Tampere University

Jesse Lyytikäinen

CONVERTING TECHNICAL DOCUMENTATION INTO DITA

An Ethnographic Study of a Conversion Process

Faculty of Information Technology and Communication Sciences
Master's Thesis
September 2020

ABSTRACT

LYYTIKÄINEN, JESSE: Converting Technical Documentation into DITA – An Ethnographic Study of a Conversion Process
Master's Thesis
Tampere University
Master's Programme in English Language and Literature
September 2020

Digital content management and structured documentation systems have become increasingly popular within companies that deal with considerable amounts of technical documentation. Implementing a structured documentation system often requires converting old content into the new system, which can lead to issues caused by the two differing systems. The purpose of this thesis is to study what kind of issues may arise during such a conversion process and how they could be prevented. The conversion process in this thesis revolves around a type of technical document, a technical specification. This study is done in cooperation with Valmet Technologies Oy.

The theoretical framework of this study consists of information management, single sourcing, and markup languages, which are all prominent concepts within the field of technical communications. Information management and its related theories of information architecture and information design are used in this thesis to provide multiple viewpoints from which to approach the conversion process and its potential issues. Single sourcing and markup languages provide background for the technical concepts discussed in this study.

Ethnographic research methodology, namely participant observation, is used in the study to gather data regarding the conversion process from the personnel at Valmet Technologies Oy. This data is processed into field notes which are then analyzed for possible issues in the conversion process. Alongside field notes, artifacts procured from the work environment are also used to support the analysis. The analysis is done by using an information analysis model and thematical analysis.

The results display that the conversion process of the technical specification faces technical issues and workflow issues. Technical issues include issues such as documentation output format and problems with a database integration. Workflow issues include problems with having multiple writers in the writing process and with increased workload brought by new tasks and roles for the employees.

The results of this thesis can potentially aid companies with their own conversion processes. For the field of technical communications, this study provides information on a common, yet relatively undocumented process within the field. Future research could be done on other conversion processes to increase openly available information.

Keywords: conversion, technical communications, DITA, ethnography, technical documentation

TIIVISTELMÄ

LYYTIKÄINEN, JESSE: Converting Technical Documentation into DITA – An Ethnographic Study of a Conversion Process
Pro gradu -tutkielma
Tampereen yliopisto
Englannin kielen ja kirjallisuuden maisteriopinnot
Syyskuu 2020

Digitaalinen sisällönhallinta ja rakenteinen dokumentaatio ovat kasvattaneet suosiotaan yrityksissä, jotka käsittelevät huomattavia määriä teknistä dokumentaatiota. Rakenteisen dokumentointijärjestelmän perustaminen vaatii usein vanhan sisällön konvertointia uuteen järjestelmään, mikä saattaa johtaa ongelmiin kahden erilaisen järjestelmän välillä. Tämän tutkielman tarkoituksena on tutkia, millaisia ongelmia tällaisen konversioprosessin aikana saattaa ilmetä ja kuinka niitä voidaan ehkäistä. Tutkielmassa käsitelty konversioprosessi rakentuu teknisen dokumenttityypin, teknisen erittelyn, ympärille. Tutkielma on toteutettu yhteistyössä Valmet Technologies Oy:n kanssa.

Tutkielman teoreettinen viitekehys koostuu teknisen viestinnän alan käsitteiden teoriasta, kuten informaatiohallinnasta, yksilähteistämisestä ja merkintäkielistä. Informaatiohallintaa ja siihen liittyviä informaatioarkkitehtuurin ja informaatiosuunnittelun teorioita käytetään tutkielmassa tuomaan useita eri näkökulmia konversioprosessiin ja sen mahdollisiin ongelmiin. Yksilähteistäminen ja merkintäkielet pohjustavat tutkielmassa käsiteltäviä teknisiä käsitteitä.

Tutkielmassa käytetään etnografista tutkimusmenetelmää, erityisesti osallistuvaa havainnointia, aineiston keräämiseen Valmet Technologies Oy:n henkilöstöltä. Aineisto prosessoidaan kenttämuistiinpanoiksi, jotka analysoidaan mahdollisten ongelmakohtien löytämiseksi konversioprosessista. Kenttämuistiinpanojen ohella myös työympäristöstä hankittuja artefakteja käytetään analyysin tukena. Aineiston analyysi toteutetaan informaatioanalyysimallin ja temaattisen analyysin avulla.

Tulokset osoittavat, että teknisen erittelyn konversioprosessissa on teknisiä ja työnkulkuun liittyviä ongelmia. Teknisiä ongelmia ovat muun muassa dokumentin ulostuloformaatin aiheuttamat hankaluudet ja tietokantaintegraatioon liittyvät haasteet. Työnkulkuongelmia ovat esimerkiksi kirjoittajien lukumäärästä johtuvat ongelmat sekä uusien työtehtävien ja roolien aiheuttama kuormitus työntekijöille.

Tutkimuksen tulokset voivat auttaa yrityksiä ratkaisemaan vastaavia konversioprosesseja. Teknisen viestinnän kentälle tutkimus tuo tietoa yleisestä, mutta vähän dokumentoidusta prosessista. Tulevaisuudessa tutkimusta voitaisiin tehdä vastaavista konversioprosesseista, jotta näistä saataisiin enemmän avointa tietoa.

Avainsanat: konversio, tekninen viestintä, DITA, etnografia, tekninen dokumentointi

TABLE OF CONTENTS

1 INTRODUCTION

Digital content management is an essential area of the field of technical communications (Isohella 2011, 26). Even though many definitions of technical communications seem to put emphasis on creating efficient and reliable information products (for example Tekom 2020; TCBOK 2020a), it is apparent that a field specializing in conveying information aptly and efficiently would require sophisticated systems to create and control the information used in these information products. Managing information products, such as technical documentation, is a challenge even with already established and functioning systems and procedures. As technologies and methodologies evolve and are sometimes replaced with completely new ones, so do the tools and methods that are used to manage information products. When the time to adopt a new technology or a method comes, the information products from the previous system often need to be made compatible with the new system. This is what is referred to as content conversion, the topic of this thesis.

In most cases of content conversion when occurring within the field of technical communications, several themes are relatively prevalent. These include concepts such as content management systems and modular writing. Content management systems (CMSs) are tools that are used in organizing and handling content, such as documentation (TCBOK 2020b). Component content management system (or CCMS) is a particular type of system in which the content is divided into smaller components which are then used to build information products (ibid.). This type of approach is closely related to modular writing, which is a method where content is written as independent components, or modules, that can be used in various situations instead of just one particular case (Ament 2002, 206). All of these themes are also prominently present in this study.

Most often conducting a conversion process is not a plug-and-play operation where the old content is simply transferred to a new content management system and evaluated whether or not it is a good fit. Instead, testing, reviewing, and evaluating is required from multiple different perspectives: what tools are available, who uses the tools, why is the system change done, what are the costs, will the new system be more efficient or not, how will it affect the personnel, does the documentation fit the system itself, and plenty more.

My aim in this study is to chart potential issues of a content conversion process by following and participating in a conversion process project. By doing so, I seek answers to the research questions of this study. The research questions are as follows:

- What kind of issues may occur during a content conversion process when transferring material to a content management system?
- What measures can be taken to ensure a fluid conversion process?

The motivation behind this thesis is to provide more openly available information on content conversion. Surprisingly, previous research on content conversion or migration was not as abundant as I initially thought. In the field of technical communications, the conversion process is often done from a linear documentation system into a modular system. A typical instance is converting existing documentation from file formats of common text editors to DITA, which is the architecture often used when producing technical documentation. This can be observed from, for example, Koppanen (2018) who investigates restructuring information and content migration, both of which are parallel themes with those of this thesis. As Koppanen’s (2018) thesis and this study indicate, companies are still adopting modular writing and DITA, which is often the tool used to achieve modular approaches, into their documentation workflows despite the relatively old ages of the method and the technology. Thus, researching the potential issues that can be encountered during a conversion process of old documentation is a crucial area of study that could help companies that are on the verge of adopting these methods and technologies to be more aware of the hurdles that can occur during such a process.

My hypothesis is that the conversion process will encounter multiple different issues throughout the project's timeline. I anticipate that these issues will be varied in nature: in a project such as this, there is an almost infinite number of variables that can be sources of potential problems, some potential examples being the massive scope of the technical documentation to be converted, the number of people producing the documentation, and the introduction of a new system and a way of producing technical documentation. These factors alone can create a multitude of different problems in terms of how the technical documentation and the personnel producing it are accommodated into the CCMS and its use.

Before proceeding further, I find it crucial to define and clarify the term *issue*. *Oxford English Dictionary* (OED) defines *issue* as "a problem or difficulty with a service or facility; a failing in any system, esp. regarded as a matter to be resolved" (OED, s. v. *issue*). This definition serves as the foundation for how *issue* is used in this study to describe problems and obstacles the conversion process faces. In more precise terms, *issue* in this study is something that prevents or hinders the conversion process and its adoption to fit the current workflow. In addition to *issue*, the terms *problem* and *obstacle* are also used synonymously.

This study investigates a case from Valmet Technologies (henceforth also Valmet), a globally acting Finnish technology company specializing in multiple different fields, such as pulp, paper, and energy industries (Valmet 2020a). In these areas Valmet provides a variety of products and services which range from maintenance services to automation solutions and all the way to entire pulp mills and power plants (ibid.). This thesis is centered around the pulp and energy department of the company which provides services and products to pulp, power, and heat producers, as well as equipment and whole facilities related to these business areas (Valmet 2020b).

The study centers around the conversion of a type of technical document, a technical specification. The technical specification in this study is a document type that contains technical data and project specific information about the parts and processes of a boiler. These kinds of documents are issued to Valmet's customers with every boiler project. Valmet wants to expand the use of their CCMS by using the technical specification as an example to test the CCMS's capabilities. The technical specification is also used as a testing ground for a database integration which, if successful, would allow for a more efficient creation process of the said document.

In terms of theoretical framework of this study, I use the field of technical communications as the starting point by making use of ideas from Haramundanis (1998) and Markel (2004), for example. From there, I proceed to investigate the concepts of information management, information architecture, and information design through Brown (2003), Rosenfeld et al. (2015), and Rockley (2003), among others. Ament (2002) is crucial work in the discussion regarding single sourcing, and Priestley et al. (2001) when examining DITA, both of which are essential concepts in terms of the conversion process discussed in this study.

As for the methodology, I use ethnography and participant observation to gather data relevant to the conversion process and its potential issues. This is done by participating in the project regarding the conversion process as an employee and also by observing the process as a researcher by gathering data on the potential problems the project and the conversion process faces. These observations are documented as field notes, which serve as the foundation for the analysis from which I will report any relevant findings. In addition to field notes, I use artifacts such as templates for the technical specifications, several older technical specifications from past projects, and meeting memos in the analysis to provide additional data sources to the study. The templates and the older technical specifications are also referred to as technical specification documents.

Analysis in this study is done by using thematical analysis to find themes from the field notes and meeting memos. The technical specification templates and the older specifications are analyzed with the help of Rockley’s (2003, 311) information analysis model to chart potential issues and answers from the target documentation itself. The results from these are then used to formulate a cohesive picture of the technical specification and its relationship to the conversion process.

The structure of the study is as follows: chapter 2 contains the theoretical framework, which includes the concepts of information management, information architecture, and information design. Moreover, single sourcing, DITA, and, finally, content conversion are also examined in chapter 2. Chapter 3 explains the methodology of this study with ethnography and participant observation being in the limelight. In addition to methodology, chapter 3 also discusses the data and artifacts used in the analysis, which in turn is reserved for chapter 4. Finally, chapter 5 is dedicated to discussing and summarizing the study.

