# User-oriented Metadata in Single Sourced Documentation:

# A Case Study of Trimble Solutions Ltd User Guides

Maija Piirto
University of Tampere
School of Language, Translation and Literary Studies
Master's Programme in English Language and Literature
Master's thesis
September 2016

**Tampereen yliopisto**
**Englannin kielen ja kirjallisuuden maisteriopinnot**
**Kieli-, käännös- ja kirjallisuustieteiden yksikkö**

PIIRTO, MAIJA: User-oriented Metadata in Single Sourced Documentation: A Case Study of Trimble Solutions Ltd User Guides

Pro gradu -tutkielma, 75 sivua, 6 liitesivua
Syyskuu 2016

---

Tämä tutkielma tarkastelee metadatan käyttöä yksilähteistetyssä dokumentointiympäristössä, sekä dokumenttien tuottajien että käyttäjien näkökulmista. Tutkimuksen tarkoituksena on tarkastella, miten metadataa käytetään digitaalisten aineistojen yhteydessä. Tutkimuksessa keskitytään erityisesti metadatan erilaisiin käyttötarkoituksiin eri yleisöjen näkökulmista.

Teoreettinen viitekehys koostuu kahdesta osasta: ensimmäisessä osassa tarkastellaan yksilähteistämistä, joka on teknisen viestinnän alalla yhä yleistyvä dokumentointitapa. Yksilähteistetyssä dokumentointiympäristössä tiedon hallinta ja järjestäminen on tärkeää, joten toisessa osassa tarkastellaan metadataa, eli tiedon merkitsemistä. Metadataa kuvataan sen erilaisten luokitusten, tarkoitusten ja laatuarviointien kautta.

Tutkimuksessa metadatan eri käyttötarkoituksia tarkastellaan tapaustutkimuksessa, jonka kohteena on ohjelmistoyritys Trimble Solutions Ltd. Tutkimusmetodeina ovat sisällönanalyysi, eli informaation arviointi sen sisällön ja tarkoitusten kautta. Tutkimusaineistona käytetään käyttäjille suunnattuja digitaalisia käyttöohjeita ja niihin liitettyä metadataa. Käytössä olevaa metadataa tarkastellaan sen käyttötarkoitusten kautta ja tämän lisäksi sisällönanalyysin keinoin tutkitaan, millaista metadataa ohjeisiin voisi lisätä, jotta se palvelisi paremmin ohjeiden käyttäjien tarpeita ja jotta metadatan käyttöä voitaisiin entisestään tehostaa.

Tutkimuksessa käy ilmi, että tällä hetkellä metadataa käytetään pääosin informaation hallintaan sen kirjoittajien näkökulmasta. Tämän käyttötarkoituksen lisäksi todetaan, että metadatan käyttäminen informaation kuvaamiseen ja tiedonhaun tehostamiseen auttaisi käyttäjiä löytämään tarvitsemansa tiedon nopeammin ja tehokkaammin. Analyysin tuloksena luodaan käytössä olevan metadatan rinnalle lista metadataelementeistä, joita käyttämällä ohjelmiston käyttäjien tiedonetsintää voitaisiin tehostaa. Tutkimuksen piiriin ei sisällytetä luodun metadatan laatututkimusta, mutta teoriakehyksessä esiteltyjen laatukriteerien avulla on jatkossa mahdollista suorittaa käyttäjä- ja asiantuntijatutkimuksia esitetyn metadatan laadusta.

Avainsanat: tekninen viestintä, metadata, yksilähteistäminen

# Table of Contents


1 Introduction ........ 1

2 Single sourcing ........ 5

2.1 What is single sourcing? ........ 6

2.2 Implications of single sourcing ........ 9

3 Metadata ........ 15

3.1 What is metadata? ........ 15

3.2 Categorizations of metadata ........ 21

3.3 Evaluations of metadata quality ........ 27

4 Material and method ........ 31

4.1 Trimble Solutions and the documentation process ........ 31

4.1.1 Trimble Solutions ........ 31

4.1.2 Darwin Information Typing Architecture (DITA) ........ 32

4.1.3 Trimble Solutions documentation process ........ 34

4.2 Case study materials and method of analysis ........ 36

4.2.1 Trimble Solutions user guides ........ 36

4.2.2 Content analysis as the method ........ 37

5 Analysis of case study metadata ........ 40

5.1 Metadata for information management purposes ........ 41

5.2 Metadata for resource description and information retrieval purposes ........ 43

5.2.1 Information type ........ 47

5.2.2 Application ........ 53

5.2.3 Subject area ........ 54

5.2.4 Task ........ 59

5.2.5 Role ........ 63

5.3 Metadata quality criteria for future research ........ 65

6 Conclusions ........ 69

7 Works cited ........ 73

Attachments ........ 76

Attachment 1 ........ 76

Attachment 2 ........ 79

# List of figures

Figure 1: Ann Rockley's (2003) illustration of how ownership of content changes between a traditional authoring environment (top) and a single sourced authoring environment (bottom). ..... 12

Figure 2: Topic with the <Information type> value **Background**. ..... 49

Figure 3: Topic with the <Information type> value **Instructions**. ..... 50

Figure 4: Topic with the <Information type> value **References**. ..... 51

Figure 5: Topic with the <Information type> value **Examples**. ..... 52

Figure 6: Topic with the <Application> value **CPP**. ..... 54

Figure 7: Topic with the <Subject area> value **Import & Export**. ..... 56

Figure 8: Topic with the <Subject area> values **Printing** and **Import & Export**. ..... 57

Figure 9: Topic with the <Subject area> value **Analyses**. ..... 58

Figure 10: Topic with the <Task> value **Creating**. ..... 61

Figure 11: Topic with the <Task> value **Deleting**. ..... 62

Figure 12: Topic with the <Task> value **Calculating**. ..... 63

Figure 13: Topic with the <Role> value **Basic user**. ..... 64

Figure 14: Topic with the <Role> value **Administrator**. ..... 65

Figure 15: Example of a topic dealing with the task of exporting analyses. ..... 79

Figure 16: Example of a topic dealing with the task of printing. ..... 80

Figure 17: Example of a topic that deals with the use of report templates. ..... 6

# List of tables

Table 1: Metadata elements divided according to the user of the specific elements, as well as the purposes the metadata serves. ........ 70

Table 2: Values of the metadata element <Information type> and guidelines on when to use each value. ........ 76

Table 3: Values of the metadata element <Application> grouped according to the solutions to which they belong. ........ 76

Table 4: Values of the metadata element <Subject area> and guidelines on when to use each value. ........ 77

Table 5: Values of the metadata element <Task> and guidelines on when to use each value. ........ 78

Table 6: Values of the metadata element <Role> and guidelines on when to use each value. ........ 78

# 1 Introduction

In a society where the amount of information produced by institutions, communities and individual people is ever growing, the challenges of finding and obtaining the relevant information in a particular situation are growing as well. Digital information is starting to become the norm, especially when talking about materials that companies and communities produce to aid their users in the tasks they are attempting to accomplish. When the amount information is massive, it becomes essential to focus on ways of representing information in a way that helps users find what they are looking for as fast as possible.

The fact that more and more companies are producing their help and support materials, most often user guides, in a digital form instead of the traditional printed form, presents the producers of these materials with new challenges, as well as new opportunities. The producers of user guides, technical communicators, are at the heart of making information as clear and useful as possible, helping the users to the best of their abilities. Understanding the importance of organizing information is vital in this process, and an important aspect of an efficient organizational system is the ability to structure information in a way that it can be retrieved at any given time.

The answer of many companies to the challenges presented by the managing of massive amounts of information has been to adopt a single sourced way of documenting. Single sourcing is a method of writing re-usable information by producing content in modular form, which basically means creating non-hierarchal and non-sequential content modules instead of linear texts, complete documents or printed books from start to finish (Ament 2003, 6). These modules of information are stored in a database, and from this single source, documents can be assembled together in various ways for various purposes (Ament 2003, 3). In order for single sourcing to be possible, pieces of information must be accurately marked with metadata. This means marking documents with specific kinds of tags to describe the modules in a document, in order to identify and specify texts in an efficient way. Using metadata is essential in keeping information in order and making it easy for the producers of information resources to manage their data, as well as for the users to search the resources for relevant information. *Information resource* is a term that is used throughout this study to refer to documents and modules of information.

The use of metadata originates from the library and information sciences, where it has traditionally been used for organizing information in library catalogues (Zeng and Qin 2008). Since the term's emergence in the 1960s, the use of metadata has expanded to other fields, and as Muriel Foulonneau and Jenn Riley (2008, 289) state "the term *metadata* has become a symbol for digital information organization and management". In the field of technical communication metadata is most often used for the convenience of the writer or the documentation system, focusing on how metadata elements are used to manage the documentation process. When technical communicators produce documentation in a single sourced documentation environment, using appropriate and useful metadata becomes vital. A documentation database may consist of thousands and thousands of documents, modules and topics, which makes searching for the appropriate information extremely laborious. As more and more companies, and therefore also their technical communicators, are focusing on more user-oriented ways of producing documentation, metadata usage should reflect this change. There are several different uses for metadata, but one of the most common today is that of enhancing information retrieval. When users search full texts using keywords, the result can be a massive list of completely unrelated topics that have no relevance for the specific user or their current situation. Instead, using user-oriented metadata should be utilized for the benefit of the users, making searching for specific topics easier, faster and more pleasant.

The goal of my study is to research how metadata could be approached and utilized from the point of view of both technical communicators and users, with a special emphasis on the latter. In order to make metadata more efficient for companies and communities with large, single sourced databases, it is important to analyze metadata from the following two aspects:

- What purposes does the metadata used serve at the moment?
- What type of metadata serves users best and efficiently helps them find what they are looking for?

When discussing the issues of single sourcing and metadata in this study, I attempt to intertwine the viewpoints of the producers, as well as the users of information resources.

I will approach the issues of single sourcing and user-oriented metadata usage on the basis of a case study on the user guides of a Finnish based software company called Trimble Solutions Ltd. Trimble Solutions' way of producing documentation is moving towards a completely modular way of writing, and because of this transition towards single-sourcing, the structure and organization of information is crucial. I will analyze the user guides to see how texts are currently marked with metadata to organize texts in the documentation database, as well as the ways metadata elements could be used to make the search function as useful to the users of the software as possible. At Trimble Solutions, the structure of documentation is determined by the principles of Darwin Information Typing Architecture (DITA), which is a structured methodology for the production and organization of reusable technical information (Priestley 2001, 152). This methodology provides tools for the creation of audience-specific content that can be clearly marked with XML-form metadata elements (see Chapter 4.1.2).

Extensive studies have been made regarding metadata usage for the management of information resources, and on the factors affecting the quality of metadata. Xavier Ochoa (2014), for example, provides a good account of various manual and automated metadata quality studies. However, hardly any literature can be found on the practical ways that metadata could be used for assisting users in performing their tasks. Therefore, I would argue that a case study on the subject of user-oriented metadata could provide an interesting addition to the theoretical research on information retrieval. In the field of technical communication, the needs of the users is a subject that is always under close inspection, and I believe that understanding the purposes that metadata is used to fulfil could provide important insight into how metadata usage could be enhanced.

As theoretical basis, I will thoroughly explain the concept of single sourcing in Chapter 2, what kinds of effects producing information using this method has in the field of technical communication, as well as what kinds of new challenges the writing of re-usable, modular texts poses for the technical communicator. The meaning of metadata is essential in managing information, specifically in single sourcing, so the term, which, in its most simple form is, language describing language, is explained in Chapter 3. I will describe the history of the term, as well as present the various purposes for which metadata can be used. In its essence, metadata that is created should always be useful, but the characteristics that make a metadata element useful is an issue that is viewed very differently by many scholars. I will also describe a model for producing good, usable metadata and present some of the ways that the quality of metadata could be, and has been, evaluated.

In Chapter 4 I provide some background for my study, as I describe the company, Trimble Solutions, and their documentation system, as well as present the materials of my case study, which are the user guides that are produced for the software at Trimble Solutions. In Chapter 4 I will also briefly describe what Darwin Information Typing Architecture is, because the structuring principles of DITA are the basis for documentation production at Trimble Solutions. In Chapter 5 I will analyze the contents of the case study materials by utilizing the theoretical frameworks presented in Chapter 3. I will present the metadata that is currently used at Trimble Solutions and the purposes it serves and after this I will explain what purposes metadata should serve to best accommodate the needs of the users. To finish off my analysis I will briefly outline how the quality of metadata could be assessed. The final chapter, Chapter 6, draws the results of my study together and provides some topics for further research.

# 2 Single sourcing

In the field of technical communication, single sourcing has become more and more popular, and this new way of creating information has brought with it drastic changes to the paradigms of writing and producing documents (Ament 2003). In a single sourced documentation environment the focus of technical communication has moved away from traditional, linear writing towards modular content development, changing the entire relationship between the writer and the document (Carter 2003). Michael Hughes (2002, 276) argues that this change stems from the fact that technical communicators creating user-centered materials no longer transfer knowledge from subject matter experts to users, but rather reinterpret technical information by creating new knowledge. According to my own experiences as a technical communicator, this process of reinterpretation is constantly ongoing, as, instead of documents that are produced and published once, documents are written, modified and combined with previous materials and possibly documents produced by coworkers or predecessors.

This reinterpretation of knowledge is done by utilizing single sourcing, which is basically a method of producing re-usable information that is stored in a single source, a database. Instead of writing one document for one specific purpose, meant to be read in a linear order from the beginning to the end, writers produce modular content, pieces of information that can then be used for multiple purposes. In this chapter I will present the basic principles of single sourcing, the advantages and challenges it poses to technical communicators and the companies in which they work, as well as the implications that working in a single sourced environment has on the work of technical communicators. I will not deal with the history of single sourcing, as it has established such a firm position in the field of technical communication that the historical background is not relevant from the point of view of this study.

## 2.1 What is single sourcing?

Simplistically saying single sourcing means “writing information once and using it many times” (Rockley 2001, 189). Ann Rockley (2001, 190), however, also points out that this does not mean simply copying and pasting the information into another source, or modifying it to fit various needs, thus creating multiple sources, but new documents are created by “referencing” information modules into the document, or drawing needed modules from a database to form the document. In this study I will use the term *module* to refer to a piece of information in a document, such as a section, paragraph or a sentence that is a stand-alone content piece that “make[s] sense in any document format or reading sequence” (Ament 2003, 4–5).

The move from writing complete documents to creating reusable modules is, according to Rockley (2001, 190), making a change in the way technical communicators communicate. This paradigm shift can be seen in the four levels of single sourcing:

- Single sourcing level 1: Identical content, multiple media
- Single sourcing level 2: Static customized content
- Single sourcing level 3: Dynamic customized content
- Single sourcing level 4: Electronic performance support system

On the first level, materials are derived from the same identical content to produce, for example, a paper and an online guide, with hardly any attempt to accommodate the differences between the media. Using the same content and presentation without modifying them to fit the media may bring serious problems when it comes to usability and effectiveness, since, for example, images used in a paper guide may not be relevant online and because the information in a paper guide is often used in a different situation than the information in an online help.

On the second level, information is designed more effectively and the materials are customized according to the needs of the user, the type of the material and the media. There is a core of content (the single source), onto which technical communicators add customized content: additional information about different products of a product family, specific details for specific audience, or perhaps different images for different information products. This content is static, meaning that only the technical communicators are able to modify it.

