The Influence of Syntactic Quality of Enterprise Process Models on Model Comprehension

Merethe Heggset[1], John Krogstie[1] and Harald Wesenberg[2]

[1]Norwegian University of Science & Technology - NTNU, Norway
merethhe@gmail.com, krogstie@idi.ntnu.no

[2] Statoil ASA, Norway
hwes@statoil.com

Abstract. In this paper we report on the use of process modelling in connection to the quality system of Statoil, a large Norwegian oil company, in particular on the aspects found necessary to emphasise to achieve the right quality of the models in this organisation. Based on investigation of usage statistics and user feedback on models, we have identified that many have trouble in comprehending some of the models. Many of these models have poorer syntactic quality than the average syntactic quality of models of the same size. An experiment with improving syntactic quality of one of these models is reported here. Further work is needed to look in more detail on the interplay between levels of quality for overall effect of enterprise models.

1 Introduction

Statoil is a Norwegian oil company with more than 23 000 employees and around the same number of external contractors. The company operates in 36 different countries and have in the last decade been using enterprise modelling in order to structure their vast amounts of organizational knowledge and information. They report to have achieved a fair amount of success with enterprise modelling in its corporate management system [13] where workflow models are used extensively to communicate requirements and best practices throughout the enterprise. The enterprise model functions as a common point of reference for the entire organisation, ensuring the quality of a large number of work processes and communicating requirements and best practices throughout the company. The models are used daily in large parts of the organization, and are a significant contributor in reducing operational, environmental and safety risks. As an example, the important SIF-index (Serious Injury Frequency) which counts the number of incidents per million work hours has been reduced from 6 to around 0.8 in the period since the models where introduced. Every week Statoil employees and contractors perform approximately 2 million work hours. That said, the process models are

only one approach to risk mitigation. One also experience that the process models could be utilized even better.

A lot of research has been done in the field of enterprise process modelling, as well as on the subject of how to evaluate model quality [3, 6, 7, 8]. However, as stated by Moody [6], many of these methods suffer from a lack of adoption in practice. While the main goal of applying such frameworks in practice normally is providing a detailed evaluation of model quality in a specific case, it can also give indications of the usefulness of the framework.

From the start of the current modeling initiative, Statoil has been aware of the need to balance different levels of quality of the models. According to [10, 13] Statoil have found that it is useful to differentiate between at least three dimensions of model quality: Syntactic quality (how well the model uses the modelling language), semantic quality (how well the model reflects the real world) and pragmatic quality (how well the model is understood by the target audience), which are core dimensions of SEQUAL framework on quality of models and modelling languages [3]. SEQUAL builds on early work on quality of models, but has been extended based on theoretical results [6, 7, 8] and practical experiences [3, 4, 5] with various versions of the framework.

This paper present some of the results from a case study on the use of enterprise process models in Statoil, in particular looking upon model usage, quality issues of existing models, and in particular how better syntactic quality can influence the pragmatic quality of models. The main research question we have investigated in connection to this paper is" How do syntactic quality of a model influence on the pragmatic quality".

In section 2 we describe the Statoil quality system in more detail, before we in section 3 describe experiences from evaluations of the current models, and an experiment on improving the syntactic quality of existing operational models. Discussion of results and ideas on further work are found in section 4.

2 Case Environment - Statoil Quality Management System

The enterprise model is realized through the Statoil management system. The Statoil Book [12], which is the foundation the management system is built upon, describes the management system as "the set of principles, policies, processes and requirements which support our organisation in fulfilling the tasks required to achieve our goals". It defines how work is done within the company, and all employees are required to act according to relevant governing documentation. The Management System consists of three main parts:

- Process models in ARIS, the modelling solution from which all governing documentation (GD) is accessed by the end users.
- Docmap, used for handling and publishing more detailed textual GD
- Disp, a tool which supports the process of handling applications for deviation permits in cases where compliance with a requirement is difficult or impossible to achieve.

The three main objectives of the Statoil Management System are

1. Contributing to safe, reliable and efficient operations and enabling compliance with external and internal requirements.
2. Helping the company incorporating their values, people and leadership principles into everything they do.
3. Supporting business performance through high-quality decision-making, fast and precise execution and continuous learning.

GD describes what is to be achieved, how to execute tasks, and ensures standardisation. Each process area has governing documentation in the form of documents and/or process models, accessible from the ARIS start page. A three-level process model structure is developed. On the bottom level, the so-called workflow diagrams, contains BPMN models [9] on the descriptive level. The quality system contains around 2000 BPMN models at this level, qualifying the case to be an example of BPMN in the large [2].