2 CONTENT MANAGEMENT IN TECHNICAL DOCUMENTATION

This chapter and its sections discuss the theoretical framework of the study. First, the field of technical communications is examined to provide a view of the larger context in which this thesis exists. This covers the concepts of technical communications and technical documentation. After that, the second section and its subsections delve into information management, information architecture, and information design. The third section discusses concepts of structured documentation and DITA. The fourth section examines the central theme of this thesis: content conversion.

2.1 Technical communications and technical documentation

Before diving deeper into the main concepts used in this thesis, it is important to provide some necessary background, particularly in terms of technical communications and technical documentation, to gain a better understanding of the context surrounding these concepts. Finding an exact definition for technical communications is challenging since the field and its boundaries can often be blurry. Society for Technical Communication (2020) provides a definition that is divided into three characteristics, and by fulfilling one or more of these characteristics, any act of communication can be considered belonging to the field of technical communications. The definition by the Society for Technical Communication (ibid.) is as follows:

- Communicating *about technical or specialized topics*, such as computer applications, medical procedures, or environmental regulations.
- Communicating *by using technology*, such as web pages, help files, or social media sites.
- Providing *instructions about how to do something*, regardless of how technical the task is or even if technology is used to create or distribute that communication.

The technical specification documents examined in this thesis fall under the first two characteristics, since the purpose of these documents is to convey technical and contractual information, and it is almost entirely done within digital media. Although the definition by the Society for Technical Communication is remarkably broad, it still conveys the essence of what technical communications is about: conveying information, often about a specialized subject or in a specialized way. Markel (2004, 4) also notes similar attributes when discussing technical communications: essentially, technical communications is creating something, often a document, that explains concepts or instructs the reader how to achieve something.

Another term that coexists with technical communications is technical documentation. Isohella (2011, 30) notes that technical communications is so product-centered that it is often referred to as technical documentation, which in most cases is the final product of technical communications. Haramundanis (1998, 1) defines technical documentation as "…any written material about computers that is based on fact, organized on scientific and logical principles, and written for a specific purpose". This definition of technical documentation is restricted to its context of computer documentation, but with slight alteration it becomes more universally applicable. By substituting the word *computers* with, for example, *technical topics*, Haramundanis' definition becomes more general while still remaining true to the original quote. Technical documentation can be categorized into three types: marketing, reporting, and instructing documents (ibid., 3). However, dividing document types into only three categories severely narrows the scope of potential technical documents. For example, the technical specifications examined in this study includes marketing material and reporting material. Thus, it seems that it is also entirely possible that documents may include aspects from all three of types of technical documentation.

Yet, from the perspective of this thesis, the most important aspect of technical documentation is not included in the previous definition. Haramundanis (1998, 1) also mentions that technical documentation is as equally concerned about the final product, which is the documentation, as it is about the initial process of creating the said documentation. This statement is especially important since it ties technical documentation as a term to the processes around it and not just the final product itself. These processes can vary according to the organization but there are several essential details that have to be considered when preparing technical documentation. Markel (2004, 33-42) suggests that when creating technical documentation, matters such as audience, purpose, budget, schedule, and the available documentation tools are factors that need to be considered. A similar list is made by Haramundanis (1998, 24):

- Know your subject
- Know your reader
- Know the rules
- Know your tools

"Know the rules" in this context means an appropriate knowledge of language as well as writing skills (ibid., 24). One detail to note about Haramundanis (1998) and Markel (2004) is that despite the time gap between the literature and this thesis, the core ideas of technical documentation have not changed.

Out of these lists by Markel and Haramundanis, two aspects are especially important from the perspective of this thesis, namely documentation personnel and tools. Although documentation personnel are not directly mentioned in either of these, they are implied since they are the people who often need to consider the items mentioned in the two lists. These two factors, personnel and tools, are closely intertwined since writers need tools to write documentation, and tools to some degree dictate how the documentation process advances (this will later be discussed in more detail in sections 2.3 and 2.4). Naturally, several of the aforementioned aspects, such as budgeting and scheduling, have secondary interest especially when assessing the results of the study. However, the main interest lies in charting issues that may emerge from the interaction of the personnel and the tools.

2.2 Information management

As observed in the previous section, technical communications and technical documentation are mostly concerned with creating information products such as technical documents. However, as previously mentioned in the Introduction, Isohella (2011, 26) notes that digital content management has become a key aspect in technical communications as well. Thus, concepts such as information management, information architecture, and information design need to be explored to gain a better grasp of the aspects that affect the conversion process.

Information management can be considered "…the conscious process by which information is gathered and used to assist in decision making at all levels of an organization" (Hinton 2006, 2). Essentially, it is a systematic approach to handling information with the purpose of gaining some sort of a benefit from it, whether it be monetary savings, easing the workload of employees, or making other processes in the organization more efficient.

In another approach, information management is a method of organizing and controlling information with the aim of trying to tackle four main obstacles: information overload, digital rot, content and transaction management, and multiplicity of formats and media (Brown 2003, n.p.). To better understand the concrete purpose of Brown's definition of information management, it is important to examine these obstacles further. Information overload is a problem where the increasing amount of data starts to cause complications as the large amount of information becomes difficult to organize and navigate efficiently (ibid.). As for digital rot, Brown (ibid.) explains it as something that is often caused by ignorance: expecting a digital system to function autonomically and assigning no one to maintain its contents can sometimes lead to less desirable outcomes, such as messy information infrastructure and lost data. When discussing content and transaction management, Brown (ibid.) describes it as controlling and keeping track of the lifespan of the stored information all the way from first drafts to final versions. Lastly, Brown (ibid.) introduces multiplicity of formats and media, which means managing the same information and data in different forms or different information and data within the same platform.

Both of these definitions of information management offer different approaches to the term. The definition by Hinton (2006, 2) is more focused on an over-arching explanation whereas Brown's (2003, n.p.) definition has a more practical perspective. Nevertheless, from these two definitions a key aspect can be gathered: organizing information with a certain objective in mind. Being aware of this factor is important from the point of view of this thesis, since it is closely related to the conversion process. Organizing, structuring, and storing information with the intentions of using it to improve aspects of an organization is something that can be applied to both, information management and the conversion process this thesis addresses.

2.2.1 Information architecture

Whereas information management is almost like a philosophy of how all the information within an organization is treated, information architecture tends to mean how the information is structured within the organization. Creating technical documentation produces massive amounts of information within an organization or an institution: internal documents, user documentation, different versions and iterations of the same documents – the list goes on. The amount of information can quickly become overwhelming if left unchecked. As discussed in the previous section, this is what is called information overload. When faced with information overload, Rosenfeld et al. (2015, n.p.) suggest the following:

> What is needed is a systematic, comprehensive, holistic approach to structuring information in a way that makes it easy to find and understand—regardless of the context, channel, or medium the user employs to access it.

This is essentially what information architecture is about: organizing information on a macro level to ensure that when information needs to be found, it is found quickly and efficiently.

However, information overload is not the only issue information architecture attempts to solve. Rosenfeld et al. (2015, n.p.) mention that information architecture is also used to make information more easily reachable by creating multiple ways of access to the information. This can be done in several ways. For example, Rosenfeld et al. (ibid.) discuss transferring physical media to digital media, which can very well be considered as providing an additional channel from which to access the information. In similar fashion, making the same digital information available on multiple different digital platforms is an example of providing multiple channels for information – for instance, information browsed on a smart phone screen requires a different approach than information browsed on a personal computer.

Information architecture is an essential part of this study since it provides a followable framework with a clear objective in mind. Fenn and Hobbs (2014, n.p.) explain that "…when practiced, [information architecture] is most often solution focused and applies models of research, organization and feedback to understand and explore the system or systems in which the problem exists". Since the data in this study needs to be transferred between two different systems, it is crucial to have a degree of system level thinking to mitigate potential problems that could arise when transferring content between the two systems.

2.2.2 Information design

While information architecture is more focused towards planning and structuring whole systems and their operations, information design is more concerned with the content itself. Coates and Ellison (2014, n.p.) note that information design can have multiple meanings for different people. One definition they (ibid.) give is that "Information design is the defining, planning, and shaping of the contents of a message and the environments in which it is presented, with the intention to satisfy the information needs of the intended recipients". Pettersson (2002, ix) addresses information design similarly to Coates and Ellison by mentioning characteristics such as "… analysis, planning, presentation and understanding of a message – its content, language and form".

However, this thesis utilizes another approach to information design. Since the sample content that needs to be converted at Valmet already exists and there are writers who are responsible for its content, there is currently no need to examine the content from a message perspective, for instance, by evaluating the readability of the language or the retrievability of the information. In this study, information modelling provides another perspective to information design. Rockley (2003, 310-311) explains information modelling as a set of analysis procedures that encompasses the information needs of a project, for example. These analysis procedures are used to formulate models from which information products are then crafted according to the requirements of the situation (ibid., 311).

Rockley (2003, 311) lists two analysis types that provide the foundation for information modelling: audience analysis and information analysis. Of these two analysis types, especially information analysis is crucial from the perspective of this thesis. According to Rockley (ibid., 311), information analysis is done to find issues from information products related to the following points:

- Repetitious information
- Similar information
- Potential missing information
- Multiple outputs (…)
- Multiple formats (…)
- Multiple audience requirements
- Information product commonality

The first three points are relatively simple: instances of repetitious, similar, and missing information are sought from the analyzed information product. The objective of this type of analysis is to minimize the amount of unnecessary information clutter and to fill any existing gaps in the examined material. By multiple outputs Rockley (ibid., 311) means charting the target outputs to where the information will be published. Outputs are essentially the platforms and file formats in which the information is delivered. These can be, for example, PDF files, web pages or paper copies, among countless other possibilities. Multiple formats refer to multiple ways of presenting the information, for example, the same information being in a list format or in a table format (ibid., 311). The information can also have multiple audience requirements, for example in terms of market area, assumed pre-existing knowledge of the audience, and even the output format of the information. Lastly, information product commonality refers to how much of the same information is present in other information products.

Rockley's (2003) information analysis formula provides a fitting framework for the information design aspect of this thesis. The technical specification documents are examined with several of Rockley's points in mind with the primary focus being on multiple outputs, multiple formats, and multiple audience requirements, and the secondary focus being on repetitious information, similar information, and information product commonality (these will be discussed in more detail in section 4.2.1). Potential missing information is not included in the scope of this thesis, since as previously mentioned, there are writers who are responsible for the content of the specifications.

Rockley's information analysis model is used for multiple reasons in this thesis. Firstly, and most importantly, the information analysis formula focuses on how the content is used and how it functions in its contexts. This is especially important, since as mentioned before, neither the message nor the visual presentation of the content are focal points of this thesis. Instead, the information analysis formula allows more focus to be channeled towards the structure and use of the content itself. An argument could be made that structure is a part of the visual presentation. However, since the content is organized with the information structure in mind instead of its visual appearance, it is apparent that structure takes a primary role instead of the visual appearance in the analysis of the technical specification documents. Furthermore, as mentioned by Ament (2002, 4), the content management method known as single sourcing, and thus by proxy DITA, enforces the idea of separating the content from the format. Secondly, the information analysis formula is targeted towards environments that utilize single sourcing as their content management method (Rockley 2003, 310). Single sourcing is also in use at Valmet to some extent and a part of this study is to transform already existing material to follow the basic principles of single sourcing. This and the theme of single sourcing will be discussed in more detail in section 2.3.

2.2.3 Information management, architecture, and design in this study

As can be noted from the previous sections, it is evident that information management, information architecture, and information design are all terms with as many definitions as there are people researching the topics. Overlap within these terms is not rare as can be noted from definitions by Brown (2003) and Rosenfeld et al. (2015). However, to avoid this overlap of terms within this thesis, information management is regarded as a higher-level concept which includes both information architecture and information design within its boundaries, as depicted in Figure 1.

