Jukka Siltakorpi

IMPROVING DATA QUALITY IN A PROCESS DOCUMENTATION SYSTEM: REQUIREMENTS AND BENEFITS; A CASE STUDY

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF INFORMATION TECHNOLOGY
2021

ABSTRACT

Siltakorpi, Jukka
Improving Data Quality in a Process Documentation System: Requirements and Benefits; A Case Study
Jyväskylä: University of Jyväskylä, 2021, 61 pp.
Information systems science, Master's Thesis
Supervisors: Seppänen, Ville & Waltari, Outi

This case study was aiming to find how data quality could be measure and improved in a process documentation system. Data was gathered using existing documentation and a survey. Data quality has not previously seen much focus in this context, and therefore new metrics for measuring data quality in a process documentation system were created. Three different quality metrics were identified during this study, steps for data quality improvement were proposed and basic requirements for improving the quality were set. The current data quality level was defined based on the created metrics and ways to improve it were identified. Based on the requirements for the improvement work, potential costs and benefits that the data improvement work could cause were also listed.

Keywords: Process modeling, Process architecture, Data, Quality.

TIIVISTELMÄ

Siltakorpi, Jukka
Datan Laadun Kehittäminen Prosessidokumentaatiojärjestelmässä: Vaatimukset ja Hyödyt; Case tutkimus
Jyväskylä: Jyväskylän yliopisto, 2021, 61 s.
Tietojärjestelmätiede, pro gradu -tutkielma
Ohjaajat: Seppänen, Ville & Waltari, Outi

Tämän tapaustutkimuksen tavoitteena oli selvittää kuinka datan laatua voitaisiin mitata ja kehittää prosessidokumentaatiojärjestelmässä. Dataa kerättiin olemassa olevaa dokumentaatiota tutkimalla, sekä kyselyn avulla. Datan laatu ei ole aiemmin saanut juurikaan huomiota tässä kontekstissa, minkä johdosta täysin uudet metriikat datan laadun mittaamista prosessidokumentaatio järjestelmässä varten luotiin tämän tutkimuksen aikana. Tutkimuksen aikana tunnistettiin kolme laatumittaria, askeleet datan laadun kehittämistä varten esiteltiin, ja vaatimukset datan laadun kehittämiselle määriteltiin. Datan laadun lähtötaso määriteltiin metriikoiden perusteella ja vaatimukset ja toimenpiteet kehitykselle määriteltiin. Vaatimusten perusteella myös mahdolliset hyödyt ja kustannukset kyettiin arvioimaan.

Asiasanat: Prosessimallintaminen, Prosessiarkkitehtuuri, Datan laatu.

FIGURES

TABLES

SISÄLLYS

1 INTRODUCTION

The effective management of business processes within an organisation can lead to increased quality and efficiency, which can be used to gain competitive advantage over rivals (Browning & Eppinger, 2002). However, the number of processes and process models in large organisations can be very high and the organisations often have difficulties with effectively managing all their business processes (Browning & Eppinger, 2002). To better understand these processes, they are often visualised as business process models (Aldin & de Cesare, 2011). These models can be representations of either current business processes in an organisation or of a desired future state of a business process (Aldin & de Cesare, 2011). Process models can be used for various purposes and in their paper Bandara, Gable and Rosemann (2005) identify four different purposes that process modelling is often used for. These are model-based identification of process weaknesses, adapting best business practices, design of a new business blueprint, and end-user training (Bandara et al., 2005). Creating flowcharts and process maps has been around since the beginning of Taylorism, but process modelling is a more disciplined, mature, and scientific method, with an increased focus on modelling specifically *business* processes (Bandara et al., 2005). The models create a structure that depicts the actions within a process as well as interactions between different processes, and the design and analysis of this structure is covered by the field of business process architecture (Eid-Sabbagh, Dijkman & Weske, 2012). Business process architecture is important for organisations to properly manage as organisations can have thousands of process models and seeing the big picture can be difficult (Eid-Sabbagh et al., 2012). Additionally, while the individual process models can be sound, the relations and the entity that they form together might not be on the same level, which is where business process architecture can be used to identify these types of hidden issues (Eid-Sabbagh et al., 2012). Process architecture is commonly seen to be a part of enterprise architecture, which is used when referring to the entire structure of an organisation and is often divided into application architecture, technology architecture and business architecture, which includes process architecture (Rohloff, 2005; Zarvic & Wieringa, 2006; The Open Group, 2020 a).

Oca, Snoeck, Reijers and Rodríguez-Morffi (2015) state that in the current scientific literature on the topics of business process modelling and business process architecture there is no generally accepted definition or framework for defining business process modelling quality. Instead, most studies focus on how to use certain modelling tools or languages (Bandara et al., 2005). The studies that do address quality mostly focus on pragmatic or empirical aspects, such as model understandability or readability (Oca et al., 2015), which can be useful when creating individual models, but when creating several models that are connected to each other, as organisations often do, other aspects, such as managing the entity formed by the individual models, should be taken more into consideration when estimating quality (Eid-Sabbagh et al., 2012). Bandara et al. (2005) also find that there have been no studies on how to evaluate the success of process modelling project and present their own suggestion for success evaluation criteria for a process modelling project.

As mentioned, the studies that do touch on the quality of process models do mostly only focus on the individual process models and do not take process architecture into consideration (Bandara et al., 2005). While few studies on the process architecture quality do exist, such as the paper on formal conceptualization of process architecture by Eid-Sabbagh et al. (2012), the business process architecture structure that the models can create, and the quality aspects of that structure have not gained much attention in the scientific literature. This can make it difficult for organisations to evaluate the level of their business process architecture, as there are no clear criteria to meet. This lack of generally accepted levels of quality and methods to evaluate success also creates an issue for organisations when attempting to utilise business process architecture, which is that having no scientifically proven standards or methods on quality management and success measurement means that organisations need to develop these themselves when starting a modelling project or attempting to measure the quality of existing models in other aspects than readability (Oca et al., 2015). This is potentially a task that requires a lot of resources and can even then be difficult to successfully accomplish (Oca et al., 2015).

As there is a major lack of studies on process architecture quality, there are also no studies on the data quality topics around process architecture. When looking at studies on the topic of data quality in general, some data quality attributes that are most commonly used in the literature are accuracy, completeness, consistency, and timeliness (Batini, Cappiello, Francalanci & Maurino, 2009). Most of these attributes can be defined in the context of process architecture, but in this study, the focus will be on accuracy and completeness of data. According to Arts et al. (2002) these two are the most commonly used data quality attributes, and they were also seen to be the most appropriate ones for this study. Accuracy is a measure of how close the data value is to a value that is considered to be correct. i.e., how close the data is to reality, and completeness is the measure of to what degree a data collection includes the corresponding set of real-world objectives (Batini et al., 2009). The difficulty in defining the quality of data in a database that contains process architecture related data then comes from the question of how to measure the quality?

That is where this study comes in. This study aims to find potential ways for defining measuring points that can be used to estimate the quality of process architecture related data in a database. Additionally, the goal is to find ways for improving and maintaining the quality of data in the process architecture of an organisation and to find out the different benefits that an organisation might gain from improving process architecture related data quality as well as identifying potential costs that might be caused by the actions required to improve the data quality.

This is a case study where a telecom company utilising business process architecture is studied. To realize the goals of this study, the potential data quality measuring points are defined and used to measure the initial level (level 1) of business process architecture in the company. A target level (level 2) will also be set. After this, the potential effects of moving from level 1 to level 2 will be estimated and from those effects, the potential benefits gained from improving the level of business process architecture can be measured. In addition to measuring the benefits, the requirements for the improvement are also identified. To help determine if the case study method is an appropriate approach, the four questions that Benbasat, Goldstein and Mead (1987) present are used:

1. Can the phenomenon of interest be studied outside its natural setting?
2. Must the study focus on contemporary events?
3. Is control or manipulation of subjects or events necessary?
4. Does the phenomenon of interest enjoy an established theoretical base?

As this topic requires a natural setting with no need for the manipulation of test subjects or events while lacking a strong theoretical base, it is safe to say that the case study approach is an appropriate choice for this study. The reasoning for using a case study approach will be further explained in chapter three. Benbasat et al. (1987) also propose five different methods for collecting data in a case study and state that preferably two or more of these methods are to be used in a case study. Out of the five proposed methods, two were selected to be used in this study: documentation, and physical artefacts. Documentation refers to any written material, from notes to official reports (Benbasat et al., 1987). In this study, the documentation largely consists of process documentation, such as official process models. With physical artefacts Benbasat et al. (1987) refer to any devices, outputs, or tools. The tool that will be looked at in this study is called ARIS, which is a process documentation tool developed by Software AG and is being used by the case company. Interviews are also one of the data collection methods suggested by Benbasat et al. (1987), but I this study these are replaced by a user survey. As for the main steps of carrying out this study Eisenhardt's (1989) paper describes the process of building theory from case study research and provides eight steps for the process of carrying out a case study that will be used as a guideline. The steps presented in Eisenhardt's (1989) paper will be slightly modified to better suit this study. The steps for building theory in a case study and how those steps are applied in this study, as well as the data collection methods are explained in more detail in chapter 3, where the case and case company will also be described in more detail.

The first step of theory building in a case study includes defining the research questions (Eisenhardt, 1989). Three research questions were defined for this study:

1. How can the quality of data in a process documentation system be evaluated?
2. How to improve the data quality in a process documentation system?
3. What are potential benefits and cost points of improving the data quality in a process documentation system?

These questions were used to guide the study and provide it with a clear direction and goals, as the focus then was on finding answers to these three questions.

For organisations, this study can provide help on how to design, measure and improve their overall business process architecture, which is a question that will become more and more relevant as time goes on and even more companies begin to utilise process modelling, as moving towards a more process-oriented organisation has been the ongoing trend for nearly two decades (Eid-Sabbagh et al., 2012; Dijkman, Vanderfeesten & Reijers, 2011; Aguilar-Savén, 2004). As for the scientific community, this study can act as innovation for new studies and motivate researchers to create more universal theory on how the quality of business process architecture affects the gained benefits or even create a framework for estimating the quality of business process architecture, as those have been lacking in the scientific literature for a long time (Bandara et al., 2005; Oca et al., 2015).

Literature used in this paper was gathered using AIS eLibrary, IEEE Xplore and Google Scholar. AIS eLibrary is a research paper repository relevant for the field of information systems, IEEE Xplore is a library for scientific literature in the fields of electrical engineering, computer science and electronics and Google Scholar is a search engine used for finding scientific literature. Keywords and phrases used to find papers were for instance "Business process architecture", "Process modelling quality", "data quality", "data quality improvement" and "Process architecture data quality".

The second chapter of this study includes a literature review on the main topics of this study, such as business process modelling, process architecture, and quality. The key concepts will also be defined. Then we will move on to chapter three where the case and the methods used for data gathering are described in more detail. In chapter four, the data that was gathered using the methods described in chapter three, will be analysed and the potential requirements and benefits that could be gained from developing the level of an organisations process architecture will be estimated. While estimating the potential benefits, the potential cost points caused by the suggested actions will also be identified. Then, in chapter five, the findings are summarised, and the research questions are answered. In the sixth and final chapter the results and their implications as well as the limitations of this study are discussed and also topics for future studies are suggested.

2 LITERATURE REVIEW

The purpose of this literature review is to create an understanding of the current state of academic knowledge on topics such as business process modelling, process architecture and data quality. By doing so, it can be ensured that there are no misunderstandings on what the different key concepts used in this study mean (vom Brocke et al., 2015). To gather the literature used in this study, the literature search checklist for information systems research proposed by vom Brocke et al. (2015) was used to guide the process of gathering literature, but it was not fully followed in every area, especially since it assumes that the review is conducted by two or more researchers, which is not the case with this study. In their checklist, vom Brocke et al. (2015) present steps that should be taken before, during and after the literature search.

The first step is taken before starting the actual search process and it includes developing an understanding of the subject matter, justifying the literature review, and defining the scope (vom Brocke et al., 2015). Justification for the literature review in this study was based on the need to create an understanding of the current literature and definitions and to find possibilities to build on current literature. The scope is defined based on the table presented by vom Brocke et al. (2015) and presented in table 1. It was determined to be an iterative process, where sources are gathered from bibliographic databases and search engines. In table 1., the methods used in this study are shown in grey. Coverage was selected to be representative, meaning that many of the selected papers in some form typify a larger body of publications (vom Brocke et al., 2015). This is mainly done by using publications with a high number of citations as well as existing literature reviews as sources. The number of citations would indicate those papers to be acknowledged as important contributes to the topic and therefore have more impact on field. For techniques, keyword search and backward search were selected, meaning that literature was found by searching directly for it, for example, in Google scholar using keywords and also by using the studies that have been cited in the papers found through keyword search (vom Brocket et al., 2015).

