SUBJECT ANALYSIS THEORIES AND THEIR APPLICATION TO GEOGRAPHIC SUBJECT METADATA FOR ELECTRONIC THESES AND DISSERTATIONS IN SOUTH AFRICAN UNIVERSITY LIBRARIES

by

KHOMOTSO AMANDA MAPHOPHA

Student No: 7103670

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy and Literature

in the subject of

Information Science

at the

UNIVERSITY OF SOUTH AFRICA, PRETORIA

SUPERVISOR: DR L. M. CLOETE

February 2022

DECLARATION

Name: Khomotso Amanda Maphopha

Student number: 7103670

Degree: Doctor of Philosophy in Information Science

Title of the thesis:

Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations in South African University libraries

I declare that the above thesis is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I submitted the thesis to originality checking software and that it falls within the accepted requirements for originality.

I further declare that I have not previously submitted this work, or part of it, for examination at Unisa for another qualification or at any other higher education institution.

_____ **February 2022**

Signature Date

ABSTRACT

The study investigated the way in which theories of subject analysis facilitate the creation of effective geographic subject metadata for ETDs. The subject analysis practices of metadata creators, who contribute geographic subject metadata to ETD repositories in South African university libraries, and the underlying theoretical aspects were explored.

Subject metadata is highly significant in facilitating the discovery of the subject content of ETDs. This study focused on geographic subject metadata as a type of metadata that helps to search for information related to a specific geographic locality. Geographic subject metadata facilitates the discovery of information resources and, in this way, it is vital in allowing a different form of subject access – i.e. apart from the common, topical approach. This type of metadata is useful in bringing together all information on a particular locality.

A mixed methods research methodology was employed in the study. The study used the explanatory sequential mixed method research design, with three data collection methods. A survey questionnaire and interviews were used to investigate the experiences and perspectives of the subject metadata creators in South African university libraries. Content analysis was conducted to examine the ETD metadata records for the extent and nature of the use of geographic subject metadata. For the first, quantitative part of the study, a survey questionnaire was used, in order to include more respondents in the collection of primary data.

A qualitative study was conducted by means of interviews, with a sample of metadata creators, so as to allow deep probing into the approaches followed by the metadata creators during subject analysis for the purpose of subject metadata creation. Quantitative/qualitative content analysis was employed to study the ETD records.

The Statistical Package for the Social Sciences (SPSS) was used for the quantitative data analysis. Basic descriptive statistics was used to summarise and present this data meaningfully, so as to make the underlying information easily comprehensible.

NVivo10 software assisted in organising and exploiting the qualitative research data. Thematic analysis facilitated detailed explanations of the data obtained from the participants' responses to subject analysis approaches, metadata creation procedures and other general contextual influences. For the content analysis phase, Excel spreadsheets were used to perform analysis functions; to manipulate the extracted ETD records; and to perform the data analysis.

The findings of the quantitative survey, qualitative interviews and the content analysis revealed a variation in the existing metadata practices for theses and dissertations and the approaches followed in subject analysis and ETD geographic subject metadata creation. Both the quantitative and the qualitative findings signified inconsistent practices and varied perceptions of the inadequacies of existing ETD geographic subject metadata. The records analysis confirmed the inconsistencies in the current approaches to ETD geographic subject metadata creation. The qualitative interview findings provided in-depth explanations that demonstrated that subject analysis theory is generally considered as important to improve efficiency. However, the level of knowledge of the theoretical basis was generally found to vary, while not always being explicit, which suggested that subject metadata creators do not apply the theory adequately.

The findings revealed a variation in the existing metadata practices for theses and dissertations and the approaches followed in subject analysis and ETD geographic subject metadata creation and the level of knowledge on the theoretical impact. Increasing knowledge and awareness of the importance and impact of theory will help to improve ETD subject metadata creation and the discoverability of ETD content.

Based on these research findings, recommendations were made for increasing knowledge and awareness of the importance and impact of theory, which will help to fill the existing knowledge gaps impeding on optimal ETD subject metadata creation and the discoverability of the ETD content. Appropriate metadata analysis practices – based on appropriate theoretical principles – will help to improve the discoverability of ETD content to the broadest user population possible.

Additionally, the research findings provided information on the support policies and guidelines that support collaboration among different role players and the coexistence of geographic subject metadata created from hybrid sources.

The theoretical approaches in the literature and the practical implications revealed in the research findings are recommended for wider implementation in related contexts, in order to guide subject analysis and the process of creating ETD geographic subject metadata.

Keywords: Subject analysis theories, electronic theses and dissertations, ETDs, boundary objects theory, subject metadata, geographic subject metadata, university libraries, ETD repositories, South African university libraries, sequential explanatory mixed methods, library metadata creators

ACKNOWLEDGEMENTS

I am grateful for the support received from the inception of this study up to the completion stage.

I want to express my gratitude to the following individuals and organisations:

- My supervisor, Dr Linda Cloete, for her guidance and patience in the production of academic work of integrity and value. Thank you for taking me through this project.

- All the research participants, my colleagues at different South African university libraries, who participated in this study. I heartily thank you for your time and willingness to share information.