On the third level, the information is customized according to the users: information modules are stored in a database that users can access through user profiles or by identifying the type of information they need. This level requires extensive research on users' needs, so that the technical communicators are able to build dynamic information models that meet the users' needs. At Trimble Solutions, this is the desired level, and it is achieved through cooperation between the documentation, usability and training teams, as well as the help of this study. This issue of creating customized information at Trimble Solutions will be examined a bit further in Chapter 4, in which I present the documentation system and process flow at Trimble Solutions.

Taking the third level even further, the fourth level of single sourcing delivers the user the information they need, maybe even before they know they need it. On this level, the users are definable and their needs and requirements are known, and because of the high level of user integration, this system is not suitable for products that have a wide variety of different users. The electronic performance support system recognizes the users' tasks and the information they need at any given point, and in order for this to work, very specific information models must be built, so that the appropriate information is offered to the user at the right time.

Rockley's model is quite comprehensive, because it describes the different ways single sourcing can be utilized in various companies, depending on their resources and the needs of their users. The role of technologies and tools is significant in each stage, since they enable the dynamic building of information that meets the users' requirements. Jason Swarts (2010, 146) shares Rockley's view on the importance of understanding the needs of the users, as he argues that single sourced writing is effective only when technical communicators are writing in a situation in which the roles, relationships, and objectives of their readers are known and can be formalized through rules that govern the assembly of the content that is created.

In addition to Rockley's model of the four levels of single sourcing that is based on the changes brought on mostly by new technologies, Joe D. Williams (2003, 321) presents two major approaches to creating single sourced documents. One way is to create a document using desktop publishing software and then reformatting the document for the desired purpose (for example, an online help or a paper manual), with or without making any changes to the content. The other one is to create a single source of information, from which desired documents are derived, and this single source can be, for example, a database in which multiple files are stored and then compiled together to create the document for a specific purpose. These approaches can be used as separate approaches, or both can be used simultaneously. The major benefits of these single sourcing methods are efficient production and improved consistency of documents, which, in turn, enables writers to produce documents with higher quality at a lower cost than by using traditional writing methods, according to Williams (2003, 321).

Locke Carter (2003, 318) views single sourcing as a technology itself, but his use of the term *technology* is very broad, meaning a "technique that is codified, implemented, taught, and instilled with organizational values". Using this definition and broadening the meaning of a technology from simply the tools used, his view is understandable. Describing single sourcing as a technology does indeed capture the strictly structured nature of creating reusable content within specific, prescribed guidelines that have been set for technical writers by the organization for which they work for. This definition, however, is not widely used by other scholars or technical communicators, since although technologies and software tools are used to automate many parts of the single sourcing process (Ament 2003, 181), it is structured and modular writing that leads to the production of coherent and usable information (Ament 2003, 1; Rockley 2003, 351). Perhaps a better description about the nature of single sourcing comes from Ament (2003, x), who calls it a “methodology, not a technology”. A single sourced writing process is not only a new combination of technologies, but a whole new way of using and organizing information.

According to Ament (2003, 1–5), single sourcing is a method for developing modular content in one source or a database, from which content is assembled into materials in different formats for multiple audiences and purposes. This modular content is produced using content-based information development tools that leverage markup languages, usually SGML (Standard Generalized Markup Language) or XML (Extensible Markup Language), which encode metadata about the information into the text. These enable technical communicators to produce information on the module level, instead of the document level.

At the heart of single sourcing is indeed structured writing, meaning the “practice of writing content following structured writing guidelines or models” (Rockley 2003, 350), and this is basically what makes single sourcing possible. Technical communicators write using specific templates, document type definitions, which determine the content that is included in a particular document and in which order, as well as the places where content can be reused. Writing according to a specific model ensures consistency, regardless of who wrote it, simultaneously freeing the technical communicators in their everyday work to focus on the information content instead of the form. Using modular writing is also the basis for improved quality control, as Filipp Sapienza (2007, 83) notes, because recurring text modules are written and modified only once, ensuring the accuracy of the information wherever it may occur.

## 2.2 Implications of single sourcing

Williams (2003, 321) describes the change that single sourcing approach brings to documentation quite nicely, when he says, “[i]n a nutshell, the paradigm shift is from document-oriented to object-oriented”, referring to the paradigm shift that takes place when technical communicators start to view information and information products with the ultimate object in mind, instead of focusing on sections and chapters. Rockley (2001, 191) notes that single sourcing requires the technical communicator to write according to a structured model, always keeping in mind the multiple ways the information can be used to ensure the content is reusable.

Single sourcing significantly impacts technical communicators' roles and tasks, as Rockley (2001, 192) argues, most notably because "the process of creating customized single source materials separates the creation of the input (content) from the output (media or information type)". A writer must write with reuse in mind, often simultaneously creating different information types for multiple audiences and media. One of the major advantages of single sourcing concerns the task of updating or correcting materials: as mentioned before, these are done only once to the core material, providing notable time and cost savings for the company. Time spent on tedious mechanical work can "now be spent on creating new material and creating innovative changes in information delivery" (Rockley 2001, 192).

Rockley (2003, 350) emphasizes the advantages of single sourcing in the work of technical communicators when she notes that using standardized forms of writing enables writers to concentrate fully on the content, and this applies to the everyday work of writers. However, according to Carter (2003, 319), integration of single sourcing into the writing process is forcing writers to evolve, as their job description no longer entails merely the producing of texts, but also knowledge of layout and publishing. At Trimble Solutions, technical communicators are very much involved in the creation of these essential models that govern the structure of documents and thus the entire writing process. When technical communicators are responsible for the creation of content, as well as the form of that content, as the situation is at Trimble Solutions, I would argue that the skill sets of writers should cover fluent use of essential tools and software and modular markup languages, as well as knowledge about information types and information design. A good grasp on information organization is important especially when producing content in a single sourced environment. The job descriptions of technical communicators are significantly diversified, because single sourcing “enables technical communicators to become knowledge, content and information designers, because knowing how to single source means knowing how to write, design, create, and (re)use content for various media” (Eble 2003, 345).

So as well as creating cost and time savings for the company, single sourcing provides new opportunities for technical communicators to diversify their job descriptions from traditional linear writing more towards information design. This diversification becomes evident from TCeurope’s[1] project TecDot-Net (TCEurope TecDot-Net 2005), in which some of the most important areas of knowledge and capabilities of technical communicators were studied and listed. According to the findings of the TecDot-Net project, core competencies of all technical communicators should include documentation planning and information development, structuring information, usability, and understanding and using tools[2]. These core competencies are, in my opinion essential for technical communicators, whether working at Trimble Solutions, or at any other company operating in a single sourced documentation environment. Some of the major effects single sourcing has on the traditional writing teams and the role of the individual writer are illustrated in Figure 1, which presents Rockley’s (2003, 352–353) view of the changing role of ownership when writing in a single sourced authoring environment.

[1]TCeurope (Technical Communication in Europe) is a non-profit organization that brings together six independent national technical communication societies in Europe. TCeurope aims to represent European technical communication professionals (tceurope.org).

[2]In addition to the TecDot-Net project, in 2015 a project consortium was established to develop TecCOMFrame, which stands for Technical Communication Competence Framework. This European-wide project aims to develop a framework for standard curricula based on a common academic qualification and competence, in order to establish more higher education programs in technical communication (conferences.tekom.de/european-academic-colloqium/teccomframe/about/).

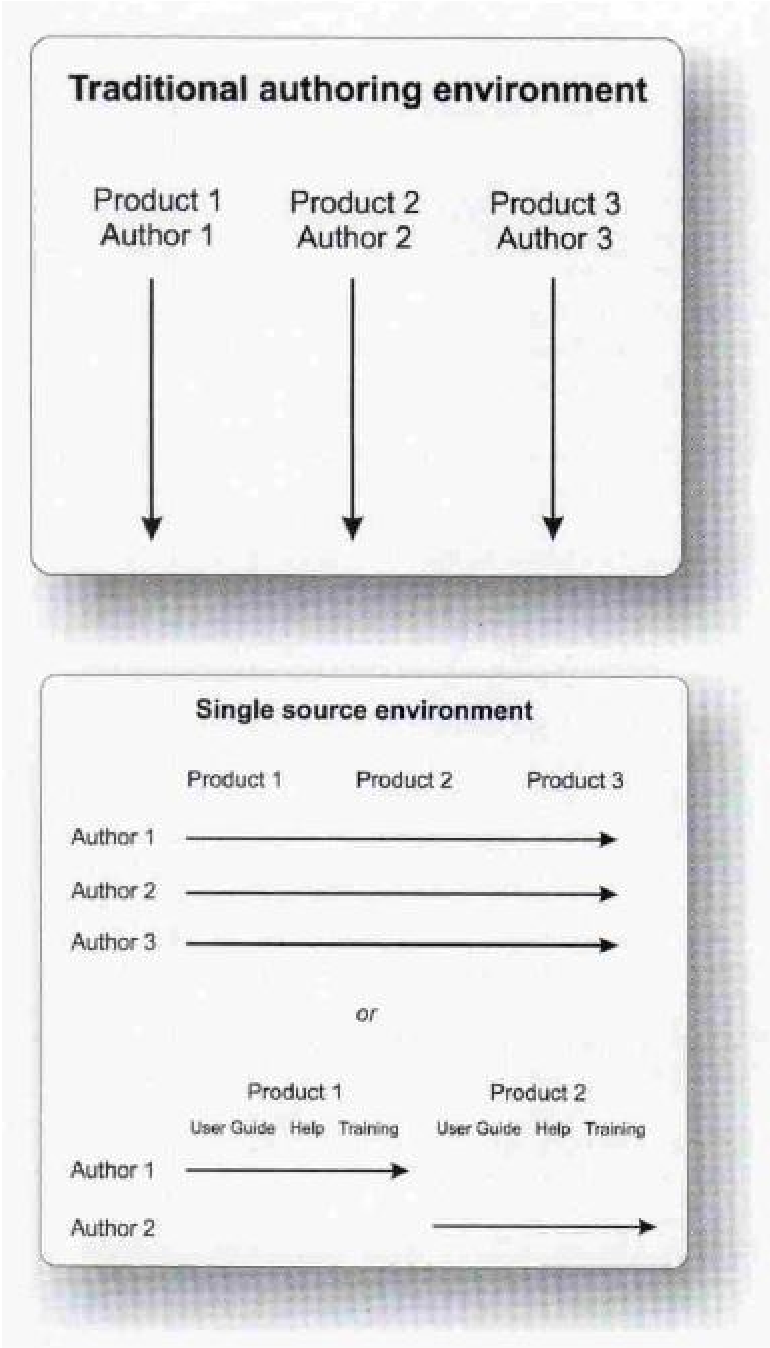



*Figure 1: Ann Rockley's (2003) illustration of how ownership of content changes between a traditional authoring environment (top) and a single sourced authoring environment (bottom).*

In a traditional authoring environment, one writer is responsible for the documentation of one product, and these documents are produced linearly from top to bottom, meaning, from the beginning to the end. In a single sourced environment, however, one instruction or description of a product is no longer one document produced by one writer, but the information contained in that document is broken into modules, individual pieces of information that are modified so that they can be reused for another description of the product, or for an instruction of another product. The individual modules can be produced by different writers at different times, so the final document is pieced together from the modules, and the result is no longer the work of a single author.

One writer in charge of the documentation of one product, however, is still possible in a single sourced environment, as Figure 1 shows, but it differs from the traditional way in that the produced contents are used in multiple information products, such as a user guide, help and training material. At Trimble Solutions, this second example of single sourced authoring environment applies: one technical communicator is mainly responsible for producing all the documents for one product, from which the contents for multiple outputs are then compiled.

Michael J. Albers (2003, 335–342) holds a similar view of single sourcing to Rockley, when he rejects the traditional "craftsman model of [text] production, with each person handcrafting their own piece". Single sourcing is a complex way of writing, and although Albers (2003) agrees with Rockley's (2001, 192) description when she says, "[r]ather than narrowing the scope of what writers do, single sourcing actually increases the scope", he sees specialization as the answer to the growing complexity of technical communicators' tasks.

JoAnn T. Hackos (2002) notes that moving away from the traditional role of a writer, expertise in, for example, information design, is becoming more and more important for technical communicators, since compiling consistent and useful documents from a single source relies on good information management and design. Many scholars agree that one of the major challenges when implementing a single sourced authoring environment is that the adoption of new ways of creating documents may cause severe resistance against change in writing teams, especially with the more senior technical communicators who are used to the linear way of writing, in which one writer is clearly the author of one document (Carter 2003; Eble 2003; Rockley 2003). Based on my own experiences as a technical communicator, I would argue that this resistance will fade as writers are trained to work according to the new policies, and as new writers enter the field of technical communication, having already studied the principles of single sourcing.

Another major challenge of single sourcing is caused by the fact that technical communicators should always start from the viewpoint of the users and their needs, but at the same time write with reuse in mind. Hackos (2002, 319) argues that in modular writing, the predetermined structure frees the writer to concentrate on communicating the relevant information to the users. She does not, however, address the additional challenge brought on by reuse: the writer does not necessarily know where the modules will be reused in the future, or what the needs of the future users will be. Because of reuse, single sourcing may in some cases result in documents that function on a general level, but the instructions may not be specific enough to meet the users' specific needs, or the needs may have changed. At Trimble Solutions, this is not such a major issue, since the user type remains basically the same, and where the users' needs differ, these differences are accounted for by marking specific parts of the instructions for specific users. This is done by using metadata to describe and mark specific modules, and this issue is examined in the following chapter.

# 3 Metadata

Traditionally, metadata has been talked about as “data about data” (Foulonneau and Riley 2008; Hider 2012; Zeng and Qin 2008), which is a compact description that can be applied to virtually every field. As can be seen from this characterization, metadata is a very broad concept, and more detailed descriptions of metadata are produced by various instances that use it, based on their needs. I believe that it is useful to understand the different facets of the term to be able to form a clear picture of the meaning of metadata in the field of information resource description and retrieval, and its implications in the field of technical communication. The term *information resource* has been used by Philip Hider (2012, 1) in reference to “all resources that, in the context of their description, are primarily intended to inform”. Within the scope of this study, information resource is used mainly to refer to electronic documents and information modules.

In this section, I will first present a brief overlook of the history of the term metadata and the ways it has been used by different researchers and practitioners over time and after this, some standards relating to the creation of metadata are discussed. Then, I will focus on metadata within the scope of my study, meaning the use of metadata for the management and retrieval of information resources. Definitions of metadata are often focused on the functions metadata performs, so finally, I will discuss some widespread models that describe how metadata can be classified according to its purpose and function. The quality of metadata is crucial in supporting effective access to information resources, so the final discussion of this chapter regarding usable and effective metadata is important for the purposes of my case study later on in my analysis.

## 3.1 What is metadata?

As Debbie Olson (2009, 18) notes, the definition of metadata as “information about information” is one that most people are probably familiar with, but I would argue that it lacks explanation and actual application of the concept in practice. A much more detailed description is provided by the National Information Standards Organization (NISO 2004): “Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage an information resource.” This is a definition that gives some context and meaning to the broad concept, linking together the main functions and applications of metadata.