TABLE 1 Definition of the search scope (Based on vom Brocke et al., 2015, p. 214)

1	Process	Sequential		Iterative
2	Sources	Citation indexing services	Bibliographic databases	Publications
3	Coverage	Comprehensive	Representative	Seminal works
4	Techniques	Keyword search	Backward search	Forward search

During the search process the main things to take into consideration are trying alternative search approaches, using justifiable search techniques and parameters and applying appropriate criteria for inclusion and exclusion (vom Brocke et al., 2015). During the literature search of this study the alternate approaches consisted mainly of using different search terms and trying different services for finding literature, but no further changes to the approaches were made. The selection criteria for the literature used in this study was based on two factors. Firstly, an initial check was done in order to determine if the publication was accessible and the meta data of the publication were satisfactory. This meta data refers to for example where the publication was published (such as the journal or conference proceedings), publication year and the number of citations (vom Brocke et al., 2015). As for the number of citations, the numbers of average citations for each period are based on the numbers by Pourmirza, Peters, Dijkman and Grefen (2017), where studies from 1994-1999 have an average of 33.77 citations, studies from 2000-2009 have an average of 33.12 citations and studies published between 2010-2015 have an average of 10.89 citations. Two studies were also used in this literature review, where the publication date was before 1994, but both of them had well over 6000 citations and were accepted for use in this literature review. For studies published after 2015 the average number of citations is not listed by Pourmiza et al. (2017), and in this literature review the studies published after 2015 are not evaluated based on the number of citations. In this literature review, the number of citations a paper has is based on the number provided by Google Scholar. The criteria for inclusion are listed in more detail in table 2. If the initial check was passed the publication was then selected for further evaluation, where the content was read in detail and its relevance to the research questions of this study or to the methods used in this study was evaluated. If the publication was found to contain information relevant to this study, it was then chosen to be included. These criteria are based on the ones that were used by Pourmirza et al. (2017) in their systematic literature review but are modified to better fit the scale and purpose of this study.

TABLE 2 Criteria for inclusion of publications in this study

Criteria	Description
Initial evaluation	
Accessibility	The publication must be freely accessible for everyone
Language	Publication must be in English
Where has it been published	Publication should be published in a well-known and trusted location
Publication year and number of citations	Study must be published after 2015 or have equal amount or more citations than the average number of citations publications from that period have
Further evaluation	
Relevance to this the research questions or methods of this study	Publication must provide information relevant for the research questions of this study OR information relevant regarding the methods used in this study

2.1 Business process modelling and process architecture

This chapter is written in order to create an understanding of what is a business process, what are business process models and what does business process architecture mean. This is done to find out what is the current status of scientific literature regarding these topics, which must be done in order to avoid conflicts and to make sure that the definitions that are used in this study are in line with the existing literature (vom Brocke et al., 2015).

2.1.1 Processes and business process modelling

One of the main concepts in this study is business process modelling, sometimes also simply referred to as process modelling (Oca et al., 2015). These models can be described as diagrams that are expressed in a more or less formal visual language, which often includes boxes that are interconnected with arrows (Krogstie, 2015). The word process itself refers to “a series of actions that you take in order to achieve a result” (Cambridge Dictionary, 2020), meaning that a process consists of several smaller actions, subprocesses, that are completed in order to achieve a common goal. In the scientific literature, there seems to be a rather common understanding of the definition for business processes and process modelling. When comparing literature reviews by Aguilar-Savén (2004), Bandara et al. (2005) and Oca et al. (2015) on the topic of business process modelling their definitions are very similar to each other and none of them mention there being any conflict in the literature that they had used. According to these definitions, business processes are the combination of activities and the different interactions that happen between people and/or systems during these actions (Aguilar-Savén, 2004; Bandara et al., 2005; Oca et al., 2015). Riemer, Holler and Indulska (2011) note that it is precisely this focus on the human actions rather than the actions performed by machines that separates business process modelling from other modelling types, such as data modelling. Additionally, business process modelling is used not only to better understand and visualize these processes but also to better understand the entity that the processes create by interacting with each other (Oca et al., 2015; Bandera et al., 2005; Aguilar-Savén, 2004). It is also good to note that according to Gordjin, Akkermans and Vliet (2000) a common misuse for process models is that they are used as business models, which they are not, as business models should usually be more generic models whereas business process models can be more detailed and contain more precise information of a specific process. And while there is a general understanding of the main concepts there are also areas where some conflict does exist, such as abbreviations. An example of this would be the use of BPM, which would in most cases refer to Business Process Management (Ko, Lee & Wah Lee, 2009) but in some papers, such as Aldin & de Cesare (2011), BPM is used to refer to Business Process Modelling.

2.1.2 Process architecture

Individually modelling the processes of an organisation can lead to issues, such as the processes not functioning well together or difficulties in identifying which processes need the most support (Green & Ould, 2005). The analysis and design of the structure that the business process models create is covered by process architecture (Eid-Sabbagh et al., 2012) and process architecture is also mentioned by Green and Ould (2005) as a viable solution for managing the processes in an organisation. Winter and Fischer (2006) have used the ANSI/IEEE standard 1471-2000 as definition for architecture and the definition is that architecture is the "fundamental organisation of a system, embodied in its components, their relationships to each other and the environment and the principles governing its design and evolution". From this we can see that architecture is a system that is formed not only by its different parts and how those parts interact, but also by the principles that have been created to govern the design and evolution of the architecture in question. Dijkman, Vanderfeesten and Reijers (2011) define process architecture as an organized overview of business processes, and they also add that the relations of those processes as well as the guidelines used to determine how the processes must be organised. They also clarify that all the business processes do not need to be modelled before the implementation of process architecture, as the missing processes can be modelled at a different stage (Dijkman et al., 2011). The exact outcome of the business process architecture in an organisation can vary depending on the design approach that was selected to be used. There are many alternatives for this approach and for example in their literature review Dijkman et al. (2011) identify five different design approaches, such as goal-based, action-based and object-based approaches. The main difference between these approaches is how the business processes are organised into groups and for example in the goal-based design approach the business goals and the relations between those goals are designed first and the rest is largely built around these goals (Dijkman et al., 2011). In their paper Dijkman et al. (2011) discovered that in practice these design approaches are often implemented as a mixture and in many cases none of them are fully implemented.

Koliadis, Ghose and Padmanabhuni (2008) identify six purposes for process architecture: (1) Establishing a shared understanding, (2) harvesting component reuse, (3) constructing implementations, (4) evolving system structure, (5) analysing high-level design and (6) managing complexity. In their paper Bandara, et al. (2005) were able to identify four different purposes that process modelling is often used for. These are model-based identification of process weaknesses, adapting best business practices, design of a new business blueprint, and end-user training (Bandara et al., 2005). From these different purposes for process architecture, we can see that there is no one clear purpose for its use and even researchers are not unanimous on what it could be used for as the purposes by Koliadis et al (2008) and Bandara et al. (2005) differ from each other. The purposes do still also have similarities, such as being rather high-level concepts and helping in implementation of process improvements (Bandara et al., 2005; Koliadis et al., 2008). Dijkman, Vanderfeesten and Reijers (2016) say that business process architecture is meant to address the need for guidelines that aim for consistent and integrated collections of process models. Traditional approach to managing different architecture in an organisation has usually been architecture management, where the drivers for architecture are the architects themselves and the users are experts in architecture (Winter, 2014). The study by Winter (2014) compares this architecture management approach with a newer architectural thinking approach, where the main differences are that in the architectural thinking approach, the idea is to involve the business specialists in the place of architects as much as possible so that they are in driver's seat.

Another important aspect of process modelling is the tool that is used by the organisation to model the processes as well as analyse the process architecture. According to Suárez, Sánchez and Villalobos (2015) the tools can often set limitations on analysing the models and especially severe limitations can occur with tools that do not enable the analysis of the relationships between the processes, meaning that process architecture cannot be fully utilised using those tools. In their study on the topic of collaborative process modelling tools Riemer et al. (2011) find that there is a very limited number of studies on the topic of process modelling tools. In their study they state that tools that enable a collaborative approach to modelling could increase the modelling efficiency by enabling better communication and effectively combining the knowledge of the people participating in the modelling process (Riemer et al., 2011).

2.1.3 Enterprise architecture

Traditionally in the scientific literature on the topics of business processes and business process modelling, the focus has largely been on individual processes and not on the interrelations between those processes (Eid-Sabbagh et al., 2012). This view is supported by Koliadis, et al. (2008), who state that there is a lack of common and practical standard on how to approach enterprise-wide business process architecture. There have been attempts to create standards and recommendations for process architecture, such as the 22 questions by Koliadis et al. (2008), but those have not been taken to wider use, and cannot be referred to as a common standard for process architecture. It seems that often in the scientific literature process architecture is not discussed as its own area but as a part of enterprise architecture (see e.g., Winter & Fischer, 2006; Barros & Julio, 2011; Gonzales-Lopez & Bustos, 2019). Enterprise architecture is usually used when talking about the structure of an entire enterprise but sometimes it can also be used when only referring to the analysis and documentation of the structure of the enterprise (Zarvic & Wieringa, 2006). While the approach where process architecture guidelines and quality standards are created on existing enterprise architecture frameworks can lead to some solid solutions for process architecture, it can also lead to a situation where the created guidelines are on a more generic level rather than providing something that would be specifically tailored towards process architecture needs. This is because enterprise architecture is divided into sub architectures, which often include some variants of business architecture, application architecture and technology architecture (see e.g. Rohloff, 2005; Zarvic & Wieringa, 2006; The Open Group, 2020 a.). In this division, process architecture would be located under business architecture along with organizational architecture, information architecture, and the business model (Rohloff, 2006). This division is visualized in figure 1. There are also deviations from this view on how enterprise architecture is constructed and for example Winter and Fischer (2006) divide enterprise architecture into business architecture, process architecture, integration architecture, software architecture and infrastructure architecture, raising process architecture to a higher level. Later on, Fischer, Aier and Winter (2007) have developed this division a bit further, dividing enterprise architecture into product/service architecture, metrics architecture, process architecture, and information/data management.

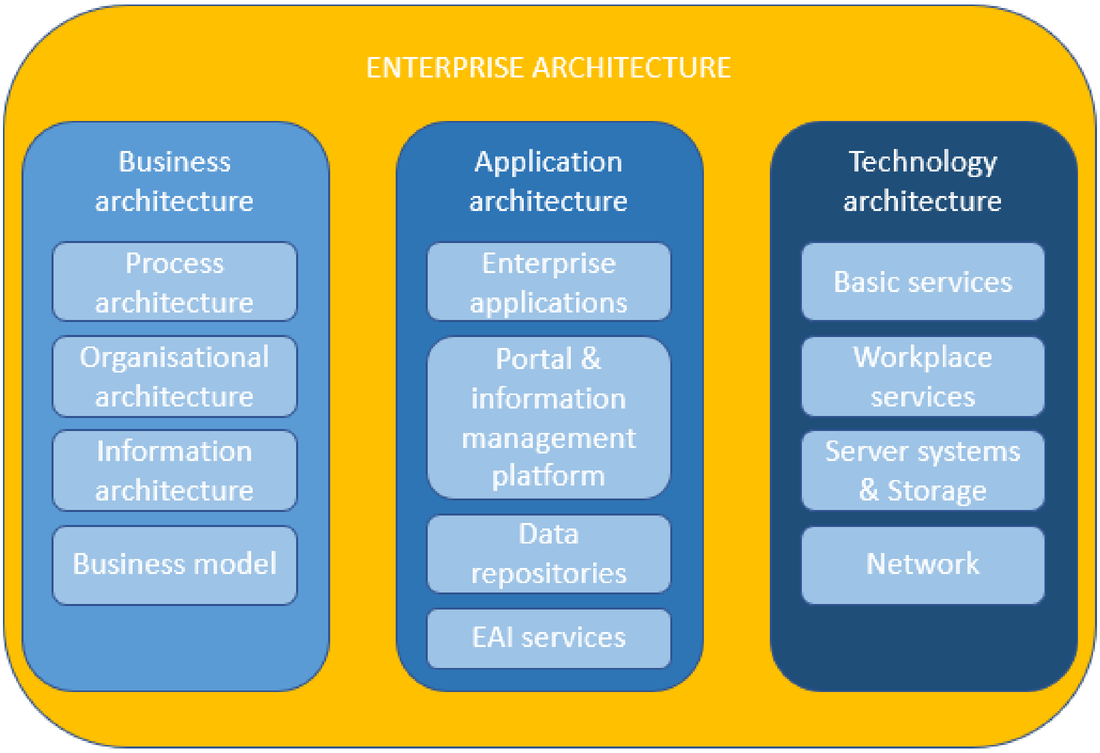

FIGURE 1 Enterprise architecture framework (Based on Rohloff, 2006)

There have been much more studies regarding enterprise architecture than process architecture and there are also plenty of widely known and used frameworks, such as the Open Group Architecture Framework, also known as TOGAF, that was initially created in 1995 for developing enterprise architectures within organisations and has been included in several studies that have focused on the topic of enterprise architecture (see e.g. Tang, Han & Chen, 2004; Urbaczewski & Mrdalj, 2006; Zarvic & Wieringa, 2006; Buckl, Ernst, Matthes, Ramacher & Schweda, 2009; Taleb & Cherkaoui, 2012). The TOGAF framework (The Open Group, 2020 b.) could perhaps be, at least on some level, also utilised when developing architectures other than enterprise architecture and it could provide some ideas on designing a process architecture as well. This is supported by the statement from The Open Group that an architecture framework (i.e., TOGAF) can be used to develop a broad range of architectures (The Open Group, 2020 b.). Though it should be kept in mind that the TOGAF model is mainly aimed for developing an enterprise architecture (The Open Group, 2020 b.) and therefore it could just as well be that it is not well suited to be used for process architecture development.

2.2 Business process model and process architecture quality

Now that a clear understanding of the basic process architecture related terminology used in this study has been set, we can take a look at how the quality of business process models and process architecture is defined and how to measure the quality of data that is stored in the process documentation system, where all the process models and related data are stored. The aim is to find out if common quality metrics for measuring process architecture related data quality already exist in the scientific literature and, if there are several metrics, to find out what they have in common. If there are no quality metrics or measurements for quality that could be seen as commonly accepted, then the aim would be to find out why and if some other areas, such as enterprise architecture, have such metrics and if those metrics could be applied to business process models or architecture.