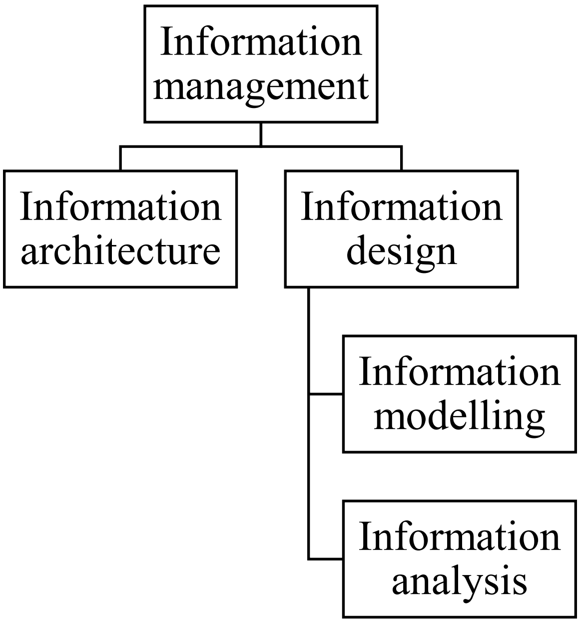

Figure 1 Information hierarchy.

Since the purpose of this study is to chart and address issues of a conversion process, the exact natures of which are currently somewhat shrouded, it is crucial to have a set of potential themes that can be used as starting points. The hierarchy of information terms in Figure 1 was created to provide a more overarching perspective from which potential issues could be observed, in addition to providing clarity to how this study manages and classifies these themes in relation to one another.

Information architecture and information design are used to organize and track information, which is why they are essential topics to consider in any content management instance. In the context of this thesis, both of these concepts are extremely valuable, since they are crucial components in securing a coherent and functioning information structure for the conversion process. For example, information architecture is employed when organizing files and information objects within Valmet's content management system and other parallel systems. As for information design, a practical instance would be assigning metadata (or data about data) to these files and information objects, which in turn helps, for example, filtering the said files and information objects.

Understanding all of these three degrees of information is essential since they are all intertwined with each other. In the context of this study it would be extremely difficult to attempt to approach the individual issues of one level in isolation since, for instance, decisions on information management level could reflect to the actions done on information architecture and information design levels and vice versa.

2.3 Single sourcing and modular content

The concepts explored in the previous sections serve as a theoretical backdrop to single sourcing and modular content. Whereas information management, information architecture, and information design are more like ideologies or guidelines to follow, the concepts introduced in this section are some of the tools and methodologies used in technical communications that can be used to strive towards a more fluid organization and use of information.

There are several ways to approach document creation, especially in an enterprise context. Every approach is affected by multiple factors: who is writing, who is the recipient, what is being written about, what kinds of tools the writers have, how much documentation is being handled, and so forth. One of these approaches is called single sourcing. The objective of single sourcing is to centralize content into one source from where it can be accessed and further processed into appropriate documents according to the documentation needs of the moment (Ament 2002, 3). This way all content in a single source system is easily accessible and reusable. Rockley (2003, 307) provides a similar definition which includes the mention of a single location where information is stored while also adding that the information stored there consists of information objects instead of files. The key difference between files and information objects is scope. For example, this thesis is contained within a one single file. One information object would instead be an individual section, paragraph or an image within this thesis. This type of content is also known as modular content (Ament 2002, 11).

Figure 2 displays the core idea of modular content. Information objects (also referred to as modules) are located in a content management system or in a database from where they are taken and assembled to fit the current information product needs.

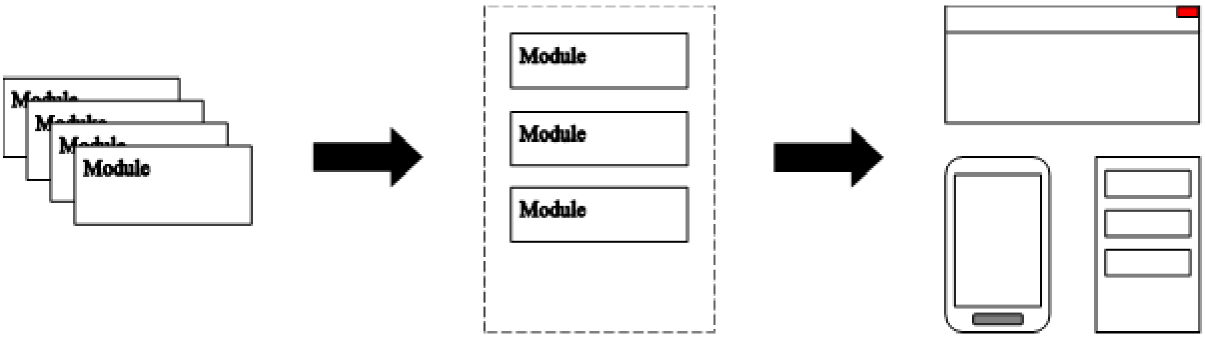

Figure 2 The basic principle of modularity.

Once the initial product is assembled, it is then rendered to fit the desired output format, whether it be a digital format such as a web page or a physical format such as a paper copy. This is an abstract model of how single sourcing operates: even though single sourcing is a methodology and not strictly a piece of technology as mentioned by Ament (2002, 1), it still requires hardware and software so it can be implemented and executed. Ament (ibid., 188) calls these technologies development tools which are divided into three categories: authoring tools, conversion tools, and content management systems. The purpose of authoring tools is to allow the creation and development of the documentation (ibid., 188). Conversion tools are used to convert the documentation to the desired output format. Although conversion tools and the topic of this thesis, content conversion, share the mutual term *conversion*, they are ultimately different entities. Content conversion happens when converting content into information objects which are then used by the conversion tools to transform the content into an output format. Finally, content management systems are data storages that are used to store and manage content, such as information objects, with the help of metadata (ibid., 188). Despite the fact that technologies and methods surrounding them change with time, the basic functions and principles Ament (ibid., 188) lists have remained intact, although instead of SGML and XML, more specialized document-markup languages, such as DITA, have been introduced and adopted into the technical communications industry. Document-markup languages and DITA will be discussed in more detail in section 2.4.

Naturally, there is no one single way of how single sourcing can be implemented and operated. However, there are guidelines and suggestions that can be used to assess what kind of a single sourcing operation is required. Rockley (2003, 308-310) introduces three levels of single sourcing (see Figure 3).

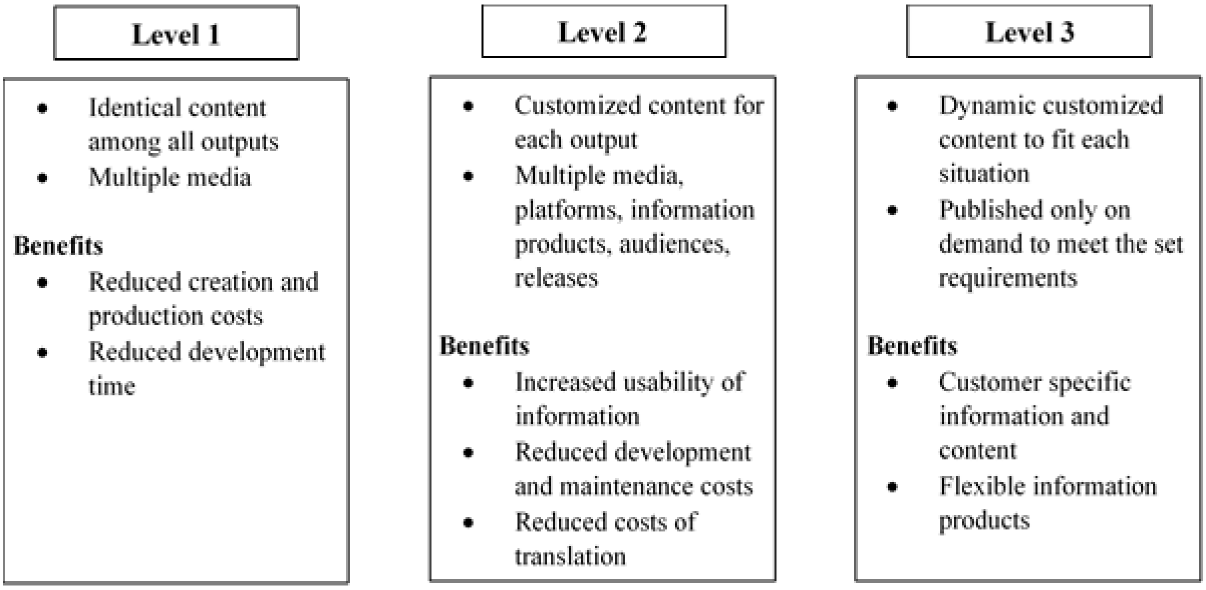

Figure 3 Levels of single sourcing according to Rockley (2003, 308-309).

In level one, the single sourced content is identical throughout all platforms it is published in (ibid., 308). Level two allows for more content customization, which leads to a more diverse portfolio of information products and to benefits such as cost reductions in information development and maintenance as well as translation (ibid., 309). Finally, level three enables the use of dynamic and customized content, which is flexible content often crafted and customized entirely according to the customer's needs (ibid., 309). The target content in this thesis is somewhere between the levels two and three by Rockley's (ibid., 309) standards since each technical specification contains unique project data and every one of them is individually tailored to each project.

As briefly mentioned in the previous paragraph and in Figure 3, utilizing single sourcing can have several benefits. Benefits mentioned by Rockley (2003, 308-309) are similar to those listed by Ament (2002, 8), which include three main areas: time and money, document usability, and team synergy. Most of these benefits are the result of reusing already created content (ibid., 8). Reusing the same content either in identical form or as altered to fit the required context eliminates the need to rewrite information that already exists, thus saving time and money when creating new information products (ibid., 8). Reuse and how it is achieved will be discussed in section 2.4. As for document usability, single sourcing increases it by guiding the content to be more universal, especially from an output perspective, meaning that it can be used in both paper and digital publications, for example (ibid., 9). As for team synergy, single sourcing encourages cohesive actions regarding for example document templates and writing guidelines (ibid., 10). Essentially, this means that all personnel working on the content follow the same principles in terms of how the content is written and how it is handled to ensure that the system works optimally (ibid., 10).

The benefits of single sourcing documentation are many, but the methodology is not without drawbacks. Issues can arise particularly during the implementation process of a single sourcing system, which ultimately may drive companies off from adopting the methodology. Fraley (2003, 58) notes that issues such as choosing the proper software to suit the needs of a company and its personnel can initially be problematic due to different needs and requirements of multiple departments within the company. This kind of an intermediate solution can lead to further consequences when the system is adopted: the system may have too many or not enough features, increased maintenance issues, and an increased need for user support. Fraley (ibid.,

58) also lists more general issues that can influence the implementation and use of single sourcing, such as scheduling issues, and inexperienced and limited personnel. Problems of this kind can prolong the implementation schedule as well as cause surprising increases in monetary affairs.

Single sourcing is also in use at Valmet. Several departments have already transferred some of their documentation to a single sourced modular system and some are conducting trials to verify if the system is beneficial for them. The purpose of single sourcing in this thesis is to provide guidelines which to follow when converting content between Valmet's content management systems. The effects of single sourcing will be discussed in more detail in chapters 4 and 5.

2.4 Markup languages and DITA

In order to properly enable modular documentation and some of its functions, such as content reuse, a document-markup language is required. Markup languages utilize tags to assign additional information to different elements of content, most often text (Priestley et al. 2001, 352). This additional information can serve multiple purposes such as acting as a set of rules how the tagged element is displayed (font, size, color, and so forth), acting as an identifier so that only particular information is displayed, and also as a tool for searching information (ibid., 352). This thesis relies on two markup languages: XML and DITA since their use at Valmet has already been established. The main focus is mostly on DITA, whereas XML's role is more of a supportive one.