As I mentioned in Chapter 1, Marcia Lei Zeng and Jian Qin (2008, 4–6) state that the origins of metadata use stem from the library and information sciences, where the organizing of information objects was, and still is, crucial. This organizing was highly structured and governed by standards, such as the *Anglo-American Cataloguing Rules*, Second Edition (AARC2) and MARC (*Machine-Readable Cataloging*), which were primarily used for library catalogues. These catalogues included bibliographical data about the items and the relationships between them, as well as made the sharing of information between libraries possible. In the Internet era, then, the methods used to organize the information contained in these catalogues were adopted to the categorization of web-based resources. In the 1990s a standard for describing web content was created by the Dublin Core Metadata Initiative (DCMI), an organization dedicated to supporting innovation in metadata design. This standard was called the Dublin Core Metadata Element Set, and it was established to enhance retrieval among the chaotic amount of web resources (Haynes 2004, 5; Zeng and Qin 2008, 6). After the launching of the still ongoing and developing Dublin Core Metadata Initiative, “a metadata movement soon spread rapidly … across research, educational, and governmental institutions, as well as businesses, and many organizations” (Zeng and Qin 2008, 6).

In the late 1990s, the number of metadata projects increased and metadata became widely recognized, and as a result, the use of metadata established a firm position in the information environment. Haynes (2004, 6) notes that as standards were being produced, they were taken to use by suppliers of software applications, which incorporated metadata into their products. Metadata was used for managing and tracing information resources in the content management systems of these software applications, as they were handling more and more extensive amounts of information resources in an increasingly dynamic environment. This notion is especially important from the point of view of my study, as my case study specifically deals with metadata relating to software applications produced by Trimble Solutions.

The wide range of disciplines using metadata has resulted in various understandings of the term, and this variety becomes apparent also in Foulonneau and Riley's (2008, 8–9) notion that "metadata can appear under different labels in a variety of types, formats and origins". Foulonneau and Riley include various terms under the broad definition of metadata, each of which fulfils different needs of different users. They mention terms such as tagging, especially popular in social networking sites, which means assigning a kind of a keyword for a digital object, or annotation, which means a person attaching his own opinion or perspective about a digital object, usually to allow interpretation later on. Both of these terms have to do with a person's subjective view of an information resource, but more importantly, this variety of terms and their uses, according to Foulonneau and Riley, means that metadata is not a rigid set of information, but it can be "adapted to a more modular structure, with more potential for reusability". This notion draws an important link between metadata use and single sourcing, which is the basic framework for this study.

I agree with Haynes (2008, 6–8) as he argues that no universally applicable definitions of metadata exist, but the exact meaning depends on the context and the community in which metadata is used. Foulonneau and Riley (2008, 5) aptly point out that "metadata is subjective in nature", which, in my opinion is true for both human and machine created metadata. Although computer generated metadata, meaning metadata that is created by electronic applications with minimal or no human involvement (Zeng and Qin 2008, 8–9), is mostly automatically produced, choices have been made by information management experts to decide which metadata an application will gather and create, just as choices are made by writers assigning more free form metadata to an information resource.

Continuing on the discussion of the subjective nature of metadata, it is important to note that, as mentioned above, metadata can perform various tasks, and these tasks are viewed differently by different scholars. Michael Day (2001, 11) summarizes the range of operations metadata can be used to support as follows: uses of metadata include “resource description and discovery, the management of information resources and their long-term preservation”. This is in line with the categories presented by many other scholars, as listed by Olson (2009), and with the types into which, for example, Robin Wendler (1999, 4) divided metadata almost three decades ago: descriptive, structural, and administrative metadata. During the last couple of decades, then, metadata has been assigned to a range of categories, more or less following this basic division, but with each scholar having their own point of focus – usually the purpose that metadata has (Haynes 2004, 13). Focusing on the purpose of metadata is quite a natural starting point, since, as Jane Greenberg (2005, 22) notes, “*functionality* is the principal reason for metadata” (italics original). These categorizations according to the function and purpose of metadata will be discussed in more detail in Chapter 3.2.

Kurt Ament (2003, 18, 183) provides us with an alternative term to metadata, but the principle is the same, as he brings up the concept of conditional text, an important feature of single sourced writing. Conditional text is a way of electronically marking sections of a document for a specific purpose, meaning that it enables the marking of elements for conditional display, after which these sections can be turned on or off as needed. A section of a document can be marked, for example, for expert users only, and this section can be turned off when compiling the standard users’ version. Although Ament does not mention the term metadata, the principal behind the use of conditional text is producing information about information, and thus it can be considered to fall under the broad definition of metadata used in this study.

Foulonneau and Riley (2008, 13) also insist on quite an obvious-sounding notion when they say that “metadata that is created should be useful”. This is not always such an easy thing to put into practice, because the costs of creating and maintaining metadata are high (Wendler 1999, 4). Cost issues might focus the attention of a metadata project to the actual mechanical application of metadata elements, leaving the laborious research and planning stages to the sidelines. As I will explain later on in Chapter 3.2, the actual production of metadata is only one step at the end of a list of stages that need to be considered when creating metadata for information resources. In a documentation project, metadata creation for each information resource can be, as Amanda J. Wilson (2007, 16) notes, the most expensive part, as it requires "both organizational and subject expertise to describe an object and its context for use". The creation and maintenance of appropriate and useful metadata requires a lot of manual labor by skilled information professionals. This process of creating useful metadata is described in more detail in Chapter 3.2, and in Chapter 5.2 I will explain how I use the process framework as a basis for the second part of my analysis.

Due to the complexity of assigning appropriate metadata to information resources, attempts at keeping the costs lower usually result in organizations adopting a suitable metadata standard to guide the process. As noted earlier in this chapter, the number of institutions and initiatives dealing with metadata projects is immense, and according to Zeng and Qin (2008, 7, 15), this has resulted in a number of overlapping metadata standards, as they have been created by communities for their own specific purposes. Haynes (2004, 13) goes on by noting that "one of the main drivers for the evolution of metadata standards within each community is the use to which the metadata is put – its purpose". Because there is basically no limit to the type or amount of information resources that can be described by metadata, different standards have been created as guidelines according to, for example, subject area, material format or functional area (Zeng and Qin 2008, 7; Wendler 1999, 4).

When starting to create metadata in a company or a community, the use of some kind of guideline is crucial, since if metadata is not consistent, it is not useful. Foulonneau and Riley (2008, 13–18) note that choosing a suitable metadata standard from the array of existing standards is not a straightforward task, because the decision must be done by considering the institution adopting the standard, the different purposes of various standards, the technologies available, the information resources to which the metadata will be attached, as well as the objectives and the resources of the process itself.  I will not discuss the details of the selection process further here, since the decision and the process are unique in each situation and with each institution or community, and, more importantly, the purpose of my study is not to select a metadata standard to be used at Trimble Solutions.

Understanding the meaning of a structured and quite strictly governed standard is, however, important, because as Hackos (2002, 120) notes, if individuals would be able to add their own metadata to documents, the information would not be readily accessible to others unaware of the intentions of the original author. Hackos emphasizes the importance of a standard set of terms from the point of view of electronic resources and their content management systems, since, if the metadata is not consistent, the resources become incoherent and disorganized, and the content management system will not be able to manage the resources in a logical way.

Hackos (2002, 120) does not promote the use of a metadata standard, but the use of any consistent set of terms and elements and she goes on to argue that when attempting to describe the resources in your content management system, “[a]greeing on the categorizing and selecting the right words to label [your information resources] will be the most difficult task”. Adopting an existing metadata scheme that has been created to serve the particular needs of a particular community of users may provide a solution for this selection. The National Information Standards Organization (NISO 2004) defines metadata schemes as follows:

> Metadata schemes are sets of metadata elements designed for a specific purpose, such as describing a particular type of information resource. The definition or meaning of the elements themselves is known as the semantics of the scheme. The values given to metadata elements are the content. Metadata schemes generally specify names of elements and their semantics.

A metadata scheme is often defined by using an established metadata standard, such as the Dublin Core Metadata Element Set, from which the suitable set of elements is selected to be used in describing content. Instead of, or, in addition to, adopting terms from a standard, a metadata scheme can also be put together by an institution for their own internal usage, which means creating a controlled vocabulary. A controlled vocabulary is basically a set of specifically selected terms that "must be constructed and maintained with specific subject area(s) in mind" (Chu 2010, 54). Haynes (2004, 152) notes that using a controlled vocabulary ensures more consistent retrieval, because limiting users to a preferred term choice frees them from having to think of what synonyms might describe the concept being searched for. Creating a specific controlled vocabulary requires extensive knowledge on the subject area, as well as on the users, but if done right, it can provide users with much more subject-appropriate descriptions of information resources, compared to some of the more general standards that may not be as specific.

## 3.2 Categorizations of metadata

According to Stewart Whittemore (2008, 94), information resources in a database can be put into categories only by using metadata, which then enables the retrieval of unique resources when composing a complete document consisting of pieces of information. Metadata is crucial for this retrieval, because without it, the pieces of content would be lost in the database and thus become "unusable because forgotten, forgotten because undifferentiated". Information retrieval, as well as information representation, are indeed central concepts used by various scholars (Chu 2010; Glushko 2013; Hider 2012) when discussing the purposes of metadata. From the point of view of my case study, these concepts provide an essential link between single sourcing and the use of metadata, since in my analysis I attempt to find ways that metadata could be used to enhance information retrieval.

However, in addition to information retrieval, metadata can be used to fulfill a number of other purposes, and one of the most prominent models describing the purposes of metadata comes from Haynes (2004). Haynes's (2004, 15–17) five-point model discusses the importance of metadata by assessing the purposes that metadata is used for in different situations. This five-point model will be used as the basis for my analysis in Chapter 5, as I will review the metadata of the case study materials.

The five purposes presented in the five-point model are as follows:

1. resource description
2. information retrieval
3. management of information resources
4. documenting ownership and authenticity of digital resources
5. interoperability and e-commerce.

The first purpose, *resource description*, is, according to Haynes (2004, 64), “the most fundamental of all metadata purposes”. He goes on to note that adequate description is a way of identifying resources and distinguishing between them so that they can later be retrieved and used. The level of description deemed accurate enough depends on the actual resource and the context of use, and the metadata elements used can describe either the resource itself (such as <title> or <author>), or its relation to other resources. The second purpose, *information retrieval*, is closely linked to the first purpose, since resource descriptions and indexing terms are used in efficient information retrieval systems (Haynes 2004, 80). Using, for example, Dublin Core metadata elements to describe an information resource can provide important semantic distinctions for terms used in searches: Haynes (2004, 84) provides an example of using the metadata element <author> to distinguish between the author Green (meaning the name of the person) and ’green’ meaning the color green.

*Management of information*, the third purpose, deals with information lifecycle and the documentation processes that have taken place during the lifecycle of an information resource. Metadata use can be applied to records management, content management and library management systems, as well as to preservation management, which is particularly important regarding digital materials, since metadata can be used to “provide an avenue for describing the format and technology of a resource, aiding its management and recovery” (Haynes 2004, 116).

The fourth purpose deals with *rights management, ownership and authenticity*. This purpose, once again, links to the previous purpose of records management, dealing with the way metadata can be used to demonstrate the authenticity of an information resource, as well as document the ways it has been managed. Being able to provide proof of the authenticity of an item may have great impact on its value and meaning, particularly when it comes to official or legal records. Rights management metadata, as Haynes (2004, 113) notes, was developed to “protect the intellectual property rights associated with digital resources and … to allow for the different types of transaction that take place in creating and distributing electronic resources”. In order for rights management to work, data must be exchangeable between systems and this factor brings us to the fifth and final purpose, *interoperability and e-commerce*. Metadata that is applied in a consistent way, following a common metadata standard, enables the exchange of information and sharing of information resources between systems (Haynes 2004, 145). Haynes (2004, 142) points out that “e-commerce depends on exchange of data and the ability to process data received from one system and then pass it on to another”, meaning that interoperability is one of the key enablers of e-commerce and the use of metadata is what allows this transfer of information.

It is important to note that metadata elements can be used in various ways and they can fulfil more than one purpose presented in Haynes’s model. Whereas Haynes approaches the issue as a whole more from the point of view of the creators of the metadata, Hider (2012) focuses more on the ways in which metadata can help users achieve their goals. Relating to the first point of Haynes’s five-point model, information retrieval, Hider (2012, 18) discusses a categorization suggested by the International Federation of Library Associations and Institutions that is based on the tasks that users may carry out when they need to access useful information resources. These basic user tasks are:

- to find
- to identify
- to select
- to obtain.

The term *find* is used here in a conceptual sense, meaning the point where the user discovers the existence of a resource, with the goal of ultimately *obtaining* the resource itself. Between these two tasks, on the basis of the information provided, the user either *identifies* the resource as the one being sought for or *selects* it on the basis of its being useful and having the attributes that the user is looking for. Elaine Svenonius (2000, 20) adds a fifth function to this list of user tasks: to *navigate.* This function, by which she means a user's attempt "to find works related to a given work by generalization, association, and hierarchy", is applicable to a wide variety of digital resources, including the digital materials discussed in my case study in Chapter 5. Instead of finding and obtaining a specific item, navigating through a list of topics related to a specific subject may help a user become familiar with the subject area as a whole. I will come back to this in Chapter 5.2.3.

Although this list presented by Hider relates quite specifically to bibliographic records, it can be used to categorize the use of all kinds of metadata, since, as Ochoa (2014, 64) notes, the use of good and effective metadata does not concern only the internal characteristics of a metadata element, but also the needs of users in practice. This user-centered view is also promoted by Heting Chu (2010, 208), as she argues that "the objective of representing information is to help users save time and effort in getting the information they need". Her views regarding the functions of metadata are made quite clear, when she claims that "[m]etadata … is ultimately intended to facilitate the representation of digital information so that it can be more effectively retrieved later" (Chu 2010, 46). Chu's scholarly focus is on information representation and retrieval, and her arguments are a good example of how very differently metadata can be thought of in different fields of study; as a very general term encompassing many forms of information management or quite a specific tool for a specific purpose.

As I have previously mentioned, in order for technical communicators to be able to produce high-quality documents, the needs of the users must be known and kept firmly in mind, and so also in my analysis, a user-centered framework must be used. Usability is a factor that should be taken into account when creating information resources, including metadata that is used to describe and manage these resources. The International Organization for Standardization (ISO 1998) defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”. Usability is essentially one way of measuring system performance and user satisfaction in the context of human-computer interaction (Rowley & Hartley 2008), and, as Jacob Nielsen (1993, 24) notes: “Usability … basically is a question of whether the system is good enough to satisfy all the needs and requirements of the users”. The way in which usability may be used in evaluating the quality of metadata is discussed later in Chapter 3.3.

Robert J. Glushko (2013) expands on the notion of creating usable information resources when he describes the *process* of describing resources purposefully. His model is meant to be used across different domains, so he uses the term *resource description* instead of the term *metadata*, which is the term that I have chosen to use in the framework of this study. This model can be quite nicely linked to the specific domain of technical communication through Haynes’s (2004) five-point model, as the purposes of resource descriptions are also strongly emphasized by Glushko. Glushko’s (2013) model will be used as the framework for my analysis in Chapter 5.2, in which I analyze what kind of metadata could be used to describe the information resources of my case study in a more user-oriented way.

Glushko (2013, 148–149) states that identifying the scope and the possible purposes of information resources is the starting point of the process, as well as determining the focus, which means selecting which ones are the primary resources and which ones are the corresponding resource descriptions. In the case of my analysis, this distinction is quite straightforward, since the information resources and their corresponding metadata elements are separated in the user interface of the documentation system, but Glushko's model could be applied to a variety of different fields in which it may not be as easy to distinguish between the actual resources and their individual descriptions.

