

MASTER'S THESIS

Business Process Model Quality: a validation of quality classifications on real-world business process models

Dhondt, R.H.P. (Randy)

Award date:
2020

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 09. Sep. 2021

Open Universiteit
www.ou.nl



Business Process Model Quality: a validation of quality classifications on real-world business process models

Program: Open University of the Netherlands, faculty of Management, Science & Technology
Master Business Process Management & IT

Course: IM9806 Afstudeeropdracht Business Process Management and IT

Student: R.H.P. Dhondt

Student number:

Date: 24-06-2020

Graduation supervisor: Dr.ir. Irene Vanderfeesten

Second assessor: Dr.ir. Guy Janssens

Version: 1.0

Status: Final



Abstract

Business process models, or process models, are used in many organizations and have been used increasingly in the past decades. Business process models are used to gain understanding of business processes and have various applications. This research focusses on the quantification of understandability of business process models, by using quality metrics and threshold values based on structural complexity of a process model. A case study has been used to gather business process models and assess the understandability of the models, by calculating a literature based quality classification. In turn, these results were discussed with professional process model users, to validate the usefulness of this quality classification. This research concludes, that the applicability of a quality classification, solely based on structural complexity can be a good indicator for understandability, but notational knowledge and visual layout cannot be neglected.

Key words

Business process models, process models, quality classification, quality metrics, threshold values, quality framework, understandability, complexity

Summary

Business process models, or process models, are used in many organizations and have been used increasingly in the past decades. Business process models are used to gain understanding of business processes and have various applications. For example, for process improvement, documentation, redesign or automation purposes.

In current literature, various methods have been identified for defining quality of business process models by assessing their complexity. These researches are limited in their practical applicability due to the lack of enough empirical validation or validation on real-world cases. Therefore, this research focusses on finding and validating quality classification models from previous researches on real-world business process models.

Literature research found that various quality initiatives have been proposed in literature, using guidelines and quality measures. In order to create a classification however, it is required to be able to objectively assess a model. Therefore, for this specific research, a subset of quality metrics and corresponding threshold values which were derived from the BPMIMA framework, were used and assessed. There was chosen to use a subset of all metrics, due to the lack of enough empirical validation of some metrics within the framework.

A case study was used to gather business process models in the BPMN 2.0. notation, from a real-world case organization. These process models were then assessed using the subset of quality metrics and threshold values from the prenoted framework, leading to an ordered list based on the degree of understandability. Subsequently, a selection of models were reviewed by professional process model users by creating an ordered list based on the degree of understandability. To assess the value of the quality classification and evaluate the equalities and differences between the rated lists, semi-structured interviews were held with the professionals.

The results of this research identified quality classification based on structural complexity is only partly validated by the process models used in the case organization. The quality metrics and threshold values are limited to model complexity, but the moderating effect of process model layout and personal factors as notation knowledge cannot be neglected in a quality classification. Although the use of only model complexity is an indicator for process model understandability, it is not the sole factor and the usefulness is limited due to the prenoted factors. In further research, it is advisable to further investigate the effect of visual layout and notational knowledge.

Samenvatting

Bedrijfsproces modellen, of procesmodellen worden veel gebruikt in hedendaagse organisaties. Deze modellen worden gebruikt om inzicht te krijgen in de bedrijfsprocessen teneinde van onder andere proces verbetering, documentatie, herontwerp of automatisering.

In de huidige literatuur zijn diverse methodes geïdentificeerd voor het bepalen van kwaliteit van procesmodellen, door een beoordeling op basis van de complexiteit. Deze onderzoeken zijn echter gelimiteerd door een gebrek aan voldoende empirische validatie of validatie op basis van echt gebruikte procesmodellen. Daarom richt dit onderzoek zich op het identificeren en valideren van een model voor het classificeren van kwaliteit van echt gebruikte procesmodellen.

Een uitgebreid literatuuronderzoek heeft aangetoond dat er vele verschillende kwaliteit initiatieven zijn die gebruik maken van modelleer richtlijnen en kwaliteitsindicatoren. Om een classificatie te kunnen maken van de kwaliteit, is het echter noodzakelijk dat de modellen objectief beoordeeld kunnen worden op bepaalde structurele eigenschappen. Daarom is er binnen dit onderzoek gekozen om gebruik te maken van een subset van metrieken en drempelwaardes van het BPMIMA framework. Er is specifiek gekozen voor een subset, omdat niet alle metrieken en drempelwaardes voldoende empirisch waren gevalideerd.

In dit onderzoek is gebruik gemaakt van een case study voor het verzamelen van modellen in de BPMN 2.0. notatie uit een casus organisatie. Vervolgens zijn deze modellen beoordeeld door het gebruik van de metrieken en drempelwaardes uit het BPMIMA framework. Dit heeft geresulteerd in een lijst met een bepaalde volgordelijkheid, gebaseerd op de mate van begrijpelijkheid. Daarna zijn professionele procesmodel gebruikers gevraagd om eenzelfde volgordelijke lijst te maken van een selectie van de modellen. Er is vervolgens met deze gebruikers dieper in gegaan op de overeenkomsten en verschillen tussen de resultaten van deze lijsten.

De resultaten van dit onderzoek laten zien dat een kwaliteit classificatie op basis van structurele complexiteit maar deels wordt gevalideerd. De kwaliteit metrieken en drempelwaardes zijn gelimiteerd tot structurele complexiteit, maar het modererende effect van de layout van een procesmodel en persoons gerelateerde factoren, kunnen ook niet worden genegeerd. Ondanks dat structurele complexiteit een beperkte indicator is voor het meten van begrijpelijkheid, is het niet de enige factor. De praktische waarde is daarom beperkt. Voor toekomstig onderzoek wordt er daarom aangeraden om het effect van layout en notatie kennis verder te onderzoeken.

Table of Contents

Abstract.....	ii
Key words.....	ii
Summary.....	iii
Samenvatting.....	iv
Table of Contents.....	v
1. Introduction.....	1
1.1. Preface.....	1
1.2. Exploration of the topic.....	1
1.3. Problem statement.....	1
1.4. Motivation and relevance.....	2
1.5. Main line of approach.....	2
2. Theoretical framework.....	3
2.1. Research approach.....	3
2.2. Implementation.....	3
2.3. Results and conclusion.....	5
2.4. Purpose of the follow-up research.....	10
3. Methodology.....	11
3.1. Conceptual design.....	11
3.2. Technical design.....	12
3.3. Data analysis.....	14
3.4. Methodology reflection.....	14
4. Results.....	16
4.1. Understandability classification of the Business Process Models.....	16
4.2. Evaluation of the Business Process Models.....	19
5. Discussion, conclusion and recommendation.....	23
5.1. Discussion.....	23
5.2. Conclusion.....	24
5.3. Recommendation for practical use.....	25
5.4. Recommendation for future research.....	25
Bibliography.....	26
Appendix I: Literature search process.....	28

Appendix II: Search query strings	29
Appendix III: Literature starting point	30
Appendix IV: Quality metrics	33
Appendix V: Selected quality metrics	35
Appendix VI: Interview protocol	37
Appendix VII: Used process models.....	39
Appendix VIII: Quality indicator values from all models.....	57
Appendix IX: Interview results	58

1. Introduction

1.1. Preface

This research is a thesis for the Master's program Business Process Management & IT of the Open University and focusses on the quality of business process models. The goal of this research is to validate literature-based quality classifications on real-world business process models which are used in practice.

1.2. Exploration of the topic

Business process models, or process models, are used in many organizations and have been used increasingly in the past decades. Business process models are used to gain understanding in business processes, for purposes such as communication, analysis, redesign and automation (Dumas, Rosa, Mendling, & Reijers, 2013). When used correctly, high quality business process models assist in improving business activities to reach specific organizational benefits (Meyer & Claes, 2018).

Quality of business process models can be defined in various ways, and depending on the purpose of the model, one quality type can be of more importance than another. In current literature there are mainly three types of quality defined (Dumas et al., 2013).

- Syntactic quality – defines whether the model conforms to the agreed modeling language;
- Semantic quality – defines the quality of the model compared to the real-world execution;
- Pragmatic quality – defines the usability of the process model for the users.

This research will focus on the pragmatic quality of business process models for communication purposes by assessing the understandability of the model.

In order to assess the understandability of process models, research has been conducted to gain understanding in this topic. The literature review of Dikici, Turetken, and Demirors (2018) distinguishes two indicators for understandability of a process model; process model factors and personal factors. They have also concluded that most research has been focused around process model factors. For example, its complexity, modeling notation, modularity and visual layout. Specifically around structural complexity, research has been performed resulting in frameworks or classification methods to assess the understandability of a process model. This research will focus specifically around structural complexity related factors and other process model factors influencing understandability will be out of scope.

1.3. Problem statement

In current literature, various methods have been identified for measuring quality of business process models by assessing their complexity (Sánchez-González, García, Mendling, & Ruiz, 2010; Sánchez-González, García, Ruiz, & Mendling, 2012; Sánchez-González, Ruiz, García, & Piattini, 2011). These researches however, indicate a limitation in their practical applicability due to the lack of enough empirical validation or validation on real-world cases. Therefore, this research focusses on validating quality classification models from previous researches on real-world business process models with the following research question:

“To what extent do real-world business process models validate quality classifications in literature?”

To answer this central research question, the following sub-questions have been defined and will be answered by the literature study in Chapter 2.

1. How is pragmatic process model quality defined?
2. What are relevant process model quality classifications in current literature?
3. What process model quality classification method will be selected for validation?

The next sub-questions will guide the empirical part of the research, of which the methodology will be described in Chapter 3 and the results in Chapter 4.

4. What process models can be used for the validation?
5. How do the process models score on quality according to the quality classifications?
6. To what extent does the quality classification match the perception of process model users?

1.4. Motivation and relevance

Pragmatic quality of business process models is important for organizations in order to utilize the benefits of using process models. Especially when organizations use the models for communication purpose, it is important that the model is understandable and the interpretation is unambiguous. A practically useful quality classification framework can be used by organizations to set quality standards or develop guidelines to improve the overall quality and/or reduce complexity.

From a scientific perspective, this research contributes to the empirical validation of a framework by using real-world process models. This aids in future improvement or development for frameworks to classify process model quality.

1.5. Main line of approach

The first three sub-questions will be answered through literature research which is performed in Chapter 2. This results in a selection of quality classifications which are empirically validated throughout the rest of this research report. The method of research is described in Chapter 3 and the results are discussed in Chapter 4. Chapter 5 contains the discussion and conclusion which answers the remaining sub-questions. This chapter also contains the reflection of the research and the research method.

2. Theoretical framework

2.1. Research approach

In Chapter 1, a total of six research questions were defined which will be answered during this research. The following three sub-questions will be answered in Chapter 2, by an extensive literature research;

1. How is pragmatic process model quality defined?
2. What are relevant process model quality classifications in current literature?
3. What process model quality classification method will be used for validation?

For the selection of quality classifications in research question three there is a dependency on the result of research question two. Therefore, no specific searches were conducted for research question three. In order to answer the first two research questions, a literature research strategy has been defined to work in a structured and reproducible way. Since the thesis supervisor has provided ten basic papers for this particular research, these papers were used as a starting point for the topic exploration and literature research. In search for relevant literature to this research the Open University and Google Scholar database were searched according to a literature search strategy. This search strategy is based on Saunders, Lewis, and Thornhill (2015) who state a strategy should include defining the search parameters, the search terms and phrases, the online databases and search engines and the selection criteria.

2.2. Implementation

The defining of the search parameters and online databases and search engines, can be found in Appendix I: Search parameters and databases. The search terms and phrases were formulated after reading the initial ten papers handed out by the research supervisor, based on the main topics which were found in the literature and classified as relevant to the research questions. This resulted in the following search terms in Table 1 and Table 2. In the aim for being as complete as possible, wildcards were used after word stems in the search query's to pick up different word forms as well.

Research question 1: How is pragmatic process model quality defined?

Main topics (AND)	
Process model	Pragmatic quality
Alternative terms (OR)	Business process model
	Quality
	BPM
	Understandability
	Complexity
	Comprehensibility

Table 1: Search terms RQ1

Research question 2: What are relevant process model quality classification frameworks in current literature?

Main topics (AND)			
Alternative terms (OR)	Process model	Quality	Classification framework
	Business process model	Understandability	Classification
	BPM	Complexity	Framework
		Comprehensibility	Threshold
			Assessment

Table 2: Search terms RQ2

The goal for the search queries was to reduce the amount of literature to an amount which can be evaluated to select only relevant literature for this research. At first, the search strings were created in order to search both the Open University and the Google Scholar database. Since Google Scholar does not allow truncation in the search strings, and the goal of searching Google Scholar was a wider search to check for possible missing relevant literature, truncations were excluded and searches were limited to title searches. The resulting search strings can be found in Appendix II: Search query strings. Since Google Scholar does not have the ability to limit search query's to title an abstract, only the title was used. This due to the massive amount of results when this is not limited (respectively 18,100 and 18,300 results).

Even these limited search queries resulted in a large amount of articles. Therefore, an initial screening was done based solely on the title. The title of the article had to be related to either research question in order to be selected for further assessment. The results of this step can be found in Table 3.

Research question and source	Search results	After title screening
RQ1 – Open University	419	24
RQ2 – Open University	17	4
RQ1 – Google Scholar	286	16
RQ2 – Google Scholar	22	3

Table 3: Search results

From this point, there were duplicate titles in the preliminary list. After excluding the duplicate titles, 37 unique titles remained, including the ten starting articles. The remaining titles were then assessed on relevance, value and sufficiency.

Each article was assessed on relevance by reading the title, abstract and the conclusion of the found articles and determining the relevance to either sub-question. In case the title and abstract were determined to be closely related to the research subject and either sub-question, the article was included for further review.

The value of each journal article has been assessed by using an impact factor, which is a calculated number determining the amount of citations received of a journal by other journals. WebOfScience (2019) calculates the impact factor for journals and also determines the median impact factor per category. In this research the lowest median impact factor for the categories (Business, Engineering, Computer science) is used as a minimum impact factor for journals to be used. In this research only articles from journals with an impact factor of minimum of 1.4 were deemed sufficient. In addition, Google Scholar does not have the option to limit the search results to peer-reviewed articles.

Non peer-reviewed articles however were excluded. Conference proceedings were assessed based on the authors due to limited access to WebOfScience Conference Proceedings Citation Index of the Open University. Only authors from conference proceedings will be included, who have published other articles within the research area which met the minimum impact factor of 1.4.

To determine whether sufficient literature was read and included for use in this research, the set of academic literature was assessed on being sufficient to answer the defined sub-questions. In addition, Google Scholar was used to check whether no other possible relevant papers were missing, which were not included in the Open University library.

One article was not accessible through the Open University database or other free public databases, this article was excluded. This resulted in 20 selected articles for the literature review. Appendix III: Literature starting point shows the 37 articles and the justification of the selected and discarded articles.

2.3. Results and conclusion

2.3.1. How is pragmatic process model quality defined?

In current literature there is no consensus about the definition of quality of process models. Meyer and Claes (2018) find that authors prefer different approaches to process model quality, which causes a lack of consensus, unambiguity and confusion. This lack of consensus is supported by the structured literature review (SLR) conducted by Moreno-Montes de Oca, Snoeck, Reijers, and Rodríguez-Morffi (2015), who state that only one fourth of the studies used in the SLR use precise definitions for process model quality. Although there seems to be little consensus on the exact definition of 'quality', according to Meyer and Claes (2018) there seems to be a common understanding of the terms syntactic quality, semantic quality and pragmatic quality, derived from the SEQUAL framework, which proposes pragmatic quality as the usability of the process model for mainly communication purposes. Syntactic quality is defined as the correctness according to the notation, and semantic quality is defined as the extent to which the model matches the real-world process. Therefore, syntactic quality and semantic quality are inferior to pragmatic quality for this research and only pragmatic quality will be used.

The SLR by Meyer and Claes (2018) and Moreno-Montes de Oca et al. (2015) conclude that pragmatic process model quality is one of the most common researched quality types. However, a multitude of definitions for pragmatic model quality are identified in different researches (Figl, 2017; Meyer & Claes, 2018; Moreno-Montes de Oca et al., 2015; Sánchez-González, García, Ruiz, & Piattini, 2013). Terms as understandability, comprehensibility, complexity and pragmatic quality are used interchangeably and lack a clear consensus. Even though different terms are used, the common finding is that most researches which use either term, focus on whether a process model is understood by the people who read it. Therefore, *understandability* is used in this research to define the pragmatic process model quality.

Understandability

For this research to have a common agreement on the exact definition of understandability of process models, the following definition based on findings in the SLR of Reijers and Mendling (2011) and Dikici et al. (2018) is used as it is the most commonly used definition. "Process model understandability defines to what extend a reader can easily understand and correctly interpret the information of a process model". In the current scientific body of knowledge many researches have focused on quality indicators by researching quality dimensions ranging from cognitive factors to modelling language.

There is however, no common agreement on how to measure understandability. In addition, Meyer and Claes (2018) state that many quality dimensions appear to be too abstract to be objectively measured. Due to this lack of consensus, it is therefore hard to determine what quality dimensions are used to measure the understandability of a process model (Figl, 2017; J. Mendling et al., 2007; Jan Mendling, Strembeck, & Recker, 2012; Moreno-Montes de Oca et al., 2015).

In the SLR of Dikici et al. (2018) an integrated framework of process model understandability was created, as can be seen in Figure 1. In general, process model factors and personal factors can be distinguished as significant influencers of process model understandability. Process model factors however, are more extensively researched than personal factors or a combination of personal- and process model factors. In addition, most research has focused on objectively measurable indicators to evaluate the quality of process models.

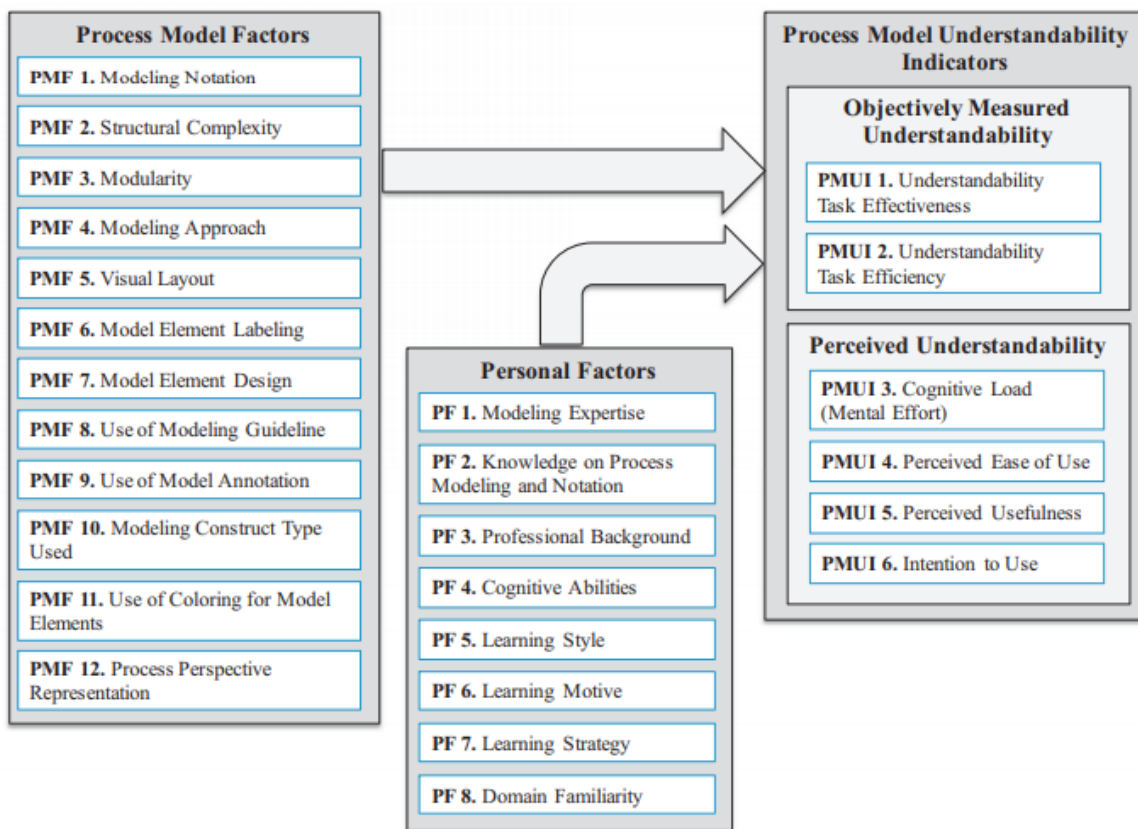


Figure 1: An integrated framework of process model understandability (Dikici et al., 2018)

More specifically, a common find is the negative effect of high structural complexity and size on the understandability of process models (Dikici et al., 2018; Polančič & Cegnar, 2017; Sánchez-González et al., 2013; Sánchez González, García Rubio, Ruiz González, & Piattini Velthuis, 2010). It is however important to notice that a mixed set of terms are used interchangeably defining this aspect of the process model factor. Sánchez-González et al. (2010) refer to a more general term, structural measures, or internal attributes of a process model. Whereas Dikici et al. (2018) use structural complexity as a generalization to define the structural measures.

In addition, Meyer and Claes (2018) distinguish size and complexity as model related metrics. Within this research, quality metrics will be used as all measurable model related factors, thus including complexity and size as can be seen in Figure 2.

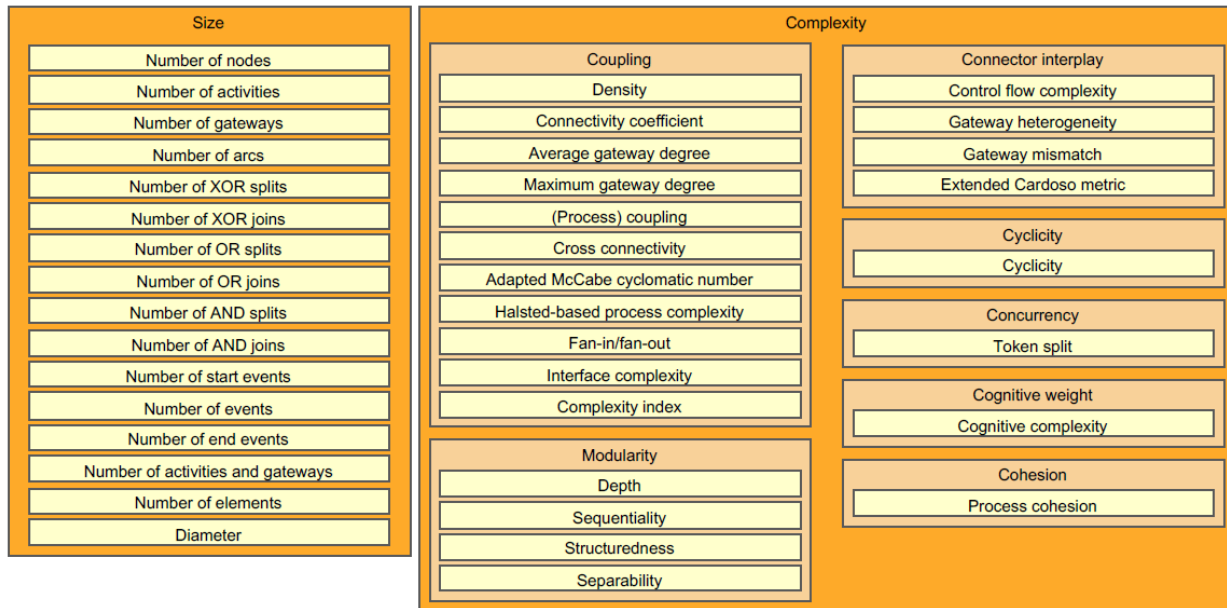


Figure 2: Model related quality metrics

Although Reijers and Mendling (2011) conclude that personal factors have a strong influence on the understandability of process models, this research focusses on the objective measurement of quality. The focus on objectively measurable indicators, can guide modelers into making adjustments to the model, to make it more understandable. The adjustment of personal factors in order to improve understandability is harder and more vague. Personal factors will therefore be excluded from this research. Dikici et al. (2018) also found that most studies focusing on modelling notation have no consensus over the effect of different modelling notations on the understandability of process models. In addition, it costs more effort to change modelling notation of existing process models than revising structural complexity. Therefore, pragmatic quality, or understandability, of a process model is defined as to what extent a reader can easily understand and correctly interpret the information of a process model.