The enterprise process model is created according to a set of rules for structuring and use of notation that has evolved over the years of model development of use [10, 11]. Using the Splunk tool[1] one can capture how often a certain page or model is accessed and how users navigate through the enterprise model. 12 out of the 20 most used models represent safety critical processes, i.e. they are either classified as Safe work (a sub-category of Operation and Maintenance) or belong to the Safety process area.

When designing diagrams in the enterprise model, requirements in TR0002 - Enterprise structure and standard notation [11] shall be met. In [1], we provided a mapping of the Statoil modelling requirements on their version of BPMN [9] from TR0002 to SEQUAL. In the next section we will in particular look upon the current syntactic quality issues of models (including lacking conformance to naming and labelling guidelines which in [1] was listed under empirical quality).

3 Influence of Syntactic on Pragmatic Quality

During the end of 2013 and the beginning of 2014, a user survey was conducted in Statoil in order to better understand users' experiences and opinions related to the management system and governing documentation including the process models. The survey uncovered challenges regarding understanding of some of the models (pragmatic quality). Although a large proportion of user feel that the governing documentation is easy to understand, others report issues of vagueness and ambiguity.

One of the main purposes of the document TR0002 [11] is to ensure a high syntactic quality in the models made. The document provides an overview of the allowed symbols and naming conventions. The degree of syntactical correctness was first measured on seven workflow models. In the user survey, respondents were asked to give examples of processes that were interpreted differently within their department/unit. This list of processes was used as a basis when selecting models for evaluation. Due to a high number of models listed, not all could be evaluated. The following criteria were applied when selecting models:

[1] http://www.splunk.com/view/splunk/SP-CAAAG57

1. The process is directly mentioned by respondents in the user survey as a cause for misunderstandings and different interpretations
2. The total number of nodes and edges in the model is larger than 20.
3. The model is one of the 100 most used workflow models.

The rules are annotated according to the symbol or aspect they are related to, i.e.:

Table 1 : Syntactic quality measurements

Model	**Size**	**Breaches**	**SYN**	**AVG**
Apply for and evaluate work permit	21	7xN2, 2xG2, 2xG3, 2xN2, CA3	0,55	0,87
Prepare isolation plan	23	CA3, G2, N2	0,89	0,89
Project control	24	12xN4, N2,E1, CA2, CA4, G2, G3,SF1	0,48	0,82
Execute mechanical completion	30	2xN4, 4xN7, 4xE1, 3xW, 4x N2, 2xSF1, G2, G3	0,37	0,80
Set, verify and approve isolation	30	2xN2, 2x SF1, CA4	0,87	0,80
Safety incident	39	E1, 7xN2, 3xN2, 3xG2,2xG3, SP2, 4xW	0,58	0,78
Commissioning and handover of systems	46	2xE1, SP1, SP2, 2xSF1, 16x N4, 2xN2, 5xG2, 5xG3, 3xT1	0,39	0,78

- N: Naming conventions
- T: Task
- OT: Optional Task
- G: Gateways
- SP: Collapsed Sub Process
- CA: Collaboration Activity
- SF: Sequence Flow
- W: Wrongly used concept

The size of the model is equal to the total number of nodes (symbols) and edges (arrows). After measuring the syntactic quality (SYN) of the seven selected workflow models, they were compared to other models of a similar size. The criteria used when choosing models for comparison were the same as the criteria listed above, except for criteria 1 which was inverted - only models without direct mentions were included. For each of the "troublesome" models, the three models closest in size from the top 100 list, that also fit the set criteria were evaluated. The results are summarised in table 1, indicating errors the types found in the bullet list below.

- N1: Names on symbols and expressions shall be formulated in singular form
- N2: Avoid terms with more than four words if possible