XML (or Extensible Markup Language) is the markup language DITA is based on, which is why it shares some basic concepts with XML. Before examining DITA in more detail, it is beneficial to formulate a basic knowledge of XML. Fawcett et al. (2012, 3-4) describe XML as a universally adopted means of "represent[ing] low-level data" and as "a way to add metadata to documents”. Some of the most important factors of XML are its readability by computers and humans, and its ability to be universally functioning with different software and systems (ibid.). Due to this universal applicability to different situations, XML is a desirable option in many instances where data transfer takes place between different systems. Priestley et al. (2001, 352) note that already at the turn of the millennium XML was a popular choice within the technical communications community when developing structured information.

Despite the popularity and universal applicability of XML, it alone could not solve and satisfy all issues faced by technical communicators. Due to this, a more specialized architecture, known as Darwin Information Typing Architecture (or DITA), was created to provide answers to existing problems, such as information reuse and information delivery to multiple media (Priestley et al. 2001, 352). DITA is “an XML-based architecture for authoring, producing, and delivering technical information” (ibid., 352). According to Day et al. (2005, n.p), it was initially developed internally by IBM to replace their previous solution to handling technical documentation. Later, after the initial development by IBM, DITA was given to the Organization for the Advancement of Structured Information Standards (OASIS) in 2004 to further develop and maintain it (Kay 2007, 30).

In practical terms, DITA “…is used for writing and managing information using a predefined set of information structures…” (Applen and McDaniel 2009, 203). These information structures are used differently according to the content that is being created, since each of them have a unique ruleset that describes the kind of markup that can be used. Additionally, these information structures can be modified to suit the needs of any given information product through a process called specialization (ibid.). Kimber (2017, n.p.) also lists specialization as one of the important aspects of DITA.

Another essential feature of DITA is the ability to reuse content. As briefly mentioned in section 2.3, reuse is a massive factor in reducing costs and increasing efficiency when creating technical documentation. Reusing content with DITA is relatively simple since publication structures are separate from the actual content: when a document is published, the content is compiled from smaller units known as topics into a map, which can be considered to be only a collection of links that represents the structure of the document (Kimber 2017, n.p.). At its core, this is the same principle as previously demonstrated in Figure 2. This way content written once can efficiently be used in multiple different documents and contexts without having to rewrite it every time it is used out of its original context.

DITA has additional features that make reusing content even more efficient. One of the most crucial ones is conditional processing. Conditional processing means that content is filtered during publishing to display only the items that need to be shown under the set conditions. This is achieved by assigning metadata to pieces of content where it functions as an identifier that tells the publishing software whether or not to display those particular items. DITA offers a basic set of common attributes by which text is filtered: audience, platform, product, revision, and generic property attributes (Bellamy et al. 2011, n.p.). With conditional processing, the same content can be used in multiple situations, which in some cases eliminates the necessity to write multiple topics for the content. For example, a product with different audiences can be conditionally processed to suit the different needs of different audiences as illustrated in Figure 4:

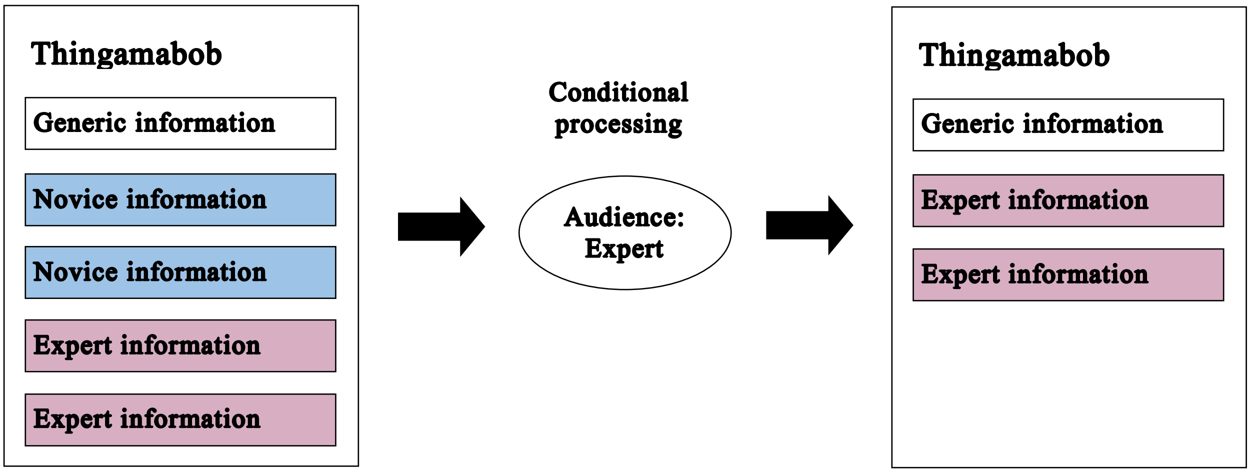

Figure 4 Illustration of conditional processing.

As can be observed from Figure 4, there is only one source where all product related information is located. Audience specific information is tagged to belong to the respective audience, in this example *novice* and *expert*, which means that when publishing with the condition set to display the information for expert audiences, only the tagged information for the expert audiences will be shown in the output document. So, instead of writing multiple source topics, writers can combine the information of multiple products to just one topic, which can then be conditionally processed to show only the relevant information when publishing.

Database integration is another feature where DITA and XML can be utilized to aid with documentation processes. A nontechnical definition for *database* from *A Dictionary of Computer Science* (DoCS) is: "a collection of data on some subject however defined, accessed, and stored within a computer system" (DoCS, s. v. *database*). In the case of this study, connecting a database containing project and product information to the CCMS would allow for a faster documentation process. Essentially, the integration would function by feeding information from the database into a library in the CCMS, from where specific information could be linked with the help of XML and DITA into the modules that require it. Figure 5 demonstrates the basic concept of how the integration would function:

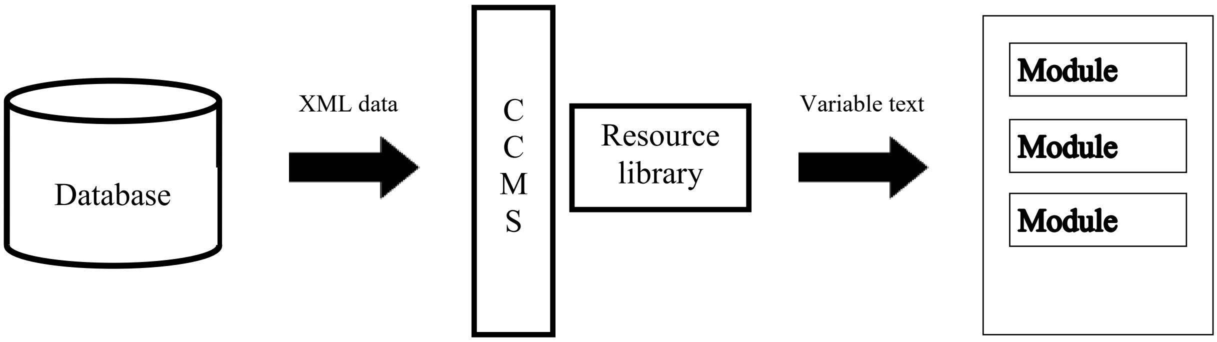

Figure 5 Database integration: flow of data.

First, as depicted in Figure 5, the data is either sent or fetched from the database into the CCMS where it is assigned to a document specific resource library. This library is where the variable parameters are stored. The data is then fed to match each parameter, after which the variables are then updated into the document itself.

Since the CCMS utilizes DITA, which is an XML-based architecture, and XML, as mentioned by Fawcett et al. (2012, 4), is a fairly universal language for "passing data", forming a link between the content management system and a database will not be hindered by a mutually illegible language. If the database integration is a success, it would allow, for example, data regarding different parts unique to the project to be fed automatically into the project specific documentation instead of inputting all of it manually. On a document level, this is achieved by using variable text. Put simply, variable text functions by creating a resource file where the variable parameters are stored. In this case, these parameters can be different parts, for instance *part1*. Then, these parameters are given values: in this example *part1 = cooling fan*. Now, when the contents of the document and the resource library are linked, *cooling fan* appears whenever *part1* is used in the text. However, were the value of *part1* changed to something else, for example *cooling fan deluxe*, all mentions of *part1* would change appropriately. This type of variable text can thus be used to reduce the amount of potential manual labor if the database integration is a success.

DITA is a crucial factor in this thesis since it is the markup language used in the new content management system at Valmet. As can be seen from features such as content reuse and conditional processing, DITA serves similar purposes as the previously discussed content management method, single sourcing. In fact, DITA in this case is one of the technologies that enables the method to be used in the first place: as previously mentioned in section 2.3, even though single sourcing is a methodology and not strictly a piece of technology (Ament 2002, 1), there still needs to be tools that enable the features single sourcing strives to achieve. The purpose of single sourcing and DITA in this context is to enable an environment which can facilitate the technical specification documents as well as enable the two previously discussed features, conditional text, and database integration, into these document templates.

However, implementing single sourcing and DITA is not an operation without drawbacks as it comes with costs: not only does the software cost money, it also affects how people work. This means that the personnel that are transferring to DITA and single sourcing often have to be trained to be able to operate in the new system. Many drawbacks of single sourcing can also be applied to the implementation of DITA due to the intertwined nature of the two concepts. Because of this, issues mentioned by Fraley (2003, 58) regarding single sourcing, such as inexperienced and limited personnel, and scheduling issues also fit the DITA implementation process.

2.5 Content conversion

In Valmet's case, to fully utilize the features DITA and single sourcing provides, the content has to be converted from its old format, DOCX, to DITA. Samuels (2014, n.p.) mentions that conversion means "…the processes that take your content pre-DITA and apply DITA tags to it until it is fully structured and valid DITA markup". Conversion is a part of a larger whole known as adoption which includes all processes and procedures that are required for content to be transformed from unstructured to structured, such as "…tool selection, content strategy, planning, filling roles, implementing reuse, setting up publishing, and a lot more" (ibid.). Currently, DITA has already been adopted at Valmet to cover the content that most benefits from it, but now Valmet is looking to expand its use further, one example being the technical specifications this thesis addresses. This means that the groundwork for DITA is established, which means that the tools and practices are already in use to some extent.

Another term that exists along *conversion* and *adoption* is *content migration*. Dye (2007, 26) describes content migration as transferring large quantities of digital content into a new CMS. All of these terms, *conversion*, *adaptation*, and *content migration* bear similar connotations but despite these similarities, each of them is still a unique term that describes a specific action. *Adoption* is more concerned with adapting to a whole new system and technology, whereas *conversion* leans more towards the practicalities of transferring content to a CMS. *Content migration*, however, appears to be a larger scale operation when it comes to moving content. This study uses the term *conversion* due to the different connotations of scale in the terms: both *adoption* and *content migration* imply a more large-scale operation, whereas *conversion* appears to refer to a more limited operation, which is more fitting to the context this study.

In the case of this study, the conversion process itself is done manually by following a simplified version of a content conversion strategy as proposed by Bellamy et al. (2011, n.p.). The reason why the strategy by Bellamy et al. (ibid.) is not used in its full form is that it assumes that there is little to no foundation present in terms of modular writing. However, due to Valmet's adoption of DITA, some of the steps in the strategy can be omitted and focus can be diverted towards more relevant steps. Below is the simplified and adapted version based on the strategy by Bellamy et al. (ibid.):

1. The content is evaluated and divided into topics.
2. Each topic is processed through an XML editor, which is used to assign the content with DITA tags.
3. The topic is reviewed and edited if necessary.
4. The topic is saved into the content management system.

Since the content is in DOCX format it is fairly efficiently converted to DITA due to mutual XML properties, thus often requiring only a quick review and few minor edits after the process.