2.3.2. What are relevant quality classifications in current literature?

To improve business process model quality or understandability, an abundance of initiatives have been proposed in literature. Most of the initiatives are based on guidelines or quality measurements, some of which are incorporated into a framework (Dikici et al., 2018; Meyer & Claes, 2018). The identified frameworks, can generally be classified into broadly oriented frameworks and more specific frameworks focusing on specific elements of process model quality. However, as Moreno-Montes de Oca et al. (2015) conclude, no generally accepted framework can be determined. In turn, this led to attempts in merging findings into state-of-the-art frameworks as the CPMQ framework of Meyer and Claes (2018) and the BPMIMA framework of Sánchez-González, García, Ruiz, and Piattini (2017). These frameworks have in common that they have identified indicators for measuring quality. Besides identifying indicators, it is important to provide guidance as to when a model is of high(er) or low(er) quality based on these indicators.

Therefore, to provide guidance in determining process model quality based on quality metrics, threshold values can be used. By using threshold values for the set of metrics, the quality can be objectively measured (Sánchez-González et al., 2010). In addition, threshold values give the opportunity to make a classification, based on different assessment values rather than a binary result. Different attempts in defining threshold values have been proposed by various authors using various methods (Jan Mendling, Sánchez-González, García, & La Rosa, 2012; Sánchez-González et al., 2010; Sánchez-González et al., 2012; Sánchez-González et al., 2011; Yahya, Boukadi, & Ben-Abdallah, 2019).

Yahya et al. (2019) propose a model for evaluating process model quality based on a so called, fuzzy logic. By using the tool BP-FuzzQual, a business process model can be assessed on the various indicators and their threshold. These thresholds are a result from a previous study, where data-mining methods and fuzzification have been used.

Jan Mendling, Sánchez-González, et al. (2012) and Sánchez-González et al. (2012) both use an Area Under the Curve (AUC) and Receiver Operating Characteristics (ROC) method, for defining thresholds of selected measures. In their study Jan Mendling, Sánchez-González, et al. (2012) calculate threshold values from an error probability measure perspective, and connecting the threshold values to modelling guidelines from their previous 7 Process Modelling Guidelines research. Sánchez-González et al. (2012) propose thresholds from a gateway complexity perspective, resulting in a classification for six quality metrics. The threshold values were in turn evaluated in a controlled experiment resulting in a first attempt of empirical validation. The authors indicate more empirical research should be done to validate the threshold values.

Sánchez-González et al. (2010) proposed threshold values, by using the statistical Bender method. The threshold values are calculated for 17 indicators which have been previously defined by Sánchez González et al. (2010). In turn, a multi-class classification on a scale of one to four has been calculated, ranging from very inefficient to very efficient. In Sánchez-González et al. (2011) only a selection of the defined threshold values were tested by using them upon a real world hospital process model, due to the absence of some process model elements. This results in some threshold values being validated, and some threshold values which are not validated in practice. The research concludes that using the threshold values helps in enhancing the understandability of process models, but require more empirical validation. These threshold values were also used in the BPMIMA framework, of which a key goal is to evaluate measurements of process models (Sánchez-González et al., 2017).

To conclude this sub-question, four relevant quality classifications have been identified of which two are incorporated in a framework.

2.3.3. What quality classifications will be used for validation?

To select what quality classification will be used, the decision has to be made what quality metrics and threshold values will be used, for the classification of process models based on its understandability. Various structured literature reviews conclude, an abundance of structural measures or quality metrics can be found for different types of quality (Meyer & Claes, 2018; Polančič & Cegnar, 2017; Sánchez González et al., 2010). But it is important to consider that not all of the identified metrics are empirically validated. Dikici et al. (2018) even report conflicting findings of empirical results in their literature review. For example, some investigated studies indicate no significant relation between structural complexity and understandability.

Whereas other studies indicate a significant moderating effect of structural complexity on the understandability of a process model (Dikici et al., 2018). This is supported by Meyer and Claes (2018) who identify a number of validated predictors for quality metrics, but advises future research to focus on a more elaborate view on the empirical validity. In two literature reviews by Polančič and Cegnar (2017) and Sánchez-González et al. (2013) quality metrics have been identified which are empirically validated to be a quality metric of understandability. The results are slightly different, but an overlap has been found in 24 quality metrics found in Appendix IV: Quality metrics. The list is an overview of which quality metrics have been empirically validated in either one or more researches.

BP-FuzzQual

The used indicators however, are not all empirically validated and are not all found in the previously defined list of indicators. Although this research appears to be the most recent research, the researchers do not appear in any literature review with this, or previous studies. Therefore, the threshold values from this study will not be used in this research or further review.

Jan Mendling, Sánchez-González, et al. (2012)

Although the practical usability appears to be high due to their connection to practical modelling guidelines, it focusses more on recommendation rather than classifications. In this research, the thresholds are limited to one threshold value, so the classification is limited to a binary classification rather than a multi-class classification. Therefore, this classification method will not be used in this research.

Sánchez-González et al. (2012)

Only six quality metrics were used in this research, limited to gateway related indicators. Although the indicators and threshold values used, were deemed useful, not all indicators were empirically validated. Therefore, due to the limited amount of indicators and the limited empirical validation, this classification method will not be used.

BPMIMA

The BPMIMA framework, as proposed by Sánchez-González et al. (2017) appears to be the most recent framework, which uses quality classifications which are well substantiated in literature for evaluating process models. Unfortunately, not all quality metrics used in this research are extensively empirically validated as concluded by Polančič and Cegnar (2017). Therefore, a sub-set of the 17 identified quality metrics and threshold values will be used, based upon the empirically validated indicators in Appendix IV: Quality metrics to be validated in this research as a quality classification. This results in ten quality metrics and threshold values which can be seen in Table 4 to conclude this sub-question. For example, a gateway mismatch of less than or equal to one, indicates a very efficient understandability, whereas a gateway mismatch from two to and including six, indicates a rather efficient understandability. A detailed description of quality metrics used, can be found in Appendix V: Selected quality metrics.

Quality Metric	Understandability threshold values			
	1. Very inefficient	2. Rather inefficient	3. Rather efficient	4. Very efficient
Gateway mismatch	>29	16	6	=<1
Depth	>4	2	1	=<1
Coefficient of connectivity	>1.7	1.1	0.6	=<0.4
TNSF	>72	49	34	=<20
TNE	>20	12	7	=<2
TNG	>17	10	5	=<0
NSFG	>40	22	11	=<0
CFCxor	>30	17	8	=<1
CFCor	>9	4	1	=<0
CFCand	>4	2	0	=<0

Table 4: Quality metrics and threshold values used for validation

2.4. Purpose of the follow-up research

The follow up of the research, focusses on answering the following research questions which will be answered in the empirical part of this research:

4. What process models can be used for the validation?
5. How do the process models score on quality according to the quality classifications?
6. To what extend does the quality classification match the perception of process model users?

3. Methodology

3.1. Conceptual design

The main goal of this research is to evaluate the chosen classification model (Table 4) , which is based on previous researches, and assess its practical value to a real-world scenario. To evaluate the classification model, real-world business process models are required to determine their relative classification, based on their calculated metric values. Hereafter, the classification has to be evaluated in order to gain insight into the practical applicability of the classification model. Therefore, two types of information are required; documentation of business process models and evaluative information. The information is required to origin from a real-world scenario, one or more companies using business process models is the most preferable source of information. In order to define the conceptual research design, the research onion from Saunders et al. (2015) was used, which defines four stages in research design. Defining the methodological choice, strategy, time horizon and techniques and procedures (3.2). Considering the main research goal and question, this research focusses on testing a theoretical proposition. According to Saunders et al. (2015) this research can therefore be classified as a deductive research. Furthermore, this research can be classified as an evaluative study since this research aims to find out how effective the selected classification is, but also why.

Methodological choice

Saunders et al. (2015) state the methodological choice can be either a quantitative research design or a qualitative research design. To decide upon the best method for this research, it is important to consider the goal of this research which is to evaluate theory findings in practice. Since the evaluative information is also required to give a more in-depth insight into the classification, a quantitative research method is expected to be insufficient and a qualitative research is preferable. However, quantitative work has to be performed in calculating quality metrics and categorize the results prior to a qualitative evaluation. Therefore, a sequential explanatory mixed methods research is preferred starting with quantitative calculations and followed by a qualitative evaluation.

Strategy

Saunders et al. (2015) defines the strategy, or plan to answer the research question, as the methodological link to the methodological choice and the method of collecting and analyzing data. The following strategies are identified as possible qualitative approach strategy's;

- Archival and documentary research,
- Case study,
- Ethnography,
- Action research,
- Grounded theory,
- Narrative inquiry.

As the research question requires an evaluation of the extent to which the selected classification matches the users perception, and thus there is an existing theory, a case study is the most preferable method (Saunders et al., 2015). In case study research, Yin (2014) distinguishes four case study methods.

All methods have been taken in consideration. Due to time limitations and manageability, a holistic single case study was deemed to be the most appropriate for this specific research.

Time horizon

According to Saunders et al. (2015) it is important to consider whether the research design requires a snapshot of a specific time (cross-sectional), or that it should be a series of snapshots over a period of time (longitudinal). For this specific research, a diary perspective is not required as the purpose is not to research change in process models over time. This, in combination with a time constraint result in a cross-sectional time horizon as the preferable time horizon.

3.2. Technical design

As the goal of the case study is to assess the practicability of a classification model, the case organization is required to use business process models, preferably for communication purposes. The gathered models can subsequently be used as input to create a classification, Hereafter, the results have to be evaluated by process model users. Thus, this research requires two types of data.

3.2.1. Business Process Models

Gathering the process models is a key element in this research. The business process models can be considered as a secondary non-text data source according to Saunders et al. (2015). When selecting the models, it is important to consider the suitability of the models for this research in terms of validity, reliability and coverage (Saunders et al., 2015). Therefore, the following selection criteria of the process models have been created, partly answering sub-question four;

- A minimum of 15 process models will be selected for further review. Although this is not statistically significant, it is a time consuming task to calculate model metrics. A minimum of 15 models aims to ensure a decent spread in the results for evaluative purpose,
- The process model images need to be of good quality and readable, preferably only digital process models will be used which are assessed by the researcher,
- The selection of process models will be done by randomly selecting a set of different models by an external person,
- The selection of process models will only include models in the BPMN (2.0) notation, as most research has been performed in this modelling notation,
- Draft models will be excluded for this research.

After the models have been collected and the metrics have been calculated, a classification will result from the calculations to use for further evaluation. Since the literature gives no clear explanation of how to classify the models according to the quality metrics and no quality metric is defined as of more importance than another, the average score per model of all quality metrics will be used to judge the understandability.

3.2.2. Evaluation by process model users

According to Saunders et al. (2015) the most frequently used method of interviewing in evaluative research is a semi-structured interview. Due to the nature of this interview, it gives the researcher the possibility to allow discussion into specific contents. The selection of this data collection method is supported by the requirement to assess the classification and gain insight on the reasoning of this classification.

The initial thought was to use semi-structured interviews for creating a ranking of the selected business process models based on the understandability. In addition, the reasoning behind the questioning will be elicited from the interview participants.

For efficiency reasons to interview a larger group of people and the possibility for discussion among the participants upon the topic, a focus group was deemed useful. The focus group should consist of a minimum of four persons. Two of which are experienced BPMN modelers and two who have less, to no experience. This way, the focus group attempts to represent a group of users from a real life scenario.

3.2.3. Methodological change due to COVID-19

Unfortunately, due to the global pandemic caused by COVID-19, data collection through a face-to-face focus group was not possible. In the researchers recent experience in online working, it became clear that group discussion through an online medium is hard and does not yield optimal results. Therefore, the choice has been made to interview the participants solely and through an online session.

Due to the shift of empirical validation through interviews instead of a focus group, changes to the original methodology were required. The initial method, to create a sequential order of all the models ranging from easy to understand, too hard to understand seemed inappropriate in an interview. To cope with this problem, two ways of validating the classification were identified and compared.

The first option was to send all eighteen models to the interview participants and score the understandability of each model on the same scale as the threshold values, ranging from one to four. The second option, was to create a selection of the gathered models, and let the interviewee create a relative order based on the understandability of the models. Eventually the second method was chosen due to several factors. Firstly, the preparation time and effort for the interviewee is reduced since the amount of models to take in and assess is lower aiming for a better response. Secondly, the models can be discussed in more depth, as there are less models to talk about, matching the research purpose. Lastly, it was expected that discussing fewer models through an online interview, keeps better focus and yields better results.

The structure of the interview consists of two parts to answer the sub-question *“To what extend does the quality classification match the perception of process model users?”*. The first part of the interview will be performed prior to the actual interview. As previously discussed, the selected models from the classification are sent to the interviewees prior to the interview, with the question to create an order ranging from high to low understandability. The results will then be e-mailed to the researcher one day before the interview.

The second part of the interview, is an online meeting where the actual semi-structured interview will be performed. The semi-structured interviews consist of the following six questions:

1. What makes a business process model understandable in your opinion?
2. Why did you place this model in its current position in regard to the other models?
3. What structural properties have the most impact on understandability according to you? And why?
4. What could make this model more understandable in your opinion?
5. What do you think of the theoretical classification in relation to your personal classification?
6. How do you think this calculation can be improved so that it better matches your perception of understandability of the models?

In addition to the prenoted questions, introductory questions were asked to assess the level of experience of the interviewee. A detailed interview protocol can be found in Appendix VI: Interview protocol.

3.3. Data analysis

The first source of gathered data will consist out of a minimum of 15 business process models used by the case organization. The selected process models will each be analyzed by calculating the previously defined quality indicators in 2.3.3. In order to calculate the exact indicator it is required that there is an unambiguous definition of the indicator. As multiple interpretations are possible according to different literature sources, an overview of the indicators used can be found in Appendix V: Selected quality metrics. As can be seen in the appendix, the quality metrics are based on the structural properties of a business process model. In some cases multiple structural properties are used to calculate a quality metric. For example, coefficient of connectivity is calculated by dividing the total amount of nodes by the total amount of sequence flows from a model. In order to ease calculations and reduce the probability of errors, an Excel table will be created to automatically calculate the quality metric values. To prevent errors, the calculations in the Excel sheet will be checked by a selected third party. This third party is required to be a BPMN expert, and be knowledgeable of the quality metrics.

After the classification is done and the process models are ranked, a classification by the participants of the group interview will be created, without knowing the framework classification. First, the ranking of the classification model will be judged by the participants. Secondly, the deviation between the rankings from the framework and the interviews are shown to the participants after which will be zoomed in on the commonalities and deviations and the reasoning. The analysis of the interview data will be built upon a thematic analysis method. The essential purpose of this approach is to search for themes or patterns that occur across the data set (Saunders et al., 2015).

The total of data analyses is greatly built upon the deductive explanation method as proposed by Yin (2014). As Saunders et al. (2015) describe, deductive explanation building can be used to build an explanation by testing and refining a predetermined theoretical proposition. In this case, the classification model as proposed by literature findings.

3.4. Methodology reflection

3.4.1. Validity

Internal validity

Saunders et al. (2015) state that internal validity, or measurement validity is the extent to which a scale or measuring instrument measures what it is intended to measure. To maximize the internal validity, the composition of the focus group will consist of process model users with different levels of experience and knowledge. In addition, the empirical validation of the thresholds values from the quality metrics used are limited, but this research proposes to contribute to the empirical validation of the thresholds.

External validity

The external validity defines the generalizability of the results to all relevant contexts (Saunders et al., 2015). This research only uses a single case study, with four interviewee participants. This could pose a threat to the external validity of the research results, due to the possible lack of generalization to other research settings. However, this research is part of a larger research group, in which the same validation is strived.

This contributes to the external validity of this, and the other researches. Personal factors have also not been accounted for in this research, and could explain a different interpretation for the research results. An attempt to minimize this threat is made in the participant selection by using both experienced and non-experienced model users and creators.

Construct validity

According to Saunders et al. (2015), construct validity is defined as the extent to which the measurement tools actually measure what is intended. This mainly subjects to the topic of understandability. By forming the interview questions based upon literature found definitions, this research attempts to enhance the construct validity.

3.4.2. Reliability

The reliability of the research is possibly affected by the participants involved in the creation of the process models which are being analyzed. Therefore, in the selection of participants, the creators of the selected process models will not be included. To enhance the reliability, the selected process models and the interview notes, which are reviewed by the participants, will be included as an appendix. Also, the calculation of the process model metrics will be checked by an expert and the metrics result from literature findings. In addition, a research log will keep track of the steps and decision taken in the research, by the researcher.

3.4.3. Ethical aspects

Ethical aspects will be taken into account by anonymizing the process models, the participants of the focus groups and the notes of the focus groups when requested by the case organization. Although this possibly threatens the transparency of the research, the original results will be saved in a safe and non-public place. For the selection of the person checking the calculations, the case organization will be consulted.

4. Results

4.1. Understandability classification of the Business Process Models

The first sub-chapter of the results will discuss the calculation and results of the classification, followed by the evaluation of the process models.

4.1.1. Gathered Business Process Models

Since the goal of this research is to validate process models used in real life, a Dutch semi-governmental organization has been selected as the organization used in this case study. In this organization BPMN 2.0 models are used as the set standard for process architects and business analysts in documenting and improving their business processes.

For the selection of the process models, one process architect was asked to deliver a minimum of 15, randomly selected, BPMN 2.0 models from their repository which represent real business processes and were no draft models. As the modelling tool used is Enterprise Architect, all models received were digital images and of good readable quality. This resulted in a total of 20 process models created by different process architects. In two process models however, multiple pools connected by message flows were used and thus were not deemed useful for this research as this is not accepted in the BPMN 2.0 notation. This left 18 useable process models which can be seen in Appendix VII: Used process models.

The organization requested to treat the process models confidentially. Therefore, for all models incorporated in Appendix VII: Used process models, all references to the case organization have been removed.

4.1.2. Calculations based on understandability threshold values

During the calculations defined in the methodology it became clear that some process models contain “complex OR” gateways, which are not mentioned in any of the quality metrics and their corresponding literature. As the structural properties of the “complex OR” gateways are closely related to the “XOR” gateways, and it is highly likely to contribute to the perceived understandability in the same way. Therefore, “complex OR” gateways were counted as “XOR” gateways. The Excel table with all quality metric calculations can be found in Figure 3.

Quality metric	Process model property	Value	
(GM) Gateway mismatch	XOR Outdegree	3	0
	XOR Indegree	3	
	OR Outdegree	3	1
	OR Indegree	2	
	AND Outdegree	0	0
	AND Indegree	0	
			1
Depth	n.a.	1	1
(CNC) Coefficient of connectivity	Sequence flows	26	
	Nodes	52	
			0.5
(TNSF) Total number of sequence flows	n.a.	26	26
(TNE) Total number of events	n.a.	9	9
(TNG) Total number of gateways	n.a.	3	3
(NSFG) Number of sequence flows from gateways	n.a.	6	6
(CFCxor) Control flow complexity XOR	Fanout(s)	3	3
(CFCor) Control flow complexity OR	Fanout(s)	2	3
(CFCand) Control flow complexity AND	Split(s)	0	0
	Input		
	Calculated value		
	Calculation used		

Figure 3: Quality indicator calculations

After all the quality metrics were calculated for each process model, the data was transposed, listing the score for all models together in one table. To reduce errors in copying values, a Visual Basic script was written to copy the data from the separate tables, each corresponding to a process model, to the cumulative table. The results from all quality indicator calculations and the corresponding Visual Basic script can be found in Appendix VIII: Quality indicator values from all models.

From this point, the quality indicator values can be used to calculate the understandability threshold values from Table 4. Again, for the purpose of reducing errors and manual handling, an Excel formula was used to determine the threshold value corresponding with the quality metric value. The following formula was used to calculate the values, which was also cross checked by an independent professional.

$$=IF([QMValue]<=[ThresholdValue4],4,IF(AND([QMValue]<=[ThresholdValue3], [QMValue]>[ThresholdValue4]),3,IF(AND([QMValue]<=[ThresholdValue2], [QMValue]>[ThresholdValue3]),2,1)))$$

The result of the calculated threshold values can be seen in Table 5.

Qual. metric	Process model																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
GM	4	4	3	4	3	3	3	4	3	3	3	3	3	3	4	3	4	3
Depth	4	2	1	4	4	1	1	4	2	1	1	1	2	1	4	1	4	1
CNC	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
TNSF	3	4	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4
TNE	2	2	2	3	3	3	3	4	4	3	4	3	3	2	4	3	4	3
TNG	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3	3	3
NSFG	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3	3	4	3
CFCxor	3	3	3	3	3	2	3	4	3	2	3	3	3	3	3	3	4	3
CFCor	2	4	4	4	4	4	2	4	4	4	4	4	4	4	4	4	4	4
CFCand	4	2	2	4	4	4	4	4	4	4	4	4	4	2	4	4	4	4

Table 5: Calculated threshold values per model

4.1.3. Classification of the Business Process Models

This paragraph contains a description of the classification process for the gathered process models. As previously described, the classification of the models is done by assessing the calculated threshold values per quality metric for each model as seen in Table 5.