Data quality can have several definitions (Arts, Keizer & Scheffer, 2002). Arts et al. (2002) provide two different definitions that can be used for data quality, that are a bit different from each other due to their context. The first definition that Arts et al. (2002) present is the definition by the International Standards Organisation, who define quality as “the totality of features and characteristics of an entity that bears on its ability to satisfy stated and implied needs”. The second definition is in the context of a medical registry, where Arts et al. (2002) define data quality as “the totality of features and characteristics of a data set, that bear on its ability to satisfy the needs that result from the intended use of data”. Arts et al. (2002) say that data quality should be assessed from the perspective of the person using the data. Wang and Strong (1996) also emphasise the user aspect in their study, where they define data quality as “Data that are fit for use by data consumers.” When reviewing literature on the topic of data quality Arts et al. (2002) found that often the definitions for data quality were ambiguous and sometimes the terms used for describing data quality were inconsistent even within a study. They did, however, find that the two attributes that were most frequently used with data quality were “accuracy” and “completeness” (Arts et al., 2002). Data accuracy is defined as “the extent to which registered are in conformity with the truth” and data completeness as “the extent to which all necessary data that could have been registered have actually been registered” (Arts et al., 2002). Batini et al. (2009) also note that the data quality research field is still evolving and the connections between data quality and process quality have not yet received much empirical evidence.

2.2.1 Business process model quality

There have not been many studies on the topic of quality aspect in process modelling and the studies that do focus on the quality aspect, mainly focus on the understandability of the process models (Oca et al., 2015). Vanderfeesten, Cardoso, Mendling, Reijers and van der Aalst (2007) find in their study that currently organisations model and design their business processes without the aid of metrics. They also note that many similarities have been identified between business process modelling and software engineering, which has been shown to greatly benefit from quality metrics, suggesting that the metrics used in software engineering could possibly be used to develop metrics for business process modelling (Vanderfeesten et al., 2007). First of the five metrics that according to them are currently being applied in software engineering but could also be applied when modelling workflows is coupling, which refers to the number of interconnections among the different modules of the model (Vanderfeesten et al., 2007). However, this metric was found not to be very reliable or informative (Vanderfeesten et al., 2007). The second metric presented by Vanderfeesten et al. (2007) is cohesion, which refers to how coherent the parts of a model are with each other. The result is calculated and used to support the business process designer when selecting the best design among a number of alternatives (Vanderfeesten et al., 2007). This cohesion method can provide numerical data, but it seems that its practicality can depend on the model. Third metric proposed for measuring quality in workflow models is complexity, which is used to measure if the design of the model is simple and understandable (Vanderfeesten et al., 2007). This method has been studied the most and it is suggested to be measured by adapting McCabe's cyclometric number which would then be used to measure the number of independent paths in the model (Vanderfeesten et al., 2007). The fourth metric proposed by Vanderfeesten et al. (2007) is modularity which refers to the number of models a module is split into. For this metric there was no existing way to measure it in process models and Vanderfeesten et al. (2007) also state that this method is likely not very useful when measuring process model quality. The fifth and final metric that Vanderfeesten et al., 2007 suggest for measuring process model quality is the size of the model, which could be measured by counting the number of activities in a model. Vanderfeesten et al. (2007) end the list by stating that the use of these metrics in business process modelling have not been studied much and that in different studies the same metrics have sometimes been used to measure different things, meaning that there is no common understanding, even in the academic world, on how to measure the quality of business process models (Vanderfeesten et al., 2007). While these metrics can be a good start, even Vanderfeesten et al. (2007) admit that they are not ideal. In the study Vanderfeesten et al. (2007) also do not provide any empirical data to back up their suggestions. It can also be argued that while many of these measurements can provide some idea of the current level of quality, having data on for example modularity or size of models can also lead to incorrect conclusions, as there can be reasons why different approach was taken in some areas than others.

Another way of looking at how to improve the quality of business process models is not to purely focus on how to measure the quality of the models but to also keep in mind that by improving the initial act of process modelling will very likely have a positive impact on the quality of the models (Benbasat et al., 2005). In their multiple case study, Benbasat et al. (2005) find several success factors for process modelling that they divide into two categories that are project-specific factors and modelling related factors. They also provide success measures for process modelling, such as project efficiency, user satisfaction and model quality (Benbasat et al., 2005). This focus on the actual act of process modelling isn't very common but it can provide good results and for an organisation looking to improve the quality of their process models and related data, a look into how the processes are modelled could provide positive results.

2.2.2 Business process architecture and enterprise architecture quality

In their literature review Oca et al. (2015) find that true quality assurance for business process architecture would require a quality system that would consist of e.g., a coherent set of quality policies, quality objectives and quality metrics. Despite the need of these quality systems, there are no instructions on how they should be developed (Oca et al., 2015). There is also a slight lack of empirical results, as only 57% of the studies that were included in the literature review by Oca et al. (2015) had performed any kind of empirical validation, meaning that it is not easy to determine which studies offer proposals that would also work in practice. Another issue is the fact that most studies only produce intangible knowledge, meaning that many of the studies do not produce any guidelines despite the need that exists for them (Oca et al., 2015). One exception to this would for example be the study on process modelling success factors by Benbasat et al. (2005) that was mentioned in the previous chapter. While the study does focus on the factors that affect the success of a single modelling project, that is also an important part in ensuring the process architecture quality since by ensuring that all modelling projects have the required support and follow the same rules and guidelines it is also ensured that all new models that then become a part of the structure in the database are created according to the common rules. The study itself is a multiple case study (Benbasat et al., 2005), meaning that it does provide empirical results, which, as previously mentioned in this chapter, is lacking in the field of process architecture (Oca et al., 2015).

While the number of studies on the topic of business process modelling is not very high and there are quality related gaps in the knowledge, the total amount of studies on the topic is still not that low either, at least when compared with business process architecture, where the main issue is not just the low number of quality related studies, but a general lack of studies on the overall topic. It seems that, as mentioned before, instead of researching process architecture as its own topic, the main topic of studies is enterprise architecture, and process architecture is only seen as a part of the enterprise architecture. This means that while there are process architecture specific solutions and guidelines, such as formal conceptualization of process architecture by Eid-Sabbagh et al. (2012), these studies specifically created for developing process architecture are very few and far between. Enterprise architecture consists of multiple areas in addition to process architecture, and it is usually divided into business architecture, application architecture and information architecture (see e.g., Rohloff, 2005; Zarvic & Wieringa, 2006; The Open Group, 2020 a.). While the studies on enterprise architecture can provide good guidelines, such as the enterprise architecture maintenance process by Fischer et al. (2007) and the Barros and Julio (2011) method for supporting business process architecture and business process design, that can aid in creating guidelines for process architecture, these methods are still aimed specifically to target the goal of enterprise architecture, which is to create a unified IT environment with tight links to business side of organisation as well as its strategy (Minoli, 2008, p.9), meaning that they are not really designed to support process architecture issues and needs.

2.3 Business process maturity models

Business process maturity models (BPMMs) are used to assess and improve capabilities within an organisation (Van Looy, De Backer, Poels & Snoeck, 2013). There are several different versions of BPMMs that have been developed over the years and some can contain quite different things (Van Looy et al., 2013). The BPMMs often include ways to evaluate the current state of the organisation by presenting descriptions for different stages of process maturity (Van Looy et al., 2013). Although there are several different BPMMs, for example van Looy et al. (2013) studied 69 different BPMMs in their study, with plenty of variation in their content, not much is known about the validity of different models, as the studies on the topic of BPMMs is often focused on creating theory on how to design BPMMs or creating a new specific BPMM (Van Looy, 2013). While the BPMMs themselves most likely won't be able to provide quality measurements directly, the models could possibly be used to set goals for process architecture that should be worked towards. By having these goals, it should be easier to set measurements and quality standards as there would be a clear goal to achieve.

One popular model for assessing process maturity is called Process and Enterprise Maturity Model (PEMM) (Hammer, 2007). It offers definitions for different levels of process maturity in four different areas that are the performers, owners, infrastructure and metrics (Hammer, 2007). Power (2007) found it to be a solid tool, with clear targets to drive for, but with some weaknesses, such as potentially being too complex for some business audience members and there also was no proven connection between the maturity levels described in PEMM and actual business performance (Power, 2007). Another widely used maturity model is the Capability Maturity Model Integration (CMMI), and in their study that included 35 different businesses Gibson, Goldenson & Kost (2006) found that six categories where performance was improved and the organisations participating in the study agreed that the increase in performance was due to CMMI. The categories and the median improvement that was gained in them were cost (34%), schedule (50%), productivity (61%), quality (48%), customer satisfaction (14%) and return on investment with a value of 4:1 (Gibson et al., 2006). Both of these maturity models have potential to provide good foundations for creating goals and key performance indicators (KPIs) for process architecture quality and ways to measure those indicators. For example, in the PEMM there is only one section that can be seen to be in direct connection to process modelling, and that is the documentation section. The documentation section focuses on how the processes are documented and on the highest level the description is that "An electronic representation of the process design supports its performance and management and allows analysis of environmental changes and process reconfigurations" (Hammer, 2007). This can provide the main goal for the process architecture, meaning that the KPIs should be designed to support it. Other areas of PEMM can also provide additional targets where high quality of process architecture can make the goal easier to achieve. An example of this would be the highest maturity level in the behaviour area, where the description of the level is "Performers look for signs that the process should change, and they propose improvements to the process" (Hammer, 2007). Here process architecture could have the goal to enable this by providing the tools for the performers to inspect the process, create an improved version of the process and share that improved version with other stakeholders.

3 THE CASE STUDY APPROACH

Studies that are carried out as case studies, revolve around cases that are studied in order to build theory (Eisenhardt, 1989). This type of theory building is a research strategy used to gather empirical evidence from one or more cases with the aim of creating theoretical constructs and propositions (Eisenhardt, 1989). Case studies usually focus on observing particular instance(s) of a phenomenon and often utilise several data gathering methods (Eisenhardt & Graebner, 2007). One basic feature of case studies is that they are used to study phenomena that are difficult or even impossible to study under laboratory circumstances because laboratory experiments isolate the phenomenon from the real-world context (Eisenhardt & Graebner, 2007). Case studies are often also used to study phenomena that do not yet have a strong theoretical foundation (Benbasat et al., 1987). This makes it a good complement to more common research method of deductive reasoning, where existing theory is tested (Eisenhardt & Graebner, 2007).

According to Williams (2007) three most common methods used to conduct research are quantitative, qualitative, and mixed methods. Quantitative approach is commonly used in studies that mainly focus on numerical data, whereas qualitative approach is commonly used in studies that focus on textural data (Williams, 2007). The third option, mixed methods, means using a combination of both numerical and textural data (Williams, 2007). When preparing this study, it was identified that both numerical, as well as textural data would be required. Numerical in the form of current completeness and accuracy and textural in the form of a survey, which is why the third approach presented by Williams (2007) was chosen. Relying mainly on the large amount of numerical data gained from the database would have provided information on the current situation and allowed the creation of measurement points for data quality and focusing purely on the numerical data could have provided more in-depth results on the measurement points. Textural data in the form of survey that included open ended questions was still seen to be important to find out the potential causes for the current state of the data quality. Purely relying on textural data would have meant not being able to identify ways to measure the quality, leaving us with no information on if the data quality is lacking and how to measure the potential improvements. Due to this it was decided that a mixed methods approach, where these two would be combined, was the best option for this study.

In addition to presenting ten key characteristics of case studies, Benbasat et al. (1987) also present four questions that a researcher can ask themselves to judge if a case study is an appropriate approach. These questions are used to justify the use of case study approach in this study. The four questions reflect on the main use purposes of a case study and they are 1. Can the phenomenon of interest be studied outside of its natural setting, 2. Must the study focus on contemporary events, 3. Is control or manipulation of subjects or events necessary, and 4. Does the phenomenon of interest enjoy an established theoretical base (Benbasat et al., 1987). For the first question and second questions, the answers are that the phenomenon cannot be studied outside of its natural setting because the focus is on creating empirical data and building theory based on the contemporary situation of the case company. Benbasat et al. (1987) say that if a natural setting or a focus on contemporary events is needed, then case methodology is clearly useful. Additionally, since there is no strong theoretical base existing for this topic, as proven in chapter 2, the natural setting of a case study can be an excellent setting for building theory (Benbasat et al., 1987). Benbasat et al. (1987) also add that if the control or manipulation of subjects or events would be required then the case study approach would not be suitable. For this study however, this is not an issue as the aim is to study the phenomenon in the natural setting without intervening. Based on this, it is safe to say that the case study approach is a good an appropriate approach for this study.

3.1 Case introduction

Only one organization was studied during this case study. This is a limitation for the case as the findings cannot be tested and confirmed to be generalizable. According to Gerring (2004), this does not mean that the study could not be carried out, because while cross-unit case studies should be used when testing theory, single-unit case studies still good for generating theory (Gerring, 2004), which is the purpose of this study. The case company in this study has been utilising process architecture for several years. As a tool, they are using a system called ARIS, which is a repository-based process modelling tool develop by Software AG. ARIS enables the organisation to store models and objects in a common database from which anyone in the organisation is able to view them. Plenty of additional information can also be added to the models and the objects within the models. The issue that the organisation is having is that with many models the information added to the model is missing or identified, or at least suspected, to be no longer valid. In some cases, the model itself can be outdated and no longer represents the current state of the process. Many models have also become obsolete but have not been removed from the database.

