responses were equally challenging and provocative, and also criticized the scientific method applied. In hindsight, we very much welcomed all the feedback we received and we still see this debate as a prime example of a healthy and fertile debate between industry and academia – especially because such conversations more often than not are absent [8].

What we learned from this episode are two things: First, making research insights more relevant requires a thorough re-write and re-publication in more accessible and readable forums, and proves to be a very worthwhile activity for academics. Second, sparking (and not necessarily winning) a debate is in itself an extremely useful activity as it sparks imagination, critical analysis and reflection – both on the side of the contributors and the recipients. We have certainly learned from this episode and continue to attempt as often as possible to convey our research not only within our scientific forums but also decisively outside of this community.

2.2 Application in practice: the US Department of Defense

After the paper generated some interest in the BPM practitioner community, we were invited to speak at industry conferences, including the 2007 Transformation + Innovation conference in Washington, DC. The CTO and Chief Architect of the US Department of Defense's Business Mission Area was the keynote speaker at this event and talked about a practical issue: The several hundred information systems in the department were documented in various proprietary languages and notations, making systems integration challenging and training onerous. Was there a way to design and implement a standard-based notation to describe the department's processes? The conference chairman facilitated a behind-the-scenes meeting, which led to an invitation to present our findings in Washington.

The brief presentation led to the initiation of a project to define the smallest usable set of BPMN constructs for the DoD's Business Mission Area, accompanied by a style guide that would help modelers develop process models in a uniform fashion. The main driver for this style guide was the disambiguation of process fragments that could be represented in BPMN in more than one way, for example branching moments that could be represented either by using a gateway or by using conditional sequence flows. The findings from our original paper guided the selection of modeling constructs, while the design of the style guide was driven by the work on workflow patterns [14].

Once the BPMN subset and patterns had been field-tested, a question arose: The available process modeling tools did not enforce the reduced symbol pallet, much less the design patterns that had been established. What would it take to get the BPM vendor community to support the effort? We began talking to the Object Management Group's BPMN Finalization Task Force.

2.3 Application in practice: Shaping BPMN 2.0 conformance classes

The original study of BPMN models was based on version 1.2 of the BPMN standard. As our work with stakeholders in industry progressed, the Object Management Group began finalizing version 2.0 of the BPMN specification. In talking to some key stakeholders in the finalization task force, namely Robert Shapiro, Bruce Silver, and

Denis Gagne, it became clear that there was appetite to group BPMN constructs into subsets to facilitate process modeling at different levels of sophistication. Bruce Silver had proposed three levels of BPMN modeling in his book [11], Robert Shapiro was representing the interests of the Workflow Management Coalition, which needed a defined subset to tailor the XPDL model interchange format to [13], and the Department of Defense had a vital interest in anchoring the newly formulated BPMN primitives in the official standard. Through a series of meetings, the elements for three BPMN conformance classes were defined: Descriptive for simple, flowchart-like diagrams; Analytic for more sophisticated models that include event handling and messaging; and Common Executable with a focus on the model attributes that a Business Process Management System would expect. The three conformance classes became part of the official BPMN 2.0 specification [12].

3.3 Conformance Classes in Practice

Now that the conformance classes were defined, they could be designated a mandatory feature of process modeling tools that could be procured by the U.S. government. Vendor briefings were held, policy was written, and after a development period of more than three years, the BPMN Analytic Conformance class was officially adopted as the process modeling standard for the Business Mission Area. Today, an increasing number of BPMN tools support the conformance classes defined by the Object Management Group. But simply providing a defined subset of symbols in software was not sufficient to ensure its proper use in practice. Training classes needed to be developed, and style guides had to be written. This work is still ongoing today.

3 Academic research on the use of process modeling notations

We have always been proud of the impact that the paper generated in industry. Still, as academics we also envisaged to leave a footprint in the body of knowledge. How do you gauge the impact of a paper on the trajectory of research in the community? A standard way of measuring impact is by means of scientometric analysis, e.g., by examining citation statistics [e.g., 1].

The 2008 CAiSE paper ranks as the third-most cited research paper on BPMN, as per Harzing's Publish or Perish (behind a paper on the semantics of the BPMN specification [2] and Steven White's guide to modeling with BPMN [10]). The paper attracted over 130 citations in the five years since its publication.

Exploring the types of research that perused our findings, we find that the research inspired research across empirical, analytical and formal dimensions, on BPMN [6], other process modeling notations [3, 7] and even other research domains such as web services [9], process mining [4] or software development, amongst others [5].