As previously mentioned, the literature gives no clear explanation of how to classify the models according to the quality metrics, so the average score per model was calculated resulting in Table 6. It can be seen that the average ratings range from a score of 2.8 to a score of 3.9 which is clearly caused by the mere present of very inefficient and rather inefficient (resp. 1 and 2) threshold values. For the total amount of 180 threshold values, the score of 1 is represented only 9 times and the score of 2, 14 times. In turn, the score of 3 is the largest represented number with a total of 88 times, leaving 69 times for the score of 4. These numbers also show that the scores of 1 and 2 are rather evenly divided over all models, resulting in a relative small spread of 1.1 points. Although there is a relative small spread, a clear classification and order of the process models can be derived from the results.

Avg. rating	Frequency	Model
2.8	2	3,14
2.9	2	6,7
3	2	2,10
3.1	4	1,12,16,18
3.2	2	11,13
3.3	1	9
3.4	1	5
3.5	1	4
3.6	1	15
3.7	0	n.a.
3.8	1	17
3.9	1	8

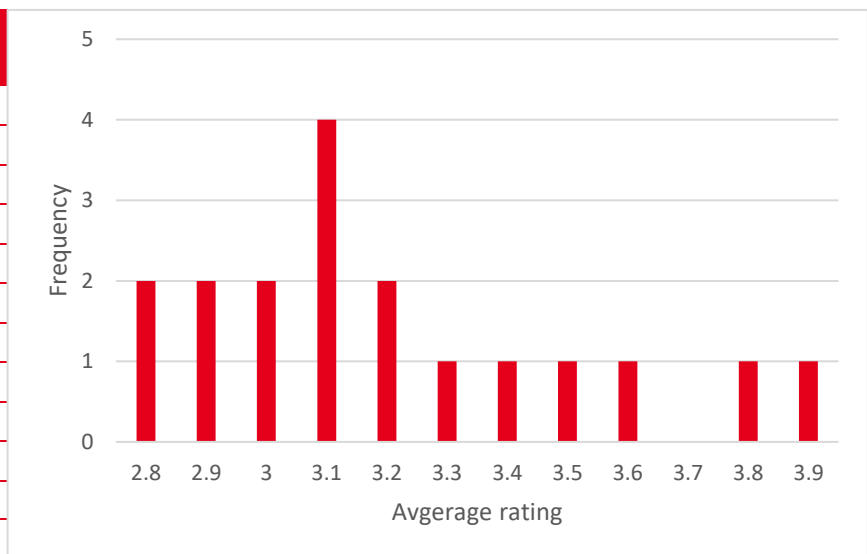


Table 6: Average rating per model

4.1.4. Classification validation

Since the calculation of the process model threshold values is a lot of work and prone to errors, an external reviewer has been selected to check this process. The reviewer was selected to be familiar with the BPMN notation, and thoroughly explained the quality metric calculation method based on the definitions found in the literature. In addition to checking the manual quality metrics calculations for all models, the results of the Visual Basic script and the Excel formula were checked as well. The reviewer found one error related to a gateway mismatch calculation.

4.2. Evaluation of the Business Process Models

The second sub-chapter focusses on the validation of the resulting classification from the first sub-chapter by discussing the method and results of the interviews.

4.2.1. Selection of the Business Process Models

The final selection of the process models to be used in the interviews, was done by an assessment of the researcher. The general idea was to select the most understandable model, the least understandable model and two models with an almost similar score from high frequency groups. With models selected from these groups, the discussion possibilities were deemed the highest to improve an in-depth discussion. However, several problems were faced with this selection method. The models with the highest rating, model 8 and model 17, were excluded from the selection. Since these models only contained activities, the scores were very similar and would probably not lead to a useful discussion. In addition, the probability of finding certain models in practice are likely to be low. This leaves model 15 as the most understandable model. As can be seen in Table 6, there were two models with the lowest rating of 2.8 and the high frequency group also contained multiple models. For this selection, a random generator was used resulting in a final selection of model 3, 1, 13 and 15.

The final selected models were then anonymized, to conform to the request of the case study organization to leave no references. Since it was not required to have knowledge about the processes the model represent, or to correctly interpret the models, all activity descriptions were removed. This results in the following four process models used in the interviews as seen in Figure 4, Figure 5, Figure 6 and Figure 7. Model 1 corresponding initial model 1, model 2 corresponding initial model 3, model 3 corresponding initial model 13 and model 4 corresponding initial model 15. This results in an order of 4, 3, 1, 2, ranging from respectively high to low understandability. This reasoning contributes to, and answers research question four.

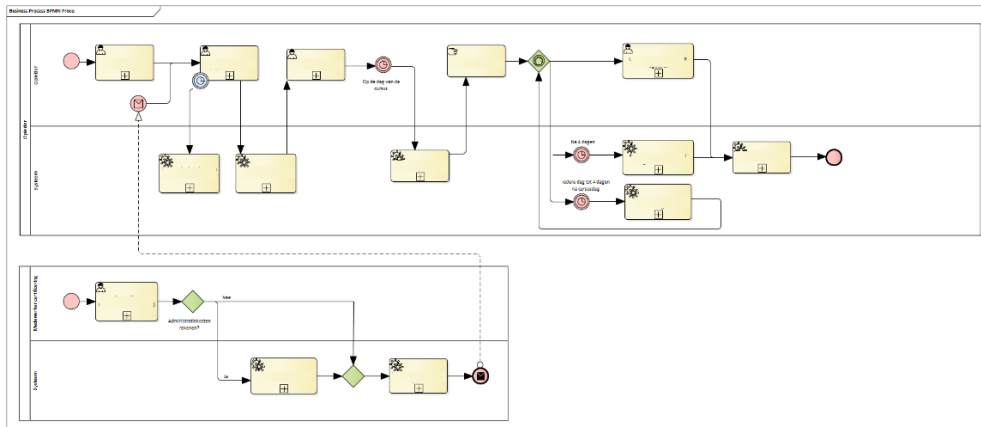


Figure 4: Model 1 used in interviews

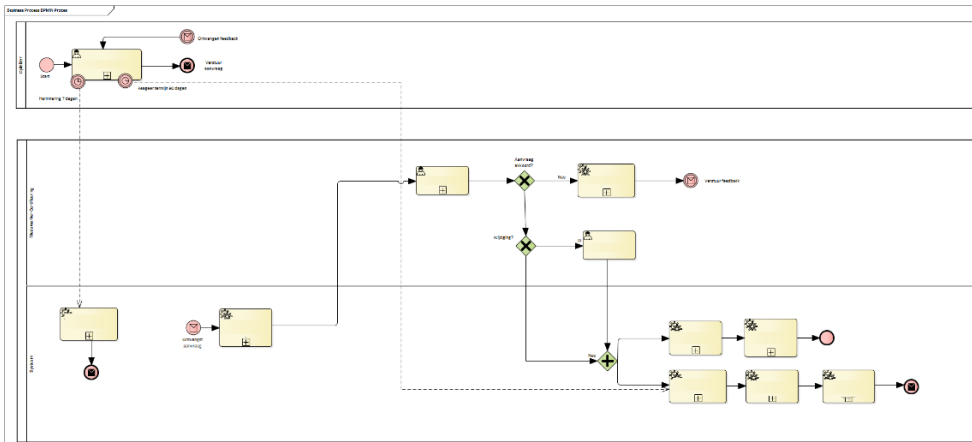


Figure 5: Model 2 used in interviews

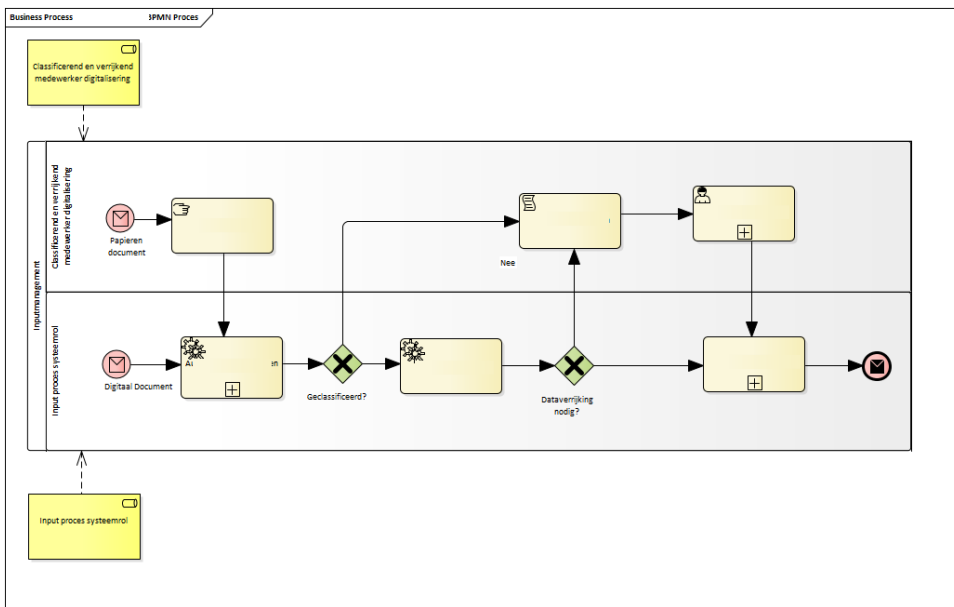


Figure 6: Model 3 used in interviews

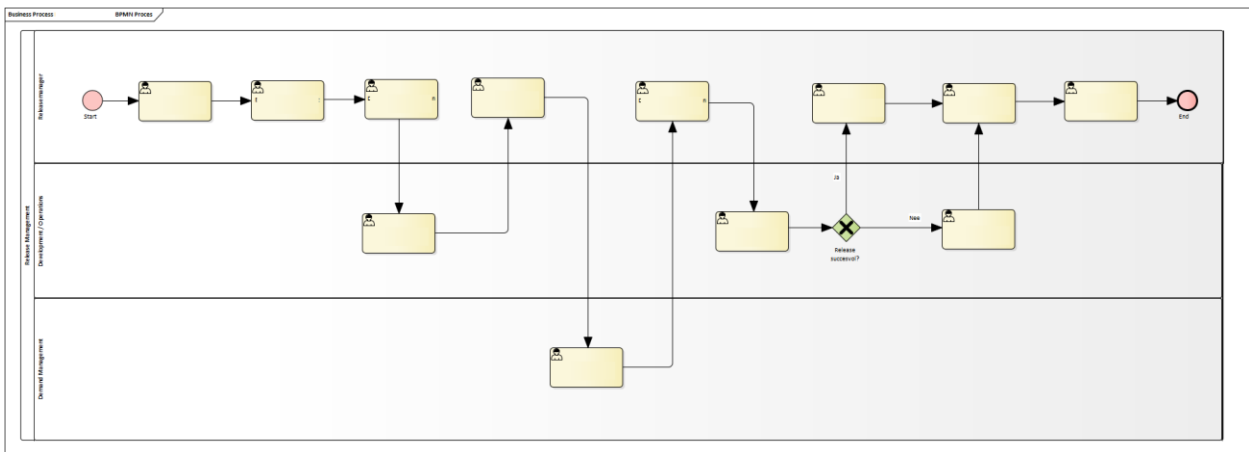


Figure 7: Model 4 used in interviews

4.2.2. Interview results

This sub-chapter is split up in two parts, firstly the results from the ranking are discussed. Secondly, the results from the interviews are discussed, mainly regarding the ranking of the model versus the ranking of the interviewee.

The results from the ranking, which were received prior to the interviews, can be seen in Model ranking results Table 7. The interviewees are shown as I1 up to I4, with their experience level as previously assessed with either a ‘-’ or ‘+’ sign. The results show that in no case, the ranking from the respondent exactly matches the ranking which was generated by the model. It should be noted however, that the model order is relative and that it is no exact score. This means the results should be interpreted with care. For example, respondent I3 has only one matching model in regard to the classification model. However, when looking at the sequence, only one model (3) causes the deviation from the classification model sequence.

		Model	I1 (+)	I2 (-)	I3 (-)	I4 (+)
Degree of understandability	High	Model 4	3	4	4	4
		Model 3	4	3	1	2
		Model 1	1	2	2	3
	Low	Model 2	2	1	3	1

Table 7: Model ranking results

The actual semi-structured interviews which were performed based on the prenoted interview protocol, were all recorded with approval of the interviewees. The recordings were in turn summarized prior to the analysis. To improve the validity of the summarized interviews, a cross-check was performed by the interviewees. The final results of the interviews can be found in Appendix IX: Interview results. Subsequently, the summaries were analyzed to evaluate the above results. Since the goal is to evaluate and validate the scorings, a mild form of coding was used focusing on the explanation of the differences.

The results from the interviews will be discussed based on three main topics which were identified in the thematic analysis method, the first two of which supported by Dikici et al. (2018); process model factors, personal factors and quality classification related factors.

These topics were deemed to be the most important in explaining the deviation between the literature based quality classification and the quality classification received from the process model users. For each topic, a summary will be given.

Process model factors

The process model factors were mentioned very often, distinguishing two main topics; visual layout factors and structural complexity. The visual layout factors were especially named in regard to model one and two, the least understandable models. The interviewees indicated that a linear flow, from left to right is preferred and partly responsible for their perceived understandability of a model.

For example, participant I1 indicates: “When I look at model 3, I can instantly see the different swimming lanes. This is clarifying, since I can directly identify the different actors. (...) In addition, this model is fairly linear. (...) Also, the arrangement of the figure plays a big part, as people like to read from left to right and from top to bottom. If the figure satisfies these factors, it feels as an understandable process model” (Appendix IX: Interview results).

Where participant I4 adds: “In my opinion, an understandable BPMN model is a model with a clear flow and a clear starting point and end point” (Appendix IX: Interview results). This is supported by a quote from model 1: “The first thing what surprises me, is two starting points in two different lanes. I don’t know whether it are two different systems, so I think this is confusing” (Appendix IX: Interview results).

The visual layout factors however, are in some cases identified to be closely related to structural complexity, as I2 mentions: “When I look at model 4, where only one choice can be made, it is fairly easy to understand since everything is linear. Then the amount of activities is not really an issue” (Appendix IX: Interview results). In addition, the use of multiple decision points or gateways in a row, or the use of various kinds of BPMN 2.0 activities were also mentioned to influence the perceived understandability. Also the use of loops and the actual amount of gateways were mentioned to be of effect on the understandability.

Personal factors

The knowledge on process modelling and notation, as a personal factor as defined by Dikici et al. (2018) was predominantly present. Even the interviewees who had experience with the BPMN 2.0 language, indicated that a lack of knowledge was in some cases responsible for negatively effecting the degree of understandability.

For example, I3 indicated: “I think models 1 and 2 are pretty much of the same level of understandability, but for the first model I think I understand the meaning of the symbols a bit better, without knowing the exact definition” (Appendix IX: Interview results). And I2 states for model 4: “I think most of my problems are due to a lack of notational knowledge” (Appendix IX: Interview results). This factor is not accounted for in the literature based quality classification, and therefore lacks the ability to take notational knowledge into account.

Quality classification related factors

Some comments were made regarding the quality classification method used, explaining the difference between the process model user classification and the literature based classification. For example, I1 states: “I think the number of gateways and number of sequence flows from gateways are most decisive in determining the understandability. However, I do think that loops are missing as a quality metrics as this can surely increase complexity and understandability” (Appendix IX: Interview results). Two other participants however, stated that gateway mismatch do not or merely affect understandability. In addition, the merge of two sequence flows in an activity rather than a join gateway, in their opinion did not negatively affect the understandability as well.

Also the scale of the quality metrics were discussed and resulted in some deviations between the perception of understandability and the quality metric rating. For example, I4 states: “When I look at the total number of gateways, which are equally rated for all models, I think this is not discriminative enough. When you see a model with five gateways, I think it negatively effects the readability” (Appendix IX: Interview results). This was supported by I2 who states the following regarding total number of gateways: “For all models the same score is given, which places the models in the same category, while I think there should be a difference. (...) In short, I agree with the indicator but not the scale” (Appendix IX: Interview results).

5. Discussion, conclusion and recommendation

5.1. Discussion

As the results have shown, the classification from the theoretical model is in no case exactly the same as the ranking from the professionals. Although these results differ, the actual deviation in the order of the models by professionals is in most cases caused by one model which is placed differently in regard to the theoretical model. The process model users validated the literature findings, that structural complexity is not the only measure for understandability. Predominantly the visual layout and notational knowledge were deemed as important factors moderating the relation between structural complexity and understandability. This research raises doubt about the applicability of using only structural related quality metrics as a predictor for understandability, as understandability is highly subjective.

Another interesting result from the first part of the empirical research, is that the real-world process models are never rated below a score of 2.8 and are largely concentrated between a score of 2.8 and 3.2. This raises questions about the possibility of generalizing the threshold scales in relation to process models used in the real-world. The process model users also indicated that indicators as depth and total number of gateways were not discriminative enough in their opinion. In addition, the indicator gateway mismatch did not always seem to influence the understandability of a model since it is allowed by the BPMN 2.0 notation, that sequence flows merge in an activity rather than a specific gateway merge.

The theoretical model used in this research was formed based upon an extensive literature research which resulted in different identified frameworks and classification methods. It became clear that in current literature there is no set standard model or framework that can be used to classify or rate models based upon understandability. This research used the BPMIMA framework with only the quality metrics that were both related to understandability and empirically validated. This selection however, could have affected the results as not all metrics from the BPMIMA model have been used. In addition, the identified quality metrics and threshold values from the BPMIMA were not clearly defined in any research paper. This led to the possibility of using different interpretations for the quality metrics, which could possibly lead to different results in the primary calculations and dispute the usability of this framework. This research attempted to mitigate this effect by finding definitions based upon previous research.

To calculate the theoretical rating, the average score of all metrics were used, as no metric was deemed to be of more importance than another. However, five out of ten indicators were gateway related indicators and therefore in a large part responsible for the average. Although the process model users indicated that gateways were a large factor in the perceived understandability, it is questionable whether the weight of gateways is correct.

Methodological reflection

The methodological choice for using a case study in this research, has proven to be a good way to validate a literature based quality classification. A case study gave the possibility to get an in depth insight, into the process models used in a real-world setting and the perception of the process model users. A weakness in this research however, is the amount of validated process models and the questioned process model users. Nor the tested process models or the interviewees can be considered statistically proven. This weakness however is partially mitigated as this research is part of a larger research setting.

Another weakness in the followed methodology is the selection of interviewed process model users. The experience level of the process model users was not tested in the same way, prior to the interview. Rather than using an objective questionnaire or set of predetermined requirements, the experience level was assessed in informal conversations by the researcher. In addition, the interviewed process model users were all working in IT related roles, but had varying experience with either process models or BPMN 2.0. This poses a possible threat to the reliability of the results.

The use of interviews instead of a focus group, due to the COVID-19 measures, had slightly changed the initial methodology. Utmost care was taken in looking for an alternative, for focus groups. The use of separate interviews rather than a group interview, is expected to have yielded more valid results. The semi-structured interviews were performed based on a repeated interview protocol. The researcher however had only limited experience in semi-structured interviews, and was in hindsight able to ask more questions based on the interviewee's answers.

5.2. Conclusion

The conclusion answers the main research question of this research.

“To what extent do real-world business process models validate quality classifications in literature?”

Based on the results of this research the identified quality classification based on structural complexity is only partly validated by the process models used in the case organization. The quality metrics and threshold values are limited to model complexity and appear to be a useful indicator in some cases, but the moderating effect of process model layout and personal factors as notation knowledge cannot be neglected in a quality classification. Therefore, the answer to this research question, contains of two parts. First, the validity of the theoretical classification based on the BPMIMA framework and second, the extent to which the classification is validated by the process model users.

All real-world process models were processed through the selected quality metrics, resulting in a score for each quality metric of each model (Table 5). Subsequently the average of all quality metrics were calculated for each model, resulting in the final score (Table 6). The use of an average for the quality metrics, resulted in a discriminative power of 1.1 points on a scale of 4 and showed clear differences in the scoring. However, the individual quality metrics mostly have a rather high rating (mainly 3 and 4) and the score of very inefficient and rather inefficient (resp. 1 and 2), is not seen very often. The discriminative power of the classification model in this specific case is present, but limited and therefore concludes a limited practical usability and generalizability.

The use of only quality metrics derived from structural complexity is an indicator for process model understandability, but it is not the sole factor and the usefulness is limited due to the impact of process model- and personal factors. Therefore, the quality classification based on process model complexity only partly matches the perception of the process model users. Process model users have indicated that other factors as visual layout, notation knowledge, and the use of many different symbols are of great influence on their perceived understandability.

5.3. Recommendation for practical use

The practical applicability of the quality classification used, is limited due to the lack of other important factors related to understandability. However, the quality metrics and threshold values are an objective way of assessing understandability, and do show the ability to discriminate models based on their structural properties.

When using the classification model in practice, it is highly recommended to evaluate the threshold values for each specific situation or organization. As the results from the case organization used show, the discriminative power of the quality metrics is probably not sufficient for practical use, as the scores 1 and 2 are merely seen.

When a scale has been identified for a specific situation, it is possible for the users to generate modelling guidelines to enhance process model understandability. In addition, the average score can be used to assess the absolute quality of a model. This could lead to the use of minimum scores, in order for a process model to be used in practice and saved in the organizations repository of process models.

5.4. Recommendation for future research

This research has shown that quality classification models in literature, based on structural complexity metrics has not been extensively researched yet. Therefore, the further validation of present frameworks using threshold values and quality metrics can be of good value to the scientific body of knowledge. Especially the effectiveness of the defined threshold values from the used quality metrics should be validated by more real-world process models, as this research has shown most metrics score in the group of 3 (rather efficient) and 4 (very efficient). In addition, it appears to be questionable whether the average of all quality metrics results in a good classification calculation. It can be of good use to investigate whether all quality metrics should have the same weight in the classification calculation, or some quality metrics should weigh either more or less.

Beside the prenoted process model factors, this research shows there is a probable moderating effect of visual layout on the perceived understandability of process models predicted by structural complexity. This effect however is currently under researched, and researching this area could lead to useful insights in strengthening a classification model.