Next, the purposes of the descriptions should be determined: are the descriptions to be used for selecting, organizing or maintaining the primary resources? After determining these purposes, the next step relates to the first point of Haynes's model described earlier: what kind of properties should be described and to what extent? Not all resources require the same level of description, so it is important to decide whether the metadata elements used should only describe the intrinsic properties of the resource, or is more descriptive, contextual information needed. After identifying what kind of properties should be described, the following step is to decide how they are described. The focus should be on who uses the metadata, so the chosen standards or vocabularies should be consistent and appropriate to the context in which the descriptions will be used. The next step, designing the form of the descriptions depends heavily on the tools and software available to the community or institution in question, so I will not go deeper into this technical stage.

After all the planning and designing has been done, the actual creation of the resource descriptions, the actual metadata elements, takes place. At this stage it is important to consider whether the descriptions should be done manually by professionals, namely technical communicators, or perhaps by users using a method of social tagging, or whether automated, computer-generated metadata could be used. The final step, evaluating the resource descriptions with respect to their intended purposes is vital, although, as Wilson (2007, 17) notes, no standard definition of metadata quality exists. Although no specific standards exist, various scholars have attempted to form definitions on what is quality metadata. In the following chapter I will present some models and aspects that have been proposed for the evaluation of the quality of metadata.

## 3.3 Evaluations of metadata quality

Before starting to discuss the quality of metadata it is essential to remember that, as mentioned in Chapter 3.1, the quality of metadata is highly subjective in nature and it is always dependent on the users and their tasks. The quality is also not static, specifically when discussing digital information resources, as information may become obsolete over time if the characteristics of the resources change, making the metadata describing them incorrect (Ochoa 2014, 66). As difficult as it is to accurately and objectively measure the quality of metadata it is, however, an important discussion, since, as Hider (2012, 77) notes, only effective and high quality metadata is essentially functional in addressing the users' information needs. In attempt to reduce the multidimensionality and subjectivity of assessing metadata quality, scholars have formed several models and criteria that could be used when evaluating the quality of metadata (Ochoa 2014, 66).

Many of the models presented in metadata literature are quite abstract and theoretical. Quite many scholars are perhaps attempting to provide too wide a framework that could be applied to a variety of fields, ultimately resulting in a list of evaluation criteria that is missing concrete definitions and characteristics, making it impossible to be applied in a practical study concerning metadata quality. The first to present a list of quality criteria were William E. Moen, Erin L. Stewart and Charles R. McClure (1998), who proposed the following list: Access, Accuracy, Availability, Compactness, Compatibility, Comprehensiveness, Content, Consistency, Cost, Data Structure, Ease of Creation, Ease of Use, Economy, Flexibility, Fitness For Use, Informativeness, Protocols, Quantity, Reliability, Standard, Timeliness, Transfer and Usability. This list is highly theoretical and no details regarding the quality dimensions was provided, so the practical applications are extremely limited. Taking into consideration the time the study was published, it is obvious that the list Moen et al. presented was not aimed at accurately evaluating complex modern digital information resources, but their model presented an excellent starting point for scholars later on.

In 2004, Thomas Bruce and Diane Hillmann (2004) presented a conceptual measurement framework for metadata quality, which has become one of the most referenced frameworks in metadata quality research. Their list of seven of the most commonly recognized characteristics of quality metadata is as follows:

- Completeness
- Accuracy
- Provenance
- Conformance to Expectations
- Logical Consistency and Coherence
- Timeliness
- Accessibility.

According to Bruce and Hillmann (2004, 243), the first criterion, *completeness*, refers to the fact that a metadata element describing the resource should be as complete as possible, and the metadata fields should be applied to the set of resources as completely as possible, meaning that using metadata elements that are rarely used should be avoided. The second criterion, *accuracy*, deals with the correctness of metadata, meaning that typographical or factual errors should be eliminated with high-quality editing. *Provenance*, the source of the metadata, can be used to evaluate quality, since knowing who created the metadata and the level of their expertise regarding the field in question or metadata standards in general may have substantial effect on quality. The modifications and transformations made to the metadata also relate to the provenance of metadata, linking this quality criterion closely to the third and fourth purposes of metadata presented by Haynes. I would argue that the management of information and verifying its authenticity are highly important factors to consider, especially in the digital age where information is easily forged and the original source of information may be overlooked. Provenance of metadata should be, in my opinion, number one on the list of quality criteria, since, if the authenticity and the origins of the information resources are not clear, the other quality factors lose their authority and significance as accurate providers of correct information.

The fourth criterion, *conformance to expectations*, relates to the point that I have discussed frequently already: metadata should fulfill the needs of the users. Metadata should help users in their tasks of finding, identifying, selecting and obtaining information resources, and metadata elements that are used should, as accurately as possible, conform to the expectations of the users. Creating a controlled vocabulary with the specific needs of the particular users in mind is a way of keeping the terminology consistent (Bruce and Hillmann 2004, 244).

*Logical consistency and coherence* means that metadata elements should be consistent with standards and concepts used in the particular field. I would argue that this factor of quality is, once again, important in assisting users in finding the appropriate information, as similar resources can be searched for using similar search terms (Bruce and Hillmann 2004, 245).

The sixth criterion, *timeliness*, is divided in two by Bruce and Hillmann: *currency* and *lag*. Problems with currency occur if the information resource changes but the metadata does not. Problems with lag appear if a complete metadata set is not available at the time the information resource is inserted into the database. The final criterion, *accessibility*, deals with the fact that "metadata that cannot be read or understood by users has no value" (Bruce and Hillmann 2004, 247). With automated information processing, problems may arise from technical incompatibility, if a system is unable to process the information in the given format. In the case of metadata that is meant for human consumption the information provided may be too difficult to understand, which results in problems with cognitive accessibility. Bruce and Hillmann (2004, 247–248) note that prioritization of these criteria is highly dependent on the particular project, specifically whether the production of the metadata relies heavily on human judgement, or whether automated metadata extraction methods are used.

In addition to the list of criteria that Bruce and Hillmann (2004, 240-241) discuss, they point out one significant factor that can have substantial effects on the quality of metadata: resource constraints. As I briefly mentioned in Chapter 3.1, metadata creation is usually a resource consuming endeavor, and important aspects, such as thorough planning may be overlooked because of budget constraints. Bruce and Hillmann (2004, 241) eloquently argue that often “[q]uality that serves outsiders is seen as unaffordable altruism”. Metadata is often viewed as something that serves the purposes of the people managing the information resources, dismissing the needs of the so-called “outsiders” of the resources, the users of a software, for example. I firmly believe that in order for metadata to be regarded high-quality, it should serve the purposes of both the users and the creators of information resources.

This framework is not a clear-cut check list for determining the quality of metadata elements, but this criteria offers an excellent base for a thorough quality study. Later on in Chapter 5.3 I will discuss what criteria could be used for evaluating the quality of user-oriented metadata at Trimble Solutions.

# 4 Material and method

In this chapter I will present Trimble Solutions more closely and describe their documentation process. I will also describe the case study materials and my method of analysis.

## 4.1 Trimble Solutions and the documentation process

This section will provide background for my study, as it describes the company, Trimble Solutions, on which I focus on in my case study. I will briefly explain Darwin Information Typing Architecture (DITA), a commonly used methodology for content production, which is also the basis of documentation production at Trimble Solutions. After this I will the present the process through which documentation at Trimble Solutions is produced.

### 4.1.1 Trimble Solutions

Trimble Solutions is a software company that provides solutions for digital information modeling in the fields of infrastructure, energy and water utilities, and civil engineering. The software offers users tools from planning to construction, as well as operation and maintenance, all of which are based on the idea of utilizing a single model that specific participants of projects are given access to. The way Trimble Solutions' software works is very similar to the method of single sourcing in documentation. For example, when an engineering company is designing a new road, instead of using plans printed on paper, the single design is located in a database, to which all changes are done. Each user has access to the design through the software on their device and all modifications are saved to the database, making the modified design instantly available to other users. This multi-user model-based way of working allows flexible interoperability between different operators within a project (Tekla Civil).

Originally the company was founded in 1966 under the name Tekla and in 2011 Tekla was acquired by Trimble Ltd, a large global technology company specializing in GPS positioning technology. In 2015 Tekla moved under the Trimble brand, and the division providing solutions for infrastructure, energy and water utilities, and civil engineering is now called Trimble Solutions. The headquarters of Trimble Solutions is located in Espoo, Finland, with offices around the world (Trimble Solutions).

In 2010 Trimble Solutions made the decision to move from product-based offerings to providing solutions, meaning that instead of offering customers single products, they now provide solutions that are comprised of customizable combinations of applications. Previously Trimble Solutions offered three different products, but now the products are separated into smaller applications, from which users are able to choose a combination that suits their particular needs. This change naturally had a massive impact on documentation as well, since instead of a single manual for each product, there would be a need to sort out the necessary instructions for the specific solution of each user.

At the time of the writing of this study, the three main solutions are Tekla GIS, Tekla NIS and Tekla Civil. Tekla GIS, Geographic Information System, is a solution for municipalities, mainly used for designing and managing spatial and geographic data. Tekla NIS, Network Information System, is a solution for electricity, water, gas and district heating utilities' operations, mainly for documenting and managing network assets. Tekla Civil is a solution for civil engineering and it is mainly used for structural designing of roads and railroads (Trimble Solutions).

### 4.1.2 Darwin Information Typing Architecture (DITA)

Darwin Information Typing Architecture, or DITA for short, is XML-based open content standard for producing and reusing technical information. According to Michael Priestley (2001, 152), DITA is an XML document type definition (DTD) that consists of a set of design principles for creating modular content that can be used to produce various outputs. The basic building blocks of DITA are modular, reusable topics, rather than documents, that are designed to separate form from content, which means that the look and form of the final deliverable is defined in a separate stylesheet (Rockley et. al 2010, 4−6). Priestley (2001, 152) sheds light on the concept of the *topic* when he states that creating independent units of reuse ”allows authors to focus on writing topics that efficiently and completely cover a particular subject, or answer a particular question, without dwelling on the various places the topic might end up being read.” A topic should consist of a piece of information that can stand on its own, so that it can be read and understood without any preceding or following information.

The most common types of topics are *concept*, *task* and *reference* (Priestley 2001, 152). Concepts are used to describe or explain what something is, tasks describe how to do something and references provide additional information on subject. All of these topic types have strict definitions regarding their structure, the order in which elements, such as the title, lists or steps, or examples may appear (Rockley et. al 2010, 37–41), and these strict structure definitions ensure that all information resources are produced in a consistent way. This not only helps technical communicators in their work as they are able to focus on the content instead of the order in which the information should appear, but, as Rockley et. al (2010, 13) note, consistency helps users find information faster, as they get used to the predictable order in which information is presented. The order in which topics appear in a final deliverable, or, document, is determined by a DITA map. DITA maps do not include any content as such, they are merely lists of links that refer to the actual topics, so a DITA map can be thought of as a table of contents (Rockley et. al 2010, 4). As the topics are only referenced to the map, they can be reused over and over again in different maps in order to produce different outputs (Rockley et. al 2010, 41).

Reuse is indeed one of the most important aspects of DITA, because topics, sections, paragraphs and even individual words can be reused (Rockley et. al 2010, 24). In addition to topic- or paragraph-based reuse, the most interesting ways of reusing content, in my opinion, are conditional and variable reuse, which are extremely useful tools when working in a documentation environment that consists of multiple products, such as is the situation at Trimble Solutions. In conditional reuse, different variants are provided for a piece of information by marking it with attributes and these attributes[3] then determine how different variants are published (Rockley et. al 2010, 27). A topic, which only pertains to expert users, can be marked with the variant "expert", so when publishing a user guide for basic users, the particular topics marked "expert" can be filtered out. In variable reuse, "a variable is set up that can have a different value in different situations" (Rockley et. al 2010, 29). This type of reuse is useful, for example, in situations where a product name is different in different regions. The name of a product can be given multiple values, so that when publishing in Finland, for example, the Finnish product name is selected from the list of possible values to appear in the final output. I will come back to how reuse is utilized at Trimble Solutions in the following chapter.

### 4.1.3 Trimble Solutions documentation process

Documentation at Trimble Solutions is produced by documentation specialists, who are professional technical communicators, in collaboration with subject matter experts (later on referred to as SMEs), meaning the engineers that are in charge of the production of the solutions and consultants, who are in charge of consulting and training the users of the solutions. Each documentation specialist is in charge of one solution, working with the SMEs and consultants of that particular solution.

Documentation is produced by using the method of single sourcing, which means that all documentation is created and maintained in a single database, from which the various information resources are compiled to form the manuals for the different solutions. The software used for documentation is Structured Adobe FrameMaker, which is a document processor and publishing

[3]In DITA, *attribute* is another term for a metadata element (Rockley et. al 2010, 28).

tools designed for creating and editing large collections of documents in a single sourced documentation environment (Ethier 2004). The documentation is produced according to the principles of structured writing, discussed earlier in Chapter 2.1, and specifically, according to the principles of DITA, discussed in the previous chapter. All documentation consists of topics: concepts, tasks and references.

The process of creating new topics about, for example, a new feature of the software usually begins with discussions between the documentation specialist and the SMEs, regarding the purpose and functions of the new feature. After this the documentation specialist designs the information structure, which means deciding what kinds of topics should be used to describe the new feature. Here the needs of the users must be considered:

- Does the feature require background information, namely, should concepts be explained?
- What types of tasks should be described and how many should be written?
- Should there be reference topics covering any additional information?

After deciding on the information structure, the topics are written using the topic templates. The basic DITA topic templates have been modified to suit the specific needs of the documentation at Trimble Solutions. Each topic follows a specific structure, which makes the documentation consistent and easy to follow, because elements appear in the same order in each topic type. During the writing process, drafts are reviewed with the SMEs to ensure that all necessary information is included and that it is correct.

When the content of a topic has been written, metadata is added to each topic. These metadata elements have been chosen to help manage the information, both by the documentation specialists and the documentation system. I will present this metadata in Chapter 5.1 and thoroughly explain what these metadata elements are and how they are currently used. Each topic is then referenced to a DITA map, which is used to control the order in which topics appear as user guides are published. The user guides are published twice a year with the new releases of the software versions. Specific guides for the pieces of software are compiled from the database (the single source) by filtering them according to the metadata with which they have been marked.

DITA enables extensive reuse, as I noted in Chapter 4.1.2, and this reuse is an important aspect of documentation at Trimble Solutions. Basic topics explaining, for example, how to use keyboard controls, are only written once and then referenced to the DITA maps relating to each piece of software. In addition to reusing individual topics, reuse is also utilized with sections of topics or even words. Individual words can be referenced to in a topic, and this can save a considerable amount of time, if, for example, the name of the solution is changed. Each time the name of a solution is mentioned in the documentation, it is referenced from a separate topic that lists all solution names. If a name is changed, it needs to be modified only once in this list and from there, the reference is updated to all other topics. This referencing feature that DITA enables will come in useful soon, because although the name of the company has already been changed from Tekla to Trimble, the names of the solutions are still Tekla Civil, Tekla GIS and Tekla NIS.

## 4.2 Case study materials and method of analysis

In this chapter I will present the materials of my use case, the user guides of Trimble Solutions software, on which my analysis in Chapter 5 is based. The method used for the analysis of the user guides is also discussed.

### 4.2.1 Trimble Solutions user guides

The Trimble Solutions user guides are currently divided according to the three separate solutions, based on the way they were previously divided into products: GIS, NIS and Civil each have their own manuals. At this time, the Trimble Solutions instruction materials consist of six manuals, one user guide for each solution, as well as one administrator’s guide for each solution. Each solution includes a HTML Help comprising of a user guide and an administrator's guide within the software itself, and in addition to this, manuals are also provided for users in PDF form, the user guide and the administrator's guide as two separate PDFs.