This has led to a situation where the database is cluttered, difficult to navigate, and it can also be difficult to tell whether a model is up-to-date or not. By definition data is "information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer" (Cambridge Dictionary, 2021). In this study, data is used to refer to object and model attribute values that most commonly are numerical, one of predefined options, or free text. It also includes the actual models and objects themselves.

The question is, what actions should the organisation take in order to improve the overall quality of the data in the database? And after the quality would be improved, how ensure that it will not over time drop back to the original level? It is also good to note that the actions required to improve and maintain the data quality would most likely require a good bit of the company resources. That's why it is also important to measure the requirements for the improvement and maintenance, so that the organisation is able to make well informed decisions on whether it is beneficial to carry out these actions or not. The detailed goals and steps to be taken in order to achieve those goals are detailed in chapter 3.3.

The results of this study can provide valuable empirical results that can benefit both researchers and organisations. As according to Oca et al. (2015) the amount of empirical results as well as tangible results on the topic of process modelling and especially the quality aspect are lacking, the results of this study can have significant value in both corporate world as well as scientific communities.

3.2 Data collection methods

Three methods were used to collect data for this study. Two of the methods were selected from the five different data collection methods that Benbasat et al. (1987) propose for use in qualitative studies. According to Benbasat et al. (1987), data from two or more of these data sources should be used to support the results of a case study. The two selected methods that are based on suggestions by Benbasat et al. (1987) are documentation and physical artefacts. Documentation refers to written material, such as memoranda or a formal report and physical artefacts range from devices and tools to outputs (Benbasat et al., 1987). Additionally, surveys were used to collect data from the users. Surveys were seen as the best way to gather data from a vast number of users. Gathering opinions and views from a large number of people was required in order to make sure that all the different user types were included. Surveys are also listed by Eisenhardt (2007) as a data collection method that can be used in case studies. These three data collection methods and their implementation in this study are described in more detail in their own sections later in this chapter.

The three data collection methods that were also suggested by Benbasat et al. (1987) but were not used in this study were archival records, direct observation, and interviews. Archival records refer to different charts and records that have been created by the organisation (Benbasat et al., 1987). These were not used as no records were identified that would have provided data relevant for this study. Direct observation refers to absorbing and noting details, actions or subtleties of the field environment (Benbasat et al., 1987). This method was not selected due to the nature of the work. Process modelling is not done on a daily basis as part of the everyday work. Additionally, it is reasonable to assume that if monitored, the process modelers might not operate the same way they normally would. As for the interviews, a decision was made to use online surveys instead. This is to avoid difficulties in organising face-to-face interviews and to give the participants more time to prepare and think about their answers. By using surveys instead of interviews, it was also possible to contact a larger number of users.

Previously in this study, during the literature review, it was found that there are no existing definitions for data quality in the context of process architecture. In the results of the literature review by Arts et al. (2002) they mention that data quality is good to define from the perspective of the person using the data. They also mentioned that two of the most common attributes related to data quality are "accuracy" and "completeness" (Arts et al., 2002). From this, it was determined that definitions for accuracy and completeness of data would be needed to be created. As the quality of data should be assessed from the users´ perspective (Arts et al., 2002), the surveys were also needed to determine what data do the users need and what is relevant for them.

3.2.1 Documentation

In their paper Benbasat et al. (1987) define "documentation" very broadly as written material that includes everything from newspaper clippings to memoranda to formal reports. They do also state that the specific data to be collected can vary quite heavily depending on the research questions and the units that are measured in the study (Benbasat et al., 1987). In this study documentation that has been used to collect data consists of the existing business process models. Data gathered from those models is then used to create a quality baseline. The older process modelling guidelines and instructions that were created internally in the case company have been used as a target that the existing models were evaluated against. The models have been created in the database over the years and will be compared with the instructions and guidelines that were in place at the time of the creation of the model. This has its challenges due to the fact that currently the database contains a large number of models where it can be difficult to say if those are meant to be official models or just drafts that have never been deleted from the database. This creates one of the limitations for this study, as the baseline is difficult to set. This means that comparing the level of quality between new and older models is challenging. This is not a limitation for the study in the sense that that was the current status of the data quality, but it can have an effect on the metrics that are presented later on.

The case company has issued new and more thorough quality policies for business process models in the company and also an updated training for process modellers to accompany the new policies. The data is still evaluated against the previous policies, as a very small number of models have been created after the change. The update to the guidelines also mainly just added new parts and provided more detailed information about the previous guidelines, meaning that the models created after the change would still need to include all the information that was required to be included in the previous guidelines as well.

3.2.2 Surveys

Originally, two surveys were created. One was for stakeholders in a manger position and the other one was for users who create the content. The surveys were created with two main goals. The first was to pinpoint the main issues that users currently have with the process modelling tool. This was to find out areas of the tool where the development is needed the most. By developing the system based on feedback by the users the usage rate would go up and especially if there would be possible connections to the possible trouble areas that are identified from the documentation then the organisation would have a clear point where the development should be started from. The second goal of the surveys was to also find what the different users and other stakeholders, such as process owners wish to gain from using the tool. As data quality is defined based on the user needs, meaning that having a clear idea as to how the different users utilise the data is important.

Dillman and Bowker (2002) list the four sources of error in sample surveys (coverage error, sampling error, measurement error, and nonresponse error) and they also provide a list of 14 actions that can be taken to reduce the risk of these errors as well as the different error types that the actions have an effect on. These actions include for example avoiding differences in visual appearance of the survey resulting from different user options, such as browser or screen size, providing the questions in a conventional and easy to understand format, and constructing the web survey so that users can scroll from question to question (Dillman & Bowker, 2002). The actions that are mentioned in the list were used as guidelines in the creation of the survey to reduce the different risks as much as possible. One difference between the survey conducted in this study and what Dillman and Bowker (2002) or other studies on the topic of conducting online surveys, such as Andrews, Nonnecke and Preece (2006) or Taherdoost (2016) is that all the people to whom the survey was sent to, work in the case organisation and therefore are guaranteed to have access to internet and tool that was used to conduct the survey. The fact that all the people who received the survey have a similar stake on the topic also means that the risk mentioned by Dillman and Bowker (2002) of only receiving responses from people with an interest on the topic should be mitigated, as all the recipients can be seen as stakeholders, even if their relationship with the system might vary.

The participants were selected with the goal of gaining information from people acting in different roles (users and managers with varying amounts of experience) and therefore having different perspectives on what the actual issues and the causes of those issues could be.

3.2.3 Physical artefacts

In this study only one physical artefact (tool) was used in data collection. The tool that was studied is used in the case company to model and store their business processes and much of the related information, such as names of process owners and process managers. The name of the tool is ARIS. The tool is also used in this study for data gathering in the documentation part of data collection, as all the business process models are stored in ARIS. This section of data collection does not focus on the content that has been created and stored in the database, but rather the tool itself, meaning that the focus is on the different functionalities and features that the tool provides. The aim with this data collection method is to see if there is something in the tool itself that makes it difficult to use or might otherwise be seen as off-putting. Examining the tool itself can provide valuable insights as to why the modelers have not always been following the official company policies. This can be especially useful when combined with the data gathered from the interviews and by developing the tool itself to better answer the needs and requirements of the users the data quality in the system can increase as the users not only create more content but also update the existing content more regularly.

The case company has two different versions of the ARIS tool in use. One is a browser-based version that is used by most modelers. In this version other company employees without the licence to make changes to process models are able to view the models. The other is a thick client version that is mainly used by the system admins and key users. As the browser-based version is the one that most modelers in the company are using, it is also the focus of this study. Data gathered from the system in the form of process models and the different data included in the models combined with the survey results is expected to point some of the key elements in the tool that should be looked at in more detail. By using the data gathered from the documentation to find the areas where the required metadata is lacking and then possibly discovering connections with the potential issues that users have based on the survey results and therefore having two different data sources backing up the findings, it could be possible to pinpoint the main issues with the tool.

3.3 Steps for carrying out the study

In her paper, Eisenhardt (1989) creates a roadmap for building theory from case study research. The roadmap contains eight steps, and it builds on previous attempts of creating such roadmaps and then expands upon them (Eisenhardt, 1989). The steps are detailed in table 3. These steps are followed during the execution of this case study, but with minor alterations so that the roadmap better suits the needs of this particular study. At the start of the research process the research questions were defined. In the second step the case was selected. Only one case was selected instead of several. This is mainly due to the lack of resources, such as time and number of investigators, and is one of the limitations of this study. The company that was chosen to be investigated was chosen based on their current situation and their high motivation to participate in the study. Once the case was selected, the data collection methods were chosen. The methods were chosen based on suggestions by Benbasat et al. (1987) with focus on gaining data from different sources while retaining flexibility. The fourth step by Eisenhardt (1989) would be to enter the field, but that is where this study slightly diverges from the roadmap. Instead step four in this study was to conduct a basic literature review to gain deeper understanding of the current scientific literature on the topic. The results of the literature review also provided evidence to support the fact that a case study approach is the correct one for this topic.

After the literature review was completed, data collection and analysis phase were started. Data collection and analysis were done simultaneously to save time as Eisenhardt (1989) suggests. The main focus of the initial data collection was to gather data that could be used to define the current level of data quality in the ARIS system. This was done by gathering information on when the last changes were made to each process model and were the mandatory model attributes added. The time since last change would be an indication of the relevance of models in the database and therefore an overall indication of the data accuracy. Mandatory attribute use levels give us an indication of how well the instructions set by the organisation have been followed, aiding us to determine the completeness of data. The use of mandatory attributes can be used to determine the completeness of data by providing information on where the attributes have been used and how many models are missing one or several of the required attributes. Data completeness referred to the extent to which the data that could have been registered, has actually been registered (Arts et al., 2002). If the mandatory attributes have not been widely registered, it means that the data is incomplete.

After the current level of quality was determined data was gathered from the current process modelling instructions and guidelines that are set at the case company. This data was combined with data gathered from the surveys to create a target level that the company would be hoping to achieve. After the data was collected it was analysed. While it was also analysed during the collection step the analysis was finished separately from the data collection. After the analysis, the hypotheses were formed and then compared with the current literature that was mainly gathered earlier in the literature review. The final step of this study was improving theoretical saturation when possible and reaching closure that way as proposed by Eisenhardt (1989).

TABLE 3 Process of Building Theory from Case Study Research (Eisenhardt, 1989, p. 533)

Step	Activity	Reason
Getting Started	• Definition of research questions	• Focuses efforts
	• Possibly a priori constructs	• Provides better grounding of construct efforts
Selecting Cases	• Neither theory nor hypotheses	• Retains theoretical flexibility
	• Specified population	• Constrains extraneous variation and sharpens external validity
	• Theoretical, not random, sampling	• Focuses efforts on theoretically useful cases – i.e., those that replicate or extend theory by filling conceptual categories
Crafting Instruments and Protocols	• Multiple data collection methods	• Strengthens grounding of theory by triangulation of evidence
	• Qualitative and quantitative data combined	• Synergistic view of evidence
	• Multiple investigators	• Fosters divergent perspectives and strengthens grounding
Entering the Field	• Overlapping data collection and analysis, including field notes	• Speeds analyses and reveals helpful adjustments to data collection
	• Flexible and opportunistic data collection methods	• Allows investigators to take advantage of emergent themes and unique case features
Analysing Data	• Within-case analysis	• Gains familiarity with data and preliminary theory generation
	• Cross-case pattern search using divergent techniques	• Forces investigators to look beyond initial impressions and see evidence thru multiple lenses
Shaping Hypotheses	• Iterative tabulation of evidence for each construct	• Sharpens construct definition, validity, and measurability
	• Replication, not sampling, logic across cases	• Confirms, extends, and sharpens construct definitions
	• Search evidence for "why" behind relationships	• Builds internal validity
Enfolding Literature	• Comparison with conflicting literature	• Builds internal validity, raises theoretical level, and sharpens construct definitions
	• Comparison with similar literature	• Sharpens generalizability, improves construct definition, and raises theoretical level
Reaching Closure	• Theoretical saturation when possible	• Ends process when marginal improvement becomes small

4 DATA COLLECTION AND ANALYSIS

The data collection essentially had two phases. The goal of the first phase was to define the current level of data quality, the second phase was done with aim of defining the desired level of data quality. Both of these also include analysing data and additionally there was a second step after the data collection that only included analysis. In the second step, the aim was to define the requirements for moving from current level to the desired level. These requirements are presented in chapter 4.3. The requirements were created by analysing the differences between the two levels and defining the actions that would need to be taken in order to move from current level to the desired level.

In their literature review Batini et al. (2009) study 13 different data quality improvement methodologies. The process was divided into assessment and improvement phases and several steps were identified in both phases (Batini et al., 2009). From both phases, the most common and best-defined steps that were seen to suit this specific topic were also carried out in this study. In the assessment phase these steps were data analysis, identification of critical areas, and measurement of quality (Batini et al., 2009). In the improvement phase the most common and best-defined steps were identification of the causes of errors, design of data improvement solutions, and improvement monitoring (Batini et al., 2009). These steps are clarified in table 4, based on Batini et al. (2009).