Bibliography

- Cardoso, J., Mendling, J., Neumann, G., Reijers, H. A., Eder, J., & Dustdar, S. (2006, 2006). *A discourse on complexity of process models*, BERLIN.
- Dikici, A., Turetken, O., & Demirors, O. (2018). Factors influencing the understandability of process models: A systematic literature review. *Information and Software Technology, 93*, 112-129. doi:10.1016/j.infsof.2017.09.001
- Dumas, M., Rosa, M. L., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of Business Process Management*: Springer.
- Figl, K. (2017). Comprehension of Procedural Visual Business Process Models. *Business & Information Systems Engineering, 59*(1), 41-67. doi:10.1007/s12599-016-0460-2
- Khelif, W., Ben-Abdallah, H., & Ben Ayed, N. E. (2017). A methodology for the semantic and structural restructuring of BPMN models. *Business Process Management Journal, 23*(1), 16-46. doi:10.1108/BPMJ-12-2015-0186
- Mendling, J., Reijers, H. A., Cardoso, J., Alonso, G., Dadam, P., & Rosemann, M. (2007, 2007). *What makes process models understandable?*
- Mendling, J., Sánchez-González, L., García, F., & La Rosa, M. (2012). Thresholds for error probability measures of business process models. *The Journal of Systems & Software, 85*(5), 1188-1197. doi:10.1016/j.jss.2012.01.017
- Mendling, J., Strembeck, M., & Recker, J. (2012). Factors of process model comprehension—Findings from a series of experiments. *Decision Support Systems, 53*(1), 195-206. doi:10.1016/j.dss.2011.12.013
- Meyer, P. d., & Claes, J. (2018). An overview of process model quality literature-The Comprehensive Process Model Quality Framework. Retrieved from <https://www.janclaes.info/pdf/DeMeyerEA2018.pdf>
- Moreno-Montes de Oca, I., Snoeck, M., Reijers, H. A., & Rodríguez-Morffi, A. (2015). A systematic literature review of studies on business process modeling quality. *Information and Software Technology, 58*, 187-205. doi:10.1016/j.infsof.2014.07.011
- Polančič, G., & Cegnar, B. (2017). Complexity metrics for process models – A systematic literature review. *Computer Standards & Interfaces, 51*, 104-117. doi:10.1016/j.csi.2016.12.003
- Reijers, H. A., & Mendling, J. (2011). A study into the factors that influence the understandability of business process models. *Systems, Man and Cybernetics, Part A: ...*, 41(3), 449–462. Retrieved from <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5628273>
- Sánchez-González, L., García, F., Mendling, J., & Ruiz, F. (2010). Quality Assessment of Business Process Models Based on Thresholds. In (Vol. 6426, pp. 78-95). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Sánchez-González, L., García, F., Ruiz, F., & Mendling, J. (2012). Quality indicators for business process models from a gateway complexity perspective. *Information and Software Technology, 54*(11), 1159-1174. doi:10.1016/j.infsof.2012.05.001
- Sánchez-González, L., García, F., Ruiz, F., & Piattini, M. (2013). Toward a quality framework for business process models. *International Journal of Cooperative Information Systems, 22*(01).
- Sánchez-González, L., García, F., Ruiz, F., & Piattini, M. (2017). A case study about the improvement of business process models driven by indicators. *Software & Systems Modeling, 16*(3), 759-788. doi:10.1007/s10270-015-0482-0
- Sánchez-González, L., Ruiz, F., García, F., & Piattini, M. (2011). Business process model improvement based on measurement activities. *ENASE 2011 - Proceedings of the 6th International Conference on Evaluation of Novel Approaches to Software Engineering*, 104-113.

- Sánchez González, L., García Rubio, F., Ruiz González, F., & Piattini Velthuis, M. (2010). Measurement in business processes: a systematic review. *Business Process Management Journal*, 16(1), 114-134. doi:10.1108/14637151011017976
- Saunders, M., Lewis, P., & Thornhill, A. (2015). Research methods for business students, 7th ed. In. Portland: Pearson.
- WebOfScience. (2019). InCites Journal Citation Reports. Retrieved from <https://jcr-clarivate-com.ezproxy.elib10.ub.unimaas.nl/JCRLandingPageAction.action>
- Yahya, F., Boukadi, K., & Ben-Abdallah, H. (2019). Improving the quality of Business Process Models: Lesson learned from the State of the Art. *Business Process Management Journal*, 25(6), 1357-1376. doi:10.1108/BPMJ-11-2017-0327
- Yin, R. K. (2014). *Case Study Research Design and Methods* (Vol. 5th ed.): Thousand Oaks, CA: Sage.

Appendix I: Search parameters and databases

Search parameters

The parameters for the literature search were based on the six parameters given in Saunders et al. (2015). In Table 8 the parameter selection is shown after which a motivation for the selected parameters is discussed.

Parameters	Selection
Language	English
Geographical area	Global
Subject area	(Business) process modelling, (Business) process model quality
Business sector	Business, Engineering, Computer science
Publication period	≥1973
Literature type	Refereed (peer-reviewed) academic journals, Conference proceedings

Table 8: Search parameters

Language & Geographical area

For this research, only English written documents will be used for further literature review. The geographical area however, is unlimited as there is no reason to exclude certain areas.

Subject area & Business sector

The subject area and business sector limits the amount of results by filtering out results which are not in the former named subject area or business sector. In both the Open University and the Google Scholar databases however, it is not possible to select the previously mentioned parameters. Therefore, for the Open University database, the 'discipline' is limited to business, engineering and computer science. In Google Scholar it is not possible to select either parameter and is therefore not used.

Publication period & Literature type

Since the basic papers, apart from one literature overview, handed out by the research supervisor did not contain any sources used from before 1973 and most sources were from after the 1990's, results from before 1973 are excluded. The one source from before 1973 did not contribute to this research and was therefore excluded. To further limit the search result to high quality sources, only peer reviewed academic journals or conference proceedings were used.

Databases and search engines

As previously mentioned, two search engines were used in this literature research. The Open University library and Google Scholar. The Open University search engine is the main search engine used for this research due to accessibility of large digital libraries and the functionality to use wildcards, Boolean operators and truncation. The Google Scholar search engine has been used for a wider search in the aim of not overlooking any relevant literature.

Appendix II: Search query strings

Research question and source	Query string
RQ1 – Open University	(TitleCombined:(((process model*) OR (business process model*) OR (bpm)) AND ((pragmatic quality) OR (quality) OR (understand*) OR (complex*) OR (comprehen*)))) AND (Abstract:(((process model*) OR (business process model*) OR (bpm)) AND ((pragmatic quality) OR (quality) OR (understand*) OR (complex*) OR (comprehen*))))
RQ2 – Open University	(TitleCombined:(((process model*) OR (business process model*) OR (bpm)) AND ((quality) OR (understand*) OR (complex*) OR (comprehen*)) and ((classification framework) OR (classification) OR (framework) OR (threshold) OR (assess*)))) AND (Abstract:(((process model*) OR (business process model*) OR (bpm)) AND ((quality) OR (understand*) OR (complex*) OR (comprehen*)) and ((classification framework) OR (classification) OR (framework) OR (threshold) OR (assess*))))
RQ1 – Google Scholar	allintitle: ("process model" OR "business process model" OR "bpm") ("quality" OR "understandability" OR "complexity" OR "comprehensibility")
RQ2 – Google Scholar	allintitle: ("process model" OR "business process model" OR "bpm") ("quality" OR "understandability" OR "complexity" OR "comprehensibility") ("classification framework" OR "classification" OR "framework" OR "threshold" OR "assessment")

Appendix III: Literature starting point

Article	Journal (IF)	Conference proceeding	Justification
Meyer, P. d., & Claes, J. (2018). An overview of process model quality literature-The Comprehensive Process Model Quality Framework.	n.a.	n.a.	This article was handed out by the research supervisor, but has not been published in either a journal or conference proceeding. The articles however is referenced 56 times, and the authors are well known within the research area. Therefore this article has been included.
Moreno-Montes de Oca, I., Snoeck, M., Reijers, H. A., & Rodríguez-Morffi, A. (2015). A systematic literature review of studies on business process modeling quality.	2.92		
Dikici, A., Turetken, O., & Demirors, O. (2018). Factors influencing the understandability of process models: A systematic literature review.	2.92		
Sánchez-González, L., García, F., Ruiz, F., & Piattini, M. (2017). A case study about the improvement of business process models driven by indicators.	2.66		
Reijers, H. A., & Mendling, J. (2011). A study into the factors that influence the understandability of business process models.	6.43		
Sánchez-González, L., Ruiz, F., García, F., & Piattini, M. (2011). Business process model improvement based on measurement activities.		x	Well known and renowned authors in this research area.
Sánchez-González, L., García, F., Mendling, J., & Ruiz, F. (2010). Quality Assessment of Business Process Models Based on Thresholds.		x	Well known and renowned authors in this research area.
Sánchez-González, L., García, F., Ruiz, F., & Mendling, J. (2012). Quality indicators for business process models from a gateway complexity perspective.	2.92		
Mendling, J., Sánchez-González, L., García, F., & La Rosa, M. (2012). Thresholds for error probability measures of business process models.	2.55		

Sánchez-González, L., García, F., Ruiz, F., & Piattini, M. (2013). Toward a quality framework for business process models.	1.43		
Yahya, F., Boukadi, K., & Ben-Abdallah, H. (2019). Improving the quality of Business Process Models: Lesson learned from the State of the Art.	1.88		
Overhage, S., Birkmeier, D. Q., & Schlauderer, S. (2012). Quality Marks, Metrics, and Measurement Procedures for Business Process Models: The 3QM-Framework.	2.59		
Nelson, H. J., Poels, G., Genero, M., & Piattini, M. (2012). A conceptual modeling quality framework.	2.14		
Mendling, J., Reijers, H. A., Cardoso, J., Alonso, G., Dadam, P., & Rosemann, M. (2007, 2007). <i>What makes process models understandable?</i>		x	Well known and renowned authors in this research area.
Cardoso, J., Mendling, J., Neumann, G., Reijers, H. A., Eder, J., & Dustdar, S. (2006, 2006). <i>A discourse on complexity of process models.</i>		x	Well known and renowned authors in this research area.
Mendling, J., Strembeck, M., & Recker, J. (2012). Factors of process model comprehension—Findings from a series of experiments.	3.84		
Polančič, G., & Cegnar, B. (2017). Complexity metrics for process models – A systematic literature review.	2.44		
Sánchez-González, L., García, F., Mendling, J., Ruiz, F., & Piattini, M. (2010). Prediction of Business Process Model Quality Based on Structural Metrics.		x	Well known and renowned authors in this research area.
Sánchez González, L., García Rubio, F., Ruiz González, F., & Piattini Velthuis, M. (2010). Measurement in business processes: a systematic review.	1.88		
Figl, K., & Figl, K. (2017). Comprehension of Procedural Visual Business Process Models: A Literature Review.	2.59		
Si-Said Cherfi, et al. (2013). Improving Business Process Model Quality Using Domain Ontologies			Aims to improve semantic quality rather than pragmatic quality.
Moody, L., et al. (2002). Evaluating the quality of process models: Empirical testing of a quality framework			Focuses on the validation of IS models and a framework related to all quality types.
Kuvykin, V. & Petukhov, M. Y. (2019). IMPROVING THE QUALITY OF PROCESS MODELS IN OIL REFINERY INFORMATION SYSTEMS			Limited to improving quality of oil refinery processes.
Carpinetti, L. C. R. & Gerolamo, M. C. (2003). Quality management and improvement: A framework and a business-process reference model			Focuses on improving processes based on customer expectations and strategic decision rather than process models.

Laue, R. & Mendling, J. (2010). Structuredness and its significance for correctness of process models			Strongly aimed at enhancing process models for automation purposes by structuredness.
Sadowska, M. (2013). Quality of Business Models Expressed in BPMN			Unable to find an impact factor for the journal and very limited references in the journal. Therefore, the article was deemed insufficient for selection.
Heidari, F. & Loucopoulus, P. (2014) Quality evaluation framework (QEF): Modeling and evaluating quality of business processes			Focuses highly on the semantic quality of process models.
Setiawan, M. A. (2013). Integrated Framework for Business Process Complexity Analysis		x	Fairly unknown author within the research area, thus unable to correctly assess the value of the paper.
Qi, Y., Qu, N. & Xie, X. (2010) Towards a preliminary ontology for conceptual model quality evaluating. In: International conference on web information systems and mining, pp. 329–334 (2010)		x	Fairly unknown author within the research area, thus unable to correctly assess the value of the paper.
Mohagheghi, P., Dehlen, V. & Neple, T. (2009) Definitions and approaches to model quality in model-based software development: a review of literature. Inf. Softw. Technol. 51(12), 1646–1669 (2009)			Focused mainly on software development modelling rather than communication purposes.
van der Aalst, W.M.P., Benatallah, B., Casati, F., Curbera, F. (20015) Using software quality characteristics to measure business process quality			Focused mainly on software development modelling rather than communication purposes.
Matook, S. & Indulska, M. (2009) Improving the quality of process reference models: a quality function deployment-based approach			Focuses on reference models to which organizations can create process models rather than process models.
Yan, Q., Melvin, L. S. & Shiely, J. P. (2005). Process model quality assessment by sensitivity analysis			Not related to this research area.
Van Mersbergen, M. (2013). A framework for business process model quality and an evaluation of model characteristics as predictors for quality			Unpublished in a journal or conference proceeding.
Rodrigues, R., Barros, M., Azevedo, L. G. & Leopold, H. (2015). An Experiment on Process Model Understandability Using Textual Work Instructions and BPMN Models			Focuses on assessing the effectiveness of process models rather than textual work.
Pavlicek, J., Hronza, R., Pavlickova, P. & Jelinkova, K. (2017) The Business Process Model Quality Metrics			Unable to access this article.
* Excluded articles from this research			

Appendix IV: Quality metrics

Quality metrics*	Sánchez-González et al. (2013)	Polančič and Cegnar (2017)
Number of nodes	X	
Diameter	X	X
Density	X	X
CNC Coefficient of connectivity	X	X
AGD Average gateway degree	X	X
MGD Maximum gateway degree	X	X
Separability	X	
Nesting depth	X	X
GM Gateway mismatch	X	X
TS Token splits	X	X
GH Gateway heterogeneity	X	X
NEDDB Number of exclusive data-based decisions	X	X
NEDEB Number of exclusive event-based decisions	X	X
NID Number of inclusive decisions	X	X
NCD Number of complex decisions	X	X
NPF Number of parallel forking	X	X
NSFG Number of sequence flows from gateways	X	X
TNG Total number of gateways	X	X
NP Number of pools	X	X
PDOPin/PDOPout Proportion of data objects as incoming/outgoing products and total data objects	X	X
TNE Total number of events	X	X
TNA Total number of activities	X	
TNSF Total number of sequence flows	X	X
CFC Control flow complexity	X	X
NDOout/NDOin Number of data objects which are input/outputs of activities	X	
Sequentiality		X
Cyclicity		X
NSFE Number of sequence flows from events		X
NMF Number of message flows		X
CLP Connectivity level between pools		X
NIMsE Number of intermediate message events		X
NEMsE Number of intermediate message events		X
TNCS Total number of collapsed processes		X
CLA Connectivity level between activities	X	X
CI Complexity index		X
ECaM Extended Cardoso Metric		X
ECyM Extended cyclomatic metric		X
Total Operational Complexity		X
SM Structuredness metric		X
CC Cross connectivity		X

P*D*S Complexity model		X
* Empirically validated quality metrics	* Empirically validated quality metrics with threshold values selected for further research	

Appendix V: Selected quality metrics

Structural indicator as quality metric	Definition
(GM) Gateway mismatch	<p>The sum of gateway pairs that do not match with each other. For example, a fork from an AND gateway, without a join later in the process model (Sánchez-González et al., 2012).</p> $GM_I = \left \sum_{c \in S_I} outdegree(c) - \sum_{c \in J_I} indegree(c) \right $
Depth	<p>Maximum nesting of structured blocs. For example, a split gateway increases the nesting depth with one, while a join gateway decreases the nesting depth with one (Jan Mendling, Sánchez-González, et al., 2012).</p> <p>Nesting depth of 2:</p>
(CNC) Coefficient of connectivity	<p>Ratio of the total number of arcs in a process model to its total number of nodes. For example, six sequence flows divided by 3 nodes gives a CNC of 2.</p> $CNC = \frac{\sum sequence\ flows}{\sum nodes}$ <p><i>Nodes:</i> Counts the number of activities and routing elements, or all elements (Khlif, Ben-Abdallah, & Ben Ayed, 2017).</p>
(TNSF) Total number of sequence flows	The sum of all sequence flows in a model (Khlif et al., 2017).
(TNE) Total number of events	The sum of all events in a model (Khlif et al., 2017).
(TNG) Total number of gateways	The sum of all gateways in a model (Khlif et al., 2017).

(NSFG) Number of sequence flows from gateways	The sum of all sequence flows originated from a gateway (Khlif et al., 2017).
(CFCxor) Control flow complexity XOR	The control-flow complexity of XORsplits is determined by the number of branches that can be taken (Cardoso et al., 2006). $CFCxor = fanout(a)$
(CFCor) Control flow complexity OR	The control-flow complexity of OR-splits is determined by the number of states that may arise from the execution of an OR-split construct (Cardoso et al., 2006). $CFCor = 2fanout(a) - 1$
(CFCand) Control flow complexity AND	For an AND-split, the complexity is simply 1 (Cardoso et al., 2006). $CFCxor = \sum and\ splits$

Appendix VI: Interview protocol

Prior to the interview

Before the interviews take place, the selected participants were emailed relevant background information about the interview structure and preferred preparation. The following information was sent to the participants.

Thank you for willing to participate in an interview at <time and date of the interview> for my Master's thesis in Business Process Management & IT. The interview will take approximately one hour. Before this interview I would like to inform you about the interview subject, structure and preparation.

As you are probably aware of, the interview will be about the understandability of business process models in the BPMN 2.0 notation. In the attachment of this e-mail you will find examples of four business process models, which we will use in the interview. The process models are from a real organization, which are being used in practice. Although the models contain a lot of process specific information, this is not relevant to the interview and it is not required to understand the actual process. The only preparation with respect to the content is to briefly look at the four models, and place the models in order of understandability. For time efficiency, you are requested to send the order by email prior to the interview. In the interview, the order will be discussed. The following format can be used for returning the order.

Degree of understandability	High	<i>Model ...</i>
		<i>Model ...</i>
		<i>Model ...</i>
	Low	<i>Model ...</i>

Furthermore, the interview audio will be recorded to facilitate note taking and the interview is entirely voluntary. The recordings will later be summarized in interview notes and sent to you for review. All information will be processed with care and anonymized if required. Please, if you do not consent this this, let me know. In addition, it is possible to receive the research results after completion of the report.

Unfortunately, due to the corona crisis, we are limited to a digital interview. Therefore, the interview will take place using Microsoft teams. I would kindly ask you to check whether your microphone and camera are working before the interview.

Introduction

First of all, I would like to thank you for participating in this interview. Prior to starting this interview, I would like to remind you this interview will be recorded and you are free to stop this interview at any point. As previously mentioned, the interview will take approximately one hour and the interview notes, will be shared afterwards for reviewing. In addition, all information will be held confidential and your participation can be anonymized in every publication if you wish.

After the introduction, the interviewee will be asked about some background information, concerning the following points.

- Could you briefly describe your role within your current organization and what roles you have fulfilled in your prior working experience?
- To what extent are you familiar with, and do you use the BPMN 2.0 notation?

Interview core

The interviewee was asked look at the process models and order the models based on the understandability of the models. If this has been received prior to the interview, continue. Otherwise, the interviewee is asked to place the models in order.

1. What makes a business process model understandable in your opinion?
2. Why did you place this model in its current position in regard to the other models?

After this question, the term structural properties will be explained to the interviewee. However, the exact indicators defined in this research will not be named.

3. What structural properties have the most impact on understandability according to you? And why?

After this question, the model order will be discussed using the following questions.

4. What could make this model more understandable in your opinion?

When the above questions have been answered, the interviewee will be explained the theoretical background and the resulting indicators with corresponding thresholds. The results of these calculations will be shown and the interviewee will be asked the following questions.

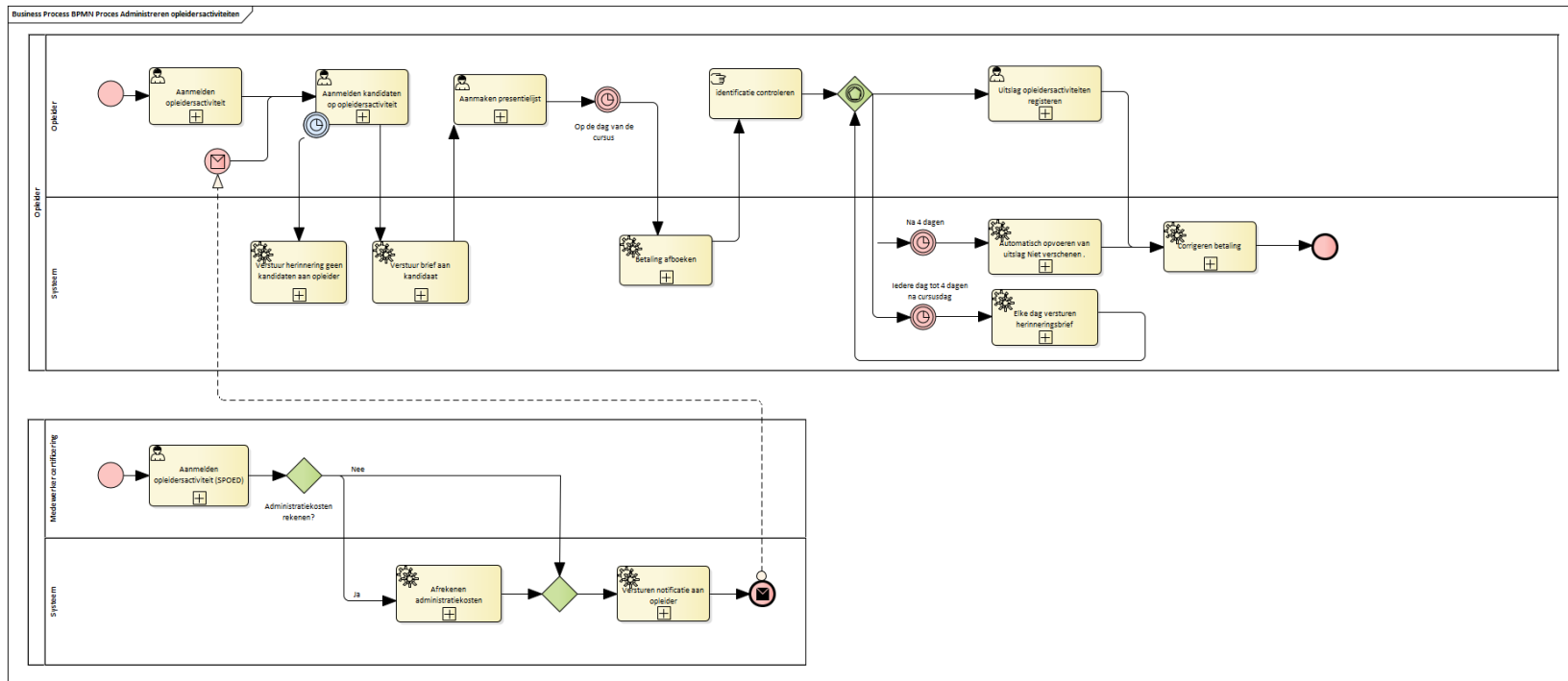
5. What do you think of the theoretical classification in relation to your personal classification?
6. How do you think this calculation can be improved so that it better matches your perception of understandability of the models?