Once the initial conversion process is done, focus can be shifted towards other topics, such as the previously mentioned conditional processing and database integration. After this groundwork is done, it is time to structure and assemble the document. This is done by linking the previously created topics into a map which holds the general structure of the document. When the structure is ready and the topics are linked to the map, the final document is ready to be published. Final step is to choose the output format which depends on whether the document will be published in a digital format or as a paper copy.

3 METHODS AND DATA

This chapter addresses the methods, namely ethnography and participant observation, used in this thesis, as well as the data gathered with the help of these methods. Moreover, this chapter provides a detailed look into the technical specification as a document type and the artifacts used as data in this study.

3.1 Ethnography

People operating in different contexts are the focal point of ethnographical research, which allows for a broad spectrum of potential research areas. Kramer and Adams (2017, 457) provide a solid definition of ethnography: "Ethnography is a qualitative research method in which a researcher—an ethnographer—studies a particular social/cultural group with the aim to better understand it". However, in the case of this thesis, the emphasis is on the data the people produce rather than the people themselves. This data is used to chart the conversion and documentation processes which are the entities this study tries to understand. Naturally, the employees of Valmet, or rather the selected group of personnel working on the project, form the social group from which the data is gathered.

Studying a social group can have numerous different approaches in terms of what kind of data is collected, how it is collected, and how the collected data is interpreted. All these factors vary depending on the research question and the approach the researcher has decided to use. Kramer and Adams (2017, 459-460) introduce different types of data that is often collected during an ethnographic research process, including field notes, interviews, and documents and artifacts. Campbell (1999, 536) also lists similar data gathering methods for qualitative research purposes in workplace contexts with the difference of adding literature review to the list and replacing field notes with observation.

Field notes are collections of observations made by the researcher while working in the field, which can also include interviews the researcher conducts during observation (Kramer and Adams 2017, 459). As noted by Kramer and Adams (ibid., 459), informal interviews can be included within the field notes. In this thesis, informal interviews are used to gather information from people participating in the conversion process. This thesis utilizes these kinds of interviews instead of more structured ones because they offer flexibility and swiftness that the reactionary nature of the collected data requires.

Although informal interviews are a vital channel of information, the main source of data in this study is in the form of field notes and artifacts collected from the work environment. Artifacts are concrete objects, such as texts in different forms, produced during day-to-day interaction within a work community (Campbell 1999, 538). Examples of artifacts are the already published technical specification documents. Field notes and artifacts are expanded upon in sections 3.3.1 and 3.3.2.

Using ethnographic methods in organizational contexts is not unheard of, on the contrary, they appear to be rather well adopted to organizational use. For instance, Neyland (2008, n.p.) mentions the growth of ethnographic research methods in organizational contexts, especially in marketing and technology development. This growth is due to "…its utility in providing in-depth insights into what people and organizations do on a day-to-day basis", as well as its flexibility and applicability to different uses (ibid., n.p.). Ethnography is also employed in the field of technical communications. For example, Hovde (2000) utilizes ethnographic methods while studying how technical writers formulate images of audiences in two different organizations. In another study, Martin et al. (2017) use ethnography to examine the effects of technical writers promoting user advocacy at their work environments. These studies illustrate that ethnography can be applied and it is often utilized in the context of technical communications. Although Hovde (2000) and Martin et al. (2017) explore themes vastly different from the topic of this thesis, they both showcase the potential of ethnographic research methods in the field of technical communications.

3.2 Participant observation

Participant observation is a method with many ties to ethnography. Gans (2010, 542) describes how participant observation and other qualitative research methods are often labelled under the term *ethnography*, which rather well illustrates the closeness of the two methods, as well as the difficulty of using the two terms since they are sometimes used almost interchangeably. However, for the sake of clarity, the differences between ethnography and participant observation need to be defined to avoid any unnecessary confusion. As previously mentioned, ethnography studies people operating in different contexts. This definition leaves open the question of *how*, *why* and *what* exactly is studied. Participant observation gives answers to the *how* but the answers to *why* and *what* are up to the researcher to decide. Participant observation can be defined as actively taking part in the daily lives of a community with the intention of trying to gain a better understanding of the said community (DeWalt and DeWalt 2010, 12). This definition is extremely close to Kramer and Adams's (2017, 457) definition of ethnography, with the exception of the researcher being an active agent involved in the community.

Participating in the everyday activities of a social group provides the researcher an insider approach which gives direct access to the people and how they operate in the field (Jorgensen 1989, 20-21). In the case of this study, my role doubles as a researcher and as a member of the project team assigned to improve the current documentation process of the technical specifications, as well as to eventually transfer the documentation into the new system. Conducting the study using participant observation allows me to directly gather data from the personnel working on the project as well as from the people who will potentially use the new system in their daily tasks.

Due to the fact that I am also working on the conversion process, I must be aware of my own actions as well. This means that I must record and consider my decision making and actions with the same degree of importance as all the other data collected via observation. However, as most of the data is gathered from other people and not from me directly, nor am I the person creating the documentation for the system, the focus remains on the other people working on the project. Nevertheless, I must be critical of my own actions during the study and I must practice a degree of self-reflection when observing and participating in the project.

The research questions and setting of this study make participant observation a fitting method for two reasons. Firstly, majority of the information is only available from the network of people working on the project or indirectly participating in it. The data required for this study is mostly based on either people's knowledge or experience. Secondly, this information is in a sense cumulative: as the project progresses, old pieces of information may reveal new findings which are used to build upon the older data. Thus, being on site for a prolonged period of time allows the information to be kept up to date throughout the project's timeline.

3.3 Data

As briefly mentioned before, the data in this study can be divided into two categories: observational data (field notes) and artifacts. As the focus of this study is in the data people produce and not directly the people themselves, individual people will be referred to according to their more general role in the project instead of their job titles in order to preserve anonymity. People from whom the data was gathered were informed that I would be gathering data for research purposes to which they gave their consent.

3.3.1 Field notes

Field notes form the backbone of the data used in this thesis as they are a common building block in ethnographic and participant observational studies. Simply put, field notes are "…used by researchers to record observations and fragments of remembered speech" (Bloor and Wood 2006, 82). In the case of this study, these notes are diary-like entries written during and after office or remote days. Field notes were gathered between October 2019 and May 2020, with some preliminary notes from August and September 2019. The field notes were analyzed to provide a cohesive picture of the issues, and to seek answers and solutions to the research questions of this study. All data for the field notes was gathered from project meetings or from smaller meetings with individual project personnel. When referring to field notes, I use the abbreviation FN and a number tag indicating the month during which they were gathered, for example, FN11 equals to field notes gathered during November. To give a concrete example of the field notes, below is a translated excerpt from field notes made in October (FN10):

> Had a meeting with the project lead. No major breakthroughs since the last meeting with the service provider. Although there were no updates, the meeting was not for naught: I introduced a new approach to the conversion: different templates could be made in the CCMS to cover different projects, for example, they could be divided according to the project type or location. New approach was noted, but as discussed, progress of other matters needs to be monitored before deciding anything.
>
> Could be an option if Word output is successful. If it is, the templates could be in the CCMS, from where the person responsible of their maintenance could fetch them when required and send them to the writers. […]
>
> Sent queries about the progress of the material sent to the service provider. Apparently, people are on vacation. Status updates most likely next week.

In addition to small text passages such as this one, the field notes also include bullet lists of thoughts and observations I considered important.

Utilizing field notes as a form of research data offers several benefits. Firstly, all kinds of communication, whether it be spoken or written, can be noted down in a relatively free form so that every potentially useful piece of information has a chance to be recorded. Secondly, as noted by Tenzek (2017, 563), collecting field notes gives the chance to acquire authentic on-location data which is "… considered critical to understanding phenomena encountered in the field".

However, field notes are not without downsides. Emerson et al. (2001, 353) note that "The ethnographer writes about certain things that seem 'significant', ignoring and hence 'leaving out' other matters that do not seem significant". This indicates that field notes are prone to human error and subjectivity, which is to be expected because reporting data based on human experience and communication can be rather difficult to address objectively. In addition, subjectivity can also be present during the analysis of the data since qualitative studies are rarely without any traces of bias. This is extremely important to keep in mind, since I am the person who creates the data through observation and analyzes the same data. However, to provide a more neutral point of view to the study, artifacts are also used as a more objective data source.

3.3.2 Artifacts

As mentioned previously in section 3.1, artifacts are concrete objects produced in day-to-day interaction within a work community (Campbell 1999, 538). Artifacts are mostly used in this thesis as supplementary material to the field notes. These artifacts are divided into two categories: technical specification documents and meeting memos.

Technical specification documents include the templates that are used as the skeleton structure when compiling a new specification, and a set of already released technical specifications. The older specifications are taken into the dataset to provide a more cohesive understanding of how the document and its contents behave and evolve through each project.

In the context of Valmet, a technical specification is a type of technical document that contains information about a specific boiler project. The information found in technical specifications can be, for example, data tables, as presented in Table 1.

Table 1 Component specifications. Yellow markings indicate variable text.

	Unit	**Value**
Width	mm	xxxx
Depth	mm	xxxx
Height (bottom – roof)	mm	xxxxxx
Cross section, free board	m^2	xxxx
Grate area, about	m^2	xxxx
Refractory (lower part)	**Unit**	**Value**
Refractory height,	mm	Xxxx
Material		LC/Sic
Thickness from the tube	mm	xxxx
Refractory (upper part)	**Unit**	**Value**
Refractory height, from bottom	mm	xxxx
Material		SiC
Thickness from the tube	mm	xxxx

Data tables often contain precise technical information, such as measurements and materials. In addition to data tables, technical specifications also include descriptive passages, such as those presented in Figures 6 and 7.

1.1.2 Internal walls

Walls of feed water pump room are made of concrete for noise protection.

In electric rooms internal walls shall be either brick, concrete or mineral wool insulated steel sandwich elements.

Walls of elevator shaft are made of trapezoidal and plain galvanized steel sheeting with PE or acrylic coating.

1.1.3 Cladding colour scheme

Colours of surfaces shall be mutually agreed form the standard colours of cladding material suppliers. Metal shade colours shall not be used.

Figure 6 An example of descriptive text regarding construction details.

2 EXECUTIVE SUMMARY

2.1 Valmet in brief

Valmet is the leading global developer and supplier of process technologies, automation and services for the pulp, paper and energy industries. Valmet's strong technology offering combined with advanced services and automation solutions improve the reliability and performance of customers' processes and enhance the effective utilization of raw materials and energy.

Valmet has over 220 years of industrial history. In 2019, Valmet's net sales were approximately EUR 3.5 billion. Valmet's head office is in Espoo, Finland and its shares are listed on the Nasdaq Helsinki.

Figure 7 An example of non-contractual descriptive text.

These descriptive passages are used to describe different parts and processes, such as the descriptions in Figure 6, as well as other descriptive text, such as the brief company introduction in Figure 7.

The scale of a technical specification is massive, meaning that its length can vary from 300 to 500 pages of information, although this includes all the appendices, drawings, and plans, which do not belong in the scope of this thesis. Thus, the analyzable material in terms of the templates narrows down to approximately 200 pages, and in terms of the already released specifications to around 800 pages. The estimates are done based on the combined number of pages from the templates and the total page count of the already released documents from which the previously mentioned information that is out of the scope of this thesis is excluded.

Currently, a technical specification is created by compiling a draft version from the templates with slight adjustments done to them according to the situation and the customer. Each of these templates cover one area of interest, such as health, safety, and environment issues, quality assurance, or construction. These areas of interest have an assigned person (or persons) who maintains the template and customizes it according to each individual project and its needs. This draft is then sent to the customer as a proposal of contents of the contract that defines what is going to be provided and by whom. Most often the customer reviews the proposal and then makes their own demands, requirements, and corrections to the document. The process is repeated until both, Valmet, and the customer, are satisfied with the contents of the specification, which is then solidified into a contract.