TABLE 4 Phases and steps, based on Batini et al. (2009)

Phase	Step	Description
Assessment	Data analysis	Analysing the data that was gathered during this study
	Identification of critical areas	Identifying the areas where development provides the most benefits
	Measurement of quality	Measure the current level of quality
Improvement	Identification of the causes of error	What are the reasons for the current quality issues with both accuracy and completeness of data
	Design of data improvement solutions	Design of a solution(s) that could solve the issues with data quality. In this study, the solutions are designed, but the implementation is not included.
	Evaluation of costs	Creating an evaluation of costs based on the improvement solutions. Can also include an estimate of the potential savings or other benefits provided by improvement
	Improvement monitoring	After the solutions for improvement have been implemented there needs to be a way to monitor the data quality to ensure the solutions are working. In this study the monitoring methods are designed, but implementation phase is not included within this study

4.1 Defining the current level of data quality in the process modelling tool

In the first step of the assessment phase the gathered data was analysed. Most of the models in the databases are created before the creation of the current set of process architecture guidelines and instructions (PAGI) that are to be followed. However, the rules for the different attributes that are to set as mandatory had not seen much change until the latest version of PAGI was released during December 2020. Until then, the list of mandatory attributes remained the same except for one attribute that was removed from the list half a year earlier. The current level is based on the use level of mandatory attributes, time since last change to a model, and model publishing rate. These can also act as good metrics in the future.

4.1.1 Use of mandatory model attributes

The list used for the model attributes is based on the previous version of PAGI that was used until December 2020. Until then the list had remained fairly unchanged except for the removal of one attribute that is also not included in this list. The attributes include information that has been defined by the case company as the most relevant information that needs to be included in every process model. These include the name of the model, model status, and person responsible. Model name needs to be descriptive so that readers will know what the purpose of the model is. Model status is selected from predefined options and can for example be "approved" or "under construction". If the model status is set as approved, two more attributes, "approval date" and "approved by", need to be filled. Person responsible is the person who is responsible for the maintenance of the model. With all this information in place the maintenance of the entire database becomes considerably easier as it becomes much easier for admins to estimate the relevance of a model and they also have a person to contact if necessary. If the attributes are not used, it means that the data in the system is incomplete as all the necessary data that has not been registered (Arts et al., 2002).

The attributes of a total of 13651 models were studied (appendix 1). 11756 of these were from the "Working" database that is meant for the models that are developed and worked on and 1895 of the models that were studied were from the "Publishing" database, where copies of models are moved to from the working database once they have been approved. A model must always be named, but if the name was set as "Untitled" the model was seen to have no name. Same goes for the Person responsible attribute in cases where "Unknown" or "TBD" were used. In these cases, the Person responsible attribute is seen to be missing. Final rule was that if a model had the words "Draft" or "Test" in its name it was not included in the study.

From the results it could be seen that even in the publishing database where approved models are stored, the name attribute was the only one that had been used in all of the models (table 5). With every other attribute the use rate is below 90%. This is a worrying sign, as all of the attributes that are listed in this study are mandatory for all models that are to be published and should always be checked by key users before moving a model to the publishing database. From this, we can determine that the publishing process could be improved.

TABLE 5 Attribute use percentages in databases

	Working	Publishing
% of models with name	97	100
% of models where status is "Approved"	16	88
% of models with approval date	14	88
% of models with approved by	14	88
% of models with model status	28	88
% of models with person responsible	23	83

4.1.2 Time since last change in a model

Model validity can be difficult to estimate for people who are not actively working with the process that the model is depicting. If the models are several years old, it can be difficult to tell if they are outdated which then causes a situation where the users can no longer be certain if the model they are looking at is actually still relevant. This can be included in different time-related data quality dimensions (Batini et al., 2009), but in this study it is included in the data accuracy dimension. The age of the model can work as an indicator of the validity of a process model, but there can be cases where a model is several years old, but has been actively updated, making it as valid as a recently created model. This is why, in this study, time from last change has been used to estimate model validity instead of using the age of a model. This does not fully remove the issues that were related with using the age of the models as an estimate for validity, but the risk of error should be lower.

The models used in this section are all located in the working database where there are several different groups, or folders, where the models are stored. In this study these groups are referred to simply as Group 1, Group 2 etc. (Appendix 1). The reason that the groups were taken into consideration is that there could be differences between different areas and if the data would be compiled into one group these differences could be lost, making the process of finding the critical areas that need improvement the most much more difficult. The data is presented in figure 2, where the models are divided into four categories within each group based on how long it has been since the last changes made to the model. From figure 2, it can be seen that there in fact is quite a lot of variation between the groups. This is also a part of the second step of the assessment phase, where the critical areas are identified. Based on figure 2, the critical areas would seem to be groups 2 and 20, as those contain models where last changes have been made more than seven years ago to 85%-95% percent of the models, but when looking at figure 3, we can see that the most critical areas actually seem to be groups 3 and 1, as those contain far more models, including ones that have not been updated during the last seven years.

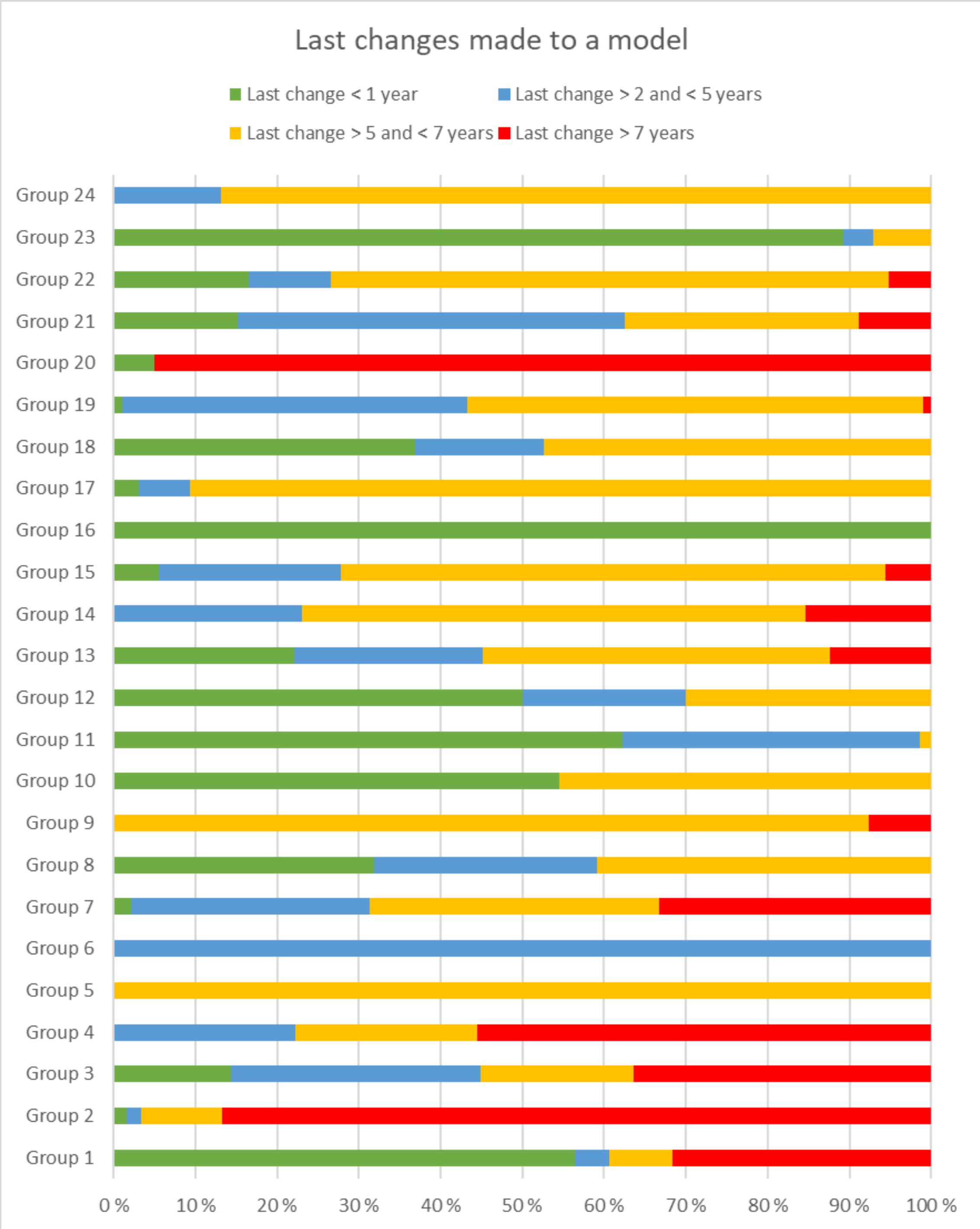

FIGURE 2 Last changes made to a model

When looking at figure 2 it is good to keep in mind that the number of models within the groups also varies and can provide a slightly distorted view. The differences between the number of models in different groups can be seen in figure 3, where the number of models can be seen in the X axis. The number of models and their age within each group can also be found in appendix 2.

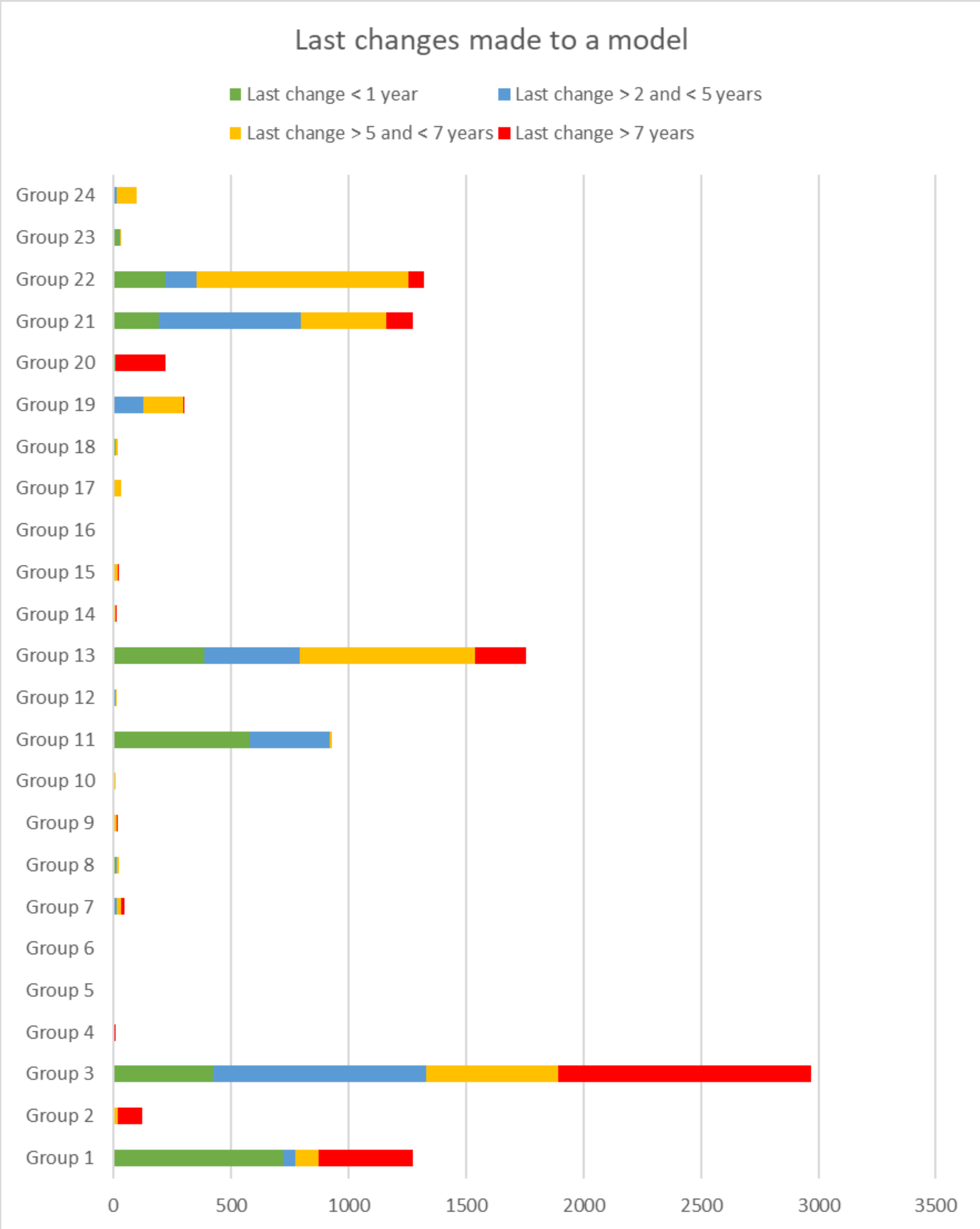

FIGURE 3 Number of models in different groups

4.1.3 Model publishing rate

The purpose for determining the number of published models in relation to the number of models within the working database is to create an indicator that could be used to estimate if the models that are created go through the process of being approved or if they are simply created in the working database and then left there. When a model is left in the working database it can be unfinished, but it can also be finished and even approved but it just hasn't gone through the publishing process for whatever reason.

From the model data that was collected when the use of mandatory attributes was studied, we can also find out how many models are published for each model created in the database. The ideal number would be 1, since whenever a model is published, it is supposed to be moved to the publishing database and a copy of the model that can be further developed, is created and added to the working database. In real life the number will never be 1, as there are always ongoing modelling projects, where several different models are created in the working database to be worked on.

From the data in table 6, we can see that there are currently 11756 models in the working database and 1895 models in the publishing database. This means that for every model in the publishing database there are 6.2 models in the working database. We can also see that in the working database there are 1853 approved models, but in the publishing database there are only 1668 approved models, meaning that at least 185 models have been approved but have not been added to the publishing database.

TABLE 6 Number of models in the databases

Database	# of models	Approved
Publishing	1895	1668
Working	11756	1853

4.1.4 Survey responses

The surveys were sent to 59 process owners and 279 users. 20 answers were gathered from users, giving us an answer rate of 7,2%. From the process owners, only two replies were received, making the answer rate only 3,4%.