Two themes have dominated the research building on our work:

(a) How suitable is BPMN for modeling certain kinds of processes?
One way that our research was continued by our colleagues was to adopt the key finding of our study (that modelers use specialized and limited subsets of the BPMN vocabulary) and examine dedicated application scenarios – which

part of BPMN do we need when we model web service interfaces? How much BPMN do we need for software development?

(b) How do modelers learn to use BPMN?
Another vein of research has started to explore another implication of our work: if modelers use different subsets of BPMN only, how could a staged approach to learning BPMN look like? Which (sets of) symbols are easier or harder to apply, and which of the symbol characteristics are more likely to introduce modeling errors or understandability problems?

The true impact of papers on the ever-evolving body of knowledge remains to be seen in the long term. There might be studies still at the planning stage that build on, extend, challenge or dispute the findings from the 2008 paper. In whatever format this work is extended, we are hoping that the study remains a fertile ground for other academics to start thinking about BPM research, even if this means that at some stage our findings will be disconfirmed and replaced with much better theory and explanation of how much language is enough.

4 Insights

In looking back at our 2008 paper, we believe there are a number of properties of the paper – and the research it describes – that offer insights to the next 25 years of advanced information systems engineering and the wider IS community.

First, at the time of writing the content of the paper – an empirical study of the use of a notation that was predominantly subject to formal and analytical research so far – was clearly a niche topic in a densely populated subject area. Both authors continue to look for such niche topics, hoping to contribute to popular debates with a different view.

Second, we learned about the importance of complementing the scientific work with other reporting styles and formats that make the findings available to and accessible for other communities that may have an interest. Means such as blogging, essaying or presenting allow academics to deliberately and decisively address different audiences beyond academia – even if that means further work. To create practical impact from a research study may take much longer than the next publishing cycle, but it can fuel the next round of inquiry.

Last but not least, one of the most important lessons is the value of feedback, and the virtue of welcoming and working with such feedback – especially the critical type. Only this way a true debate is emerging, and only through debate can we continue to identify topics that are (a) challenging (otherwise there would be no debate) and (b) relevant (otherwise debates would not become intense and fierce).

We look forward to the next twenty-five years of advanced information systems engineering research and the lessons and challenges that this era will bring.

References

1. Chen, C., Song, I.-Y., Yuan, X., Zhang, J.: The Thematic and Citation Landscape of Data and Knowledge Engineering (1985–2007). Data & Knowledge Engineering 67 (2008) 234-259
2. Dijkman, R.M., Dumas, M., Ouyang, C.: Semantics and Analysis of Business Process Models in BPMN. Information and Software Technology 50 (2008) 1281-1294
3. Lassen, K.B., van der Aalst, W.M.P.: Complexity Metrics for Workflow Nets Information and Software Technology 51 (2009) 610-626
4. Li, C., Reichert, M., Wombacher, A.: Mining Business Process Variants: Challenges, Scenarios, Algorithms. Data & Knowledge Engineering 70 (2011) 409-434
5. Monsalve, C., Abran, A., April, A.: Measuring Software Functional Size from Business Process Models. International Journal of Software Engineering and Knowledge Engineering 21 (2011) 311-338
6. Recker, J.: Opportunities and Constraints: The Current Struggle with BPMN. Business Process Management Journal 16 (2010) 181-201
7. Recker, J., Rosemann, M., Indulska, M., Green, P.: Business Process Modeling: A Comparative Analysis. Journal of the Association for Information Systems 10 (2009) 333-363
8. Robey, D., Markus, M.L.: Beyond Rigor and Relevance: Producing Consumable Research about Information Systems. Information Resource Management Journal 11 (1998) 7-15
9. Swanson, E.B.: Information System Implementation: Bridging the Gap Between Design and Utilization. Irwin, Homewood, Illinois (1988)
10. White, S.A., Miers, D.: BPMN Modeling and Reference Guide. Lighthouse Point, Florida, Future Strategies (2008)
11. Silver, B.: BPMN Method and Style, 2nd Edition, with BPMN Implementer's Guide: A structured approach for business process modeling and implementation using BPMN 2.0. Cody-Cassidy Press, Aptos, CA (2011).
12. Object Management Group: Business Process Model and Notation (BPMN 2.0). OMG Specification dtc/11-01-03. Framingham, MA (2011)
13. WfMC: XML Process Definition Language (XPDL), WfMC Standards. WFMC-TC-1025, http://www.wfmc.org, 2001.
14. van Der Aalst, W. M., Ter Hofstede, A. H., Kiepuszewski, B., & Barros, A. P.: Workflow Patterns. Distributed and Parallel Databases, 14 (2003) 1, 5-51.