As for meeting memos, they include all memos procured between August 2019 and May 2020, which sum up to 11 different files of varied file formats and lengths. These memos are notes made by someone, often the project lead, during meetings and they include information regarding the topics discussed during each individual meeting, thus serving as an important source of information for the project, as well as the topics of this thesis. The memos are referred to similarly to the field notes: abbreviation M and a number tag indicating the month during which the memos were created, for instance, M01 would refer to memos from January.

4 ANALYSIS

The main purpose of the data described in section 3.3 is to reveal the issues that the conversion process faces. In this chapter, section 4.1 investigates the observation data collected throughout the project, whereas section 4.2 examines the technical specification documents and memos in greater detail.

To form an overview of the conversion process, I produced a timeline illustrating the progression of the project based on field notes and artifacts gathered throughout the ten-month period. The timeline presented in Table 2 illustrates important events of the project, such as identified issues, discovered solutions, and actions taken. I created this timeline to provide an illustration of the project to aid in formulating a cohesive picture of the project as a whole.

Table 2 Project timeline.

	August	September	October	November	December	January	February	March	April	May
DOCX output										
Database integration										
Licenses										
Training										
Workflow/roles										

Legend:

Issue identified	Solution proposed	Actions taken	Unknown/idle

The timeline is a chronological representation of the project month by month. The leftmost column lists the issues discovered during the study. The progress of these issues is color-coded in the table according to their status at the time. I will discuss each of these issues located in the leftmost column in detail in the subsections of sections 4.1 and 4.2. Section 4.3 examines the proposed solutions and actions taken towards solving these aforementioned problems.

Additionally, it must be noted that this is not a complete reproduction of the progress of the project: this thesis only reports its development from the perspective of the research questions of this thesis.

4.1 Observation data

The analysis in this section is based on the data I gathered by attending meetings remotely and on-site between August 2019 and May 2020. Prior to choosing participant observation as the method for gathering data in September, I had already started taking notes earlier in August, which means that some of the earlier entries were not as focused as the later entries in terms of their content and approach. Nevertheless, the earlier entries provide crucial information about the inception of the project and thus serve an important function in the study.

The process I used to analyze the field notes follows the principles of thematical analysis. Braun and Clarke (2012, 57) define thematical analysis as "a method for systematically identifying, organizing, and offering insight into patterns of meaning (themes) across a data set". I read through the gathered data with the purpose of trying to find and identify potential problems the conversion process encounters. After analyzing the data and finding a number of obstacles the conversion process faces, I realized that they could be divided into two relatively clear-cut categories: technical issues and workflow issues. Technical issues comprise of technical difficulties, such as issues with software, and other problems with the available tools and services. Workflow issues focus on problems that the personnel in the process face, in addition to some issues regarding resource management, such as those relating to time and money.

4.1.1 Technical issues

Multiple different technical issues arose during the project ranging from relatively small inconveniences within the CCMS to larger system scale issues that cannot be simply fixed overnight. This section explains the technical issues the project encountered, in addition to reporting their progression and development. The most prominent problems occurred with the output file format the technical specification requires and its clash with information management and the centralized principle of single sourcing (see sections 2.2 and 2.3), as well as the difficulties of establishing a connection between the CCMS and the database containing project and product information (section 2.4).

Output file format

One of the first technical issues that arose was the lack of DOCX output from the CCMS (FN08). Most often all documents published from the CCMS are in PDF or HTML formats, which explains the current situation: there has been no need to transform the content into DOCX in larger quantities. However, this project is one example that would benefit from such an update, although from a wider perspective, implementing the DOCX output would only be useful in very niche situations where post-publish editing is a necessity.

The reason why this new output format would be useful is the writing process of the technical specifications: once the initial proposal for the technical specification is done, it is sent to the customer, who then reviews and edits the document according to their own needs and standards. After the customer has done their review, the document is sent back to Valmet for further inspection. This process is repeated until both parties are satisfied with the result (see section 3.3.2). The key aspect here is editability: DOCX is readable with common text editors, such as Microsoft Word, and it also allows the use of features most text editors have, such as tracking changes done to the document.

During the first project meeting where the document conversion project was introduced, several writers voiced concerns about the CCMS's web editor and its lack of features, such as tracking changes done to the document (FN08). Technically, the web editor has this feature, but it is not as sophisticated as the one found in Microsoft Word in which technical specifications are currently prepared. This and other features requested by the writers, such as proofreading utilities (which are also present in the editor, although, in more rudimentary form) amplified the need to have the previously mentioned DOCX output available.

Since the content in the CCMS is in the form of information objects instead of traditional files, as discussed in section 2.3, the content needs to be processed and published into a readable file format, as it otherwise cannot be accessed without the necessary permissions to the software and the repository in which the information objects are located. PDF is often the output format of choice due to the universal applicability and readability of the file format. This way the information objects can be compiled into a single readable file that cannot be edited. DOCX instead is of more curious nature since it allows for post-publish editing, which in this case is both a positive and a negative feature, depending on the perspective. From the writer's perspective, it is a positive feature since it allows them and the customer to freely edit the document (FN08). However, from the perspective of the CCMS, publishing to DOCX creates suboptimal circumstances since an editable output format interferes with some of the principles of information management. As previously discussed in section 2.2, Brown's (2003, n.p.) list of issues that information management tries to mitigate contains four items: information overload, digital rot, content and transaction management, and multiplicity of formats and media. Out of these four items, information overload and content and transaction management are at risk in this instance. Firstly, information overload: having an editable version or potentially multiple versions of the document after it has left the CCMS creates undesired information clutter. Secondly, content and transaction management become much more arduous tasks when editable versions are relocated from the system to, most often, personal hard drives of the writers. Both of these are undesired side effects of such a file format change, and it must be weighed whether the gained benefit from the file format change overrides the negative impact of these aspects.

Notions about single sourcing and content management systems by Ament (2002, 191) are also essential. As introduced in section 2.3, single sourcing is a content management method where the content is stored in a single location from where it can be accessed, processed, and reused into whole documents (Ament 2002, 3). The problem here is that when the content is processed into DOCX it remains editable after it has been published. Publishing to DOCX causes negative consequences, such as loss of control after the document has been published and diminished reuse capabilities of the content. Firstly, version control becomes arguably a more difficult task once an editable document is published from the system since the editing is done in software that is not linked with the CCMS, which is designed to keep track of all the information objects within it. Secondly, although reuse capabilities are already somewhat limited to the core elements of the technical specification templates due to the project specific nature of the document type, the DOCX output prevents any further reuse opportunities simply because it is known that the content will be altered in unpredictable ways by the customer during the negotiation phase, which eliminates the need to develop reuse further within the CCMS itself. In addition, DOCX is not automatically backwards compatible with the current CCMS, which means that were the technical specifications in the CCMS to be updated with the most current information, the updates would have to be done manually.

Implementing the DOCX output to the CCMS appears to be somewhat challenging. When compared to the current needs for the feature, the cost of its implementation would prove to be relatively high. Ordering the work from the provider of the documentation software would cost a significant amount of money for a feature with niche use. On the other hand, implementing the feature inhouse would draw resources from other important tasks (FN11).

As of now, these are the two available options from which to choose if having another output format is deemed necessary.

Database integration

Connecting the project and product information database into the documentation to allow for at least a semiautomatic input of data straight into the documentation is one of the most important features of this project (FN10). This is due to the massive amount of technical data found in the technical specifications – integrating the database with the CCMS would enable a faster process of inputting technical data into the specifications with increased efficiency and accuracy. However, achieving this goal is not a simple task. As much as it brings answers to problems, such as how to quicken the creation process of the technical specifications, it brings its own issues to the table as well. Issues the database integration faces are related to system compatibility and the implementation of the database integration.

Firstly, the compatibility of the database with the CCMS: the basic principle is that information from the project database could be exported in XML format to the CCMS, where it is embedded into the documentation with the help of variable text, as discussed in section 2.4. Since DITA is an XML based architecture, the CCMS can easily interpret the XML data exported from the database. However, some compatibility issues were noted. For example, in early September during a meeting dedicated to charting the concept of the database integration, a problem was reported regarding the embedding of variable information (FN09). The problem was with the identifiers that are assigned to each piece of variable text. Assigning unique identifiers to the documentation ensures that the systems communicate properly and that the correct information is retrieved. Without proper identifiers the systems would not be able communicate, which in turn prevents data from transferring from one system to another. During the meeting, one experienced CCMS user noted that a prototype integration had been done before where the issue was first encountered and where the core of the problem was revealed. The issue here is that the number of potential variables, which all require unique identifiers, could prove to be immense. Apparently, this has been a known obstacle for some time, but this was the first time I encountered any mention of it.

Secondly, even though the XML data is easily read by both systems, there are still several obstacles regarding the communication between them. One reoccurring point of discussion in several meetings was that establishing a connection between the two systems requires resources, such as personnel with programming skills and knowledge about the operating principles of both systems (FN08, FN09). Without the proper resources, the implementation of the integration would halt, thus affecting the whole progress of the project as well.

Thirdly, in the case that the database integration is successful, another single sourcing related problem is inevitable. During a project meeting in February, it became evident that if changes are made to the information within the database after a technical specification has already been published as DOCX, these changes would have to be manually updated into the most current version of the technical specification (FN02). This is caused by the lack of backwards compatibility of DOCX to the CCMS from where the information is retrieved, which means that there is no way for the CCMS to track any changes done to the document after it has been published. This negates some of the positive effects gained via single sourcing, which according to Ament (2002, 3), are reduced costs and less time spent on writing the documentation. Thus, rewriting content or manually inputting any changed data can be seen as factors that challenge the efficiency of this process.

4.1.2 Workflow issues

In addition to technical issues, several workflow issues were also encountered during the project. Here workflow issues are considered to be problems that affect or could potentially have an impact on the personnel creating the technical specifications and vice versa.

Issues caused by having multiple writers

Several workflow issues can be traced back to having multiple writers creating the technical specifications. Having multiple writers collaborate on a piece of documentation is not by any means abnormal. However, in the case of the technical specifications, the writers are mostly engineers who in addition to their other tasks write and update the contents of the technical specifications (FN09). Essentially, this means that the writers have less time to dedicate towards the writing process, which leads to some of the issues described below.

The first issue with multiple writers is the required number of software licenses (FN08). The potential number of different software licenses the writers need is up to three licenses. Each individual writer would require access to at least the browser-based web editor, while the other two, one for a more advanced XML editor and one for managing content and publications, are not strictly as necessary for an individual writer's tasks in this case. This narrows the required number of licenses down to only one, which in turn saves a significant amount of money and time due to the reduced amount of training required. However, the writers still need to be trained to use the web editor, the use of DITA tagging, and they need to be introduced to the basics of the content management system as well. This brings another issue to surface: the ratio of effort between training the writers to use the new system and a new set of tools, and the benefit gained from it, is in imbalance. Training writers to use the CCMS and DITA tagging is not necessarily the most optimal route, especially when some of the core principles of the system appear to contradict with the needs of the document and the writers. The workflow will become unnecessarily complex if the writers are required to use the new system and the tools that accompany it for one task only. Another aspect that affects the workflow is the fact that all writers are not currently situated at the same office (FN09). Although not a major issue, it still creates potential problems with communication, training, and support.

New tasks and roles

Another workflow related problem arose from the maintenance that the documentation in the modular system requires (FN10). For example, if the documentation needs to be updated, someone with the proper software licenses and permissions needs to access the information objects from the CCMS and update the requested sections. This implies that either someone with existing licenses and permissions has to accommodate additional tasks, or new licenses must be procured either to the writers or to someone else without creating any extra workload.

Another similar issue arises when it comes to publishing since it also requires the use of specialized software as well as particular permissions. The same dilemma as mentioned above applies to publishing, too: more licenses and training are required if delegating the work to someone with existing licenses and knowledge to use the CCMS is not an option.