The ones who replied to the user survey were mainly beginners in using the system (50%), but there was also a good portion of people with basic skills (30%) and advanced skills (20%). The tool had been introduced to many new areas before the surveys and plenty of new people were trained to use it, which means that the larger number of beginner level users is actually a good representation of the current level of user knowledge.

From the answers (Appendix 3) we can see that some of the main issues are that many of the users (35%) feel that the tool is too complicated and at the same time only 35% of them felt that they were familiar with the current version of process architecture rules and guidelines that have been set at the company. This would indicate that many of the users that are having issues with the system also do not know where they are able to get support. This is further supported by the answers in question eight, where we can see that a lack of training and support is seen as an issue. Full list of the questions and answers can be found in appendix 3.

These issues that have been identified based on the survey, such as users seeing the system as being complicated, difficulties in finding support, and lack of knowledge regarding the process architecture rules guidelines, indicate that the users do not have the required information and skills to properly follow the guidelines set to ensure the data quality in the system. Although, it is important to point out that in the responses there were also many who felt that the system is not too complicated, making the results for that area a bit contradictory. This can be caused by the fact the people included in the survey had different levels of experience with the tool.

4.2 Defining the desired level of data quality in the process modelling tool

The desired level of data quality was defined based on the replies that were gathered from the surveys. Everyone in the case company with a modelling license is required to participate in modelling training, meaning that they know at least the basics of the system and should therefore be able to answer the survey. The answers were collected by using Webropol. A link to survey was sent via email and one person could only answer once. The survey contained questions such as "Do you feel that you are familiar with the current process modelling rules and guidelines that were released during Q4 2020 and know where to find them?" and "Do you feel that using (the tool) is too complicated?". The aim was to use this information to create a depiction of the desired future state of process architecture at the company and see if the different instructions and guidelines had been effectively communicated to users.

The desired level of quality was defined based on the survey results and the views of the people from the organisation, who were involved in a process architecture community within the organisation. The decisions were that the use of mandatory model attributes in the publishing database would be increased to 100% during the following six months. This was due to the fact that the quality of the data in the publishing database is seen to be more critical, as the database should only include approved models that have been checked by a key user. The use level of the mandatory attributes is already quite high in the publishing database, between 83% - 100% depending on the attribute, and the database contains far fewer models than the working database, meaning that reaching for the 100% use level is a realistic goal. For any organisation using the "dual-database" approach, a 100% use for mandatory should be set as the goal for the database that contains the approved models, as the content moved to the publishing database can be monitored and controlled by the key users. For the working database, the goals are more modest, as that part requires more work than the publishing database. This is due to the difficulty of defining weather a model is obsolete or not. If there is no data on who to contact about a model or the contact person has left the organisation, it is up to the key users to determine if the model should be kept. This is why the targets for working database are lower than for the publishing database. All models should have a name, which is why the target there is still 100%. For the status and person responsible attribute the target values are 50%. The aim is to set targets that can realistically be achieved, and in this case 100% is not realistic for the case company, as that would require massive resources to the high number of models in the database. The target for model publishing rate was set as four, meaning that for every model in the publishing database there would be a maximum of four models in the working database. This also meant that the approval related attributes should have a use rate of 25% in the working database, as the rate of the approved attributes (Status = Approved, Approval date, Approved by) usage should match the model publishing rate.

For the age of the models, setting the targets is more difficult if the organisation is unable to determine which of the models currently in the database are needed to be updated or completely obsolete. If there is no contact person for a model, it will become an extremely time-consuming job for the key users and admins to start evaluating the models. Therefore, the organisation should set steady long-term goals for renewing the models and follow the progress. As mentioned, depending on the current state and available resources, this might be a costly task, which is why it is recommended that the organisation evaluates this with other potential improvement areas and decides what is the most efficient way to allocate resources. But if the organisation is working with a process modelling tool that enables exporting data from the system, it is possible that the older models will cause inaccuracies in exported data, making it harder to rely on and use to aid decision making.

4.3 Requirements for developing the data quality in process modelling tool

By finding out the problem areas of the publishing process and renewing it based on those findings. The main areas with issues are identified by combining the data gathered from survey results with the data that was gained from examining the current state of the database. While the more specific attributes and survey questions and results can undoubtably vary between organisations, these general steps can be used regardless of the differences. The common areas where issues were identified in this case were insufficient communication/support and the tool itself, as it was seen as too complicated in certain areas. The more detailed issues are mentioned in chapter 4.3.1 Recommended actions.

Oca et al. (2015) mentioned that to have proper quality assurance, the organisation would require a coherent set of quality policies, quality objectives and quality metrics. For the case company, an initial set of quality policies have been created and it is being further developed. This work should continue, and in the future, it must be ensured that these policies ae effectively communicated to all stakeholders. The survey that was conducted during this study showed that there are people new to the system and even more experienced users are having issues with finding and staying up to date on the latest company policies. Surveys also showed that it would be good to see if there are ways to make the system, and especially navigating within the system, easier. Therefore, finding ways to simplify the system should be one of the actions to be taken in the future, even though there were contradictions in the results. These contradictions could have been caused by the different answers from new and experienced users. It can be that for new users the system seems complicated but the users that have gotten used to the system, have no difficulties in using it.

Madnick and Zhu (2006) also point out that in some cases issues can be caused because data can be misunderstood and misused. They call far effective use of metadata to fully realize the opportunities of the data (Madnick & Zhu, 2006). In the context of this case, this would be the model and object attributes that can be added, especially the mandatory ones, and ensuring that their purpose is understood not just by the readers but also by the modelers. An example of this would be the Person responsible attribute. Is it clear for everyone that this refers to the person who is responsible for the model, who might not be the person responsible for the actual business process? This comes back to ensuring efficient communication to make sure that there are no misconceptions as to how to add and interpret the data in the system.

The requirements that have been identified for increasing the level of data quality are presented in table 7. At this stage it is assumed the organisation has already created the quality policies, set the objectives for quality and the metrics on how to measure quality. As Oca et al. (2015) stated, while these are important to have, there are no templates for creating these and no scientifically proven methods to help the organisations to create these themselves either. This means that the organisations need to develop these from ground up. The information provided in this study on the required attribute data for the models, used metrics and the quality objectives can be used by organisations to get started in creating their own versions of these.

4.3.1 Recommended actions

The quality objectives were defined based on the survey results and process architect views, which will vary from case to case. The metrics that were identified during this study were using mandatory model attributes, which in general means defining a way to monitor how well the users follow the set quality policies, number of models in the publishing database in comparison to number of models in working database, which refers to the number of models that can be identified to be finalised and approved in comparison to the number of models that are under development or otherwise unfinished. The third metric was the age of the models. While old models can still be valid, it can still indicate the general level of how up to date the data is in the database, meaning that an organisation should also consider alternative ways to define the validity of model, such as the time since last changes have been made to a model. The main causes for the issues were insufficient communication to the users and the tool being seen to be overly complicated in certain areas by some users. Due to the insufficient communication, the users were unaware of the current process architecture related rules and guidelines, meaning that they cannot know for example what attribute values should be included with the models or how the modelling notations work. Additionally, they seemed to be lacking the information on how to get support, meaning that they had no connection to process architects or the local key users, who could have helped them. This situation can be seen to have to possible outcomes, where the users either continue modelling or stop the work. In the first alternative content is created and updated in the system, but the new content is not created according to the rules and guidelines, meaning that it may differ from what is wanted by the organisation. In the second alternative the modelers simply do not create anything new, as they are not confident enough to use the system. This would lead to a situation where the content in some areas would not be updated.

Based on the difference between current and desired states in data quality, requirements and recommended actions that would need to be completed in order to achieve the desired state. These are presented in table 7. Additionally, it also includes the potential benefits that the organisations can gain once those actions have been taken.

TABLE 7 Requirements for improving data quality

Requirement	Description	Suggested actions
Ensure efficient communication	Ensure that all users are aware of the rules and guidelines that have been set and are able to locate the documentation when needed. New system features and available training should also be communicated.	Ensure sufficient training for users and identify the best communication channels and methods.
Ensure simplicity of the tool	Develop the tool so that it is as simple to use as possible. Ensure that especially functionalities related to the actions that are required in set quality policies are easy to find and use.	Gather and analyse feedback from the users. Create development initiatives based on the findings and ensure sufficient resources for the development work.
Improve the quality of existing data	Improving the quality of existing data. Quality is measured based on previously defined metrics. The actions for this can require a large amount of manual labour if there are no identified ways to automate the work or handle large quantities of data simultaneously.	Identify the main areas with issues, based on either level of quality or importance of the area. Identify the cause of the issue and implement appropriate action to improve the quality.
Monitor the quality of data	Ensure the quality of new content and maintain the quality of older content	Automate as much as possible by creating scripts that are run automatically on e.g., monthly basis. This requires defining the factors that can be monitored with scripts. Define the areas that are still needed to be monitored manually.

The “Improve the quality of existing data” is a rather broad requirement, as are the suggested actions for that requirement. This is due to the fact that the more specific improvement actions depend heavily on the issues and causes of errors. For the case organisation of this study, some of the appropriate actions would be identifying areas that contain models that are for testing purposes and can be removed as well as contacting modellers about models where it is difficult to determine if the model is still up to date and request them to update the attribute information.

4.3.2 Identifying the potential costs and benefits

To evaluate the relevance of these actions, we also need to estimate the costs that each of them would cause and evaluate those costs against the benefits that the actions would provide. The potential costs are not measured by monetary value in this study as those numbers can be very different between organisations. Instead, the costs that will be discussed are for example the potential issues caused by gaps and errors in the data or the benefits that cannot be gained due to poor data quality. The benefits and costs that were identified for each requirement are also listed in table 8.

For the first requirement, ensure sufficient communication, the costs from the communication are estimated to be fairly low, especially in comparison to the benefits of good communication. In their study Gochhyat, Giri and Suar (2017) find that organizational communication has an effect on both the culture and effectiveness of the organisation. After the preferred communication channels have been identified, information sharing consumes little resources from the admins. Some suggestions for the communication methods that Gochhyat et al. (2017) presented were online networking, meetings, and group discussions. If the organisation aims to secure high-quality communication, then communication training could be organised for the admins. In larger organisations this could potentially be done in-house by utilising organisations own communication experts. Additional costs can be caused by increased number of support requests from the users as their knowledge of BART and who to contact for help increases, meaning that the workload for key users is likely to increase. Increased number of users means that more licenses are consumed, assuming that the tool is based on licenses. The organisation will need to monitor the number of available licenses and calculate how many additional licenses might be needed to obtain. The costs are still likely to remain low in comparison to the benefits that are improved data completeness and accuracy due to increased use and user knowledge. Additionally, several studies have shown that good communication increases employee satisfaction and commitment (Gochhyat et al., 2017; Rajhans, 2012; Marques, 2010; Mayfield & Mayfield, 2002).

The second requirement of ensuring the simplicity of the tool is a more complicated topic than communication. In the survey that was carried out during this study the answers on whether the tool is too complicated or not had a lot of variation. This implies that in this case the benefits gained from simplification of the system might not provide the greatest results for the cost. Costs in this case are caused by the resources consumed by the development work. These resources include the time of employees involved in development work as well as the money potentially spent on consultants. Simplification can also be done to make the work easier for admins but the need or effects of that were not studied in this paper, as this was more focused on the user aspect.

Improving the quality of the existing content was identified as the third requirement for improving data quality in the system. This will require a lot of manual work from admins as well as the people who have originally created the models, as the content needs to be evaluated individually. Especially for the admins and key users this will consume a large number of resources if the models are lacking data on who to contact about the model. This is still a necessary step as otherwise there will still be a large quantity of data that is inaccurate or where the accuracy cannot be determined, making the data unreliable in both cases. If these issues are not fixed, the reports and other information exported from the system cannot be trusted, making all other improvement actions rather redundant. If this step is done, but instead of carefully looking at the models, the organization decides to directly delete a large number of the models that are assumed to be obsolete based on the attributes, such as time of creation and last changes, there is a risk that critical information is lost that might not be possible to recover when it would be needed next time. This improvement can also slow down the creation of new content, as the time of the modelers and admins is used on improvement work instead. Dijkman et al. (2011) say that not all the processes of an organization need to be modelled for process architecture to be implemented, meaning that it is ok at times for the focus to be somewhere else than including all processes in the system.

After improving data quality, but also during the time period when the improvements are done, the organisation should monitor the quality by, for example, utilising the metrics presented in this paper. Monitoring was also mentioned as the last step in most of the data quality improvement methods studied by Batini et al. (2009). Costs caused by monitoring the quality are dependent on how much of the work can be automated. Automation would mean creating scripts that follow the main quality elements, such as attribute use, guideline compliance, and age of the models. The proper creation of these scripts will require resources in the beginning, as the requirements planning, and creation and implementation of the scripts and roles will require efforts from the admins. This is still a vital step for any data quality improvement project and monitoring as the final step was also included in all of the data quality methods that were reviewed by Batni et al. (2009). Without monitoring it is impossible to know if the improvement project has been successful and the goals have been reached (Batini et al., 2009). Monitoring the data quality during the improvement project means that the actions can be tuned during the project if it seems that there are no improvements. Implementing these monitoring methods will also make it easier to monitor the data quality in the future, ensuring that the potentially gained increases in the quality are not just temporary.

If nothing is done about the data quality by the organisation and no actions are taken to improve the quality, it can lead to a situation where the best option for the organisation is to put an end to using the system all together. This means that there will be sunken costs that cannot be recovered, but it will prevent further losses as the system might no longer in these extreme cases be providing any value for the organisation.