Closing

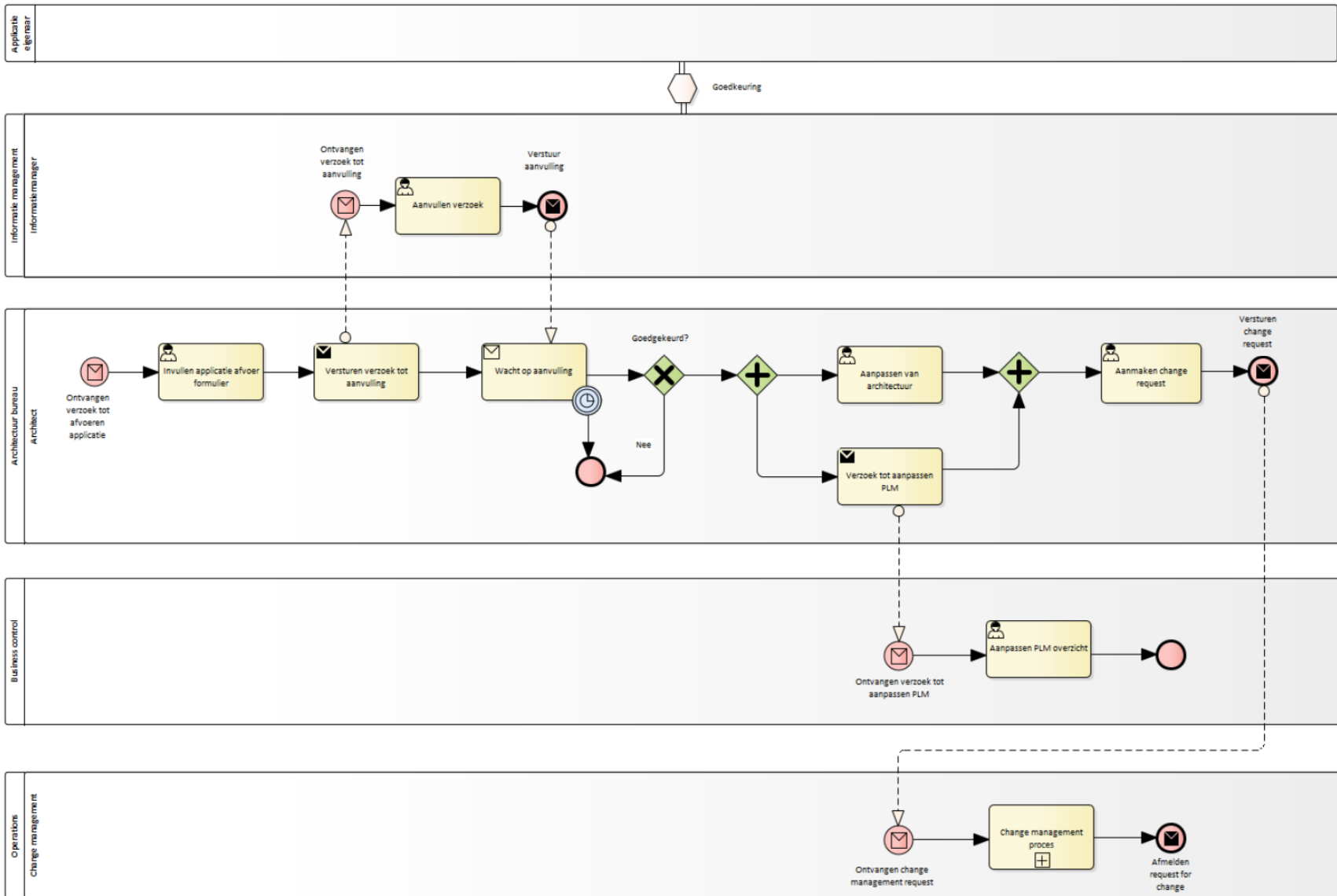
The interviewee will be thanked for participation and asked about any comments or questions regarding the interview.

Appendix VII: Used process models

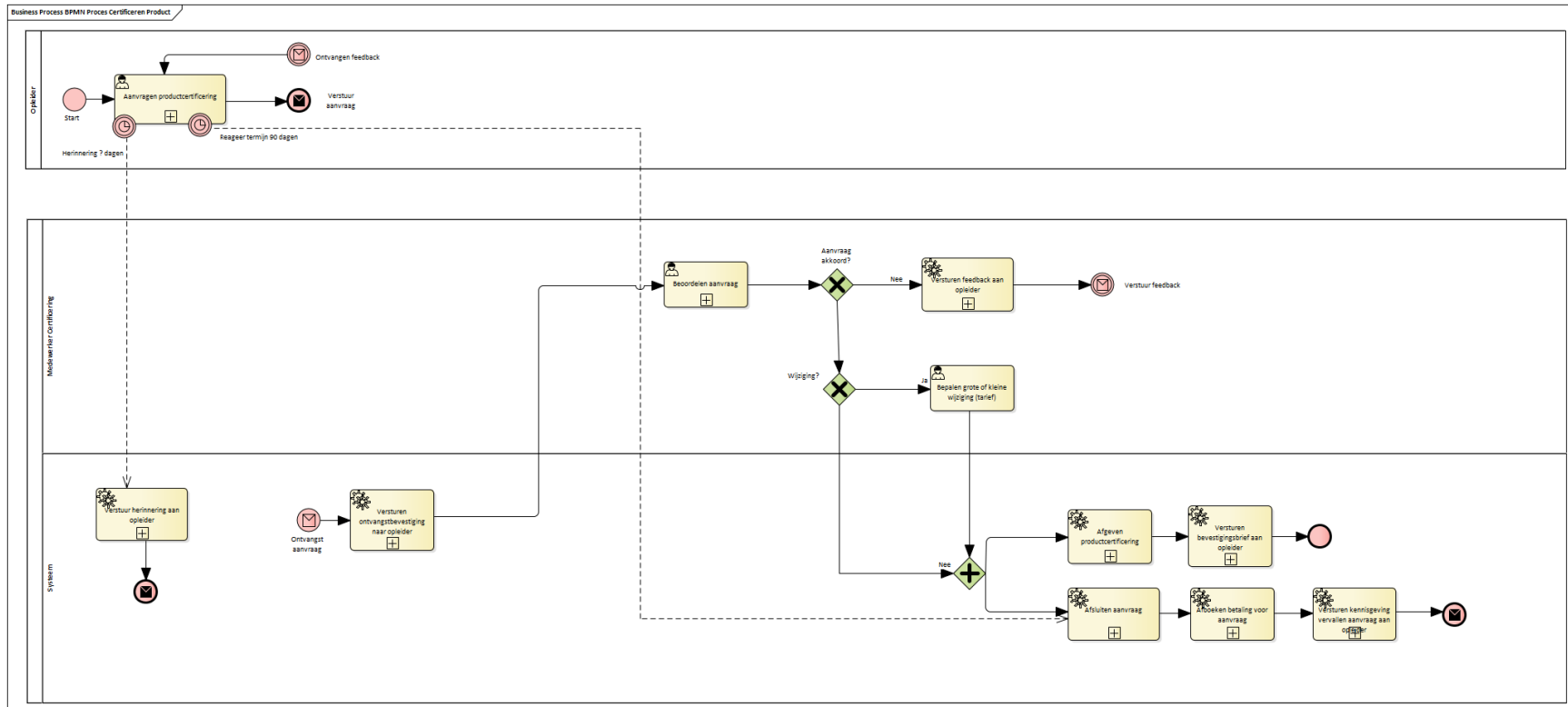
Model 1



Model 2



Model 3

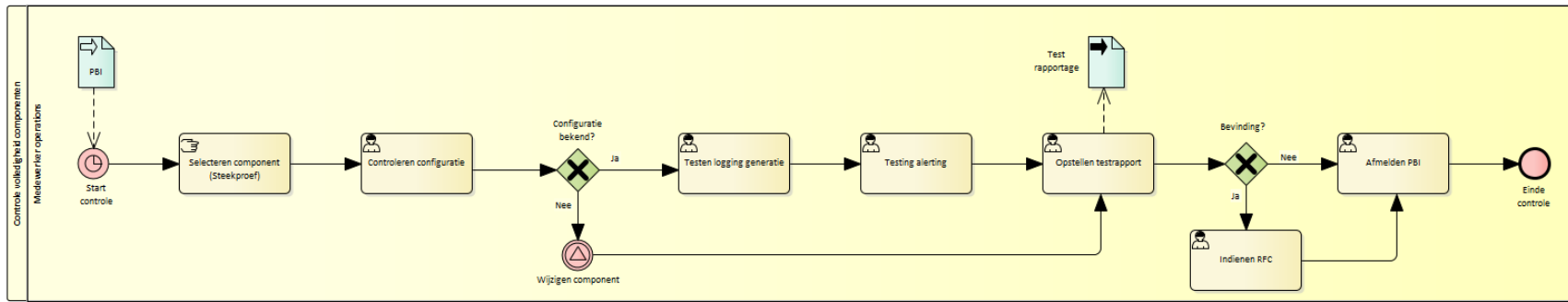


Model 4

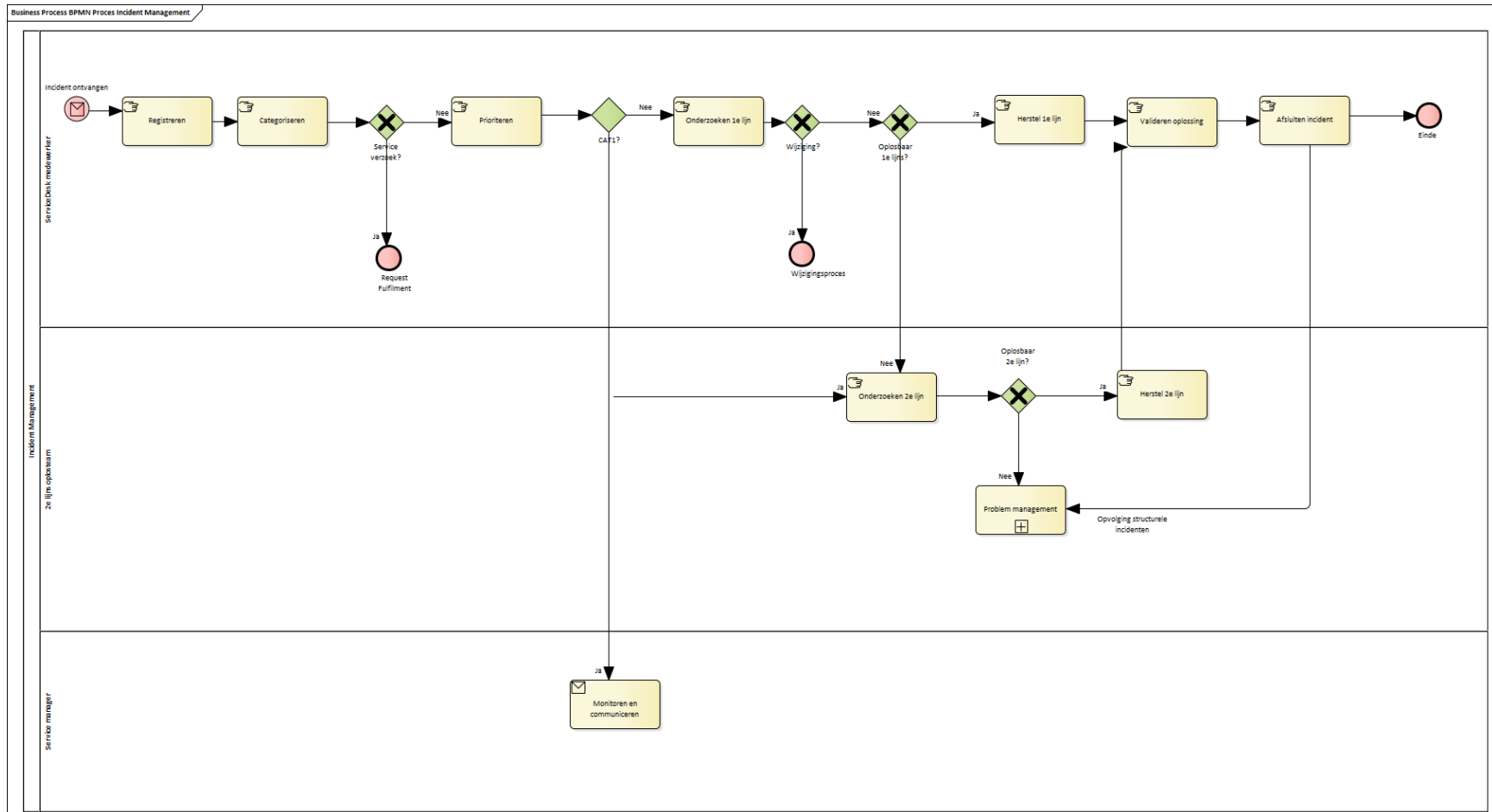
Deleted for confidentiality purposes.

Model 5

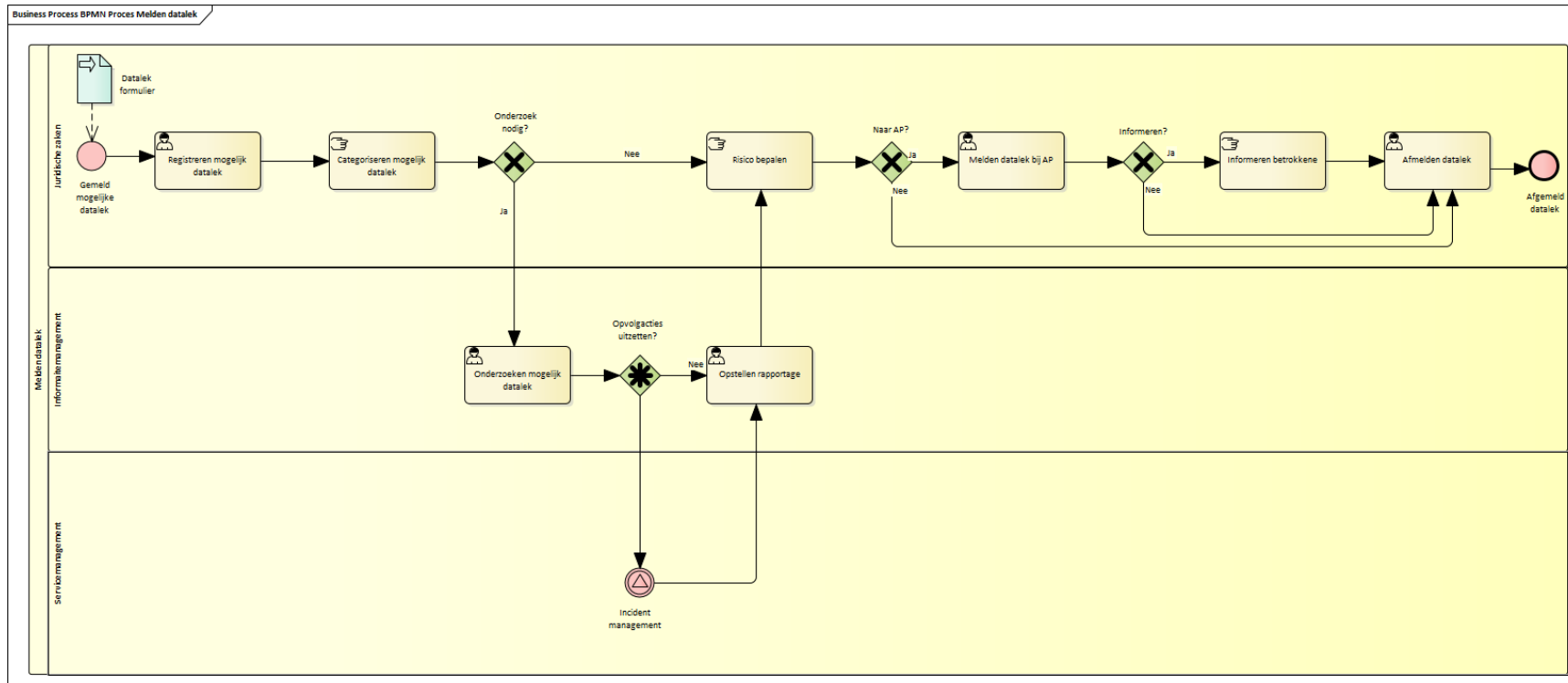
Business Process BPMN Process Controleren volledigheid bestaande component