Both of the abovementioned issues indicate that existing personnel have to adopt new roles as the persons responsible for the new tasks, or alternatively, new personnel would need to be hired to fill the existing gaps in the workflow.

4.2 Artifact data

Artifact data in this thesis consists of two types: technical specification documents (templates and older technical specifications) and other artifacts, namely the meeting memos, as explained in section 3.3.2. The technical specification documents are analyzed with the help of Rockley's (2003, 311) information analysis model, and the other artifacts are analyzed by referring to the list of technical issues and workflow issues that were examined in section 4.1 and its subsections to see how the artifacts reflect these themes that had emerged.

4.2.1 Technical specification

As described in section 3.3.2, technical specifications consist mostly of tables containing technical data and descriptive text regarding the parts, processes, and services related to a boiler project. Dividing the information into these two categories, data tables and descriptive text, helps with the analysis of the technical specification documents: tables are data that is to be input from the database through the use of DITA's ability to process variable text, as discussed previously in section 2.4. Thus, it allows me to divert more attention towards the descriptive text passages. The technical specification documents are reviewed and assessed through the perspective of Rockley's (2003, 311) information analysis model, which was introduced in section 2.2.2. This model provides a clear set of points to look out for in content that is to move to a single sourced system.

As I previously mentioned in section 2.2.2, I have divided aspects of Rockley's information analysis into two categories of aspects according to their relevance to the context of this study. I made this division based on how each of these aspects affect the conversion process: primary aspects are crucial in terms of their impact on the conversion process, whereas secondary aspects are not as necessary. Nevertheless, I consider them to be useful, especially if the conversion process is a success, since they provide additional information that can be used to streamline the process further, for example in terms of reusing content. Below are the lists of primary and secondary aspects:

Primary

- Multiple outputs
- Multiple formats
- Multiple audience requirements

Secondary

- Repetitious information
- Similar information
- Information product commonality

Having read through the technical specification documents, I evaluated their content from the perspective of the beforementioned six aspects to find potential issues or other factors that could affect the conversion process, such as how compatible the contents of the specifications are with DITA and the CCMS.

Multiple outputs. As was previously discussed in section 4.1.1, the output file format of the document proved to be a major issue based on my analysis of the observational data. This also proved to be the case during the analysis of the technical specification documents. Solely from a content perspective, the technical specification documents would only require a PDF output from the CCMS due to how documents are usually published from the system as read-only formats. However, a DOCX output would be a necessity in this case since the review process between Valmet and the customer requires a file format that is readable and editable by both parties. Thus, it would be necessary to have two publication outputs available in the CCMS, one for editing (DOCX) and one for the final product (PDF). Yet, as discussed in section 4.1.1, this approach would clash with the basic principles of single sourcing. This is a major problem since it undermines the effectiveness of the already established single source system as well as its associated tools.

Multiple formats. In terms of multiple formats, the analysis did not reveal any relevant findings from the technical specification's perspective. Information was portrayed aptly and efficiently, and there were no unnecessary overlaps or repetitions. However, investigating whether or not there are multiple portrayals of the same information in other information products at Valmet and how this kind of information reuse could be utilized is something that could be investigated at some point in the future.

Multiple audience requirements. Since each technical specification is tailored individually to each project, it is evident that multiple audience requirements are present. DITA's conditional processing capabilities can be utilized here to filter and narrow down the existing content. As was discussed in section 2.4, common attributes used in conditional processing in DITA are audience, platform, product, revision, and generic property attributes (Bellamy et al. 2011, n.p.). Of these attributes, audience is the most useful one, since pre-existing conditions based on the customer's market area are already in use in different information products at Valmet. This eliminates the need to create new market area conditions, thus reducing unnecessary information clutter. However, other new conditions would need to be created to help with further text processing. For example, a conditional attribute based on the project type according to Valmet's role in the contract could be assigned to either the product category or to generic property attributes. Already with these two condition attributes, the generic text templates could be processed to fit a number of different projects depending on their market area and project type and the combination of the two. Some preliminary filtering is already done manually in the current iterations of the technical specification's templates by using color coding to indicate whether or not a table, paragraph or an individual word, for example, is mandatory, optional or variable (see Table 1 in section 3.3.2). A system that does this semi-automatically could be implemented by using DITA's conditional processing tools, although this would mean that a large number of unique conditions would need to be created specifically for this purpose, which could bloat the already high number of existing conditions.

Repetitious and similar information. The analysis for the secondary interests revealed that repetitious and similar information had minimal presences within each document. However, when compared to each other, similar information can be found from all of the documents. This is due to the use of templates which provide a foundation for the technical specifications. These templates contain the basic structure and content of a specification which is then later edited by Valmet and the customer to suit the requirements of any given project.

Information product commonality. Information product commonality did not reveal anything strictly relevant from the perspective of this study. However, it did reveal that a further analysis regarding the reuse of the technical specification documents' content in other project documentation is something that should be considered investigating further. This is especially the case if the decision to convert the technical specification into the DITA format is made, because then the reuse capabilities of the single sourced system and DITA could be fully utilized in multiple information products.

As for the content of the technical specifications, I would argue that it would fit well into a modular system, since most of it is recurring text that can be found from multiple previous specifications. Some minor editing would be required to make the text more generic by removing references to specific pages, tables, and figures, for example. In addition to removal, they could be replaced with more generic references so that the intention to direct the reader would remain but the reference itself would be slightly more generic. This would make the text more universal and thus more easily reusable. However, as discussed in section 3.3.2, because the technical specifications need to be available for editing to both, the writers at Valmet and the customer, it becomes evident that even though the content fits the system in a vacuum devoid of any external factors, the modular system does not suit it when factors such as the editing requirements are also accounted for.

4.2.2 Meeting memos

I analyzed the meeting memos through close reading and assessing whether or not they have any relevant content in terms of the project or the technical specification. Put briefly, the analysis revealed that these artifacts mostly repeated what was already noted during participant observation. For example, several meeting memos verify the issues regarding the DOCX output, complications with the database integration, and the issues regarding necessary items and training required to succeed in the conversion process. In terms of the DOCX output, concerns regarding its necessity and the available options to implement it were expressed in several memos (M08 and M03). One memo also offered insight regarding the issues surrounding the database integration, such as those concerning the available resources, as well as its obstacles and achievability (M03). As for items and training, memos from September (M09) contain data regarding the same workflow issues (the number of software licenses and the training the software requires) that were discussed in section 4.1.2, although in greater detail by accounting for factors such as prices and purposes of the licenses.

Even though there were no new revelations, the artifact analysis overall provided something equally or even more important: confirmation and validation. Having multiple sources of data stating the same issue helps with assessing the project as a whole. Although having different streams of data state the same thing does validify issues regarding the document conversion in some cases, it must be noted that repetition in itself does not mean that a finding is somehow more important. These notions are important from the perspective of participant observation. DeWalt and DeWalt (2010, 179) suggest the following for checking the validity of the collected data: “Checking informant accuracy whenever possible by checking against more objective data (logs, texts, records) and other informants’ narratives”. In this instance, the data collected from the artifacts can be cross-referenced with the data gathered through participant observation to see whether the results align or not, which they do in this case. Validifying the data through more objective means is crucial for the reliability of the study and its results. Ultimately, the results are reported through my own subjective experiences and interpretations as an observer, but having slightly more objective data, such as the artifacts, stating similar information aids with the reliability of the claims I have made throughout the analysis.

4.3 Potential solution

As previously noted in this chapter, the technical specifications and the new documentation system have some discrepancies that do not function well with each other, such as the need to convert the document to DOCX to be edited further after it has left the CCMS. To combat this issue and to offer more flexibility, I introduced a potential solution during a one-on-one meeting with the project lead in mid-October (FN10). The idea was to utilize the modular platform as a slightly more sophisticated template builder instead of a primary documentation tool in this case (FN10). Effectively, this would mean that the specification templates are moved to the CCMS, where the templates could be conditionally processed to fit any given project in terms of market area and project type, for instance. In addition, data could be input to the templates from the database containing project, product, and part information, were the database integration to be implemented. Then the templates would be published as DOCX to the writers, who then proceed to review and add necessary changes before sending the proposal to the customer for reviewing and editing.

In terms of technical issues, the template solution would offer more acceptable, although still not ideal, circumstances to the problem with DOCX. The modular documentation system could be used as an effective tool to do the initial processing on the technical specification. Yet, this poses a problem with storing the documents. A CCMS is a platform where the content, such as documentation, is created but also saved. This is done to aid traceability, organization, and to prevent unnecessary information clutter, all of which are important factors in information management and information architecture, as was discussed in section 2.2. However, removing editable and not yet finalized documents from this system compromises all of the previously mentioned aspects, which in turn is not ideal from a content management perspective.

As for workflow issues, the template solution would remove the need for additional software licenses and thus additional training. The elimination of both of these factors is a desirable action in terms of workload and cost efficiency. Since the templates exist in the modular system only to receive data and to be filtered, there is no need for the writers to have access to the documents at this stage, unless the templates require updates or corrections. Requests to update or correct the templates could be directed to the person(s) responsible for managing the templates in the modular system.

However, in order for this solution to work, some of the previous issues need to be solved before proceeding any further. Firstly, an operational DOCX output is required so that editable templates can be published from the system. Secondly, the connection between the database and the CCMS also needs to be operational for this solution to work properly. Once these issues have been solved, the implementation of this solution is possible.

Although this proposed template solution does contradict some of the theory examined in chapter 2, for example, Ament's (2002, 3) notions on single sourcing and Brown's (2003, n.p.) descriptions of information management, it currently appears to be the most potential solution to approach the document conversion without completely overhauling the ways the technical specifications are crafted, edited, and used.

5 DISCUSSION

In this study, I explored the potential issues of content conversion when transferring technical documentation into a component content management system. This was achieved by following and participating in a development project by Valmet Technologies, who wanted to expand the use of their modular documentation system as well as to chart its limits and capabilities. This study revolved around a type of technical document, a technical specification, and its conversion process to suit the new CCMS. Simultaneously, the conversion process was also used as a testing ground for other features, such as the previously discussed database integration.

The goal of this study was to answer the two research questions presented in the Introduction:

- What kind of issues may occur during a content conversion process when transferring material to a content management system?
- What measures can be taken to ensure a fluid conversion process?

Answers to both of these questions are inevitably highly context dependent due to the high number of unique variables that each conversion process faces. However, in addition to case-specific answers this study provides, some general conclusions can also be made.

Regarding the first research question, as discussed in the analysis in chapter 4, case-specific issues proved to be of two qualities: technical issues and workflow issues. Technical issues included the implementation of DOCX output and a database integration. Workflow issues consisted of having multiple writers and the appearance of new tasks and roles. These indicate that problems a conversion process might encounter can be of various natures: technology, resources, and how work is conducted may be some of the areas that need to be considered when conducting a conversion process in any organization. As for the second research question, having a holistic understanding of the setting of the conversion process will help in preventing issues from emerging. In the case of emerged issues, context awareness regarding the available resources and knowledge of the available tools can also, in addition to prevention, be used to craft solutions, as was the case with the proposed template solution presented in section 4.3.

The analysis of the technical specification documents did not necessarily reveal any additional issues within the document type itself, but it did provide some additional information on how the document and its content could be used more effectively if the conversion process is successful. Based on the analysis, the technical specification as a document type is to some extent compatible with the CCMS in terms of its contents. An important takeaway from this analysis is that some parts of the technical specification's content fit well into DITA's conditional processing capabilities, although accomplishing such a feat would be somewhat challenging due to the number of potentially unique conditions required by each project. Additionally, it must be noted that although some aspects of DITA, such as conditional processing, function well with the technical specifications, other crucial features of DITA do not. As discussed in chapter 4, one such feature is the capability to reuse content. Reuse would only be functional in a more rudimentary form, which is already possible to accomplish with other means than DITA, such as using and updating the existing DOCX templates.