TABLE 8 Costs and benefits of the requirements

Requirement	Costs	Benefits
Ensure efficient communication	Time used to define the optimal communication methods and time used to create the additional communication materials.	Quality of content created by users will increase as they are more likely to work according to the policies. This not only increases data quality, but also reduces the amount of monitoring required by admins.
Ensure simplicity of the tool	Uncertain if the results will have significant impact on the data quality. Designing and implementing the solutions will most likely consume plenty of resources.	Following the quality policies will be easier for users, making them more likely to act according to the policies. Additionally, this simplicity might attract new users from different areas, helping to fill some gaps in the data.
Improve the quality of existing data	Requires a lot of manual work from both admins as well as users. Will take a long time to reach the targets. Some areas might not have any contact persons remaining in the organisation, meaning that the admins have to decide what to do with the remaining models.	Increased data accuracy. If this is not done, the data that is added to the system later on cannot be fully utilised as the old content will make the reports unreliable.
Monitor the quality of data	Development costs for designing and creating the automated scripts that would create the reports. Alternatively, the data could be monitored manually. In that case the costs would be caused by the continual manual work.	Increased confidence in that the data in the database is up to date. This is also likely to make the database easier to navigate and find content in, as there will be much less obsolete content

5 SUMMARY

The aim of this study was to find out how to improve data quality in a process documentation system and what are the potential requirements and benefits from the improvement. As described in the introduction chapter, organisations are moving towards a more process-oriented culture, which requires proper process documentation. Despite the fact that this has now been known for a while, quality aspects of process architecture have not been studied much, and especially the quality of data that is related to the process documentation has not been studied before. This study was carried out as a case study and was largely based on the steps for conducting a case study by Eisenhardt (1989). During the first step, the research questions were formed to support the work. Three questions were created:

1. How can the quality of data in a process documentation system be evaluated?
2. How to improve the data quality in a process documentation system?
3. What are potential benefits and cost points of improving the data quality in a process documentation system?

In this chapter the answers to those questions are summarised as well as how those answers were reached.

The steps for the data quality assessment and improvement were based on the literature review by Batini et al. (2009), where they study data quality improvement methodologies. The process is divided into two phases, assessment and improvement. The steps that were included were some of the most commonly included ones in the different methodologies included in the literature review that were also seen to best suite the exact topic of this case study. The steps included in assessment phase were data analysis, identification of critical areas, and measurement of quality. In the following improvement phase the included steps were identification of the causes of errors, design of improvement solutions, evaluation of costs, and improvement monitoring. These steps were also followed in this study, with the exception of improvement monitoring, as that is a more long-term step to be continuously maintained after the other steps have been completed.

5.1 Evaluating the current level of data quality

For the first research question, it was recognized based on the literature review that there was no existing method for measuring the quality of process architecture related data. This meant the measurements for the quality were defined during this study. The two quality aspects that were considered were the accuracy and completeness of data. The measurement points were selected to be ones that would provide a good overall view to these to aspects. The eventual measurement points were 1. Use of mandatory attributes, 2. Time since last change in a model, 3. Model publishing rate. Mandatory attribute use rate signals how much of the required data about the business processes is included in the system and can be used to estimate data completeness. Time since last change indicates the relevance of the models and helps in determining the accuracy of data in the system. Model publishing rate provides information on data accuracy and on data completeness by showing the number of models in Working database in relation to the Publishing database.

The data was collected from the system and from a user survey and in the data collection and analysed according to the steps of the assessment phase. The surveys were used to help identify areas where system and management processes could be developed to better suit the user's needs, improving the work quality. The critical areas were then identified based on the number of models where the last changes were made more than seven years ago.

Based on this, the answer to the first research question, how can the quality of data in a process documentation system be evaluated, is that the details vary between cases, but the organisation would need to define what data is needed by the people using the data. Then the organisation would need to define the metrics for measuring the quality. Examples of such metrics are the age of models or other data, compliance with possible rules and guidelines that have been set by the company, and the number of official, approved models in comparison to other models in the database. When designing the metrics, one thing to keep in mind is if the defined metrics only assess the data from a certain aspect, such as completeness, and if there are other aspects that should be taken into consideration as well. To make process easier to start, the most critical areas should be defined, and development should be started from these areas.

5.2 Improving data quality

The improvement phase consisted of identification of the causes of errors, design of improvement solutions, evaluation of costs, and improvement monitoring. The causes of errors in this case were determined to be largely from the lack of knowledge by the users. This result was based on the survey, where the users mentioned that they were unaware of the current guidelines and were also hoping for improvements in the communication from the admins. The improvement solution was to ensure sufficient training for users and identify the best communication channels and methods. This would help to maintain the data quality in the future, but additional actions were also proposed to fix the already existing issues in the database. These actions will vary between organisations.

The costs were not evaluated on monetary level, but instead the focus was on what can be lost and gained from not improving the data quality. The organisations can then calculate the monetary values based on their own numbers and estimates. The biggest potential costs come from increased user knowledge achieved by improving communication, improving the quality of existing data by assessing the issues areas, contacting the people responsible for those areas, such as process owners or managers, and updating the data or removing obsolete content. Third requirement for improving the data quality was consistent monitoring of data quality during and after the other improvement actions. One identified requirement was the simplification of the tool, but for that requirement the usefulness was not as clear as for the other requirements.

The target level for data quality was defined with the process architects working in the organisation. The desired level will vary from organisation to organisation, as will the measuring points, since those are based on the user needs related to the data (Wang and Strong, 1996; Arts et al., 2002; Batini et al., 2009). For the organization in this case study, the target levels were defined with the process architects, but the users' needs were taken into consideration as well. This was done by analysing the survey results. Setting the target levels is vital, as otherwise there are no goals for the project and the success of the data quality improvement cannot be properly measured.

5.3 Potential costs and benefits

Based on the actions that were recommended to be taken in order to improve the data quality in a process documentation system, to potential benefits that can be gained from those actions were identified. The points that would be causing costs during these actions were also identified, as were the costs that might occur if the actions are not taken. The main benefits of the improved communication were increased user knowledge and satisfaction. Increased user knowledge would mean less maintenance work in the future. The costs for improving communication were also low, especially when compared to the benefits. The second requirement was the simplification of the tool and for that requirement there was a bit more uncertainty on the benefits as the survey results on the topic were conflicting. This means that the potential benefits, that would in this case be the reduced risk of mistakes, and increased usage might not be achieved. At the same time the costs for development of the system can be fairly high, at least when compared with the costs that come from improving communication. But it is important to keep in mind that these results can vary between organisations, which is why it is important to involve users in the work. The third and largest requirement for improving the data quality was improving the quality of the existing data. In most cases this will require lot of manual work from both admins as well as users, making the costs for this step quite high. This step was still seen to be vital, as otherwise the work done to increase the quality of data that is added to the system after the improvement project. If the existing issues are left unfixed, it will render all other actions useless. The final requirement is monitoring the quality of data. This means monitoring the quality during the improvement project, but also continuous monitoring after the actual project has been finished. Without monitoring, the success of the actions that have been taken to improve the data quality cannot be evaluated. The costs for this monitoring are fairly low and caused by the time that the people who are responsible for the monitoring spend to monitor the data quality instead of their normal tasks. As the metrics and measurement points have been defined in earlier stages, the measurement should not cause much design costs. The monitoring can also be automated, which causes additional costs from the design and development of the scripts but will reduce the cost in the long term.

6 CONCLUSION

The results of this study can help organisations to improve the quality of data that is related to their process architecture, therefore helping them to improve their process architecture, enabling them to become more process oriented. This study also had limitations that should be kept in mind when evaluating the results. These limitations also provide topics for future studies.

As was shown in the introduction chapter, more and more organisations are moving towards a more process-oriented culture. This having well defined processes and process architecture to support the work. Much of the benefits that are hoped to be achieved with the shift towards a process-oriented organisation can be lost if the process related data is not properly maintained. This means that the process documentation can become outdated, it can be difficult to say if a process is approved or under construction, and it might not be easy to see who is responsible of which process. This is why good data quality matters. Even if it only some of the data that is not properly maintained, there is a risk that the poorly maintained areas cannot be distinguished from the well-maintained areas, making all of the data unreliable. This means that data driven decision making will not be possible in the organisation, at least based on the data stored in the process documentation system.

6.1 Meaning of the results

The results of this study consist of tangible actions that organisations can perform to improve the data quality in a process documentation system. During the study, metrics were designed, phases and steps were introduced. These phases and steps were largely based on the literature review by Batini et al. (2009). Core requirements for successfully improving the data quality were also defined, as well as the potential costs and benefits for each of these requirements. The metrics, steps, and requirements presented in this study might not fit all organisations and will most likely require some adjustments before they can be taken to use in different organisations, but they give a good base for the process documentation data quality improvement work.

6.2 Limitations of this study

The main limitations for this study were the lack of resources, especially time and number of researchers. Studying the results of the actions was out of scope for this case study, as that would have taken an additional year to properly study. This means that the effectiveness of the suggested metrics, improvement steps, and requirements could not be empirically proven.

The other main limitation for this case study was the number of organisations included. While a single case study can provide a good setting for creating theory, as was done in this study, it also limits the study in the sense that testing that theory cannot be done, which can lower the validity of the results (Gerring, 2004).

6.3 Suggestions for future studies

For future studies it is suggested that the effectiveness of these metrics is studied, as one of the limitations for this study was that the results from the proposed actions could not be measured. Studying the results of the findings of this study would provide more empirical validation for the results or it could also reveal potential improvement areas.

The second suggestion for a topic for future studies is generalisability of the findings. Only one organisation was studied during this case study, meaning that there is no information on how much the different metrics and requirements would need to be adjusted for other organisations, or if they could be used as-is.

REFERENCES

Aguilar-Savén, R. (2004). Business process modelling: Review and framework. *International Journal of Production Economics, 90(2),* 129-149.

Aldin, L. & de Cesare, S. (2011). A literature review on business process modelling: new frontiers of reusability. *Enterprise Information Systems, 5(3),* 359-383.

Andrews, D., Nonnecke, B & Preece, J. (2003). Electronic Survey Methodology: A Case Study in Reaching Hard-to-Involve Internet Users. *International Journal of Human-Computer Interactions, 16(2),* 185-210.

Arts, D., Keizer, N. & Scheffer, G. (2002). Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study and Generic Framework. *Journal of the American Medical Informatics Association, 9(6),* 600-611.

Bandara, W., Gable, G. & Rosemann, M. (2005). Factors and measures of business process modelling: model building through a multiple case study. *European Journal of Information Systems, 14(4),* 347-360.

Barros, O. & Julio, C. (2011) Enterprise and Process Architecture Patterns. *Business Process Management Journal, 17(4),* 598-618.

Batini, C., Cappiello, C., Francalanci, C. & Maurino, A. (2002). Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys, 41(3),* Article 16.

Benbasat, I., Goldstein, D.K. & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly, 11(3),* 369-386.

Bowden, C. & Galindo-Gonzales, S. (2015). Interviewing when you're not face-to-face: The use of email interviews in phenomenological study. *Internatinal Journal of Doctoral Studies, 10,* 79-92.

Browning, T.R. & Eppinger, S.D. (2002). Modeling impacts of process architecture on cost and schedule risk in product development. *IEEE Transactions on Engineering Management, 49(4),* 428-442.

Buckl, S., Ernst, A., Matthes, F., Ramacher, R. & Schweda M. (2009). Using Enterprise Architecture Management Patterns to Complement TOGAF. *2009 IEEE International Enterprise Distributed Object Computing Conference,* Auckland, USA, 34-41.

Cambridge Dictionary (2020). Meaning of *process* in English. Retrieved 25.11.2020 from https://dictionary.cambridge.org/dictionary/english/process

Cambridge Dictionary (2021). Meaning of *data* in English. Retrieved 15.01.2021 from https://dictionary.cambridge.org/dictionary/english/data

Castillo-Montoya, M. (2016). Preparing for Interview Research: The Interview Protocol Refinement Framework. *The Qualitative Report, 21(5),* 811-831.

Dijkman, R., Vanderfeesten, I. & Reijers, A. (2016). The road to business process architecture: an overview of approaches and their use. *Beta publicatie: working papers, 350*, 1-22.

Dijkman, R., Vanderfeesten, I. & Reijers, A. (2016). Business process architectures: overview, comparison and framework. *Enterprise Information Systems, 10(2),* 129-158.

Dillman, D. & Bowker, D. (2002). The Web Questionnaire Challenge to Survey Methodologists. *Online Social Sciences.* Hogrefe & Huber Publishers, 53-71.

Eid-Sabbagh, R.H., Dijkman, R. & Weske, M. (2012). Business Process Architecture: Use and Correctness. *Business Process Management, 10th International Conference, BPM 2012, Tallinn, Estonia, September 2012 Proceedings.* Tallinn, September 3.-6., 2012, 65-81.

Eisenhardt, K. (1989). Building Theories from Case Study Research. *The Academy of Management Review, 14(4),* 532-550.

Eisenhardt, K & Graebner, M. (2007). Theory Building from Cases: Opportunities and Challenges. *The Academy of Management Journal, 50(1),* 25-32.

Fischer, R., Aier, S. & Winter, R. (2007). A Federated Approach to Enterprise Architecture Model Maintenance. *Enterprise Modelling and Information System Architectures, 2(2),* 14-22.

Gerring, J. (2004). What Is a Case Study and What Is It Good for? *American Political Science Review, 98(2),* 341-354.