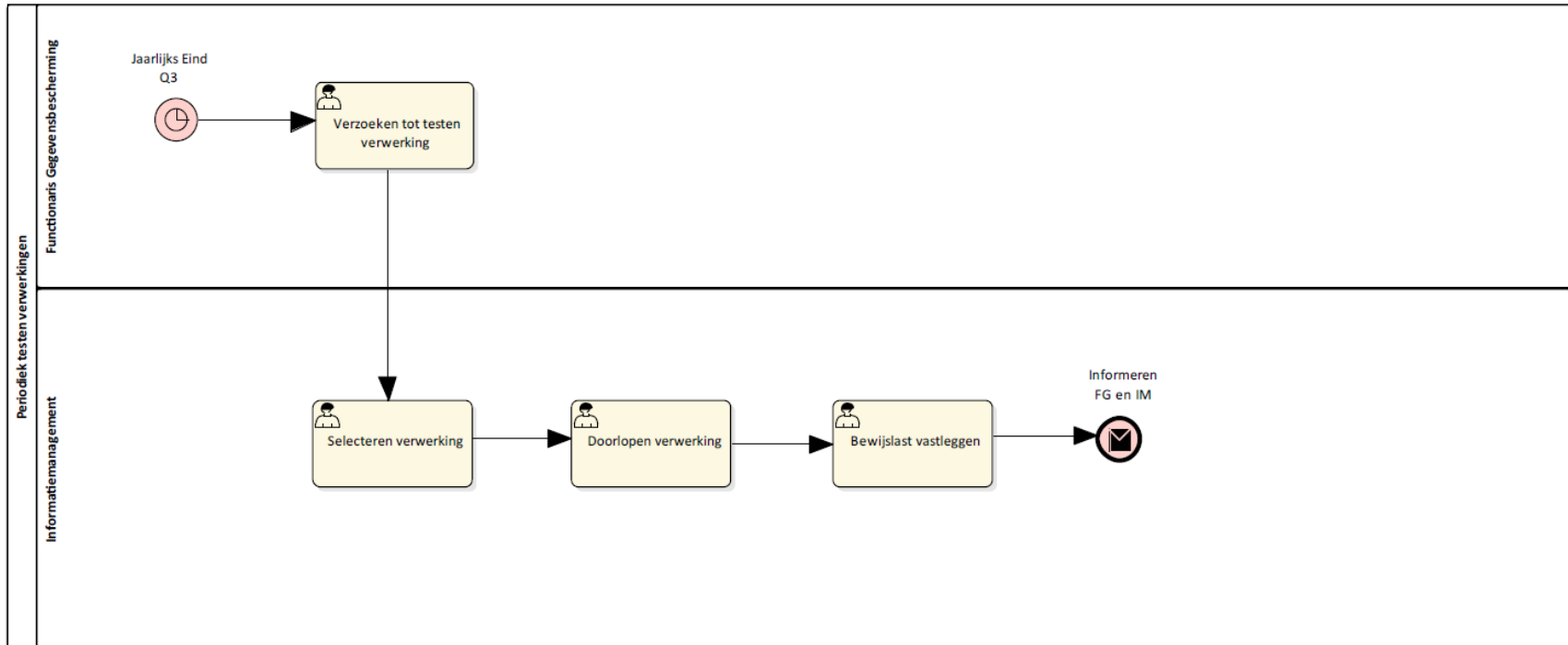
Model 6



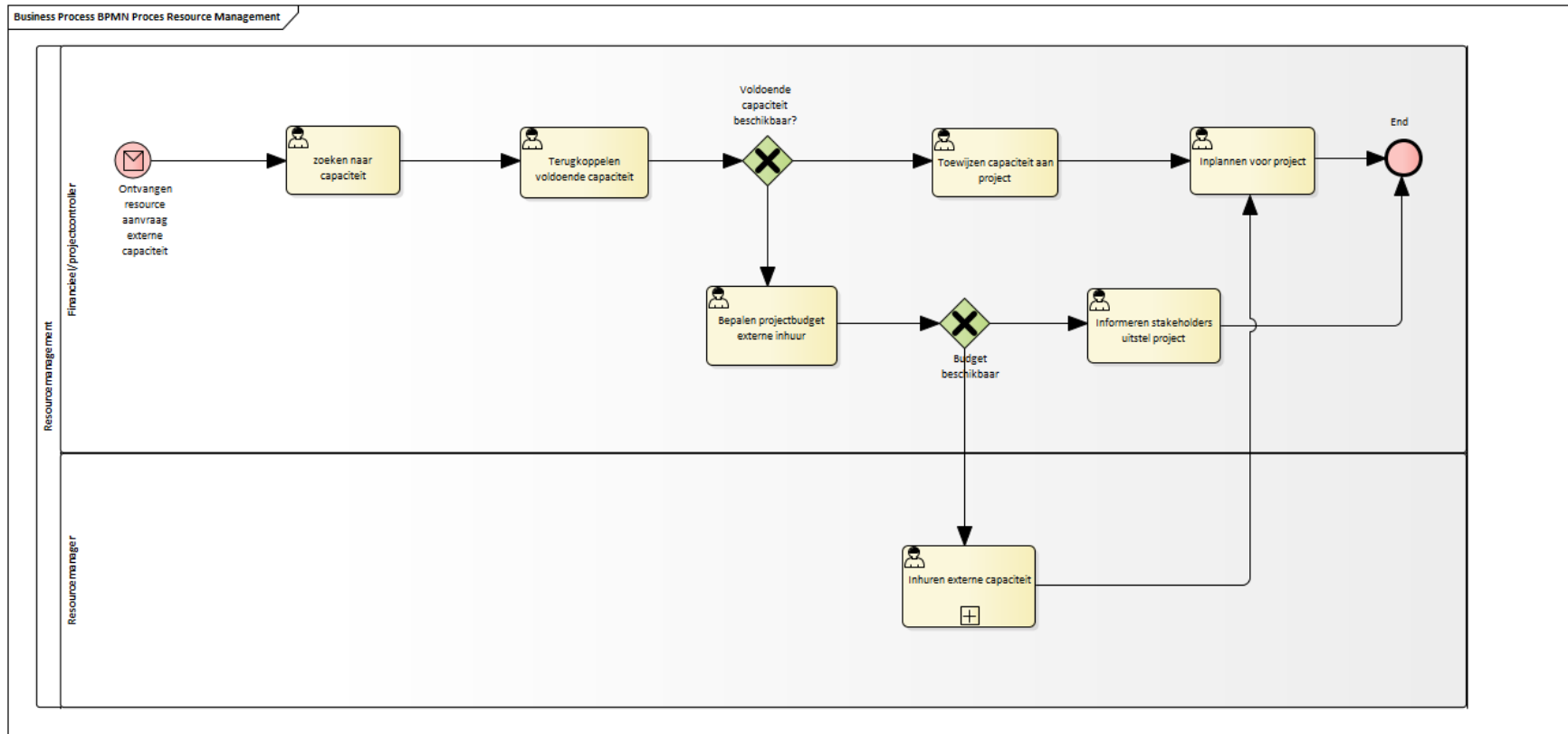
Model 7



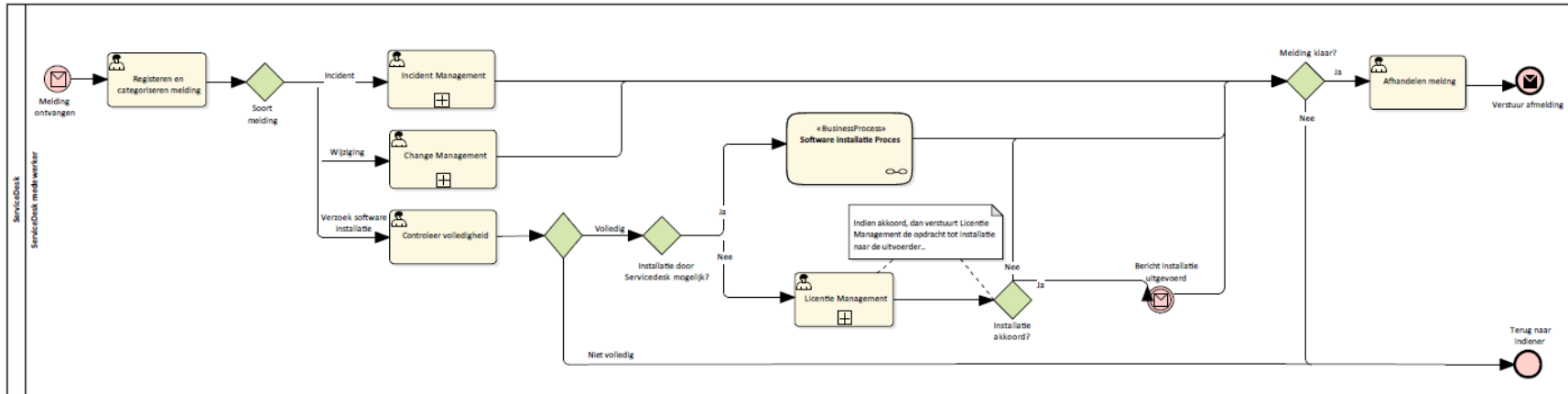
Model 8



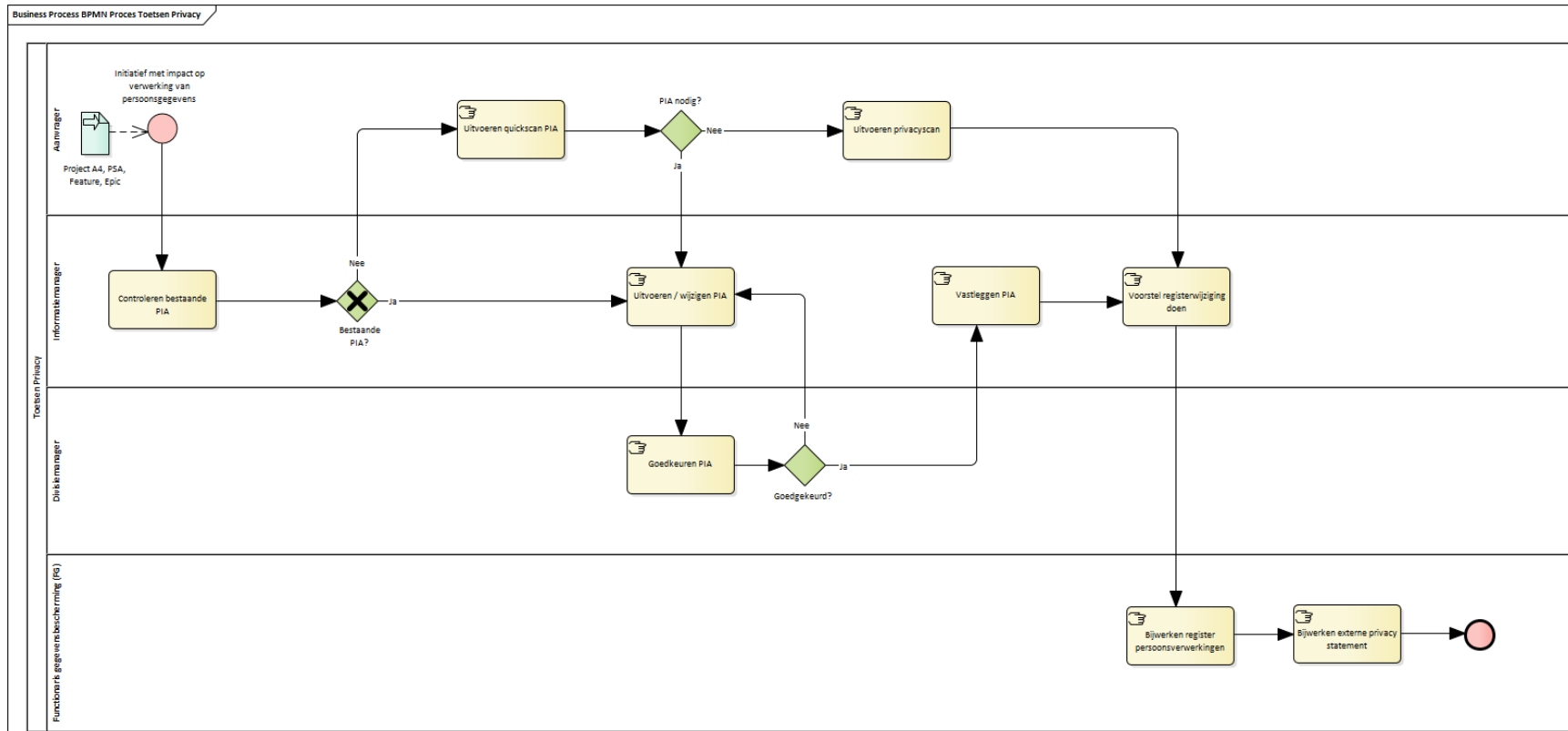
Model 9



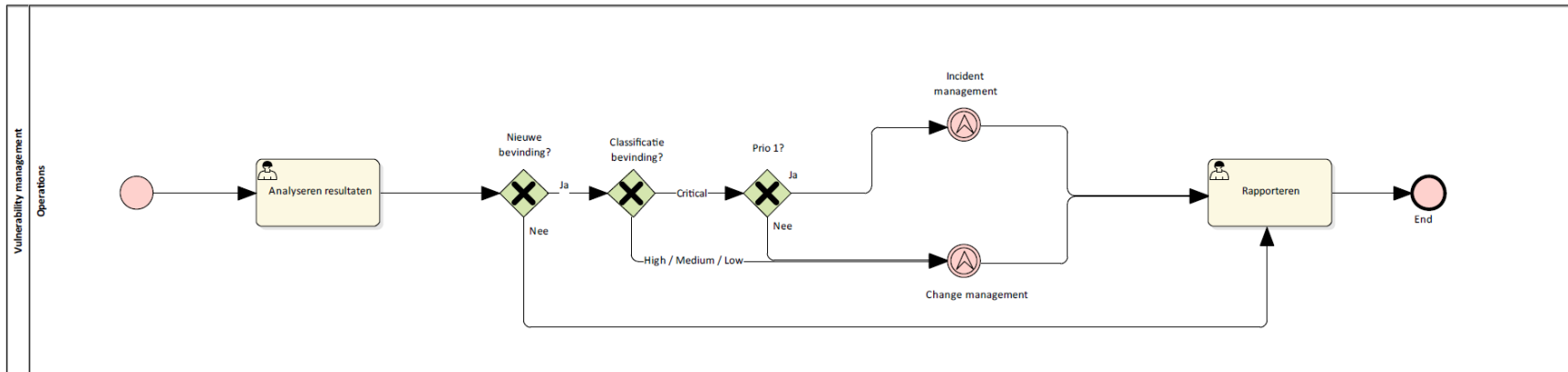
Model 10



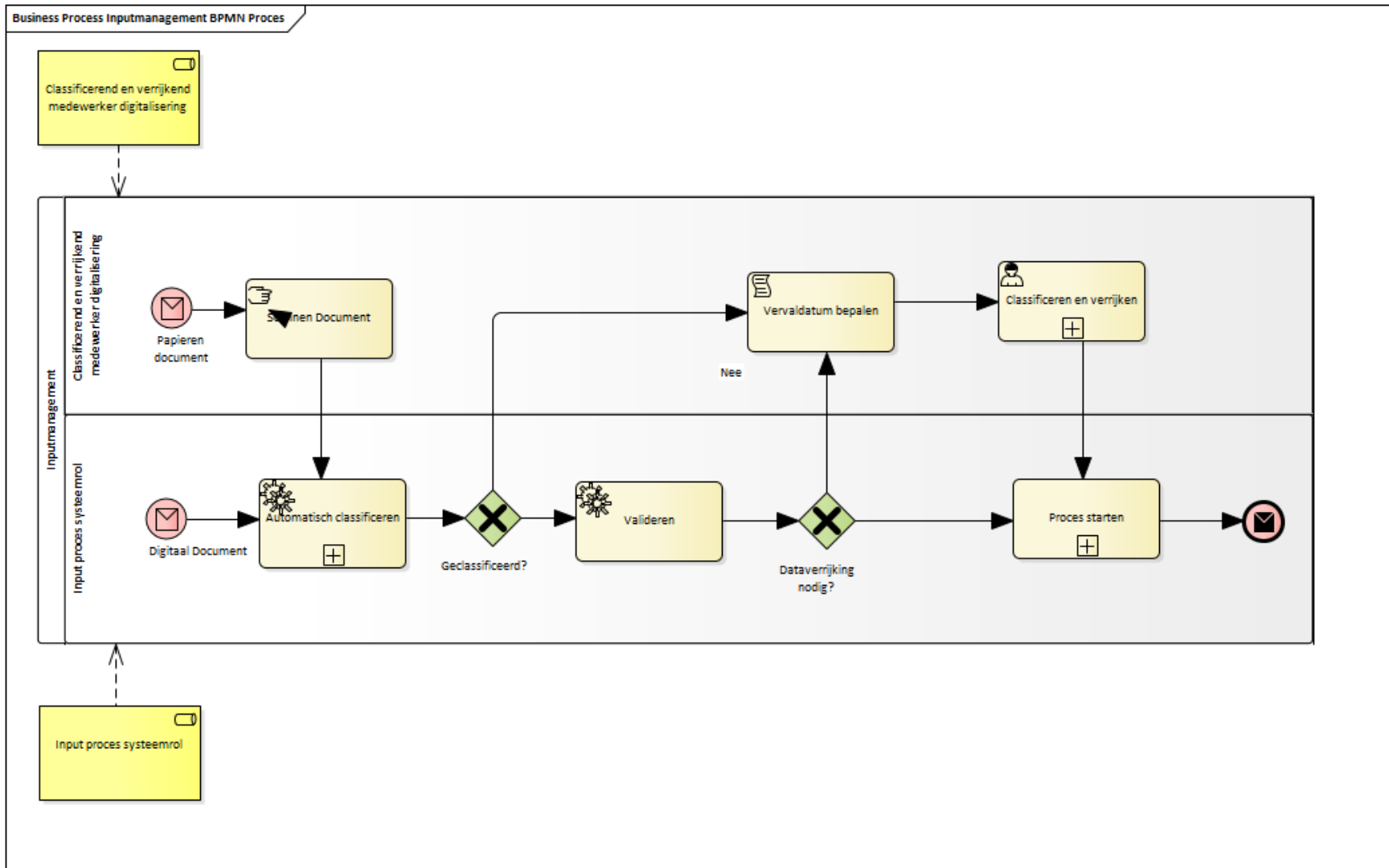
Model 11



Model 12

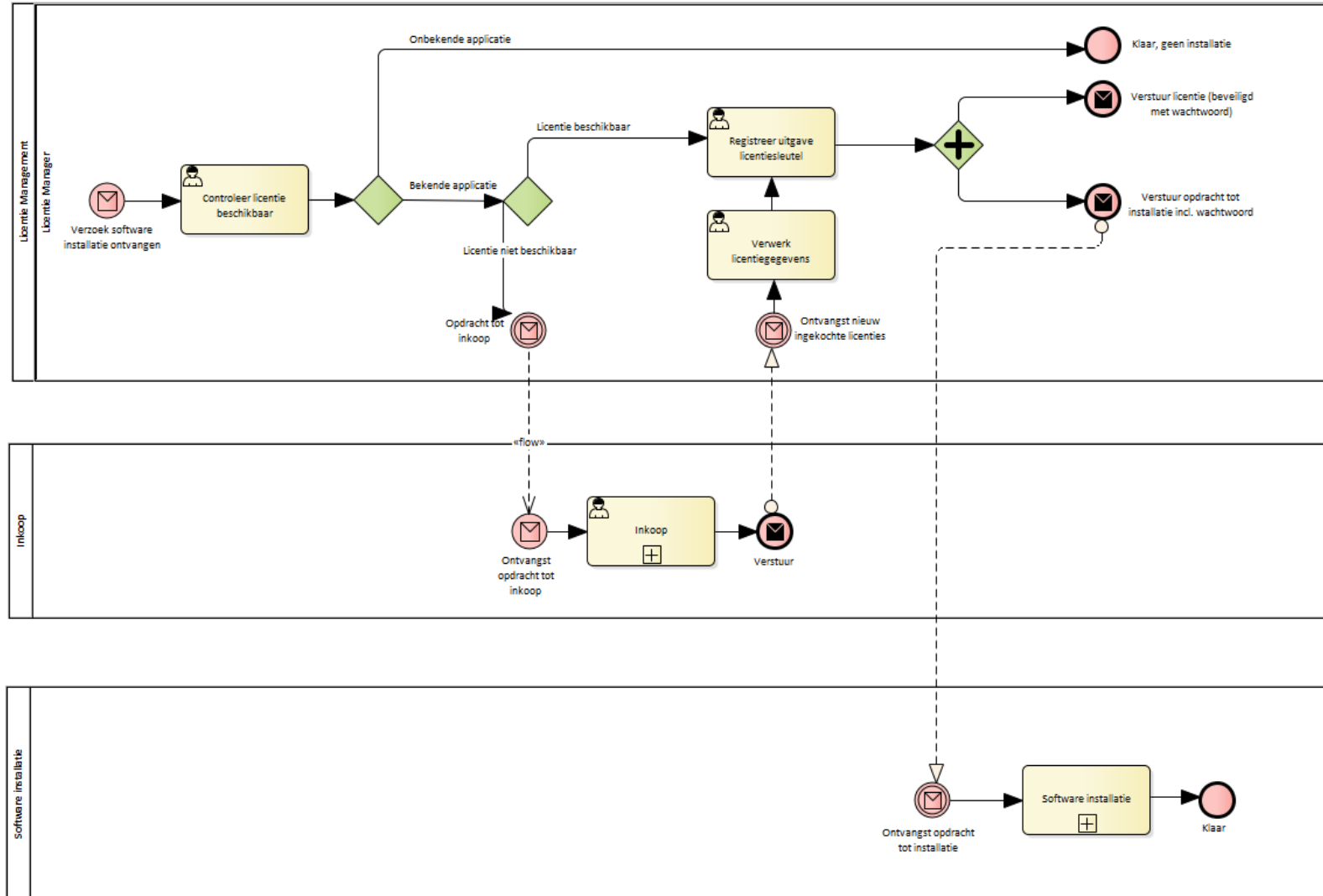


Model 13

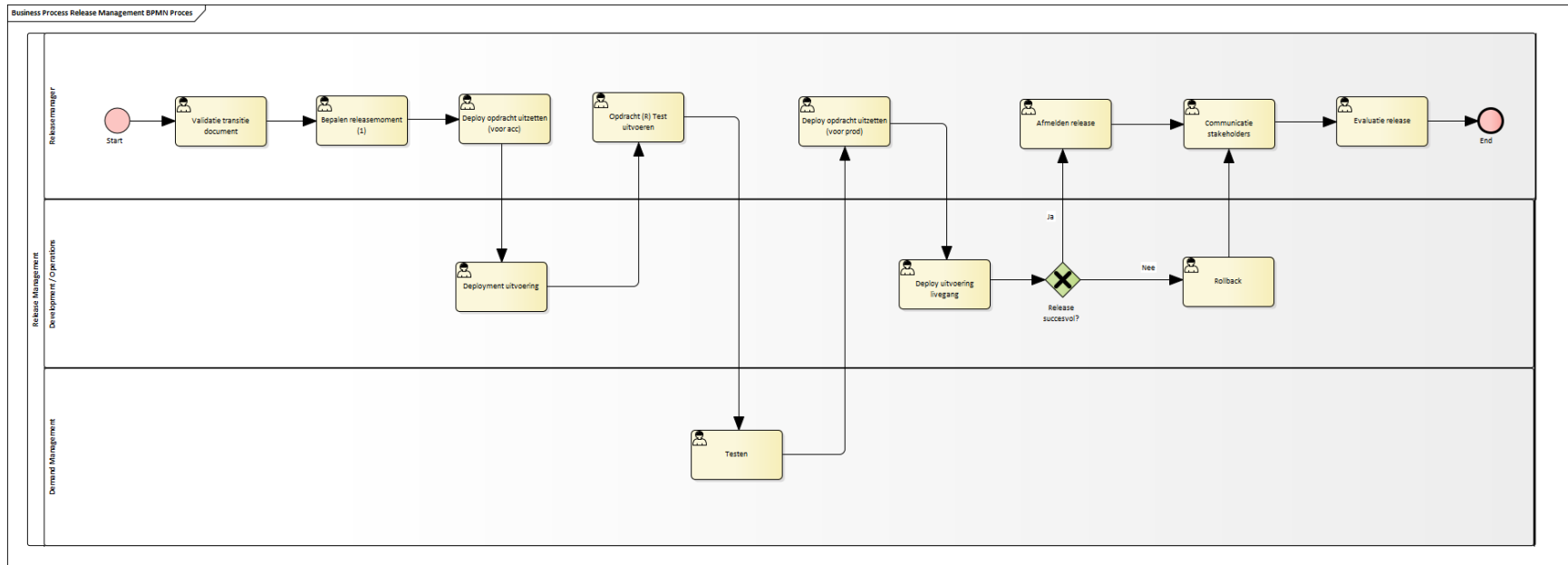


Model 14

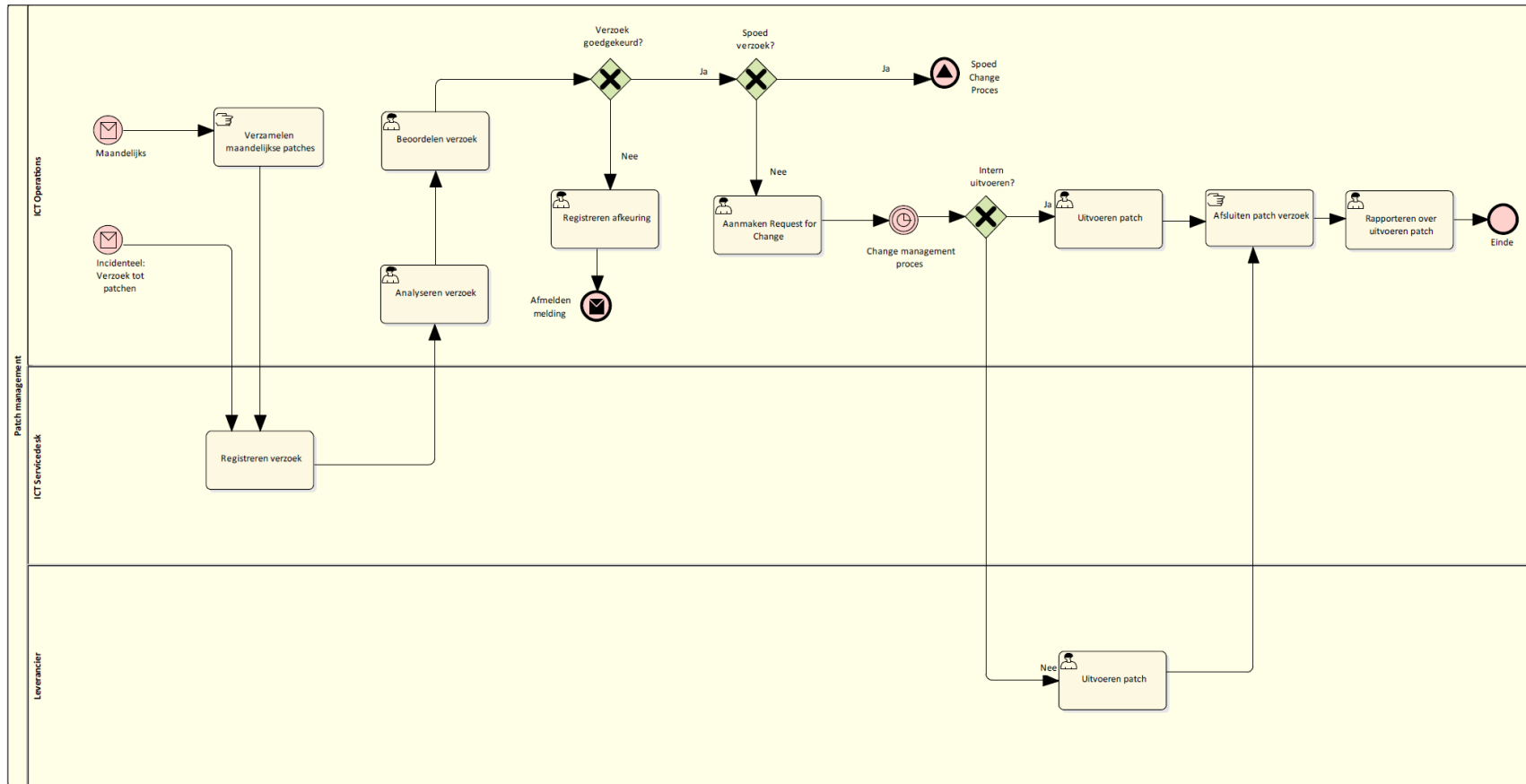
Business Process License Management BPMN Proces