All of the user guides are similar in regards to their structure. The guides begin with a description of the basic structure and functions of each software, then moving on to explaining the functions and features that are divided up according to their subject areas. The sections of the manuals start with an explanation of a concept, such as, "*What is a workflow?*", usually followed by the tasks that are related to that concept. The manuals are mostly written according to the principles of DITA, meaning concepts, tasks and references, but the manuals also still contain material from the time before single sourcing and DITA were implemented. This means that parts are still divided according to the old way of documenting, by the order that the commands and dialogs appear in the menus, and this combination of old and new makes quick browsing of the manuals quite laborious.

The intent at Trimble Solutions is to create a web-based service for users, in which all of the materials of the user guides are combined together in a single service. As more and more users are working with smaller solutions comprised of different applications, instead of one large piece of software, users need to be able to find the instructions they need without having to browse through the thousands of topics that are included in the manuals of each piece of software. Performing a simple full-text search in a database this big would result in massive lists of search results, so helping users find the relevant information to suit their particular needs is essential in order for the service to be useful.

### 4.2.2 Content analysis as the method

The method that I will use in Chapter 5 for my analysis of metadata is content analysis. Content analysis is a "systematic, replicable technique for compressing many words of text into fewer content categories" (Stemler 2001). The content and meanings of the messages of my case study materials are my main focus, so qualitative content analysis will be the most important method. The quantities of instances relating to different kinds of contents are taken into consideration, since the purpose of this study is not to describe each and every one of the information resources with a separate metadata element. Assigning a metadata element and values for each information resource would be counterproductive, because long lists of elements would only overwhelm the user. Presenting users with too many options would bring an information overload, thus resulting in a situation where users are not provided with tools for easy and fast information filtering.

So, through qualitative content analysis I will review the metadata that is currently assigned to topics at Trimble Solutions, pointing out the main purposes that they serve. After that I will determine what kind of metadata should be added to describe the case study materials, in order to best assist the users of the software. I this process I will use the general framework for the production of resource descriptions discussed earlier in Chapter 3.2: Glushko's (2013, 148–173) process for describing resources purposefully. The process included the following steps:

1. Determining the scope and focus, meaning the decisions on which resources will be described and in how much detail.
2. Considering the purposes of the resource descriptions.
3. Identifying the properties of the resources that relate to the purposes they need to serve. At this point it is useful to determine how many descriptions should be used for each information resource, and whether the resource properties that are used for the descriptions relate to the intrinsic or extrinsic properties of the resource.
4. Designing the actual vocabulary that is used for the resource descriptions.
5. Creating the resource descriptions.
6. Evaluating the resource descriptions.

The final step of the process is important, but because the evaluation of the quality and usability of resource descriptions is a massive task, it must fall out of the scope of this study. However, at the end of my analysis I will provide a basis for the quality evaluation of the metadata I will produce by determining which of the quality criteria suggested by Bruce and Hillmann (2004), discussed in Chapter 3.3, are the most relevant for the metadata elements presented in my analysis.

# 5 Analysis of case study metadata

In this chapter I will perform an analysis of the user guides of the Trimble Solutions software, using the theoretical frameworks provided by Haynes (2004) and Glushko (2013). In Chapter 5.1 I will present the metadata that is currently used at Trimble Solutions, explain their purpose and describe in detail the way each metadata element is used and the possible values each element may have. Throughout the analysis the names of the metadata elements are written in brackets, as in <metadata element>, and the possible values each element may have in bold, as in **value**. In Chapter 5.2 I will use Glushko's process for the creation of usable metadata as the basis for discovering what kind of metadata could be used in addition to the existing metadata. As I have previously mentioned in Chapter 3.1, the creation of metadata requires organizational and subject expertise, which I gained through working as a documentation specialist at Trimble Solutions. My knowledge of technical communication, my subject expertise and knowledge of the users of the software are important grounds for my content analysis and the assessment of the metadata.

The discussions on the purposes of the metadata in both Chapters 5.1 and 5.2 are based on Haynes's (2004) list of the five purposes of metadata. Haynes (2004, 149) points out that ”using the five-point model … it is likely that one purpose will predominate, although others will play a role and need to be taken into account”, and in the case of the materials of this study, the dominating purposes are points number one, two and three:

1. resource description
2. information retrieval
3. information management.

Although, it should be mentioned that Haynes's (2004, 145) purpose number five, interoperability, pertains to all metadata elements used at Trimble Solutions, since they are in all XML-form, and adhering to this common standard means that the metadata used is interoperable with different systems and software.

## 5.1 Metadata for information management purposes

At Trimble Solutions information resources are currently marked with metadata mainly to assist the technical communicators in their work. Metadata is used to keep track of the versions of topics, to manage what kinds of operations must be done to the topics before publishing, as well as to mark where each topic will be used when publishing the final documents as a new version of the software is released. Essentially this means that the main purpose of the metadata that is currently used is linked to the third point of Haynes's (2004, 101) five-point model, the management of information within the company. I have listed below the metadata elements that are used and their possible values:

- <company>: The documentation system marks each topic automatically with the value **Trimble**.
- <version>: This value is determined automatically by the documentation system. The value is the version number of the next release of the software.
- <product>: This element contains a predetermined list, from which the writer selects the appropriate value. The available values are **Civil**, **GIS** and **NIS**. When a new user guide is being published, a filter file is used to select the appropriate topics from the database to that specific user guide. So, for example, when publishing a new version of the Tekla Civil User Guide, only the topics that are marked with the value **Civil** are selected to form the final document.
- <lang>: This element also contains a predetermined list, from which the writer chooses the appropriate value, such as **fi-fi** or **en-us**. This metadata element is used to document

the original language of the topic. Tekla GIS and Tekla NIS topics are originally written in English, whereas the original language used to produce Tekla Civil topics is Finnish.

- <id>: This metadata element is a free text field. The writer gives each topic a unique name that helps the system identify individual topics. The name is usually the title of the topic, which is used purely for the benefit of the writers to help them identify the topics, but from the point of view of the documentation system, any unique name would suffice. The purpose of this metadata element is, in addition to help manage the information, also to describe the resource. As Haynes (2004, 67) points out, ”[a] fundamental part of any description will be some kind of system for uniquely identifying them”. The value of the <id> field could just as well be automatically determined by the system, for example, as a serial number, but that would make it impossible for the writers to quickly identify the topics when they are browsing the long lists of topics in the database.
- <to-translate>: This metadata element has to options as the value, **yes** or **no**. This value is at first assigned automatically by the documentation system and it is assigned individually to each section and paragraph of a topic. When new content is produced, the system automatically marks the topic with the value **yes**, meaning that the text must be translated before publishing. In addition to this, when a modification is done to an existing topic, the system marks the modified section with the value **yes**, which means that the specific section must be translated. The major advantage of producing content with DITA is that because each topic is divided into smaller sections, when a part of a topic is modified, the entire topic does not need to be translated again, only the modified section. This brings massive savings to translation costs. Then after the topic or a section of it has been translated, the value must be changed to **no**, which means that the translation has been done and the topic is ready to be published.

All of this metadata is essential for the documentation system and the technical communicators, as these help in keeping track of the information resources in the database. The elements <company> and <version> also relate to the fourth purpose of Haynes's (2004, 126–127) model, which is rights management, ownership and authenticity. These metadata elements play an important role in establishing the provenance of the information resources, as they establish that the resources have been produced by Trimble Solutions, who has the intellectual rights to the content.

Using metadata that fulfils the purposes of information management and rights management, however, is not sufficient in my opinion. This metadata serves the needs of the technical communicators quite nicely, but in order to consider the needs of the users, additional metadata should be taken into use. As has been stated many times already, resource description is a fundamental purpose of metadata (Haynes 2004, 64) and accurate resource descriptions are what aid in efficient information retrieval. Although the metadata elements presented in this chapter do describe the information resources to the technical communicators, they are not visible outside of the documentation system, and so, they are of little use to the users. In the following chapter I will propose additional metadata elements that could be used to describe the information resources so that users could easily find what they are looking for.

## 5.2 Metadata for resource description and information retrieval purposes

In this section I will present additional metadata categories that could be used to mark the case study materials so that the users of the software would be able to find, identify and obtain the appropriate instructions as easily and efficiently as possible. As noted in Chapter 4.2.2, I will form these categories by utilizing Glushko's (2013, 148–173) guidelines for the creation of usable information descriptions.

The first step in the process of metadata creation is to identify the resources that are to be described (Glushko 2013, 151–155). In this case these information resources are the individual topics of the Trimble Solutions' user guides. One topic is treated as one information resource, and the main content of that resource should be described using metadata elements. The purpose of creating these metadata categories is not to prescribe a description to each individual information resource, but to help users in navigating through the complex user guides, and to filter out irrelevant information. Providing a specific category for subject areas that include only a few information resources would result in category lists that are unnecessary long, so the scope of the metadata categories is limited to cover the most commonly used subject areas. What, then, the most common types of content are, is determined by the number of instances that a specific subject appears in the user guides, as well as by my professional knowledge of the behavior of the users and their use of the software.

The second step is to consider the purposes that the metadata should fulfill (Glushko 2013, 155–159). The aim of the analysis is to provide users with lists of metadata elements that can be used for filtering information resources, thus helping users eliminate irrelevant search results and find the relevant information more easily. This means that Haynes's (2004) purposes regarding resource description and information retrieval are key. Haynes (2004, 64) argues that resource description is the main function of usable metadata, because this purpose underpins all other purposes. He goes on to note that ”[a]dequate description is an essential prerequisite for resource discovery” (Haynes 2004, 64), meaning that when talking about information retrieval, a description of some kind is essential in order for users to evaluate the relevance of the resource that has been retrieved. Here we come back to the issue of determining the scope of the metadata, of how much should be described. When searching for a very specific information resource, a full-text search will be sufficient in most cases and the user is able to easily determine the relevance of the resource.

But, when dealing with issues that appear hundreds of times in the user guides, some categorization is needed in order to aid the retrieval of appropriate information in a given situation.

The main purposes of the metadata that I am creating are resource description and information retrieval, so the next step is to determine what kind of properties of the resources should be described in order to help fulfil these purposes (Glushko 2013, 160). Haynes (2004, 66) makes a distinction between intrinsic and applied metadata, which means that metadata may be either intrinsic, describing certain qualities of the information resource itself, or it can be applied, describing, for example, the context or the intended use of the resource. When searching for information in a user guide, I believe that users resort to using intuitive terms that are familiar to them, so describing the intrinsic properties of the resources is in most part the most efficient way of aiding information retrieval. However, when a user is browsing for information, instead of searching for a specific piece of information, using applied metadata elements may be appropriate. Some topics that deal with, for example, modifying something, may use the term *edit* instead of the term *modify*, so I suggest using an applied metadata element to indicate that both topics, editing and modifying, fall under the general subject area of modifying. I will discuss the issue of terminology a bit further in Chapters 5.2.3 and 5.2.4.

The next step is to design the actual description vocabulary. Glushko (2013, 165–166) emphasizes that here the most important factor is to focus on the users of the descriptions. The knowledge I accumulated of the users of the Trimble Solutions software as a technical communicator is essential for this point of the analysis, because I am familiar with the terms that are used by the users, consultants and the documentation team. For each metadata category, a list of terms is created that is compatible with the terms that are used in the user guides, as well as by consultants who train new users of the pieces of software. In cases where the terminology used in the user guides is inconsistent, a decision must be made on which term to use for the metadata.

According to the principles of DITA, all topics are produced by following a specific structure in order to create cohesion and ease of use. In addition to the uniform structure, vocabulary should also be consistent. At Trimble Solutions, all topics that deal with, for example, an instruction on how to do something, should be titled using a present participle form of a verb, such as ”Selecting a map”. Using terms in a consistent way is strongly emphasized at Trimble Solutions, and this should also be reflected in the metadata that is used.

The following points of Glushko’s (2013, 167–170) model, designing the form of the descriptions and creating the actual resource descriptions, are not dealt with in this study. The creation of the metadata will be done at Trimble Solutions by the documentation team, and the form of the descriptions is determined by the software that is used, Adobe FrameMaker, which means that the metadata is in XML-form. The final point, evaluating the quality of the resource descriptions (Glushko 2013, 171–173), is discussed later on in Chapter 5.3, in which I will provide some basis for a complete study of the metadata quality.

By following Glushko’s (2013, 148–173) model, I arrived at five metadata elements that could be used to help the users of Trimble Solutions software find relevant information more easily. The metadata elements are:

- <Information type>
- <Application>
- <Subject area>
- <Task>
- <Role>.

In the following chapters I will discuss these metadata elements and the values each element may have. My intention for the use of these elements is that these elements would be visible to the user in a user interface of an online help system. The available values of the elements would be listed as titles, and users could filter their search by selecting a value or values under each metadata element.

For reference I have also summarized the suggested metadata elements, their available values and guidelines on when to use each value in Attachment 1 in Tables 2−6.

In each of the following chapters I will provide some examples of topics that relate to the various values that specific metadata element may have. I will not present an example of each value, since I believe that a few concrete examples from the user guides will suffice to illustrate my analysis. In Attachment 2 in Figures 15−17 I will also provide as a summary examples on how to assign topics with the appropriate metadata elements and values, using three topics from the user guides as examples.

## 5.2.1 Information type

The basic principle of DITA is the fact that it is topic-oriented (Rockley et. al 2010, 4), which means that resources are written by using specific templates with a strictly defined structure, with one information resource as one topic that covers one piece of information. These topic templates are usually divided into three categories based on their information type (Priestley 2001, 152), a division commonly used in the field of technical communication. As I explained earlier in Chapter 4.1.3, information resources are produced at Trimble Solutions using the following three DITA topic templates:

- concept
- task
- reference.

As noted in Chapter 4.1.2, these templates dictate the structure of each resource that is produced, showing what elements can be used and in which order. *Concepts* are used to simply describe what something is and *tasks* are used to describe how to do something step by step. The *reference* template is used to provide additional information, for example, explaining a use case scenario or listing interface details. However, these template types are only visible for the technical communicators when they are producing the topics, and their purpose is to provide a clear structure so that all topics are written in a consistent way. The user may interpret the type of an information resource from the title of a resource, but I believe this information should be clearly marked for the benefit of the users, not only for the writers. Hence, I propose that information resources should be marked by using the metadata element <Information type>.

<Information type> is a metadata element that describes both the intrinsic and applied characteristics of resources. On one hand, it states the intrinsic form and structure of a resource, as it separates descriptive concepts from explanatory tasks. On the other hand, this element also determines how all of the different information resources relate to each other, by distinguishing between descriptive background information and additional reference information.

In describing the nature and context of resources the element <Information type> fulfils the first purpose of Haynes's (2004) five-point model, resource description. This description could aid the user in narrowing down his search, if he was interested in, for example, familiarizing himself on the basics of a particular topic. In this case the user could select the value **Background**, and his search for a specific subject would yield only results on explanatory topics, leaving out topics on performing tasks.

In the following list the available values for the metadata element <Information type> are listed, accompanied with explanations to the technical communicator on what kinds of topics to mark with each value. This list is summarized in Attachment 1 in Table 2.