Gibson, D., Goldenson, D. & Kost, K. (2006). Performance Results of CMMI -Based Process Improvement. *Carnegie Mellon University,* Journal contribution.

Gochhyat, J., Giri, V. & Suar, D. (2017). Influence of Organizational Culture on Organizational Effectiveness: The Mediating Role of Organizational Communication. *Global Business review, 18(3),* 691-702.

Gonzales-Lopez, F. & Bustos, G. (2019). Integration of Business Process Architectures Within Enterprise Architecture Approaches: A Literature Review. *Engineering Management Journal, 31(2),* 127-140.

Gordjin, J., Akkermans, H. & van Vliet, H. (2000). Business Modelling is not Process Modelling. *Conceptual Modeling for E-Business and the Web. ER 2000 Workshops on Conceptual Modeling Approaches for E-Business and The World Wide Web and Conceptual Modeling,* Salt Lake City, Utah, USA, October 9–12, 2000, 40-51.

Green, S. & Ould, M. (2005). A Framework for Classifying and Evaluating Process Architecture Methods. *Software Process Improvement and Practice, 10,* 415-425.

Hammer, M. (2007). The Process Audit. *Harvard Business Review,* April 2007 issue.

Hawkins, J. (2018). The practical utility and suitability of email interviews in qualitative research. *The Qualitative Report, 23(2),* 493-501.

Ko, R., Lee, S. & Wah Lee, E. (2009). Business process management (BPM) standards: a survey. *Business Process Management Journal, 15(5),* 744-791.

Koliadis, G., Ghose, A. & Padmanabhuni, S. (2008). Towards an Enterprise Business Process Architecture Standard. *IEEE Congress on Services, IEEE Congress on Services - Part I*, Honolulu, HI, 2008, 239-246.

Krogstie, J. (2015). SEQUAL as a Framework for Understanding and Assessing Quality of Models and Modelling Languages. *Encyclopaedia of Information Science and Technology, Third Edition,* 1611-1620.

Madnick, S. & Zhu, H. (2006). Improving data quality through effective use of semantics

Marques, J. (2010). Enhancing the quality of organizational communication; A presentation of reflection-based criteria. *Journal of Communication Management, 14(1),* 47-58.

Mayfield, J. & Mayfield, M. (2002). Leader Communication Strategies Critical Paths to Improving Employee Commitment. *American Business Review, 20(2),* 89-94.

Minoli, D. (2008). Enterprise Architecture from A to Z: Frameworks, Business Process Modeling, SOA, and Infrastructure Technology. CRC Press.

Oca, I.M-M., 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.

Pourmirza, S., Peters, S., Dijkman, R. & Grefen, P. (2017). A systematic literature review on the architecture of business process management systems. *Information Systems, 66,* 43-58.

Power, B. (2007). Michael Hammer's Process and Enterprise Maturity Model. Retrieved 15.01.2021 from https://www.bptrends.com/michael-hammers-process-and-maturity-model/

Rajhans, K. (2012). Effective Organizational Communication: A Key to Employee Motivation and Performance. *Interscience Management Review, 2(2).*

Riemer, K., Holler, J. & Indulska, M. (2011). COLLABORATIVE PROCESS MODELLING – TOOL ANALYSIS AND DESIGN IMPLEMENTATIONS. *ECIS 2011 Proceedings*. 39.

Rohloff, M. (2006). Enterprise Architecture – Framework and Methodology for the Design of Architectures in the Large. *ECIS 2005 Proceedings,* 113.

Suárez, M., Sánchez, M., Villalobos, J. (2015). Towards business process architecture analysis, *10th Computing Colombian Conference (10CCC),* Bogota, Colombia, 2015, 535-542.

Taherdoost, H. (2016). Validity and Reliability of the Research instrument; How to Test the Validation of a questionnaire/Survey in a Research. *International Journal of Academic Research in Management, 5(3),* 28-36.

Taleb, M. & Cherkaoui, O. (2012). Pattern Oriented Approach for Enterprise Architecture: TOGAF Framework. *Journal of Software Engineering and Applications. 5(1),* 45-50.

Tang, A., Han, J. & Chen, P. (2004). A comparative analysis of architecture frameworks. *11th Asia-Pacific Software Engineering Conference,* Busan, South Korea, 640-647.

The Open Group (2020 a.). The TOGAF Standard – Version 9.2. Retrieved 27.11.2020 from https://www.opengroup.org/togaf

The Open Group (2020 b.). Welcome to the TOGAF Standard Version 9.2, a standard of The Open Group. Retrieved 27.11.2020 from https://pubs.opengroup.org/architecture/togaf9-doc/arch/

Urbaczewski, L. & Mrdalj, S. (2006). A COMPARISON OF ENTERPRISE ARCHITECTURE FRAMEWORKS. *Issues in Information Systems, 7(2),* 18-23.

Vanderfeesten, I., Cardoso, J., Mendling, J., Reijers, H. & van der Aalst, W. (2007). Quality Metrics for Business Process Models. *2007 BPM and workflow handbook,* 179-190.

Van Looy, A., De Backer, M., Poels, G. & Snoeck, M. (2013). Choosing the Right Business Process Maturity Model. *Information & Management, 50(7),* 466-488.

Williams, C. (2007). Research Methods. *Journal of Business & Economic Research, 5(3),* 65-72.

Winter, R. (2014). Architectural Thinking. *Business & Information Systems Engineering, 6,* 361-364.

Winter, R. & Fischer, R. (2006). Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. *10th IEEE International Enterprise Distributed Object Computing Conference Workshops,* Hong Kong, China, 2006, 10-22.

vom Brocke, J., Simmons, A., Riemer, K., Niehaves, B., Plattfaut, A. & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in Information Systems research. *Communications of the Association for Information Systems, 37(1),* 205-224.

Zarvic, N. & Wieringa, R. (2006). An Integrated Enterprise Architecture Framework for Business-IT Alignment. *Proceedings of the 18th International Conference on Advanced Information Systems Engineering (CAiSE´06),* June 5-9, Luxembourg, 63-75.

APPENDIX 1 ATTRIBUTE VALUES

Working Database	15.1.2021						
Group	# of models	Name	Approval Date	Approved by	Model Status	Approved	Person responsible
Group 1	1606	1603	767	809	946	811	950
Group 2	121	121	0	0	2	0	2
Group 3	2975	2611	2	87	438	51	438
Group 4	2	2	0	0	0	0	0
Group 5	9	9	0	0	0	0	0
Group 6	1	1	0	0	0	0	0
Group 7	26	26	0	0	11	11	0
Group 8	48	48	0	0	1	1	0
Group 9	15	15	0	0	10	10	0
Group 10	13	13	0	0	0	0	0
Group 11	22	22	0	0	5	5	0
Group 12	1027	1027	76	26	618	80	507
Group 13	17	17	0	0	0	0	0
Group 14	1876	1876	250	257	504	252	120
Group 15	10	10	0	0	10	10	0
Group 16	19	19	1	1	2	1	1
Group 17	24	24	0	0	11	11	0
Group 18	32	32	0	0	1	0	0
Group 19	297	297	0	0	4	0	0
Group 20	221	221	0	0	11	0	0
Group 21	1475	1470	1	4	7	0	5
Group 22	38	38	0	0	35	35	0
Group 23	1783	1783	574	504	634	575	656
Group 24	99	99	0	0	0	0	0
TOTAL	11756	11384	1671	1688	3250	1853	2679

Publishing Database							
Group	# of models	Name	Approval Date	Approved by	Model Status	Approved	Person responsible
Group 1	1895	1895	1667	1667	1668	1668	1572

APPENDIX 2 AGE OF MODELS

Top Level Group Name	Last change < 1Year	Last change > 2 and < 5Yea	Last change > 5 and < 7Yea	Last change > 7Years
Group 1	720	54	99	403
Group 2	2	2	12	105
Group 3	425	908	558	1080
Group 4	0	2	2	5
Group 5	0	0	2	0
Group 6	0	1	0	0
Group 7	1	14	17	16
Group 8	7	6	9	0
Group 9	0	0	12	1
Group 10	6	0	5	0
Group 11	579	339	13	0
Group 12	5	2	3	0
Group 13	387	405	747	217
Group 14	0	3	8	2
Group 15	1	4	12	1
Group 16	6	0	0	0
Group 17	1	2	29	0
Group 18	7	3	9	0
Group 19	3	126	166	3
Group 20	11	0	0	210
Group 21	193	604	366	113
Group 22	219	133	903	68
Group 23	25	1	2	0
Group 24	0	13	86	0

APPENDIX 3 SURVEY RESPONSES

Question	Options	Number of answers	Percentage of answers
Please select the option that best defines your experience in BART		**n**	**Percent**
	Beginner (New to the system and still learning how to use it)	10	50 %
	Basic (Is able to create and edit models with relative confidence)	6	30 %
	Advanced (Is familiar with different functions and uses BART on regular basis)	4	20 %
Do you feel that using BART is too complicated?		**n**	**Percent**
	Yes	7	35 %
	No	13	65 %
Is there any specific point that you find to be too complex?		4	
	It just seems complex to get started with. Perhaps some 5 minute videos on how to get started would be good.		
	Compared to Aris designer, it is difficult. Since I don't use it regularly, I don't remember the functions and in Aris connect the user interface is not intuitive.		
	Connecting different process parts to others.		
	This is the first time I heard about it and I have to fill out a form to get any info about what BART is.		
Do you feel that you are familiar with the current process modeling rules and guidelines that were released during Q4 2020 and know where to find them?		**n**	**Percent**
	Yes	7	35 %
	No	9	45 %
	I have never even heard about these guidelines	4	20 %
In the future, in which channels would you hope to receive information on changes to the process modeling rules and guidelines?		**14**	
	Trainings (Name removed for privacy) arranged in LT were very informative, effective. For bigger changes it is the best way, I think. For small changes email is fine.		
	Email to process community (with link to BPM)-		
	Mail and in system and in meetingsforum for users		
	e-mail		
	In the site for BPM but very centralized located on the site and easy to find. Every change/new thing needs to be implemented with different change management activities towards the end users, i.e. those that are performing process mapping of any sort in the organisation.		
	From TC BPM, either on global och local level		
	E-mail		

	I use Bart only occasionally and I am not part of any process modelling community, so I am not directly receiving any updates related to processes.		
	Email trigger to read wiki with release notes		
	News on Entrypoint.		
	I don't know. Perhaps something I can subscribe and unsubscribe to. There is som much information and it seems very random what information you end up getting.		
	Teams		
	I'm quite new to working with the process modeling, and as such don't know how changes to the process modeling rules and guidelines have been communicated historically. But I think I'd like to have the changes be communicated through a presentation and/or a PDF document with examples of "before -> after".		
	Direct newsletter with links to the changes.		
	Information meetings		
Do you prefer to use some other tool for process modeling?		**n**	**Percent**
	Yes	2	10 %
	No	18	90 %
What do you see as the biggest pain point(s) with BART? (Can select multiple)		**n**	**Percent**
	Difficult to navigate	9	45 %
	Creating and editing content is too complicated due to the system	5	25 %
	Creating and editing content is too complicated due to the rules set by Telia	1	5 %
	Lack of support and instructions or difficulties in finding them	6	30 %
	Lack of support from the management	4	20 %
	Something else (please specify)	9	45 %
	not process owner/manager start using this tool for day-to-day modeling activities (now organization uses some other tools)		
	Hard to understand the 'copy'-rules - how to reuse activities/lane names without destroing things		
	to publish, reuse guidelines		
	its ok, more trainings would be nice		
	Difficult to know what to do when you want to state a process map as approved and connected to Telias Process map.		
	Multiple EA repositories - how to secure interfaces		
	No pain points.		

	As said above, BART will be used only in really small amount of people and the main stakeholders who the collaboration is needed wont work there.		
	Bad data quality makes it hard to know what is right or wrong to use. Not all is aligned with new as well which also makes it complicated		
How would you like to see BART developed in the future?		**19**	
	everyone in orgnization has modeling rights and knowledge in BART (for D2D activities), not oly process owners/managers		
	Autosaveimport BPMN from other process toolsEasy to understand - information FAD		
	Easy to resuse and navigate and connect to other system use in production		
	no suggestions		
	Always involve all countries		
	It has to be the Telia Standard		
	That we in L5 maps always connect the activities/tasks to what Information type is managed and in what way (created, used, retired or deleted) regarding those information types that we need to keep track of based on input from Infromation Management team to be compliant with different laws and regulations + that we connect the activities/tasks to what IT service(s) that supports the activity/tasks, i.e. what tool(s) do we use when performing the activity/task + connection to what customer facing product(s) and/or offering are supported by the activity/task. And what role or function are performer of each activity/task.		
	BART is probably a great tool, but more local support is needed for the userses		
	- should be used for all (main) processes in Telia		
	More process support and that related to tool support would be great. One training done in the beginning is not enough. Important to keep up the knowledge as well.		
	Quality assurance of models; clear scope, principles and support to fulfill those. Target setting and follow-up for process modelling. Visibility of process risks in process models (to support management of process risks).		
	Clear on where to work and put things.		
	I would lite to see BART first.		
	I hope there would be a group in Teams where you could also ask questions.		
	I need to use it more and try to learn the Telia rules on actual applications. Can not answer this now.		
	I don't know how BART has been developed earlier. I have no comment.		
	It is complete model of enterprise arhitecture, not just processes.		
	Hopefully we will find a solution, that is more supporting creating a process knowledge base.		
	Clean up and all countrys apply to the new.Big focus on visualization.		