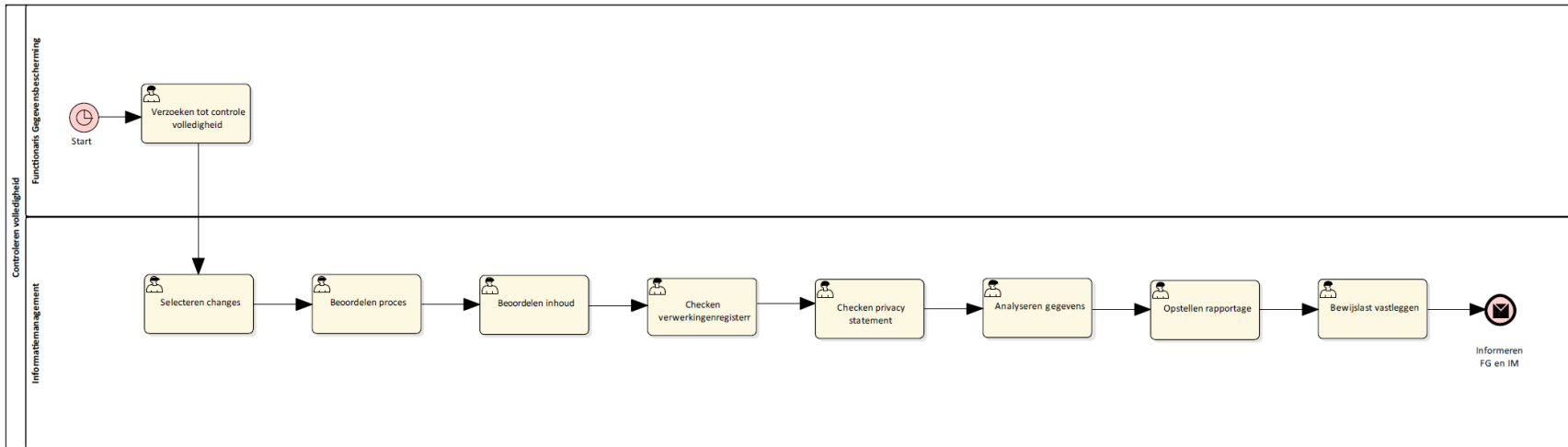
Model 15



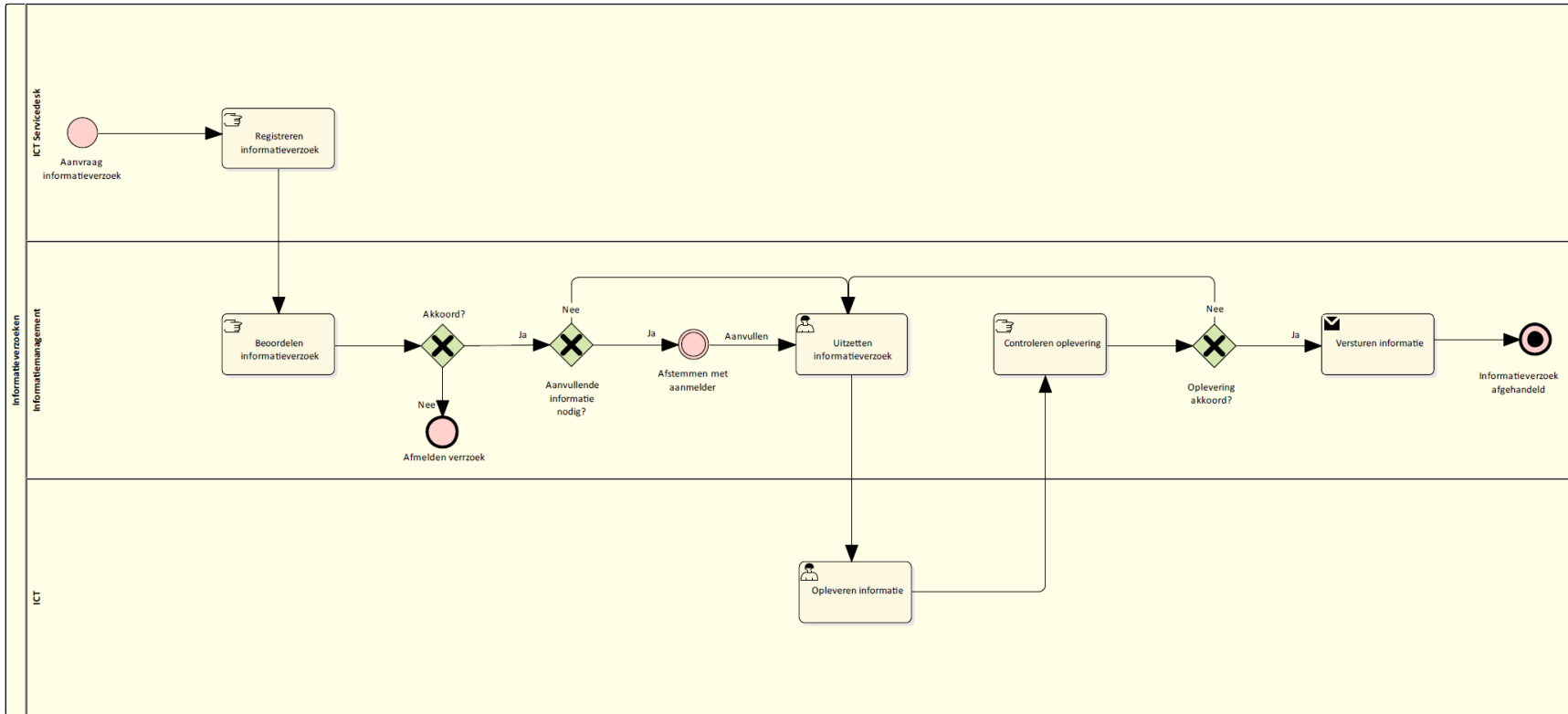
Model 16



Model 17



Model 18



Appendix VIII: Quality indicator values from all models

Copied quality metric values from models																		
Quality metric / Model number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
(GM) Gateway mismatch	1	1	2	1	2	5	4	0	2	2	3	3	2	3	1	3	0	3
Depth	1	2	3	1	1	4	4	0	2	5	3	3	2	3	1	3	0	3
(CNC) Coefficient of connectivity	0.50	0.42	0.44	0.48	0.45	0.52	0.51	0.45	0.52	0.56	0.50	0.55	0.52	0.42	0.50	0.50	0.48	0.52
(TNSF) Total number of sequence flows	26	19	20	12	13	22	18	5	12	20	16	11	12	15	15	19	10	14
(TNE) Total number of events	9	10	10	7	3	4	3	2	2	4	2	4	3	10	2	6	2	4
(TNG) Total number of gateways	3	3	3	1	2	5	4	0	2	5	3	3	2	3	1	3	1	3
(NSFG) Number of sequence flows from gateways	6	5	6	2	4	10	8	0	4	11	6	6	4	6	2	6	0	6
(CFCxor) Control flow complexity XOR	3	2	4	2	4	10	6	0	4	11	6	6	4	4	2	6	0	6
(CFCor) Control flow complexity OR	3	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
(CFCand) Control flow complexity AND	0	1	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

```

Sub Button2_Click()

    Dim WS_Total As Integer
    Dim I As Integer

    WS_Total = 18
    For I = 1 To WS_Total

        Worksheets("Totalen").Cells(3, 1 + I) = Worksheets(I).Cells(8, 4)
        Worksheets("Totalen").Cells(4, 1 + I) = Worksheets(I).Cells(9, 4)
        Worksheets("Totalen").Cells(5, 1 + I) = Worksheets(I).Cells(12, 4)
        Worksheets("Totalen").Cells(6, 1 + I) = Worksheets(I).Cells(13, 4)
        Worksheets("Totalen").Cells(7, 1 + I) = Worksheets(I).Cells(14, 4)
        Worksheets("Totalen").Cells(8, 1 + I) = Worksheets(I).Cells(15, 4)
        Worksheets("Totalen").Cells(9, 1 + I) = Worksheets(I).Cells(16, 4)
        Worksheets("Totalen").Cells(10, 1 + I) = Worksheets(I).Cells(17, 4)
        Worksheets("Totalen").Cells(11, 1 + I) = Worksheets(I).Cells(18, 4)
        Worksheets("Totalen").Cells(12, 1 + I) = Worksheets(I).Cells(19, 4)

    Next I

End Sub

```


Appendix IX: Interview results

Interview I1.

Zou je kort je huidige rol en de rollen die je voorheen hebt vervuld kunnen toelichten?

Mijn huidige rol bij het NVWA is Architect, waarbij ik met de invulling van die rol tussen de business en de IT in zit. Ik bouw zelf geen functionaliteiten, maar probeer vooral een vertaling te vinden voor processen en het applicatielandschap. Dit soort rollen heb ik meer gehad, bij diverse organisaties zoals bijvoorbeeld DICTU. Inmiddels ben ik zelfstandige, maar voordat ik zelfstandige ben geworden heb ik meer in architect en BI rollen gewerkt.

In hoeverre heb je ervaring met het gebruik van BPMN 2.0?

Ik ben er mee bekend, maar ik zou niet zeggen dat ik de syntax uit mijn hoofd ken. Ik heb er ook mee gewerkt, maar ik gebruik het op dit moment niet omdat ik geen processen aan het modelleren ben. Daarnaast heeft de NVWA gekozen voor ArchiMate, waar geen BPMN in voor komt. Vroeger heb ik ook nog wel met voorlopers van de tools en de taal gewerkt, ofwel modelleren met Splits, Joins, Ands et cetera. Ik heb er dus niet heel recent ervaring mee, maar ik ben er mee bekend en ik snap het.

Wat maakt een BPMN model in jouw ogen begrijpelijk?

Ik vind dit een lastige vraag. Omdat deze processen geanonimiseerd zijn en de processtappen leeg zijn, vind ik het dus lastig om te zeggen of ik het proces begrijp. Want dat maakt voor mij ook de begrijpelijkheid en niet bijvoorbeeld alleen de structuur of het aantal elementen. Als ik kijk naar model 3 bijvoorbeeld, geeft mij dat het gevoel van begrijpelijkheid. Het is een eenvoudig en overzichtelijk plaatje. En, wat ik hier belangrijk ik vind, daar kom ik straks op terug, is dat complexiteit naar mijn gevoel niet wordt gemedend. Want je kunt namelijk complexiteit niet weg moffelen. Daarnaast krijg ik het gevoel dat door de gele blokjes, de hoofd en bijzaken worden gescheiden, waardoor ik door kan klikken naar meer informatie.

Waarom heb je de modellen op deze volgorde geplaatst?

Afgezien van de restrictie die ik net al aangaf, dat de proces stappen leeg zijn, heb ik voornamelijk gekeken naar de structuur en overzichtelijkheid van het model. Het scheiden van hoofd- en bijzaken.

Model 4:

Dit model vind ik een flutmodel, maar het is uitiem begrijpelijk omdat het alleen maar lineair is. Ondanks dat er veel blokjes en pijltjes in zitten, vind ik dit niet synoniem met onbegrijpelijk. Er staat veel, maar het gaat bijna alleen maar van links naar rechts, met maar één aftakking. Bij mij staat begrijpelijk ook niet synoniem met goed, want ik vermoed dat dit een slecht model is. Ik kan niet precies duiden waarom, maar de zo een lineair model lijkt me bijna onmogelijk. Wat ik precies bedoel met een 'slecht' model, is dat ik het gevoel krijg dat deze niet klopt met de werkelijkheid.

Model 3:

Als ik terugga naar bijvoorbeeld model 3, zie ik bijvoorbeeld direct de swimlanes. Dit maakt het overzichtelijker, omdat ik direct zie wie de actoren zijn.

Hoewel je hierdoor soms wel in een teken technische kronkel kan komen, omdat je de blokjes in een dergelijke lane moet plaatsen. Ook vind ik dat hier het mindere aantal elementen toch wel mee speelt. Ook is dit model redelijk lineair. Daarnaast krijg ik het gevoel dat hier meer over na is gedacht, ondanks dat ik niet precies weet wat erin staat. Ook de ordening van het plaatje zelf speelt hier een grote rol in, mensen willen graag van links naar rechts en van boven naar beneden lezen. Als het plaatje hieraan voldoet voelt het als een begrijpelijker plaatje. Als je bijvoorbeeld kijkt naar model 4, had het startpunt bijvoorbeeld ook ergens midden in de plaat kunnen staan, toch doe je dat niet omdat het dan niet van links naar rechts leesbaar is. Wellicht dat het bij Arabische modellen wel anders is. Ook de grootte van de scope, vind ik wel invloed zijn. Dat is natuurlijk wel afhankelijk van wat er in de processtappen staat. Dus niet zozeer de grootte van het model, maar meer wat het proces of de proces stap beoogt. Hier zijn natuurlijk ook wel regels voor en kom je tot een bepaalde granulariteit van de stappen of het proces. Dit zijn denk ik keuzes van de ontwerper, maar dit heeft wel degelijk invloed op de complexiteit. Te grote stappen, kunnen te weinig detail bevatten en nietszeggend zijn, maar te kleine stappen kunnen weer te gedetailleerd zijn.

Model 1:

Om de een of andere reden komt dit model me over als een 'goed' model. Hier voelt het aan alsof complexiteit niet uit de weg is gegaan en het goed overeen komt met de werkelijkheid. Ik vind het geen makkelijk model, maar waarschijnlijk wel begrijpelijk als je erin duikt. Begrijpelijkheid en goed hebben natuurlijk niet altijd wat met elkaar te maken. Wat dit model wel begrijpelijker zou maken, is als bijvoorbeeld de onderste pool was 'ingeklapt' in één blokje, wat je zou kunnen uitklappen.

Model 2:

Ik kan het niet helemaal uitleggen waarom ik dit model begrijpelijker vind, waarschijnlijk is het toch de ordening van het plaatje op het eerste oog. Het heeft denk ik toch iets te maken met het lezen van linksboven naar rechtsonder. In dit model lijkt het stukje linksonder eerder te komen. Het model had denk ik begrijpelijker geweest wanneer de onderste swimming lane in het midden had gestaan.

Welke structurele eigenschappen van een model hebben in jouw ogen de meeste impact op de begrijpelijkheid? En waarom?

Dat vind ik lastig om te zeggen. Ik denk dat splits en joins in het algemeen het model complexer maken. Maar het grappige daarin vind ik wel, dat de notatie juist deze elementen gebruikt om het model begrijpelijker te maken. Ik denk dus dat wellicht de elementen die de semantiek duidelijker maken, de modellen complexer maken. Waarschijnlijk hangt het dus ook af van het type mens wat een model moet lezen. Als je kijkt naar model 4, welke ik als minst begrijpelijk heb aangemerkt, wil het niet per sé zeggen dat ik moeite heb met de elementen die erin staan. Ook als ik weer kijk naar de swim lanes, vind ik dit een begrijpelijkheid verhogend element. Maar ik denk dat opvolgende decision points de begrijpelijkheid weer verlagen, zoals in model 2.

Wat zou de modellen voor jou begrijpelijker maken?

Model 3:

Volgens mij niet zoveel eigenlijk, ik vind dit een begrijpelijk model.

Model 4:

Ook dit is bijna een grote lineaire flow, dus ik zou zeggen, dit kan bijna niet begrijpelijker. Misschien stappen samennemen als dat kan? Ik zie bijvoorbeeld dat de rol releasemanager drie dingen achter elkaar moet doen. Misschien kunnen die worden samengenomen tot één proces stap.

Model 1:

Eigenlijk wat ik net ook al aangaf, afhankelijk van wat je wil uitleggen het proces tonen. Voor mijn gevoel probeert het twee processen uit te leggen. Zoals hier is het bovenste afhankelijk van de onderste, maar de onderste helemaal niet afhankelijk van de bovenste. Het hangt een beetje van het doel af van de plaat, maar ik denk dat het te doen is om één van beide te tonen. Verder zie ik niks wat de begrijpelijkheid kan verhogen.

Model 2:

Hier hebben we het eigenlijk ook al over gehad. De swimming lanes omwisselen zou hier ten goede komen van de begrijpelijkheid en iets doen met de decision points die elkaar opvolgen. Ook hier geldt weer dat het samenvoegen van een aantal opeenvolgende activiteiten door dezelfde actor, indien mogelijk, ook de begrijpelijkheid vergroten.

<Uitleg van de interviewer over de berekeningen van de modellen, volgens de theoretische achtergrond en indicatoren. Er is hierbij ingegaan op de berekeningen en de threshold waardes voor de modellen die zijn uitgegeven aan de geïnterviewde.>

Wat vind je van de classificatie van het model ten opzichte van jouw eigen classificatie?

Ik ben natuurlijk geen rekenmachine, dus bij mij zit er een bepaalde mate van subjectiviteit in. Ik denk dat je namelijk de persoonlijke factoren niet kan uitrekenen. Al mijn jaren van modelleren in verschillende talen en het gevoel wat je krijgt bij zulk soort dingen, telt gewoon mee in hoe ik ernaar kijk. Ik heb bijvoorbeeld ook niet gecontroleerd of er wiebertjes zijn die ergens weer samen komen nadat ze gesplitst zijn. Ik denk dat de begrijpelijkheid ook bepaald wordt door een non-structurele component zoals persoonlijke factoren. Ook de ordening van het plaatje, wat niet voorgeschreven wordt door de taal, heeft hier wel invloed op. Misschien dat het verschil wat mijn oordeel ten opzichte van de berekening kan verklaren is dat ik onbewust toch een balans opmaak van de hoeveelheid elementen. Toch vind ik die proces logica die meer voorkomt in model 3, niet storend. Het is geen complexe logica, maar wel een beknopter plaatje.

Hoe denk je dat de classificatie deze beter kan aansluiten bij jouw ervaring van begrijpelijkheid van de modellen?

Ik denk dat number of gateways en number of sequence flows from gateways, voor mij wel bepalend zijn voor de begrijpelijkheid van de modellen. Wat ik hier wel in vind missen zijn de loops, ik vind dat dit de complexiteit en de begrijpelijkheid van een model zeker kan verhogen. Toch ben ik van mening dat de begrijpelijkheid niet in een model te vangen is, omdat ik er op een bepaalde manier naar kijk. Een model kan die afhankelijkheid daarin niet meenemen.

Interview I2

Zou je kort je huidige rol en de rollen die je voorheen hebt vervuld kunnen toelichten?

Ik ben informatie architect en op dit moment heb ik een opdracht bij de gemeente Amsterdam. Hier ben ik voornamelijk bezig met het uittekenen van bedrijfsprocessen, hoe die ondersteund worden door applicaties en hoe die vervolgens worden ondersteund door de infrastructuur. In vorige opdrachten heb ik meer gewerkt aan de detailuitwerking van applicaties, zowel de functionele- als de technische werking.

In hoeverre heb je ervaring met het gebruik van BPMN 2.0?

Ik heb nooit echt de notatie geleerd, maar ik denk dat ik een intuïtief begrip heb van wat de notatie betekent. Vergelijkbare proces notaties, zoals UML, die lijken er erg op, maar de componenten die in BPMN zitten en niet in Activity Diagrams zitten weet ik niet de exacte betekenis. Ik heb zelf ook nooit BPMN-modellen gemaakt.

Wat maakt een BPMN-model in jouw ogen begrijpelijk?

Het eerste wat me op viel, was dat er geen teksten in de blokjes stonden. Het eerste waar ik namelijk naar zoek is een vergelijking met de werkelijkheid. Ik probeer me dan voor te stellen dat ik er doorheen loop in één van de rollen en het werk uitvoer. Ik denk dat een model vooral begrijpelijk is wanneer ik me kan voorstellen dat ik dat in het echt doe. Ik denk dat de teksten in de activiteiten hier ook een bijdrage in hebben, deze moeten zo nauwkeurig mogelijk beschrijven wat er moet gebeuren. Ook als ik kijk naar model 4, waar één keer een keuze in zit, vind ik deze vrij simpel omdat alles gewoon lineair is. Dan maakt het aantal activiteiten ook niet zoveel meer uit.

Waarom heb je de modellen op deze volgorde geplaatst?

Model 4:

Zoals ik eerder aangaf, is dit model erg lineair. Er zit één keuze in met een beschrijvende tekst, dus ik snap wat er gebeurt in dit model. Ik zou bij wijze van spreken de teksten in de activiteiten al kunnen invullen.

Model 3:

Dit model vond ik ook best begrijpelijk, dus eigenlijk heb ik ook best wel getwijfeld tussen model 3 en model 4. Ik vond ze bijna even eenvoudig. Het enige wat model 3 wat minder begrijpelijk maakte zijn de verschillende soorten activiteiten. Bijvoorbeeld, een handje, een poppetje en een document. Ik weet zelf niet helemaal wat het verschil is. Misschien dat als de tekst erin zou staan dat ik het beter zou begrijpen. Ik vermoed dus dat het proces niet zo ingewikkeld is, maar dat kan het niet zo goed beoordelen. Daarnaast begrijp ik ook niet helemaal waarom de rol van de swimlane ook nog in het gele blokje wordt gepresenteerd. Dat lijkt me dubbelop. Kortom, de structuur vind ik niet direct complexer, maar door de andere notatie elementen is het iets complexer dan model 4.

Model 2:

In model 2 zie ik een soort gestippelde pijl van de onderste swimlane naar de bovenste swimlane, en ik weet niet helemaal wat het verschil is tussen de normale pijltjes. Misschien is dit een ander soort informatie-uitwisseling en dan vind ik het complex dat er meerdere informatie-uitwisselingen door elkaar heen lopen. Ook vind ik hier de klokjes linksboven niet helemaal duidelijk, ik vraag me af of deze bijvoorbeeld ook allebei kunnen gebeuren. Eigenlijk snap ik bij dit model niet meer wat er gebeurt en dit komt met name door de gestreepte pijlen.

Model 1:

Ook hier geldt weer dat de gestippelde pijlen voor mij verwarrend zijn, maar hier had ik er wat meer problemen mee. Bij de tweede activity heb je bijvoorbeeld een blauw klokje en een normale pijl die ergens heen gaat. Stel je volgt de pijl van het klokje, ben je daarna dan klaar? Of kun je ook nog door naar andere activiteiten? Ook zie ik dat er 2 pijltjes binnen komen, en dan vraag ik me af of allebei de voorgaande activiteiten gedaan moeten zijn voordat deze activiteit kan starten. Ook zie ik nog een ruitje met een rondje erin, dit lijkt op een keuze, maar ik weet niet precies wat dit betekent. Daarna zie ik weer klokjes met verschillende teksten, zou dit dan ook allebei kunnen? Lopen er dan 2 activiteiten naast elkaar? Hier zitten allemaal dingen in waarvan ik denk dat er verschillende interpretaties mogelijk zijn. Kortom, er zitten in model meer elementen in dit model ten opzichte van model 2, waarvan ik niet helemaal zeker ben van de betekenis.

Welke structurele eigenschappen van een model hebben in jouw ogen de meeste impact op de begrijpelijkheid? En waarom?

Persoonlijk maakt het mij niet heel erg uit hoeveel blokjes erin staan of hoe de lijntjes precies lopen. Ik denk dat elementen zoals het klokje of een ruitje met een rondje erin voor mij complexiteit toe voegen omdat ik niet helemaal zeker ben van de betekenis. Ook wanneer er meerdere pijlen een activiteit uit- of in komen, of er gebruik wordt gemaakt van verschillende soorten pijlen. Ik ga me hierdoor afvragen wat precies het verschil is en waarom het uitmaakt. Ook loops hebben hier wel impact op, niet direct met één loop, maar vooral wanneer er meerdere loops zijn. Het kan dan sneller zijn dat je niet precies meer weet waar je bent in het model. Wanneer ik bijvoorbeeld kijk naar model 1 en model 4, moet ik bij model 1 meer nadenken omdat er meer vertakkingen in zitten. Ik denk dat deze vertakkingen en loops het model dus intrinsiek complexer maken, ondanks dat je de notatie kent of niet. Model 4 daarentegen is vrij recht toe recht aan.