- **Background**: This value is used for concept topics. Concepts provide basic information about what something is and help the user understand how a feature could be used and it what kind of context. Concepts include links to the tasks that relate to that specific matter. Figure 2 is an example of a concept topic that provides background information on maintenance plans:

**What is a maintenance plan?**

Maintenance plans organize inspection works into entities for easier control.

Maintenance plan always belongs to a pre-defined maintenance theme.

Grouping inspction works into maintenance plans is based on areas. Grouping can for example be done on the basis of city districts or a separate area division done specifically for maintenance purposes.

Maintenance plan is always done for a limited period of time. Maintenance plans can be valid for example for one year, or a shorter period of time; one alternative is to have a maintenance plan for each week - and under the maintenance plan, have an inspection work for each day.

**See also**

Management of inspections (502)
What is a maintenance theme? (502)
What is an inspection work? (503)
Scheduling inspection works (529)

*Figure 2: Topic with the* <Information type> *value* **Background**.

- **Instructions**: This value is used for task topics. These topics explain what the user should know or do beforehand, how to perform the task step-by-step, and possibly what the result of the task will be. Tasks also include links to other tasks or concepts that are related to that specific task, or, if needed, links to reference topics that may provide additional information that is related to that specific task. Figure 3 is an example of a task topic that gives instructions on how to export analysis definitions:

## Exporting analysis definitions

Export an ***analysis*** or a ***data collection*** to a file. After this, you can use the exported definitions for example in another Tekla system.

**How to do it**

1. Select the command **Analysis > Data transfer > Export...**
2. In the *Data transfer - Export* dialog, in the **File** field, enter the file in which definitions are written.
3. The table lists all definitions that you can export. Select the rows to be exported. Select several rows by holding the <Ctrl> key down.
4. Click **OK**.

**Result**

The program writes the selected definitions to file in the order they are in the table.

**See also**

What is data transfer of analysis definitions? (464)

Importing analysis definitions (464)

*Data transfer dialog*

*Figure 3: Topic with the* <Information type> *value* **Instructions**.

- **References**: This value is used for reference topics. References provide additional information regarding a task or a concept. Reference topics should not contain information that is mandatory to know in order for the user to be able to perform a task or understand a concept. Figure 4 is an example of a reference topic that provides additional reference information about different types of codes:

## Different types of codes

Codes of a certain kind are in general linked to a certain application. All code names are in English and in most cases the prefix of the name tells to which application the code belongs. The prefixes are:

- *Add* belong to address and place names information.
- *Bu* belong to building information.
- *Co* belong to company and place of business information.
- *G2k* belong to Swedish property information.
- *Gen* are general codes.
- *Lpr* belong to land property information.
- *LU* belong to land use plan information.
- *Pop* belong to population information.
- *Pro* belong to property information.
- *Prs* belong to Swedish property information.
- *Rfp* belong to fixed property information.
- *Rpa* belong to street and park information.
- *Rpm* belong to permit data handled with the Permit or the Environment control applications.
- *Tra* belong to property transfer information.

Note that you cannot modify the names of the codes.

*Figure 4: Topic with the* <Information type> *value* **References**.

- **Examples**: This value is used for use cases. Use cases are examples that further explain how to perform a possibly complex task by using an actual example. Figure 5 is an example of a use case that illustrates how to perform a specific kind of calculation:

### Use case: Calculating age as condition index factor value

This is an example of using age of objects to calculate their condition index factor value.

**Before you begin**

The condition index must have been created, and a database table or class must have been attached to it. See Creating condition indices.

**How to do it**

1. In the tree pane of the *Condition index* dialog, select the database table or class for the objects of which you want to calculate an age factor.
2. Click the **Add factor...** button. The *Select condition types and attributes* dialog opens.
3. In the *Select condition types and attributes* dialog, under the *Attributes* branch, select an object property which states the purchase year, planting year, etc. of objects, and click the **OK** button.
   The condition index factor is displayed in the *Condition index* dialog.
4. In the *Condition index* dialog:
   a. In the **Operator** list, select alternative *subtraction (variable-factor)*.
   b. Under the **Variable** part, select option **Value**, and enter the current year in the field.
      For example `2010.000`.
   c. Click the **Apply** button.

**Result**

Objects now get their value of this condition index factor from the current year subtracted with their purchase year, planting year, etc. - that is, their age.

**See also**

*Condition index* dialog

*Select condition types and attributes* dialog

What is condition index factor? (323)

Use case: Defining classification as condition index factor value (324)

What is a condition index? (322)

Calculating condition index values

*Figure 5: Topic with the* <Information type> *value* **Examples**.

### 5.2.2 Application

As I mentioned previously in Chapter 4.1.1, Trimble Solutions no longer offers its customers products, but solutions. The three pieces of software that were previously separate products comprise of smaller components, applications, which are offered as combinations that form customizable solutions. This means that a user may be working with only a handful of applications and would only need the instructions regarding those specific pieces of the software. Some users may also use different applications for different projects. Because of this modular structure of the solutions provided for the users, it would be beneficial to distinguish between parts of the user guides that pertain to different applications.

Using the metadata element <Application> would help users narrow their search of information by allowing the user to select the specific application that they are using, filtering out the resources relating to all other applications. The values on this list are intrinsic descriptions of the applications, the names that are visible in the user interfaces of the solutions. It would not be beneficial to list the purpose and use of each application here, since the users of each applications will be familiar with the specific applications that they are using.

A user of Tekla Civil may use the software only for keeping track of the locations of street signs, in which case all resources regarding, for example, construction designs or soil investigations would be of no use to the user. In this case the user could choose ”Equipment” from the list of applications, effectively leaving only the instructions regarding the management of street signs visible for him. A part of the Tekla GIS software may be used by municipalities to manage, for example, applications for building permits, so selecting ”Permit” from the applications list would provide the user with the relevant information.

The various applications are listed in Attachment 1 in Table 2. As I already mentioned, I will not explain the applications further, but below in Figure 6 I have provided an example of an introductory topic that belongs to the application called Construction Project Planning:

16 Construction Project Planning (CPP)

**Introduction**

This section describes the CPP application.

The application helps you to design electrical networks. You can use it to provide authorities with cost estimates on building a new network. Use it also for network planning tasks and for providing network constructors with relevant information to build a network.

The program has the following functionality:

- Creating basic data for a work
- Calculating hours and costs for a work based on cost information
- Selecting materials for a work based on material data
- Managing calculation results and material orders

If no work is active, all the *CPP*-related menu commands are disabled except for the command **Construction > Works...**.

**Contents of this section**


- Network planning procedure (409)
- User interface of Construction Project Planning (410)
- CPP Editor (463)


*Figure 6: Topic with the* <Application> *value* ***CPP****.*

### 5.2.3 Subject area

Haynes (2004, 88) states that ”[r]etrieval performance can be improved by use of controlled terminology to describe the subject content of the resource”. As I previously mentioned in Chapter 3.1, this means that information retrieval can be greatly enhanced by using consistent terms in the metadata elements, regardless of the actual terms used in the information resources. Utilizing a controlled vocabulary to describe resources can help users specifically in situations where they are not quite sure what they are looking for. As I briefly noted in Chapter 3.2, navigating through topics relating to a specific subject is a good way of familiarizing oneself with the subject in question. Providing a list of the most common subject areas can help users when they are not searching for a specific instruction, but rather, browsing for any kind of information on a specific subject. When analyzing the case study materials, I discovered some variance between the terms that are used to describe similar subjects. Because this discrepancy between terms may prove problematic when the user is performing searches, I believe that standardizing the terms is important.

Hence, I propose using an applied metadata element, <Subject area> that could be used for information retrieval purposes. The user can select a subject area from the list and browse through all content related to that subject, concepts as well as tasks and examples. This metadata element would be useful when the user would like to familiarize himself with the subject and all that it entails, or perhaps perform further searches within a specific subject area. In addition to users, the <Subject area> element could be used for training purposes. At Trimble Solutions, consultants use the materials of the user guides in their trainings, so selecting a value from the list of subject areas would provide consultants with an easy tool for putting together a set of training materials regarding a specific issue.

As I analyzed my case study materials I arrived at a list of the most common subject areas that are addressed throughout the user guides. These subject areas cover the most common actions of the user who is working with the pieces of software. Once again, not all important subject areas are covered in this list, since users looking for information on a smaller, specific area would be able to find what they are looking for using a full-text search. This list is intended as a frame of reference, a starting point for the user who wishes to, for example, familiarize himself with different kinds of analyses that can be performed with the software.

The values of the <Subject area> metadata element serve the first and second purposes of Haynes's (2004) five-point model, as they are short and precise descriptions of the resources and as such, enhance information retrieval. The vocabulary follows the original terms of the user guides quite strictly, since these are the terms that are familiar to the users. In cases where terminology was inconsistent, I decided on the most descriptive and commonly used term.

Below are listed the possible values of the metadata element <Subject area>, along with explanations to the technical communicator on when to use each value. This element and its values are summarized in Attachment 1 in Table 4. The available values and their uses are as follows:

- **Import & Export**: This value is used for topics that deal with importing data into the system and exporting data to other systems. These topics include explanations of different kinds of file formats, what formats can be used and how, as well as instructions on how to import and export data. Figure 7 provides an example of a topic that deals with the subject area of importing and exporting:

**Importing or exporting transformation**

Export the parameters of a previously created basic or projection transformation to file, or import them from file to the Tekla Civil system. This is, for example, a useful way to transfer transformations between different Tekla Civil systems.

**How to do it**

1. Select File > Coordinate transformations > Edit transformation....
2. In the *Choose the transformation* dialog:

| | |
|---|---|
| To export | • Select transformation.<br>• Click Export to file.<br>• Enter file where to export. |
| To import | • Click Read from file.<br>• Select file to be imported. |

**See also** Available transformation types (63)

Allowing transformation for other users (62)

*Figure 7: Topic with the* <Subject area> *value* **Import & Export**.

- **Basics**: This value is used for topics that include basic information about the software. These are topics that describe the basic structure of the software, the basic menus, commands and dialogs that appear throughout the software, as well as the basic ways of moving on the map view and making selections.

- **Settings**: This value is used for topics that explain what kinds of settings can be modified to customize aspects of the software to better fit the individual needs of the users.
- **Printing**: This value is used for topics that deal with printing, explaining what kinds of printouts can be created and how. The subject area of printing may overlap with exporting, because, for example, drawings can be printed out on paper, or to a file in a specific file format. Figure 8 provides an example of a topic that would have multiple values in the metadata element <Subject area>:

**Printing drawings to files**

Print a finished drawing to a file.

**How to do it**

1. In the *Drawings* dialog, select either the command File > Print... or File > Export to CAD....The *Print* dialog opens.
   If you selected the command File > Print..., select on the *Printer* tab of the *Print* dialog the setting Print to file and select the file format in the field next to the setting.
2. Click the File... button. A directory dialog opens.
3. In the directory dialog, browse to the folder where you want to save the drawing and enter a name and file format for the file to be created.
4. Click the OK button. The program prints the drawing in the file.

**See also**

Printing drawings on paper (673)

*Print dialog*

Printing (39)

*Figure 8: Topic with the* <Subject area> *values* **Printing** *and* **Import & Export**.

- **Plans**: This value is used for topics that describe what plans are, what kinds of plans can be created and how they can be modified.
- **Networks**: This value is used for topics related to networks and network objects, and on their creation, modifying and handling.

- **Maps**: This value is used for topics that describe the using and handling of various maps.
- **Drawings**: This value is used for topics that deal with drawings, meaning parts of plans that can be printed out of the system.
- **Analyses**: This value is used for topics that deal with performing different kinds of analyses, what kinds of data can be analyzed and how, as well as how the results of analyses can be presented. Figure 9 is an example of a topic that provides information on handling analyses:

**Copying analysis**

Copying makes it easy to create a new *analysis* that is almost the same as an existing analysis.

**How to do it**

1. In the *Analyses* dialog, go to the *All* tab and select an analysis.
2. Click the **Copy...** button.
3. In the *Copy analysis* dialog, enter the name of the analysis to be created. Click **OK**.

**Result**

The program creates the copy and adds it in the tree pane.

Tips:

- To modify the copy, select it and click the **Modify...** button.
- To move the copy to another folder, drag it.

**See also** *Copy analysis dialog*

Getting started with creating analysis (409)

*Figure 9: Topic with the* <Subject area> *value* **Analyses**.

- **Reports**: This value is used for topics that deal with producing different kinds of reports, what kind of data can be included in a report and how the results of reports can be presented.
- **Maintenance**: This value is used for topics that deal with the handling of maintenance data of objects, as well as planning maintenance works.

- **Editors**: This value is used for topics that describe the use of various editors. Editors are used to modify many aspects of the software, for example presentation techniques or symbols that are used in the software.
- **Templates**: This value is used for topics that deal with various kinds of templates that can be used in the software, what kinds of templates are available, for what purposes and how they are used.

## 5.2.4 Task

As I analyzed the user guides, I concluded that the majority of the case study materials consist of instructions on how to perform specific tasks. As this is the information that users most often look for in a user guide, it would be useful to provide users with a list of the most commonly occurring types of tasks that can be performed with the software.

These applied descriptions provide context for the information resources, as they describe various tasks and the situations in which these tasks could be performed. The *-ing* form is persistent with the principles of DITA and the company style guide, meaning that the titles of tasks must be in the present participle form. Once again, I discovered some inconsistencies with the terminology that was used in the user guides. As I came across different terms used to describe the same subject, I attempted to unify the terminology by proposing that the most familiar and clear term should be used. An example of this inconsistency is the way terms relating to editing something were used in the user guides: the same action was described by using the terms edit, modify, or even change. I suggest using the term *editing*, as its meaning is, in my opinion, clear and unambiguous.

This list of values has some overlap with the values presented in the <Subject area> list, but with the difference that the <Task> element contains only specific instructions on how to do something, whereas topics marked with the <Subject area> element may be either concepts, tasks or references. In the following list I have presented the possible values for the metadata element

<Task>, as well as the explanations to the technical communicator on what kinds of tasks should be marked with which value. This list is summarized in Attachment 1 in Table 5.

- **Importing**: This value is used for tasks that deal with importing data into the software, for example, importing plans from another system or software, or adding additional data to an existing plan from another software. Soil investigation data, for example, could be imported into the software and added to an existing design plan.
- **Exporting**: This value is used for tasks that deal with exporting data from the software to another system or software. The design data and measurements of a road could be, for example, exported from the system to be used in a machine control system of an excavator on a building site.
- **Creating**: This value is used for tasks that include information about creating new plans, designs or objects, adding or inserting objects to an existing plan, setting or defining the values and measurements of an object, or, for example, forming new contract areas for the purposes of the building site. Figure 10 is an example of a task topic that explains how to create surface models:

**Creating surface models from files**

Create a surface model directly by importing data from a file.

**Before you begin** The surface, to which the file is imported, must have been created as the type TIN-model surface (data not stored in database), and it must be the active surface. See:

- Setting active surface (249)
- Creating a surface (250)

**How to do it**

1. Select Surfaces > Create surface model from files....
2. In the *Read file to surface model* dialog, give the file or directory to be imported. If you give a directory, its subdirectories are also imported.
   a If the file format is other than XYZ, Tielaitos or grid, see Defining formatting when importing surface models from files (254).
3. In the Area selection list, select the area to be imported from.
4. In the Target surface list, select the surface onto which the file is imported.

**Result**

The program imports the points and triangulates automatically.

**Problems?** The program writes a log file of each imported file. The log files are name dafter the file, with an `.log` ending. If the import is interrupted, you can continue it by moving the successfully imported files from the log file to another location.