- N3: A name shall not be a detailed description
- N4: The first letter of a symbol name shall be in upper case. All other letters should be lower case
- N5: Proper names shall start with upper case letters
- N6: The Statoil official name of a concept shall be used when alternatives exist
- N7: Abbreviations should be avoided
- T1: The title of a task shall be a verb imperative (reflecting the activity performed in order to add value), followed by a noun (reflecting the asset)
- OT1: The title of an optional task shall be a verb imperative (reflecting the activity performed in order to add value), followed by a noun (reflecting the asset)
- OT2: The use of an optional task is only allowed within a collaboration activity
- OT3: It is not allowed to connect sequence flows to the optional task symbol
- SP1: The title of a collapsed sub-process shall be a verb imperative (reflecting the activity performed in order to add value), followed by a noun (reflecting the asset)
- SP2: The collapsed sub-process symbol is drawn using a standard activity shape with a "+" attached
- CA1: The tasks grouped by a collaboration activity symbol shall not be sequenced in time or contain dependencies
- CA2: The title of a collaboration activity shall be a verb imperative (reflecting the activity performed in order to add value), followed by a noun (reflecting the asset)
- CA3: The name of a collaboration activity shall be unique and you shall not name the collaboration activity with names that have been used in the tasks that have been framed by the collaboration activity symbol
- CA4: Each of the tasks framed by the collaboration activity symbol must have a unique title, clarifying different type of activities performed by different roles
- E1: You shall define the title of a start or end event as a noun (reflecting the asset) followed by a verb past participle (reflecting the activity performed to add value to the asset)
- G1: You shall not name parallel gateways
- G2: The title of a diverging exclusive gateway shall consist of the term control (can be replaced with check, verify, evaluate or clarify) followed by a noun (reflecting the object submitted to control)
- G3: The exclusive flow shall be described through an adjective or a phrase describing the alternative flows. You shall not use yes or no when designing exclusive gateways
- SF1: A sequence flow shall have only one source and one target
- SF2: You should not use more than one sequence flow from an activity
- W: Using the wrong symbol (or similar errors)

4.1 Experiment Design and Results

In the experiment, two workflow models were selected, and changes were made to these models to increase their syntactic quality according to the guidelines described above. Participants were to answer questions related to the models in order to measure their understanding (pragmatic quality) of the models.

The original plan was to use only Statoil employees from different departments and locations as participants, but since it proved to be difficult to find enough volunteers, a student experiment was carried out in parallel. In total, 18 students and 9 Statoil employees participated in the study. In order to avoid participants answering based on personal knowledge rather than by consulting the models, the participants from Statoil were did not have first-hand experience with the modelled processes. The models selected for the experiment had a syntactic quality below average, and were found to be easily improvable by correcting mistakes according to the rules found in TR0002 [11]. The two models used in the experiment were:

- SF103 - Safety incident (see characteristics in table 1)
- OM05.07.01.03 - Reset isolation and pressurise

SF103 was also part of the syntactic quality evaluation reported above because it was highlighted in the user survey as a model subject to misinterpretations, and we focus on this below. (OM05.07.01.03 had syntactic quality on 0.72 i.e. around average).

Syntactic quality was here measured on the Norwegian versions of the models, as the experiment was conducted in Norwegian. This was decided in order to avoid language-related misunderstandings, as all of the respondents were native Norwegian speakers. When making the new versions, the models were adjusted to make the syntactic quality as close to 1 as possible. Major changes were made to SF103, as many of the errors were large, e.g. the wrong symbol was used in several cases. With OM05.07.01.03, the changes made were mostly corrections in naming of symbols and splitting of arrows.

The participants were each given two models to interpret - one original and one modified. The participants were split into four groups, and each group was given a different combination of models and questions, following a Latin square design. In addition, they were given an overview of the language notation. The participants were each given 15 questions connected to SF103, and 10 questions connected to OM05.07.01.03. When summarising the results, each wrongly answered question was given -1 points, unanswered questions were given 0 and correct answers were given a score of 1. The total number of available points for each model is the result of (number of participants x number of questions), e.g. 9 x 15 = 135 for questions to the old SF103 in the student experiment. Whereas few improvements were found on OM05.07.01.03, probably due to that the quality was sufficient; we look in more detail on SF103 below:

The overall results for SF103 are summarised in table 2. As shown, the modified version of SF103 scored much higher than the original version both in the Statoil experiment and the student experiment. Some specific questions are worth taking a closer look at, as they give insight into certain problem areas and normal misunderstandings. Question 2 stands out, as all of the Statoil participants answered it wrongly when looking at the old version of the model, and half of those looking at the new:

2. True or false: The process always starts with a safety incident occurring

Looking at the student respondents the change is even bigger: as many as 7 out of 8 that were given the original version answered the question wrongly, and only two that were given the new made the same mistake. The question is related to events, and in reality there are two possible triggers to the process. In the original version, many event-related symbols are used incorrectly, e.g. there are two cases of ”end event" symbols with sequence flows pointing out from them, and event symbols are used instead of task symbols even though the process does not start or end at these points.

Table 2 SF103 results

Experiment	**Old version**	**New version**
Statoil	33/60 p (55%)	52/60 p (87%)
Students	93/135 p (69%)	122/135% (90%)

The next critical question is number 6 (the question had three alternatives):

6. What is special about the activity "categorize, classify and decide causes"?

2 of 4 answered incorrectly when looking at the old model, while everyone answered correctly when looking at the new. This might be due to that the sub-process symbol used in the original model does not correspond exactly to the one defined in the standard notation overview, as it lacks the "+" a collapsed sub-process is supposed to have attached to it. However, this mismatch is not reflected in the students' responses - all of them answered the question correctly.