It must be noted that converting the technical specification into the CCMS is a relatively simple process of assigning DITA tags to the already existing content. Most of the issues arose from the fact that the technical specifications are not currently documents designed for the CCMS. Due to this, the technical specifications' requirements and needs from the new system clashed with the current workflow and the available resources. During the study, it became apparent that in order to make all aspects of the conversion functional – the aspects being the documentation, the personnel, and the CCMS – extensive adaptation would be required from each one of them.

In terms of the theoretical approach used in this study, using broad concepts, such as information management, information architecture, and information design, as starting points proved to be successful. As discussed in section 2.2.3, since the nature of the issues was at the time unknown, it was extremely valuable to have multiple viewpoints available. This became extremely evident during the analysis when the issues did not directly fit under one specific category. Additionally, having more specific theory included in the theoretical framework of the study also aided in directing the search of the potential issues. One such example being the concept of single sourcing discussed in section 2.3.

During the study, it became apparent that there can be conflict between theory and practice. Theories often portray an ideal and near clinical scenario of how a theory or a concept should function. An example of this in this study would be single sourcing since some aspects of the ideal model of single sourcing, such as the issues of control discussed in section 4.1.1, could not be fulfilled in this project. However, this does not indicate that applying a less-than ideal form of a theory to practice would be an inappropriate action to take. Utilizing some aspects of single sourcing instead all of them can still provide benefits, for example, some of the benefits presented in Figure 3 in section 2.3. Observing how these concepts function in practice reveals more information about their use in different contexts, which could be a potential area for further research.

As for my hypothesis presented in the Introduction, I assumed that issues of varied nature would occur throughout the project's timeline due to the high number of factors that could possibly cause them. The hypothesis proved to be correct since problems of different nature, such as those relating to technical aspects or to the problems with the workflow, were discovered. Yet, it must be considered that these are only the issues that I perceived while participating in the project. As discussed in section 3.3.1, there is a possibility that I might have missed something or accidentally dismissed potential issues due to utilizing methods that rely heavily on human interpretation.

Despite these potential problems in the chosen methods, I would consider ethnography and participant observation to be extraordinarily well fitting for the purposes of this study. Ethnography and participant observation were ideal approaches when gathering information regarding the conversion process, since they allowed me to collect data straight from the personnel directly involved with the project. However, due to my own inexperience with the methodologies, the substantial amount of information that needed to be absorbed and recorded was at times overwhelming. After a brief adjustment period it became much more manageable, which allowed me to be more focused on the quality of the recorded data. Overall, these methods appear to be extremely well fitting for research conducted in the field of technical communications.

Yet, ethnography and participant observation alone could not cover all aspects of the methodology used in this thesis, which meant that I had to employ additional tools when analyzing the data, namely thematical analysis and the information analysis model by Rockley (2003, 311). The combination of these and the previously mentioned ethnographic methods formed a surprisingly successful layered network of methodologies where each individual method attempted to complement the positives and compensate the negatives of the other methods. Due to the researcher's position in using methods such as ethnography and participant observation, it can be argued that the collected data is inherently subjective, since it is collected and reported through the perspective of an individual. To balance this subjective perspective, I chose to employ artifacts as data sources to provide the study with slightly more objective data. The analysis of the artifacts proved to be in line with the analysis of the field notes, which in turn verified the claims I made based on the more subjective data I gathered by participating in the conversion process.

The project for converting the technical specification is still ongoing as of May 2020. Currently, the plan is to experiment on the template-based solution in which the CCMS is used as a template builder instead of a primary documentation tool where the process is done from start to finish. In order to achieve this, practical solutions for the DOCX output and the database integration need to be developed, but once they have been implemented, the solution is ready to be put into practice. However, due to the effects of the COVID-19 pandemic of 2020, the project has taken a more dormant role, which in turn has postponed the development of the features required by the template solution proposed in section 4.3.

Further research would be needed on similar conversion or migration processes such as this to create a more holistic picture of these processes in the field of technical communications. Since each conversion and migration is unique with their own variables, such as the nature of the content and the available tools, it is apparent that all conversions have their individual quirks. Whereas some of the steps remain the same no matter what kind of a conversion process is in question, the unique issues of each individual conversion process have been relatively uncharted territory. Having more information available on these cases will undoubtedly prove to be useful in future conversions.

WORKS CITED

MATERIALS

FN01-05, FN08-12. 2019-2020. Field notes. Compiled from data gathered from Valmet personnel.

M01-05, M08-12. 2019-2020. Artifacts. Acquired from Valmet.

REFERENCES

Ament, Kurt. 2002. *Single Sourcing: Building Modular Documentation*. Norwich: William Andrew.

Applen, J. D. and Rudy McDaniel. 2009. *The Rhetorical Nature of XML: Constructing Knowledge in Networked Environments*. New York: Routledge.

Bellamy, Laura, Michelle Carey, and Jenifer Schlotfeldt. 2011. *DITA Best Practices: A Roadmap for Writing, Editing, and Architecting in DITA*. Upper Saddle River: IBM Press.

Bloor, Michael and Fiona Wood. 2006. *Keywords in Qualitative Methods.* London: SAGE Publications Ltd.

Braun, Virginia and Victoria Clarke. 2012. "Thematic Analysis." In *APA Handbook of Research Methods in Psychology: Vol 2 Research Designs*, ed. Harris Cooper, 57-71. Washington: American Psychological Association.

Brown, Peter. 2003. *Information Architecture with XML: A Management Strategy*. New York: Wiley.

Campbell, Kim. 1999. "Collecting Information: Qualitative Research Methods for Solving Workplace Problems." *Technical Communication* 46, 4: 532-545.

Coates, Kathryn and Andy Ellison. 2014. *An Introduction to Information Design*. London: Laurence King.

Day, Don, Michael Priestley, and David Schell. 2005. *Introduction to the Darwin Information Typing Architecture*. https://www.ibm.com/developerworks/library/x-dita1/ Accessed: 03/20/2020.

DeWalt, Kathleen Musante and Billie R. DeWalt. 2010. *Participant Observation: A Guide for Fieldworkers*. Blue Ridge Summit: AltaMira Press.

Dictionary of Computer Science. 2016. "Database." Eds. Andrew Butterfield, Gerard Ekembe Ngondi, and Anne Kerr. 7th ed. Oxford: Oxford University Press. Accessed 7/1/2020.

Dye, Jessica. 2007. "Move It or Lose It: Content Migration Strategies." *EContent* 30, 2: 26-30.

Emerson, Robert M., Rachel I. Fretz, and Linda L. Shaw. 2001. "Participant Observation and Fieldnotes." In *Handbook of Ethnography*, eds. Paul Atkinson et al., 352-368. London: SAGE Publications Ltd.

Fawcett, Joe, Danny Ayers, and Liam R. E. Quin. 2012. *Beginning XML*. Indianapolis: John Wiley & Sons, Incorporated.

Fenn, Terence and Jason Hobbs. 2014. "The Information Architecture of Meaning Making." In *Reframing Information Architecture*, ed. Andrea Resmini, 11-30. Cham: Springer.

Fraley, Liz. 2003. "Beyond Theory: Making Single-sourcing Actually Work." *Proceedings of the 21st annual international conference on Documentation* (SIGDOC '03): 52-59. New York: Association for Computing Machinery.

Gans, Herbert J. 2010. "Participant Observation in the Era of "Ethnography"." In *SAGE Qualitative Research Methods,* eds. Paul Atkinson and Sara Delamont, 541-548. Thousand Oaks: SAGE Publications, Inc.

Haramundanis, Katherine. 1998. *The Art of Technical Documentation*. 2nd ed. Boston: Digital Press.

Hinton, Matthew. 2006. *Introducing Information Management: The Business Approach*. Amsterdam: Elsevier Butterworth-Heinemann.

Hovde, Marjorie Rush. 2000. "Tactics for Building Images of Audience in Organizational Contexts: An Ethnographic Study of Technical Communicators." *Journal of Business and Technical Communication* 14, 4: 395-444.

Isohella, Suvi. 2011. *Työelämän asettamat vaatimukset teknisen viestinnän koulutuksesta valmistuneille*. University of Vaasa, Licentiate thesis.

Jorgensen, Danny L. 1989. "The Methodology of Participant Observation." In *Participant Observation,* ed. Danny L. Jorgensen, 12-25. Thousand Oaks: SAGE Publications, Inc.

Kay, Russell. 2007. "DITA". In *Computerworld* 41, 14: 30.

Kimber, Eliot. 2017. *What Is DITA?* https://www.xml.com/articles/2017/01/19/what-dita/ Accessed: 5/8/2020.

Koppanen, Iida. 2018. *Restructuring Information in Technical Documentation when Migrating to XML-based Documentation System – A Case Study for Marioff Corporation Oy*. University of Tampere, Master's thesis.

Kramer, Michael W. and Tony E. Adams. 2017. "Ethnography." In *The SAGE Encyclopedia of Communication Research Methods*, ed. Mike Allen, 457-461. Thousand Oaks: SAGE Publications, Inc.

Markel, Mike. 2004. *Technical Communication*. 7th ed. Boston: Bedford/St. Martin's.

Martin, Sarah, Nicholas Carrington, and Nancy Muncie. 2017. "Promoting User Advocacy to Shift Technical Communication Identity and Value." *Technical Communication* 64, 4: 328-344.

Neyland, Daniel. 2008. *Organizational Ethnography.* Los Angeles: SAGE Publications Ltd.

OED Online. 2020. "Issue, n." Oxford: Oxford University Press. Accessed: 5/5/2020.

Pettersson, Rune. 2002. *Information Design. An Introduction.* Philadelphia: John Benjamins Publishing Company.

Priestley, Michael, Gretchen Hargis, and Susan Carpenter. 2001. "DITA: An XML-Based Technical Documentation Authoring and Publishing Architecture." *Technical Communication* 48, 3: 352-367.

Rockley, Ann. 2003. "Single Sourcing and Information Design." In *Content and Complexity: Information Design in Technical Communication*, ed. Michael J. Albers and Beth Mazur, 307-336. Mahwah: Lawrence Erlbaum Associates, LEA.

Rosenfeld, Louis, Peter Morville, and Jorge Arango. 2015. *Information Architecture: for the Web and Beyond*. 4th ed. Sebastopol: O'Reilly Media, Inc.

Samuels, Jacquie. 2014. *What Is DITA Conversion?* https://techwhirl.com/dita-conversion/ Accessed: 5/8/2020.

Society for Technical Communication. 2020. *Defining Technical Communication*. https://www.stc.org/about-stc/defining-technical-communication/ Accessed: 02/10/2020.

Technical Communication Body of Knowledge (TCBOK). 2020a. *Definition of Technical Communication*. https://www.tcbok.org/wiki/about-tc/definition-technical-communication/ Accessed: 05/15/2020.

Technical Communication Body of Knowledge (TCBOK). 2020b. *Content Management Systems*. https://www.tcbok.org/wiki/content-management-systems/ Accessed: 05/15/2020.

Tekom Europe. 2020. *Defining Technical Communication*. https://www.technical-communication.org/technical-communication/defining-technical-communication Accessed: 05/15/2020.

Tenzek, Kelly E. 2017. "Field notes." In *The SAGE Encyclopedia of Communication Research Methods*, ed. Mike Allen, 563-566. Thousand Oaks: SAGE Publications, Inc.

Valmet. 2020a. *Valmet in brief*. https://www.valmet.com/about-us/valmet-in-brief/ Accessed: 04/07/2020.

Valmet. 2020b. *Pulp and energy business line.* https://www.valmet.com/about-us/valmet-in-brief/our-businesses/pulp-and-energy/ Accessed: 04/07/2020.