Wat zou de modellen voor jou begrijpelijker maken?

Model 4:

Het enige wat ik hier zou kunnen bedenken is het terugplaatsen van de teksten in de activiteiten. Verder zie ik geen structurele complexiteit in dit model. Er zit maar 1 branch in, en die is volledig duidelijk.

Model 3:

Hier weet ik gewoon niet wat de verschillende soorten activiteiten zijn, en of het uitmaakt dat deze verschillend zijn getekend. Verder is de flow me wel duidelijk. Wel zou ik de gele blokjes met de rollen weghalen, dit voegt niet zoveel toe. En natuurlijk de teksten toevoegen, maakt het ook helderder.

Model 2:

In dit model is het me niet helemaal duidelijk wanneer het proces start. Ik snap wel dat er gestart wordt met het proces nadat er feedback binnen is gekomen, maar het eerste blokje start snap ik niet. Verder heb ik het vermoeden dat er wat stappen worden overgeslagen, met name bij het herinnering gedeelte. Het lijkt me dat er nog wat complexiteit verstopt zit in dit blokje, en dat wanneer je dit verder uit zou vouwen het duidelijker wordt. Met name door de hoeveelheid pijltjes die er in- en uitgaan vermoed ik dat er wat complexiteit verborgen zit. De rest van het model ziet er dan een stuk simpeler uit. Het valt me ook op dat er bij medewerker certificering 2 soorten activiteiten voorkomen, 2 met een poppetje en 1 met tandwielletjes. Dit snap ik niet helemaal, omdat de tandwielletjes ook bij het systeem staan. Ik zou dus verwachten dat dit een geautomatiseerde activity is.

Model 4:

Ook hier denk ik dat ik de meeste problemen heb doordat ik de notatie niet helemaal ken. Eigenlijk een beetje hetzelfde verhaal als bij model 3. Zoals ik ook al eerder aangaf vind ik het ook vreemd dat er een activiteit is zonder uitgaande pijl, in de systeem swimlane.

<Uitleg van de interviewer over de berekeningen van de modellen, volgens de theoretische achtergrond en indicatoren. Er is hierbij ingegaan op de berekeningen en de threshold waarden voor de modellen die zijn uitgegeven aan de geïnterviewde.>

Wat vind je van de classificatie van het model ten opzichte van jouw eigen classificatie?

Ik denk dat het verschil in volgorde grotendeels te verklaren is doordat het model niet de kennis van de notatie mee neemt. In mijn beantwoording heb ik namelijk veel gewicht gegeven aan het feit dat ik door mijn gebrekkige kennis van BPMN de dingen niet begrijp. Ik kan het me wel goed voorstellen dat model 1 complexer wordt bevonden wanneer je puur kijkt naar de structurele elementen. Omdat dit model een loop heeft. Zoals in model 2, snap ik dat de gateways punten opleveren, maar die vind ik nou juist weer niet zo complex. De complexiteit zit wat mij betreft ook in het evalueren van het feit of je links of rechts afslaat, maar die staat niet weergegeven in het model.

Hoe denk je dat de classificatie deze beter kan aansluiten bij jouw ervaring van begrijpelijkheid van de modellen?

Ik zou verwachten dat de modellen op een schaal van 1 tot 4, verder uit elkaar liggen. Ik vind dit erg weinig discriminerend. Wat ik me bijvoorbeeld afvraag, als ik model 4, 2 keer zo lang maak, vind ik deze net zo begrijpelijk. Dus ik zou verwachten dat deze dan dezelfde score zou krijgen of maar marginaal afwijkt. Ook zou ik op deze schaal model 1 en model 2 al meer aftrek willen geven. Ik zie ook niet zo goed waarom dat verschil van diepte zo groot is tussen model 1 en 2. Ik vind het namelijk niet zoveel complexiteit toevoegen. Wat ik wel verschil vind maken zijn het aantal beslissingen. Je moet namelijk deze beslissingen evalueren. Voor al deze modellen wordt nu dezelfde score gegeven waardoor ze in dezelfde categorie vallen, terwijl ik wel vind dat het een onderscheid zou moeten maken. Ik denk ook dat je redelijk los moet gaan in een model om op 10 of 17 gateways te komen. Kortom, ik ben het eens met de indicator, maar niet met de schaal. Ook het aantal sequence flows from gateways, voelt een beetje dubbelop, gezien het vaak 2 of 3 uitgaande pijlen zijn. Als ik verder kijk bijvoorbeeld naar model 4, daar zit een gateway mismatch omdat de pijlen niet meer bij elkaar komen in een join. Toch vind ik dit niet afdoen aan de begrijpelijkheid, omdat ik prima snap dat maar 1 van de 2 routes gevolgd kan worden. Een join hier tussen zetten zou voor mij niks toevoegen. Bij model 2 vind ik ook dat de message flow een bepaalde mate van complexiteit toevoegd, omdat deze midden in een proces binnen komt en niet aan de start van een proces. Het roept bij mij dan ook vragen op, of de rest van het proces al gestart moet zijn, of dat het proces vanaf daar wordt vervolgd.

Interview I3

Zou je kort je huidige rol en de rollen die je voorheen hebt vervuld kunnen toelichten?

Op dit moment ben ik als informatieanalist werkzaam bij het NVWA, in die hoedanigheid ben ik betrokken bij een aantal ontwikkeltrajecten. Diezelfde rol heb ik ook gehad in het ontwikkeltraject wat hiervoor bij het NVWA heeft gelopen. Daarvoor ben ik ook als informatieanalist betrokken geweest bij andere grotere projecten, met name binnen de overheid.

In hoeverre heb je ervaring met het gebruik van BPMN 2.0?

Ik ken de notatie omdat ik het langs heb zien komen tijdens de ontwikkeling van Inspect, en vanuit de ondersteuning vanuit verschillende toolingen. Ik ken het niet zo goed dat ik het zelf toe heb gepast binnen projecten. Dus de symbooltjes ken ik vast niet allemaal, maar ik ben denk ik wel in staat om het te plaatsen en te begrijpen tot op een zeker niveau. Je ziet namelijk ook veel herkenbare symbolen toch komen uit andere talen, zoals UML.

Wat maakt een BPMN-model in jouw ogen begrijpelijk?

Dat ik kan begrijpen voor welk doel of doelgroep die het model dient. Een voorbeeld van een doel kan zijn is een plaat gebruikt die wordt voor communicatie met stakeholders. Maar ook hier is het weer afhankelijk van welke stakeholders. Als het stakeholders uit de business zijn, zou ik een minder complex model gebruiken. Dus een procesmodel mag best complex zijn in notatie, of groot zijn in omvang, als het doel is om het grote geheel duidelijk te maken. Maar als het doel is om een proces in hoofdlijnen duidelijk te maken, is een complexe plaat juist niet begrijpelijk. Voor mij zelf is een proces plaat begrijpelijk wanneer ik alle gebruikte symbolen en technieken ken, dat ik het hele plaatje zelf kan vertalen naar inhoud. Ik wil me dan niet af hoeven vragen wat er met een symbool wordt bedoeld, maar ik wil me alleen af hoeven vragen waarom een bepaalde stap door een bepaalde actor wordt uitgevoerd.

Waarom heb je de modellen op deze volgorde geplaatst?

Model 4:

Deze had ik als meest begrijpelijke omdat ik me hier niet af hoeft te vragen wat de symbolen betekenen, en er worden relatief weinig symbolen gebruikt. Los van de betekenis van de blokjes, kan ik denk ik goed inschatten hoe het proces loopt.

Model 3:

Dit model heb ik als minst begrijpelijk geclassificeerd, omdat ik uit dit model niet goed kan opmaken wat de relatie is met de architectuur services of componenten. Dat komt door de gele blokjes die boven en onder het procesmodel staan. Ik denk dat als die gele blokjes er niet waren geweest, dat ik deze hoger zou hebben geplaatst. Maar als ik het geheel beschouw, inclusief die blokjes, dan vind ik deze het minst begrijpelijk.

Model 1 & 2:

Ik vind model 1 en 2 redelijk van hetzelfde niveau, maar bij het eerste model kan ik de betekenis van de symbolen wat beter plaatsen, zonder dat ik ze precies ken. Bij model 2 had ik wat meer moeite met de relatie tussen de bovenste en de andere swimlane. Ik moest me hier meer afvragen hoe die relatie precies is en hoe deze zich verhoudt. Daarnaast zie ik ook een aantal symbolen met een envelop, en ik vroeg me af of deze niet meer een relatie met elkaar zouden hebben. Dat de ene envelop een ingaande is van een andere envelop, maar dat kon ik op het eerste gezicht niet goed achterhalen. Voornamelijk dus de eind- en begin symbooltjes, niet zo zeer het aantal, maar meer dat ik uit moet gaan puzzelen hoe de symbolen zich tot elkaar verhouden.

Welke structurele eigenschappen van een model hebben in jouw ogen de meeste impact op de begrijpelijkheid? En waarom?

Dit vind ik een lastige vraag, maar zoals ik net eigenlijk ook al aan gaf, kunnen de events het in bepaalde gevallen voor mij minder begrijpelijk maken. Met name wanneer ik dus moet zoeken naar wat de relatie tussen deze symbolen is. Toch kunnen deze events ook de begrijpelijkheid vergroten vind ik, dit omdat het door de notatie duidelijk wordt gemaakt wat er met een symbool wordt bedoeld. Dit gaat dan denk ik wel meer over de begrijpelijkheid van het proces dan van het model. Ook vind ik dat een groot model niet altijd minder begrijpelijk hoeft te zijn, als het ook een groot en complex proces is. In principe zou ik gewoon goed in staat moeten zijn om een procesmodel te lezen wanneer deze gewoon van linksboven naar rechtsonder gaat wanneer de swimlanes en notatie goed gebruikt worden en wanneer je niet te veel heen en weer gaat met de lijntjes. Wel is het belangrijk dat er consistent gebruik wordt gemaakt van de soorten activiteiten en de benoeming daarvan, anders komt dit niet ten goede van de begrijpelijkheid. Bijvoorbeeld, als er de ene keer staat “verwerk polis” en de andere keer staat er “typ klantnummer in”, zijn dat 2 verschillende manieren om iets aan te duiden. De omvang van een activiteit is dan heel anders, dus wordt er geen eenduidig gebruik van gemaakt.

Wat zou de modellen voor jou begrijpelijker maken?

Model 1:

Allereerst natuurlijk het toevoegen van de beschrijvingen zou het voor mij begrijpelijker maken, maar dat geldt voor alle modellen. Verder is het enige wat ik voor dit model zou zeggen, dat er blijkbaar een volgorde zit tussen het bovenste en het onderste gedeelte, ik zou het dan fijn vinden wanneer dit in de leesrichting staat in plaats van gestapeld. Hier wordt het natuurlijk wel een langer model van.

Model 2:

Hier vraag ik me af of alle relaties wel duidelijk en volledig zijn weergegeven. Het onderste gedeelte zit meer een soort flow in, en het bovenste gedeelte niet. Ik denk dat hier vooral een beschrijving het model duidelijker zou maken. Het voelt namelijk niet heel eenduidig aan omdat het bovenste blokje een grote losstaande activiteit is, maar dat het totaal wel weer 1 proces is.

Model 3:

Dit zit voornamelijk op het gebruik van die externe gele blokjes. Deze zou ik zeker weglaten om de begrijpelijkheid te vergroten. Het voegt immers niks toe gezien het dezelfde tekst is als dat in de actor lane staat.

Model 4:

Het is eigenlijk al een begrijpelijk model. Wel zou ik graag willen dat als er events plaats vinden, dat deze opgenomen zouden worden. Niet dat deze weg gelaten worden om de begrijpelijkheid te vergroten.

<Uitleg van de interviewer over de berekeningen van de modellen, volgens de theoretische achtergrond en indicatoren. Er is hierbij ingegaan op de berekeningen en de threshold waardes voor de modellen die zijn uitgegeven aan de geïnterviewde.>

Wat vind je van de classificatie van het model ten opzichte van jouw eigen classificatie?

Ik vind model 4 het meest eruit springen qua begrijpelijkheid, maar dat komt ook overeen zo te zien. Waardoor ik model 3 onderaan heb staan zie ik natuurlijk niet terug in de scoring, omdat de gele blokjes denk ik niet heel zwaar meewegen in het model, maar voor mij wel. Ik denk wel dat als ik deze elementen buiten beschouwing had gelaten dat het model wel anders gepositioneerd was geweest en dat de kans groter was geweest dat het overeen was gekomen. Als ik kijk naar de rest van de volgorde; 4, 1, 2, komt dat overeen. Alleen model 3 zit er dan een tussen bij mij. De scoring lijkt dus eigenlijk wel overeen te komen met mijn waarnemingen.

Hoe denk je dat de classificatie deze beter kan aansluiten bij jouw ervaring van begrijpelijkheid van de modellen?

Dat ligt dus met name bij de gele blokken uit model 3, maar ik weet niet of dit meegewogen moet worden in het model. Gezien ik nu beter weet wat het betekent, vind ik het model best begrijpelijk omdat deze eigenlijk niet heel complex is. Verder zou voor mijn gevoel het verschil tussen 4 en 3 ten opzichte van 1 en 3 anders liggen. Voor mijn gevoel zouden model 3 en 4 dan dicht bij elkaar moeten liggen. En niet zoals nu, dat model 1 en 3 dicht bij elkaar liggen. Als ik kijk naar de indicator depth, vind ik deze wel enorm verschillen tussen model 1 en 2. Op zich vind ik het wel een goede indicator, maar in dit geval vind ik het niet heel erg afdoen aan de begrijpelijkheid. Misschien dat hier de schaal anders zou moeten zijn, maar ik durf niet direct te zeggen hoe.

Interview I4

Zou je kort je huidige rol en de rollen die je voorheen hebt vervuld kunnen toelichten?

Na mijn opleiding bedrijfskundige informatica ben ik begonnen als object-orientatedontwikkelaar. Daarna ben ik al gauw ontwerper geworden en vervolgens doorgegroeid naar informatieanalist. Sinds ongeveer 6 jaar werk ik als business analist en ben ik ook 2 jaar product owner geweest.

In hoeverre heb je ervaring met het gebruik van BPMN 2.0?

Een jaar of 6-7 geleden heb ik een cursus hiervan gevolgd van Bruce Silver zelf. Ik heb hier ook een examen in gedaan en een certificaat gekregen. Dus de notatie als theorie ken ik. Ook passen we bij mijn huidige werkgever BPMN deels toe, maar UML 2.0 heeft toch wel onze en mijn voorkeur. Ik werk er dus wel mee, maar niet helemaal in de volle potentie. Ik gebruik het met name voor het uitwerken van processen, op het hoogste abstractieniveau. Met name het in kaart brengen van de interactie tussen verschillende systemen of contexten werkt hierin erg goed. Ondanks de mogelijkheid dat er met BPMN ook ingezoomd kan worden, doe ik dit liever met een activiteiten diagram in UML 2.0. Kortom, ik gebruik met name de activiteiten, decisions en begin- en eindpunten. Die gebruik ik om in eerste instantie het proces in kaart te brengen op een bepaald abstractieniveau en dat te valideren met de gebruikers, om vervolgens te kunnen verdiepen.

Wat maakt een BPMN-model in jouw ogen begrijpelijk?

Een begrijpelijk BPMN-model is naar mijn mening een model met een duidelijke flow: heldere start en helder eind. Ook het aantal componenten in een diagram dient overzichtelijk te zijn. Daarnaast prevaleer ik een begrijpelijk model boven een geheel technisch correct model.

Mijn modellen moeten begrijpelijk zijn voor gebruikers zonder al te veel technische achtergrond. Dat houdt ik altijd voor ogen bij het modelleren van processen.

Waarom heb je de modellen op deze volgorde geplaatst?

Ik heb hierbij eigenlijk het gebruikersperspectief in ogenschouw genomen. Zoals ik net aangaf, als ik modellen maak, doe ik dat vooral voor de gebruikers in de business. Als ik kijk naar de volgordelijkheid, heb ik dus met name begrijpelijkheid van de gebruiker bovenaan gezet. Hoe meer details en hoe meer gebruik je dus maakt van verschillende symbolen die voldoen aan de BPMN 2.0 notatie, hoe verwarrender het voor de gebruiker is naar mijn mening. Ik wil dat mijn procesplaten begrijpelijk zijn en dat ik niet elk symbooltje hoeft uit te gaan leggen. Dat leidt alleen maar af van de essentie van het model.

Model 4:

Ik vind dit een overzichtelijke plaat, je hebt de swimlanes, volgordelijkheid van activiteiten en een decision. Ik vind dit proces goed te volgen. Het heeft 1 start, een logische flow en een eind. Dat is volgens mij prima te begrijpen.

Model 2:

Ook hier weer een duidelijke start, en flow. Maar hier zie je al meer iconen, vooral linksboven is het druk met activiteiten en icoontjes. Dat maakt het wat lastiger, maar op zich is er nog wel een duidelijke stroom die te volgen is. Hier mis ik wel een beetje de samenhang tussen de start en ontvangst aanvraag. Dit lijkt los te staan van elkaar. Zelf heb ik de voorkeur aan een vloeiend proces. Wanneer er meerdere startpunten zijn, vind ik dat dit beschreven moet zijn of dat er 2 procesmodellen van gemaakt moeten worden.

Model 3:

Op zich vind ik 2 en 3 uitwisselbaar. Ik vind alleen die gele beschrijving verwarrend werken. Ook mis ik hier eigenlijk wel de merge. Ik zie graag merges waar de pijlen weer bij elkaar komen, na dat er splitsingen zijn. Overall is het best een klein en overzichtelijk plaatje. Maar zoals ik net ook al aangaf, vind ik het ook belangrijk dat er maar 1 start is. Deze punten maken dat ik dit model onder model 2 heb geplaatst.

Model 1:

Dit is een druk schema en het eerste wat me verbaast, zijn 2 startpunten in verschillende lanes. Ik weet niet of dit 2 verschillende systemen zijn, dus dit vind ik verwarrend. Het is voor mij onduidelijk waar het proces begint. Ook vind ik dat het systeem, wat buiten jouw systeem valt, niet in detail uitgewerkt moet worden. Dat is de verantwoordelijkheid van het externe systeem, en dat wordt hier wel weer uitgewerkt. En ik vind dat hier veel keuzes in zitten, ik kan me voorstellen dat deze plaat veel vragen op gaat leveren bij gebruikers met een niet-technische achtergrond. Daarom heb ik deze op de laatste plek gezet.

Welke structurele eigenschappen van een model hebben in jouw ogen de meeste impact op de begrijpelijkheid? En waarom?

Ik denk dat dit heel erg publiek afhankelijk is. Als je een ontwikkelteam hebt, die de notatie goed beheerst, dan kun je heel veel business rules en eisen kwijt in het model. Dit zie je ook wel een beetje terugkomen in model 1. Zelf zit ik niet meer in de aansturing van ontwikkelteams, maar meer in de hoek van business change. In dit geval ga ik niet veel detail invoeren omdat ik meer bezig ben met het groter geheel te beschrijven. Dan zijn swimlanes, activiteiten en decision points in combinatie met een duidelijke start en stop al voldoende. Als ik meer ga toevoegen, open ik deuren voor meer complexiteit en detail discussies. Dus ik gebruik bewust ook maar een beperkte hoeveelheid en soort objecten.

Wat zou de modellen voor jou begrijpelijker maken?

Model 4:

Hier zouden, op het niveau waarop ik acteer, het aantal activiteiten nog iets worden verminderd. Door bijvoorbeeld de eerste activiteiten te aggregeren. Dit zorgt voor een abstractere beschrijving. Meer eigenlijk niet.

Model 2:

Hier vind ik het linksboven erg druk, met al die time events en ontvangen en sturen. Verder vind ik het een duidelijke plaat. Waarbij het nog wel opvalt dat er weer 2 start events zijn. Kortom, ik vind dat linksboven herzien kan worden en dat het reduceren van startmomenten de plaat begrijpelijker maken.

Model 3:

Wat ik hier een beetje slordig aan vind, zijn de decisions zonder antwoorden. Met name bij die tweede dataverrijking vind ik dit nodig. Ook hier geldt weer, 2 start momenten, ik weet niet helemaal of dit nodig is. En die gele blokken, ik heb geen idee wat die toevoegen. Die teksten staan ook gewoon in de swimlane.

Model 1:

Wat dit proces in ieder geval duidelijker zou maken is aangeven waar het proces begint, 1 duidelijke start. En ik vraag me af wie de onderste lane representeert. Wie is die actor? Er is duidelijke een bepaalde betrokkenheid, maar ik maak uit het model niet wat op wat die is. Verder vind ik de gateway erg onduidelijk, het bolletje wat er in staat zegt mij helemaal niks.

<Uitleg van de interviewer over de berekeningen van de modellen, volgens de theoretische achtergrond en indicatoren. Er is hierbij ingegaan op de berekeningen en de threshold waardes voor de modellen die zijn uitgegeven aan de geïnterviewde.>

Wat vind je van de classificatie van het model ten opzichte van jouw eigen classificatie?

Syntactisch gezien zal het ongetwijfeld kloppen dat model 2 minder goed scoort. Maar op deze manier heb ik ze niet beoordeeld. Ik heb bijvoorbeeld niet gekeken naar of het klopt dat splits opgevolgd worden door een join. Toch vind ik het voor de integriteit van het model wel goed dat zoiets wordt gecontroleerd. Ik moet ook veel meer moeite doen om model 1 te begrijpen ten opzichte van model 3, omdat model 1 een veel uitgebreider model is. Er zijn bijvoorbeeld meer handelingen en meer verschillende soorten notatiewijzen die gebruikt worden.

Hoe denk je dat de classificatie deze beter kan aansluiten bij jouw ervaring van begrijpelijkheid van de modellen?

Er zit denk ik een verschil tussen of iets begrijpelijk is en of het correct is volgens de structuur. Zoals voorgaand voorbeeld over de splits en joins, wanneer de pijlen niet in een join samen komen maar op een andere manier, zou het nog prima begrijpelijk kunnen zijn. Ook als ik kijk naar de total number of gateways, die in elk model gelijk wordt beoordeeld, denk ik dat dit niet fijnmazig genoeg is. Wanneer je een model hebt met 5 gateways, denk ik niet dat het heel erg bevorderlijk is voor de leesbaarheid. Dus ik vraag me af of de schaal niet te grofmazig is. Vooral op het niveau waar ik op acteer, is iedere gateway altijd een punt van afleiding of discussie. Ik vind verder de scores ook wel wat dicht bij elkaar liggen, relatief gezien klopt het natuurlijk wel, maar ik vraag me af hoe erg je dit wil benadrukken. Ik denk dat vooral model 2 hierin wel een mindere score zou mogen krijgen.