Question 9 also got two wrong answers in the original version, and none with the new:

9. The process ends when an accident investigation is carried out

Here, some of the students are also confused: the old version lead to three wrong answers and one unsure (unanswered), whereas the new lead to only correct answers. This question is also event-related, so the reasoning is the same as for question 2.

4 Discussion, Conclusion and Further Work

The quality system of Statoil is developed supporting in particular compliance to requirements to reduce risk, an area where large improvements have been observed over the last decade. Still one find challenges with the comprehension of some of the models as described above. While the requirements given in TR0002 are quite detailed and structured, they are not always followed in practice. Measurements on syntactic quality show that syntax errors are quite common in the workflow models.

The user survey, interviews and conversations provided valuable insights into how users experience the management system. Some measures can be taken to achieve higher quality. The experiment we did gave in a way mixed results; whereas improvement in labelling and syntax appeared to improve the comprehension in one of the cases, the other case which had less severe syntactic errors initially, showed no difference, point-

ing to that good syntactic quality is useful for comprehension, but that in some cases other aspects are more important if the syntactic quality is sufficiently good.

The main threat to validity in the model quality experiment is that the number of participants was low. Hence, the data is not sufficient for proving or disproving a hypothesis with statistical significance, and the trends discovered may be coincidental. Additionally, students are not part of the target group of the enterprise model, and the findings would have greater validity if all participants were Statoil employees.

Based on the internal evaluation, updated modelling standards and tool support is being developed. When the new functionality has been implemented in full-scale, the actual effect of these changes on model quality in practice can be analysed. A new user survey, similar to the one carried out in 2013/2014 will be distributed by Statoil when these changes have been put into effect. Studying the results based on the new standards and tools and comparing them to the old may give important insight into the real value of such changes. In particular, following the implementation of the new TR0002 document in practice, and how it impacts model quality and use is an interesting possibility for future research. Another possibility is to carry out a more quantitative study, in which an experiment similar to the model quality experiment reported here is carried out in a larger scale with enough respondents to get statistically significant results.

References

1. Heggset, M., Krogstie, J. and Wesenberg. H. Ensuring quality of large scale industrial process collections: Experiences from a case study. In The Practice of Enterprise Modeling, pages 11--25. Springer, (2014).
2. Houy, Constantin, Fettke, Peter, Loos, Peter, van der Aalst, Wil M. P., & Krogstie, John. (2011a). Business Process Management in the Large. *Business & Information Systems Engineering*(6).
3. Krogstie, J.: Model-based development and evolution of information systems: A quality approach, Springer, London (2012)
4. Krogstie, J., Dalberg, V., Jensen, S.M.: Process modeling value framework. In: Manolopoulos, Y., Filipe, J., Constantopoulos, P., Cordeiro, J. (eds.) Selected papers from 8th International Conference, ICEIS 2006, pp. 309–321. Springer, Paphos (2008)
5. Moody DL, Sindre G, Brasethvik T, Sølvberg A (2002) Evaluating the quality of process models: empirical analysis of a quality framework, In Proc. 21st International Conference on Conceptual Modeling (ER'2002), Tampere, Finland, 7-11 Oct
6. Moody, D.L.: Theorethical and practical issues in evaluating the quality of conceptual models: Current state and future directions. Data and Knowledge Engineering 55 243-276 (2005)
7. Nelson, H.J, Poels, G., Genero, M., Piattini, M. A conceptual modeling quality framework. Software Quality Journal 20:201-228 (2012)
8. Price, R.,Shanks, G.: A semiotic information quality framework: Development and comparative analysis. Journal of Information Technology, 20 (2), 88-102 (2005)
9. Silver, B. BPMN Method and Style. Cody-Cassidy Press (2012)
10. Statoil: TR0002 Enterprise Structure and Standard Notation. version 1 (2009)
11. Statoil: TR0002 Enterprise Structure and Standard Notation. version 3 (2013)
12. Statoil: Statoilboken
    http://www.statoil.com/no/About/TheStatoilBook/Downloads/Statoil-Boken.pdf (2014)
13. Wesenberg, H. Enterprise Modeling in an Agile World PoEM 2011, Proceedings of the 4th conference on Practice of Enterprise Modeling, Oslo, Norway, November 2-3 (2011)