**See also** Creating a surface model from files - basics (252)

File formats when creating surface models from files (255)

*Figure 10: Topic with the* <Task> *value* **Creating**.

- **Selecting**: This value is used for topics that deal with selecting or activating objects or areas, as well as removing or releasing a selection.
- **Viewing:** This value is used for tasks that deal with viewing objects, networks, plans or maps, hiding or showing objects on the map, visualizing results of reports or analyses, filtering objects to be shown on the map.
- **Editing**: This value is used for all tasks that provide information about various kinds of editions and modifications that can be performed with the software, such as editing or modifying objects, designs or plans, marking objects or areas as incomplete or finished, cutting and pasting objects, redrawing maps or views after modifications have been done, as well as changing the values of an object.

- **Deleting**: This value is used for all tasks that deal with deleting or removing items or objects, ending or terminating a process, taking an object out of use, detaching an item, or clearing a map or a view. Figure 11 is an example of a topic that relates to the task of deleting something:

**Removing visualizations from map**

After running an analysis, results are typically shown on the map of the main window. When you no longer need them, you can remove them.

**Before you begin** Visualizations must be open in the *Legend* dialog. It opens when running analysis. You can also open it with the command **Analysis > Legend...**.

**How to do it** 1. There are two ways to remove:

| | |
|---|---|
| To hide | In the *Legend* dialog, unselect the **Show** setting of the row. |
| To delete entirely | In the *Legend* dialog, select the row and click the **Remove results** button.<br>Deleted visualizations can be shown again only by re-running the analysis. |

**See also** Hiding analysis legend from map printout (433)

What is visualization? (428)

*Figure 11: Topic with the* <Task> *value* **Deleting**.

- **Moving**: This value is used for all tasks that have to do with moving or positioning objects on a map, adjusting the position of an object on a map, as well as for the user's maneuvers of the map itself, meaning zooming or centering the map view or magnifying an object on a map.

- **Calculating**: This value is used for tasks that deal with performing calculations in the software. Figure 12 is an example of a topic that instructs the system administrator on how to perform certain calculations:

**Calculating center point of distribution substation in SOSI import**
Calculate the center point of a distribution substation.

**How to do it**

1. Define the state of the `SOSI_DS_CENTER` environment variable to be `SET`. The application creates a DistrSubstation object in the point calculated by using the SOSI object's first and third coordinate lines. The point is the midpoint of these coordinates.
2. Define the cross-reference file to connect the TEMA codes related to the distribution substation wall lines and the distribution substation point database table and class.
3. Convert the distribution substation wall lines in a separate conversion run. This reduces memory usage.

**See also** Environment variables used with SOSI import files (288)

*Figure 12: Topic with the* <Task> *value* **Calculating**.

- **Reporting**: This value is used for all tasks that deal with various kinds of reports that can be performed by using the software.
- **Configuring**: This is a value is meant for system administrators and it is used for tasks that deal with various kinds of configurations that an administrator may make to the software.

### 5.2.5 Role

According to Haynes (2004, 66), a metadata element can be used to show that a specific information resource is ”intended for use by particular (age) groups, or is designed to exclude specific groups”. When dealing with software user guides, making a distinction between groups of users is quite useful, since user guides often include very basic information for novice users, as well as very specific information for expert users.

In the case of Trimble Solutions, as I mentioned in Chapter 4.2.1, the user guides are divided into two parts: a basic user's guide and an administrator's guide. The administrator's guide includes information that is meant only for the most expert level user of the software within a company. This system administrator is the only one who has the administrative rights to make modifications to the features and configurations of the software, so the information included in the administrator's guide is quite irrelevant for the basic user. Below are listed the possible values for the metadata element <Role> and guidelines to the technical communicator on when to use each value. This metadata element and its values is summarized in Attachment 1 in Table 6.

- **Basic user**: Resources marked with this value include information for all basic users of the software, who do not have system administrator's rights. These resources are available for all users. Figure 13 is an example of a topic that deals with the task of visualizing reports and it is aimed for basic users:

**Visualizing the report of degree of readiness for masses on the map**

Visualize the location of a mass class from the report of the degree of readiness on the map. If there are several masses at that location, it is sensible to visualize only one mass class at a time.

**Before you begin** See Reporting masses' degree of readiness (240).

**How to do it** 1. In the table at the bottom of the *Report masses* dialog, select Visualize for the mass class.

**Result**

Masses are visualized using different colors depending on their degree of readiness:

- Planned masses: white.
- Completed masses: green.

If the *Profile of alignment* dialog is open, mass is also visualized there using the same color fill as on the map.

**See also** Settings for the report of masses' degree of readiness (242)

*Figure 13: Topic with the* <Role> *value* **Basic user**.

- **Administrator**: Resources marked with this value consist only of information that is meant for the system administrator. These topics include, for example, parameter tables for the configuration of the software, or database tables, which can be used for managing the way objects are presented in the software. Figure 14 is an example that describes various templates that are available for reporting and it is aimed for the system administrator:

### Report templates

The program includes the following report templates that are stored in the **program\config\report** folder. The templates can be modified and saved to a local directory. Template extension .rtf generates a .rtf file, and extension .txt generates a csv file to be used with MS Excel.

The report templates are:

- Estates report (**RepeEstatesTemplateFile.xre and EstatesReport_<language>.rtf**)
- Low voltage switches in distribution substation (**RepeLVSwitchesTemplateFile.xre**) **and LVSwitchesReport_<language>.rtf**)
- OMS service break event log (**RepeOmsEventLog.xre**) **and OmsEventLogReport_<language>.rtf**)
- OMS single scheduled service break (**OmsOutageScheduleReport_<language>.rtf**)
- OMS single service break (**OmsServiceBrekaOneReport.txt**) **and OmsServiceBrekaOneReport_<language>.rtf**)
- OMS service breaks during a selected time period (**OmsServiceBrekaSummaryReport.txt**) **and OmsServiceBrekaSummaryReport_<language>.rtf**)
- Single work report (**RepeWorkTemplateFile.xre**) **and WorkReport_<language>.rtf**)
- All works of a plan (**RepeWorksTemplateFile.xre**) **and WorksReports_<language>.rtf**)

*Figure 14: Topic with the* <Role> *value* **Administrator**.

## 5.3 Metadata quality criteria for future research

The next stage of the process of describing information resources purposefully, as noted by Glushko (2013), would be to perform a quality analysis. Because a full quality analysis cannot be fitted within the scope of this study, I will suggest what aspects of metadata quality could be focused on in an in-depth study of the quality of my analysis results. These suggestions are based on the quality criteria presented by Bruce and Hillmann (2004), presented earlier in Chapter 3.3, as their framework is widely recognized and has been used for assessing the quality of metadata by many scholars. The list of the seven quality metadata criteria was as follows:

- Completeness
- Accuracy
- Provenance
- Conformance to Expectations
- Logical Consistency and Coherence
- Timeliness
- Accessibility.

Previously in Chapter 3.3, I pointed out that *provenance* is perhaps the most important aspect of quality, and this is certainly the case with the metadata that I have described in Chapter 5.1, the metadata that is currently used at Trimble Solutions. The metadata that is used for the purposes of information management must, in my opinion, first and foremost prove the authenticity of information resources, because if the origins of the resources cannot be verified, the information itself cannot be verified either. Bruce and Hillmann (2004, 243–244) point out that the provenance of metadata is a good starting point for a quality study, and important aspects that affect provenance are, for example, knowing who made the metadata, what their level of expertise is, as well as knowing what modifications have been made to the data.

*Completeness* is an aspect of quality that would require further study, because the completeness of metadata is quite a subjective aspect. In most cases when the quality of a metadata element set is being evaluated, no definitive conclusions can be made about whether the element set used is describing the resources as completely as possible. However, assessments can always be made. For the method of these assessments Ochoa (2014, 73) suggests using metadata experts to "manually evaluate a statistically significant sample of the metadata instances according to a predefined framework or set of quality criteria". Ochoa also promotes the use of Bruce and

Hillmann's quality criteria for these assessments by experts. The metadata elements and values that I have proposed in Chapter 5.2 should, in my opinion, be assessed in relation to this quality aspect. I would suggest using subject matter experts to determine whether the metadata element set is complete enough, or should additions or eliminations be made.

*Accuracy* is another aspect that should be focused on, because as Bruce and Hillman (2004, 243) argue, ”the information provided in the values [of metadata elements] needs to be correct and factual”. This criterion must of course apply to all metadata elements, both for the ones primarily meant for the technical communicators, as well as for the ones meant for the benefit of the users.

*Conformance to expectations* is an extremely important quality factor, especially when talking about metadata that is meant to be user-centered. Bruce and Hillmann (2004, 244) point out that ”metadata choices need to reflect community thinking and expectations”, and ”[e]lement sets … should, in general, contain those elements that the community would reasonably expect to find.” This aspect is quite closely linked to the completeness of metadata, since the expectations of users are what constitute completeness: user expectations should be considered realistically, considering the resources available for a metadata creation project, because in most metadata projects it is not possible to include everything for everyone. Controlled vocabularies should be constructed using the terms and expressions that the users would use and expect to find. As I pointed out in my analysis in Chapters 5.2.3 and 5.2.4, controlled vocabularies should include the clearest and most commonly used terms for the situation in question. In order to determine what is the quality terminology the user would expect to find, extensive knowledge on the specific field is required.

The final criterion that I believe would be important for a quality study in this case is *consistency and coherence*. In high-quality metadata definitions and concepts should be consistent and in line with the terms used in the specific subject area. Bruce and Hillmann (2004, 245) specifically emphasize that ”[u]sers expect to be able to search collections of similar objects using similar criteria, and increasingly they expect search results and indicative indexes to have similar structures and appearance.” When the goal is to produce good and usable user-oriented metadata, this criterion of consistency and coherence understandably takes center stage.

Regarding the quality aspects of conformance to expectations, and consistency and coherence I believe that the most beneficial method of evaluation would be user studies. The metadata elements that I proposed in Chapter 5.2 are meant to aid the users of the software, so naturally it would be useful to analyze whether the terminology used is actually what the users would expect to find.

The five aspects of metadata quality pointed out here are, in my opinion, the most important ones regarding my case study. The two final criteria from the list presented by Bruce and Hillmann, *timeliness* and *accessibility*, are not as crucial at this point, since problems with timeliness usually apply to situations where the resource changes but the metadata does not. In this case, as the metadata is just being added to the information resources, timeliness should not be a problem, just as accessibility should not be, because the form of the metadata is currently interoperable with other systems. In the future, should situations change, these two aspects should of course also be considered to make sure that the metadata will stay current and accessible.

# 6 Conclusions

In this study I have examined how the information resources of a software company could be described in order to provide users with an effective way of searching for information and applied a framework for the creation of useful and descriptive metadata. The goal of my study was to research how metadata could be approached and utilized from the point of the users, not just the technical communicators. I analyzed the purposes that metadata serves at Trimble Solutions at the moment, as well as what kind of metadata could be used in addition to the existing metadata to better serve the needs of the users. I concluded that currently the metadata is used for the purposes of managing information resources and the authenticity and ownership of these resources. As such, the metadata is useful internally in the company, because it is utilized by the technical communicators and the documentation system. However, I believe that this is not sufficient, and in order to consider the users of the software, additional metadata should be used. After analyzing the contents of the user guides I arrived at a collection of metadata elements, all of which include a list of values that could be used to mark the topics in the user guides. These elements serve the purposes of the users, as they describe the resources enabling more efficient information retrieval. All of the metadata is in XML-form, thus ensuring an important purpose of digital resource metadata, interoperability. Table 1 below summarizes the current metadata, my suggested additional metadata, and the purposes the metadata elements serve, according to the five-point model presented by Haynes:

*Table 1: Metadata elements divided according to the user of the specific elements, as well as the purposes the metadata serves.*

| User of metadata | Metadata elements | Purposes according to Haynes's five-point model |
|---|---|---|
| **Technical communicator** | • <company><br>• <version><br>• <product><br>• <lang><br>• <id><br>• <to-translate> | 3. Management of information resources<br>4. Documenting ownership and authenticity of digital resources<br>5. Interoperability |
| **User of software** | • <Information type><br>• <Application><br>• <Subject area><br>• <Task><br>• <Role> | 1. Resource description<br>2. Information retrieval<br>5. Interoperability |

In order to help users deal with massive amounts of information, it is important to provide them with all the possible tools for easy and fast information processing. Especially for companies with large, single sourced information resource databases, using metadata to help users find the relevant information at the right time is extremely important. The process of creating user-oriented metadata is a large undertaking, and the process should be carefully planned. Although Glushko discusses his model for the creation of resource descriptions on quite an abstract level, I found it to be very applicable for the purposes of my study.

Probably the most beneficial way of using the metadata elements that I have proposed in this study would be to utilize them in a system that includes user profiles and individual sign-in. This way the basic information of the user could be entered only once and the information would be saved in their user profile. By basic information I mean the solution or applications that the user works with, as well as his role. When the user signs in to the help system, topics relating to irrelevant applications, for example, would automatically be filtered out, thus helping the user focus on the topics that are relevant to him.

Because this case study was done by a professional technical communicator, it cannot provide comprehensive results from the point of view of users. The terminology that was used in the materials of my case study was at times inconsistent and I attempted to clarify these discrepancies by creating a controlled vocabulary, a list of terms that should be used at Trimble Solutions when creating information resources and the metadata related to them. However, it is important to note that as I made these decisions, my terminology selections might have not always been in line with the views of the users or other technical communicators. Therefore, an essential topic for future research would be performing user studies to see how the users of the software view the metadata elements and their values presented in this study. Only through practical user studies is it possible to receive confirmation on whether the terminology I have chosen is truly intuitive for the users of the software. Regarding the quality of the metadata I have presented in this study, I would also suggest using subject matter experts to perform assessments on the terminology that I have used. Expert assessments together with user studies would provide important insight into the quality of my analysis.

The metadata elements that I presented in Chapter 5.2 are applicable for use at Trimble Solutions, but I believe that this study could provide a kind of a starting point for other companies and communities who wish to utilize metadata in enhancing the management and retrieval of their information resources. As I have noted throughout this study, metadata is expensive. However, companies should perhaps look at the issue from another angle, because fierce competition in the world of business requires companies to always provide their customers with the best possible value. The creation of high-quality metadata should not be viewed as an unnecessary use of resources, but as an important investment in customer satisfaction. High-quality metadata that is designed to be user-oriented will surely pay back the investment.

In the field of technical communication, especially with the rise of single sourcing, special attention should be paid to the use of metadata. Previous studies of metadata have mostly focused on the way metadata could be utilized to organize and manage information resources within a community or a company. This is indeed an important aspect, since the management of information is a crucial aspect of technical communication and the use of metadata is important. However, I believe that the purposes of metadata should be considered from a wider perspective, paying attention to the ways it could be used to facilitate faster information retrieval from the point of view of users. After all, the aim of technical communication is to find ways of helping users and to serve their needs to the best of our abilities. Understanding the ways that metadata usage could be enhanced in technical communication is, in my opinion, important, especially for technical communicators that are operating in a single sourced documentation environment that holds massive amounts of information. My case study has provided an example of how, in the field of technical communication, metadata usage could be viewed from various angles and thus possibly be improved upon.

I believe that especially in the digital age, fast and accurate information retrieval is just as important and beneficial to users as is the accuracy and clarity of information. In the field of technical communication, research on the needs of the users is a subject that should be constantly studied, especially at a time when fast developing technologies bring changes to the ways information is presented and consumed. I believe that my study has shown how the theories discussing information retrieval and the purposes of metadata could be used in a practical process of metadata assessment, as well as creation. The use of user-oriented metadata is an important subject that surely deserves more attention and further research.

# 7 Works cited


Albers, Michael J. 2003. Single Sourcing and the Technical Communication Career Path. *Technical Communication* 50, 3: 335−343.

Ament, Kurt. 2003. *Single Sourcing. Building modular documentation*. Norwich: William Andrew Publishing.

Bruce, Thomas R. and Diane I. Hillmann. 2004. The Continuum of Metadata Quality: Defining, Expressing, Exploiting. In *Metadata in Practice*, ed. Diane I. Hillmann and Elaine L Westbrooks, 238−256. Chicago: ALA Editions.

Carter, Locke. 2003. The Implications of Single Sourcing for Writers and Writing. *Technical Communication* 50, 3: 317−320.

Chu, Heting. 2010. *Information Representation and Retrieval in the Digital Age*. 2nd ed. Medford: Information Today, Inc.

Day, Michael. 2001. Metadata in a Nutshell. *Information Europe* 6, 2: 11.

Dublin Core Metadata Initiative (DCMI). Available from: http://dublincore.org/ [Accessed 20 May 2016]

Eble, Michelle F. 2003. Content vs. Product: The Effects of Single Sourcing on the Teaching of Technical Communication. *Technical Communication* 50, 3: 344−349.

Ethier, Kay. 2004. *XML and FrameMaker*. New York: Apress.

Foulonneau, Muriel and Jenn Riley. 2008. *Metadata for Digital Resources: Implementation, Systems Design and Interoperability.* Oxford: Chandos Publishing.

Glushko, Robert J. 2013. *The Discipline of Organizing*. Cambridge: MIT Press.

Greenberg, Jane. 2005. Understanding Metadata and Metadata Schemes. *Cataloging & Classification Quarterly* 40, 3–4: 17−36.

Hackos, JoAnn T. 2002. *Content Management for Dynamic Web Delivery*. New York: John Wiley & Sons, Inc.

Haynes, David. 2004. *Metadata for information management and retrieval*. London: Facet Publishing.

Hider, Philip. 2012. *Information Resource Description: Creating and Managing Metadata*. London: Facet Publishing.

Hughes, Michael. 2002. Moving from Information Transfer to Knowledge Creation: A New Value Proposition for Technical Communicators. *Technical Communication* 49, 3: 275–285.

International Organization for Standardization (ISO). 1998. *Guidance on usability* (ISO 9241–11). Geneva: ISO.

Moen, William E., Erin L. Stewart and Charles R. McClure. 1998. Assessing Metadata Quality: Findings and Methodological Considerations from an Evaluation of the U.S. Government Information Locator Service (GILS). In: ADL '98: *Proceedings of the Advances in Digital Libraries Conference*. Washington: IEEE Computer Society.

National Information Standards Organization (NISO). 2004. *Understanding metadata*. Bethesda: NISO Press. Available from: http://www.niso.org/publications/press/UnderstandingMetadata.pdf [Accessed 31 May 2016]

Nielsen, Jacob. 1993. *Usability engineering*. New York: Morgan Kaufmann.

Ochoa, Xavier. 2014. Metadata Quality. In *Handbook of Metadata, Semantics and Ontologies*, ed. Miguel-Angel Sicilia, 63−89. Singapore: Worldwide Scientific Publishing Co.

Olson, Debbie. 2009. Metadata: A Primer for Indexers. *Key Words* 17, 1: 18−20.

Priestley, Michael. 2001. DITA XML: A reuse by reference architecture for technical documentation. *Proceedings of the 19th Annual International Conference on Computer Documentation*. New York: ACM Press.

Rockley, Ann. 2001. The Impact of Single Sourcing and Technology. *Technical Communication* 48, 2: 189−193.

Rockley, Ann. 2003. It's About People, Not Just Technology. *Technical Communication* 50, 3: 350−354.

Rockley, Ann, Steven Manning and Charles Cooper. 2010. *DITA 101: Fundamentals of DITA for Authors and Managers*. 2nd ed. The Rockley Group Inc.

Rowley, Jennifer E. and Richard J. Hartley. 2008. *Organizing knowledge: An introduction to managing access to information.* Burlington: Ashgate.

Sapienza, Filipp. 2007. A Rhetorical Approach to Single-Sourcing Via Intertextuality. *Technical Communication Quarterly* 16, 1: 83–101.

Stemler, Steve. 2001. An overview of content analysis. *Practical Assessment, Research & Evaluation* 7, 1. Available from: http://PAREonline.net/getvn.asp?v=7&n=17 [Accessed 6 September 2016]

Svenonius, Elaine. 2000. *The Intellectual Foundation of Information Organization*. Cambridge: MIT Press.

Swarts, Jason. 2010. Recycled Writing: Assembling Actor Networks From Reusable Content. *Journal of Business and Technical Communication* 24, 2: 127–163.

TCeurope TecDoc-Net. 2005. Professional education and training of Technical Communicators in Europe. Guidelines. Version 1.0. Available from: http://www.tceurope.org/images/stories/downloads/projects/tecdoc.pdf [Accessed 26 April 2016]

Tekla Civil. Available from http://www.tekla.com/products/tekla-civil [Accessed 17 August 2016]

Trimble Solutions. Available from: www.tekla.com [Accessed 17 August 2016]

Wendler, Robin. 1999. Metadata in the library. *Harvard University Library Notes* 1286, July/August: 4–5.

Whittemore, Stewart. 2008. Metadata and Memory: Lessons from the Canon of *Memoria* for the Design of Content Management Systems. *Technical Communication Quarterly* 17, 1: 88−109.

Williams, Joe D. 2003. The Implications of Single Sourcing for Technical Communicators. *Technical Communication* 50, 3: 321−327.

Wilson, Amanda J. 2007. Toward Releasing the Metadata Bottleneck. *Library Resources & Technical Services* 51, 1: 16−28.

Zeng, Marcia Lei and Jian Quin. 2008. *Metadata*. London: Facet Publishing.

# Attachments

## Attachment 1

The following tables can be used for reference when adding user-oriented metadata to topics.

<Information type>

The use of this metadata element is mandatory for each topic.

*Table 2: Values of the metadata element <Information type> and guidelines on when to use each value.*

| Value | When to use |
|---|---|
| **Background** | Concept topics |
| **Instructions** | Task topics |
| **References** | Reference topics |
| **Examples** | Use cases |

<Application>

The use of this metadata element is not mandatory for all topics.

*Table 3: Values of the metadata element <Application> grouped according to the solutions to which they belong.*

| Solution | Value |
|---|---|
| Tekla Civil | • Civil Main application<br>• Terrain and map<br>• Construction support<br>• Equipment<br>• Soil investigations<br>• Structure<br>• Drawings<br>• On-Site operations |
| Tekla GIS | • GIS Basic<br>• Map handling<br>• Land use plan<br>• Address and names<br>• Street and park<br>• Permit<br>• Maintenance Management System (MMS) |

| | |
|---|---|
| | • Terrain model<br>• Spatial analysis<br>• Feedback system<br>• eSite<br>• Webmap |
| Tekla NIS | • NIS Main application<br>• Distribution Network Operation (DMS)<br>• Operation Management System (OMS)<br>• District heating<br>• Power system analysis<br>• Asset Management<br>• Reliability-based Network Analysis (RNA)<br>• Technical-Economic Planning (TEP)<br>• Construction Project Planning (CPP)<br>• Location map<br>• Maintenance Management System (MMS)<br>• Land use<br>• Webmap |

<Subject area>

The use of this metadata element is not mandatory for all topics.

*Table 4: Values of the metadata element <Subject area> and guidelines on when to use each value.*

| Value | When to use |
|---|---|
| **Import & Export** | Importing data into the system and exporting data to other systems. |
| **Basics** | Basic information about the system. |
| **Settings** | Settings used to modify and customize the system. |
| **Printing** | Printing to paper or file. |
| **Plans** | Using and modifying plans. |
| **Networks** | Handling network data. |
| **Maps** | Using and handling maps. |
| **Drawings** | Creating and using drawings. |
| **Analyses** | Handling and performing analyses. |
| **Reports** | Handling and producing reports. |
| **Maintenance** | Handling maintenance data. |
| **Editors** | Using editors. |
| **Templates** | Using and creating templates. |

<Task>

The use of this metadata element is not mandatory for all topics.

*Table 5: Values of the metadata element <Task> and guidelines on when to use each value.*

| Value | When to use |
|---|---|
| **Importing** | Importing data into the system. |
| **Exporting** | Exporting data to other systems. |
| **Creating** | Creating, adding, inserting, defining, forming, taking into use. |
| **Selecting** | Selecting, activating, releasing. |
| **Viewing** | Viewing, hiding, showing, displaying, highlighting, visualizing, windowing, filtering. |
| **Editing** | Editing, modifying, cutting, pasting, adjusting, changing, marking as, redrawing. |
| **Deleting** | Deleting, removing, cutting, ending, clearing, terminating, taking out of use. |
| **Moving** | Moving, zooming, centering, magnifying, positioning, transferring. |
| **Calculating** | Performing calculations. |
| **Reporting** | Producing reports. |
| **Configuring** | Configuring settings. |

<Role>

The use of this metadata element is mandatory for each topic.

*Table 6: Values of the metadata element <Role> and guidelines on when to use each value.*

| Value | When to use |
|---|---|
| **Basic user** | Topics meant for the use of basic user without system administrator’s rights. |
| **Administrator** | Topics meant for the use of system administrator. |

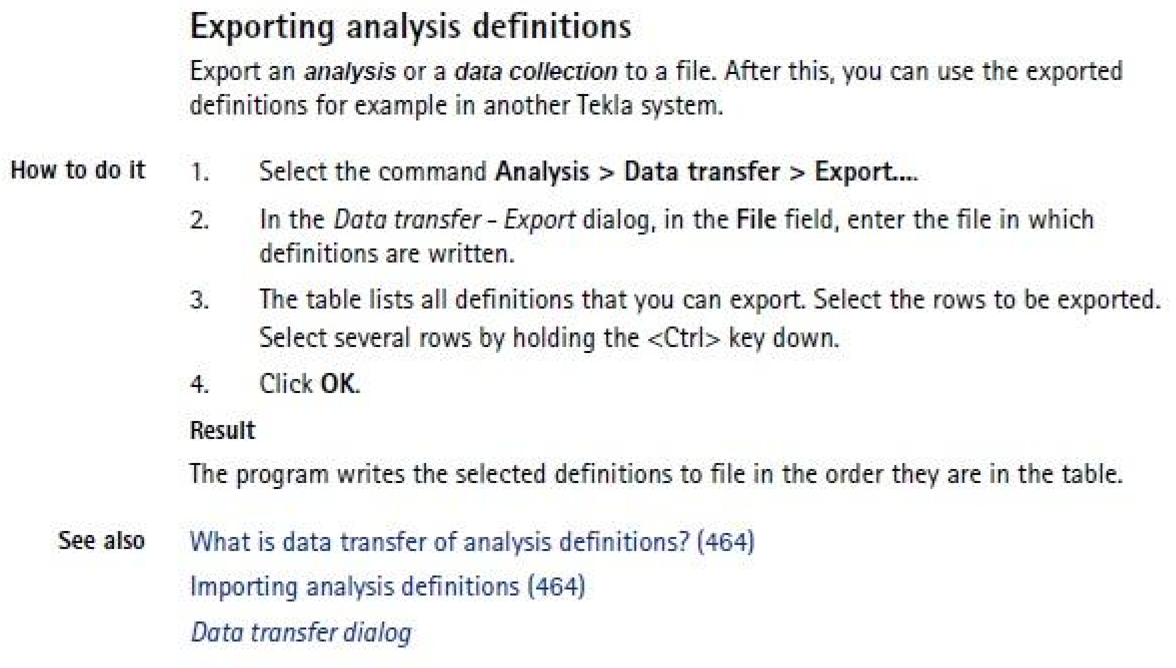
# Attachment 2

The following examples show how some of the topics provided as examples in Chapter 5.2 would be marked using the user-oriented metadata.

**Example 1**

The topic in Figure 15 would be marked with the following metadata elements and values:

- <Information type> **Instructions**
- <Application> **Spatial analysis**
- <Subject area> **Import & Export**; **Analyses**
- <Task> **Exporting**
- <Role> **Basic user**.



## Exporting analysis definitions

Export an *analysis* or a *data collection* to a file. After this, you can use the exported definitions for example in another Tekla system.

**How to do it**

1. Select the command **Analysis > Data transfer > Export...**
2. In the *Data transfer - Export* dialog, in the **File** field, enter the file in which definitions are written.
3. The table lists all definitions that you can export. Select the rows to be exported. Select several rows by holding the <Ctrl> key down.
4. Click **OK**.

**Result**

The program writes the selected definitions to file in the order they are in the table.

**See also**

What is data transfer of analysis definitions? (464)

Importing analysis definitions (464)

*Data transfer dialog*

*Figure 15: Example of a topic dealing with the task of exporting analyses.*

**Example 2**

The topic in Figure 16 would be marked with the following metadata elements and values:

- <Information type> **Instructions**
- <Application> **Drawings**
- <Subject area> **Import & Export**; **Printing**; **Drawings**
- <Task> **Exporting**; **Printing**
- <Role> **Basic user**.

**Printing drawings to files**

Print a finished drawing to a file.

**How to do it**

1. In the *Drawings* dialog, select either the command File > Print... or File > Export to CAD....The *Print* dialog opens.
   If you selected the command File > Print..., select on the *Printer* tab of the *Print* dialog the setting Print to file and select the file format in the field next to the setting.
2. Click the File... button. A directory dialog opens.
3. In the directory dialog, browse to the folder where you want to save the drawing and enter a name and file format for the file to be created.
4. Click the OK button. The program prints the drawing in the file.

**See also**

Printing drawings on paper (673)

*Print dialog*

Printing (39)

*Figure 16: Example of a topic dealing with the task of printing.*

**Example 3**

The topic in Figure 17 would be marked with the following metadata elements and values:

- <Information type> **References**
- <Subject area> **Reports**; **Templates**
- <Role> **Administrator**

## Report templates

The program includes the following report templates that are stored in the **program\config\report** folder. The templates can be modified and saved to a local directory. Template extension .rtf generates a .rtf file, and extension .txt generates a csv file to be used with MS Excel.

The report templates are:

- Estates report (**RepeEstatesTemplateFile.xre and EstatesReport_<language>.rtf**)
- Low voltage switches in distribution substation (**RepeLVSwitchesTemplateFile.xre**) **and LVSwitchesReport_<language>.rtf**)
- OMS service break event log (**RepeOmsEventLog.xre**) **and OmsEventLogReport_<language>.rtf**)
- OMS single scheduled service break (**OmsOutageScheduleReport_<language>.rtf**)
- OMS single service break (**OmsServiceBrekaOneReport.txt**) **and OmsServiceBrekaOneReport_<language>.rtf**)
- OMS service breaks during a selected time period (**OmsServiceBrekaSummaryReport.txt**) **and OmsServiceBrekaSummaryReport_<language>.rtf**)
- Single work report (**RepeWorkTemplateFile.xre**) **and WorkReport_<language>.rtf**)
- All works of a plan (**RepeWorksTemplateFile.xre**) **and WorksReports_<language>.rtf**)

*Figure 17: Example of a topic that deals with the use of report templates.*