- I thank all the university libraries that supported the study and the universities that gave permission to conduct the study on their premises.

- My institution, the University of Limpopo, for all the support and time that was availed to me to complete the study.

- Ms Lucia Geyer, who edited the thesis.

- The statisticians and everyone who contributed in different ways to the completion of this work.

I thank you all.

DEDICATION

All things were created by God and for Him. We are nothing without You, Lord. All glory and honour to You alone.

- To my husband, Edgar, and my sons, Khutjo, Tiishetjo, Lefa, Retang and Agang Maphopha, and my daughter-in-law, Tebogo, who remain strong and courageous in all situations. You are a strong team that kept me focused and encouraged. I can never thank you enough for all the sacrifices you made and the support you gave me. May God keep and bless you.

- My grandmother, Masoko, and my mother, Jeanette. I thank God who has kept you until this time. This achievement is my gift to you.

- My brothers, Lucky, Deon and Molefe and their families. You remain my team. Thank you for the support.

- This work is also dedicated to my departed father, Tselagale, and my brothers, Simeon and Evans, who never left my thoughts as I continued with this project.

- My in-laws from the Maphopha family. Your love and support kept me going. I heartily thank you.

- I also appreciate my church family at the Assembly of God, Seshego. I always left the services inspired to push on. May God expand your territory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

ETDs	Electronic Theses and Dissertations
FAST	Faceted Application of Subject Terminology
GSM	Geographic Subject Metadata
IFLA	International Federation for Library Associations
LCSH	Library of Congress Subject Headings
LIS	Library and Information Science
LRM	Library Reference Model
NDLTD	Networked Digital Library of Theses and Dissertations
NETD	National Electronic Thesis and Dissertation
MODS	Metadata Object Description Schema
SPSS	Statistical Package for the Social Sciences
UN	United Nations
Web	World Wide Web
WWW	World Wide Web

CHAPTER 1: INTRODUCTION

1.1   Introduction and background to the study

This study is aimed at contributing to development in the description of information resources to facilitate their discovery, based on the subject content. The focus point of the study is to improve the understanding of the approaches to ETD geographic subject metadata creation followed by the metadata creators in South African university libraries, with particular reference to the use of subject analysis theories. One form of description of information resources is assigning geographic subject metadata to represent the subject content. Information resources containing data of geographic importance need to be properly analysed to facilitate geographic subject descriptions. Such information resources may be available in different types of media, including electronic or digital format. The study focuses on subject representation in the context of electronic theses and dissertations.

According to Oh and Park (2018), substantial growth has occurred regarding publications addressing the topic of metadata. However, Terra, Lacruz, Bernades, Fujita and De la Fuente (2021:2) indicate that little has been researched on ETDs subject access in digital repositories. Earlier studies, including those of Tarver, Phillips, Zavalina and Kizakhethil (2015) and McCutcheon (2010 & 2011), highlight different subject metadata practices in university ETD repositories, and the efforts made to enhance the discoverability of the ETDs by using subject metadata. McCutcheon (2011:65) maintains that, despite the automated systems capabilities for harvesting metadata, fullest access to ETDs requires subject analysis.

Theses and dissertations are important primary sources of information that require appropriate subject representations of their content to facilitate their discovery and use. Different description elements are assigned to ETDs in digital format, including geographic subject metadata, so that they can be easily found and accessed. This study investigates the creation of geographic subject metadata, with focus on the analysis of the subject content of the ETDs. ETDs benefit from subject analysis to identify geographic topics and subtopics that are relevant to describe their content.

In this study, the quality of the subject metadata is highly significant in terms of maintaining the credibility of the descriptions and in determining whether the ETD metadata creators produce appropriate, descriptive subject metadata. Metadata creators at different university libraries commonly include the authors of theses and dissertations, information professionals and other library staff, including library assistants or library technicians. In their study, Maurer and Shakeri (2016) found that the author-supplied metadata was usable, but that the library staff contributed to improving the findability of the ETDs. Consistent with this finding, Terra et al (2021:2) and (Husic 2014:2) affirm that a common practice in most libraries is for the authors of theses and dissertations to submit subject metadata in the form of keywords, with metadata specialists enhancing their quality and usefulness by effecting enrichment or adding new subject metadata.

Library staff engage in subject analysis for enhanced subject descriptions. Subject analysis is a process that facilitates the determination of "aboutness". The existing theories of "aboutness" and the theoretical models that have been formulated in the context of Library and Information Science contribute the context of this study. It is assumed that librarians are trained to possess knowledge in subject analysis and the translation of their analysis into appropriate descriptive terms. Libraries across the world use controlled vocabularies to provide a system of maintaining consistency and quality subject descriptions. In a study conducted by Maurer and Shakeri (2016), it was found that the use of a controlled vocabulary – particularly the Library of Congress Subject Headings (LCSH) – increases the record discoverability of ETDs; improves access to the faceted catalogues; and provides a basis for future linked data environments. Controlled vocabularies are used in subject descriptions in the university ETD repositories. This study focuses on subject analysis and the creation of subject metadata among metadata creators working in South African university libraries.

Other studies, including the study conducted by Thompson, Liu, Duran and Washington (2019), address ETD metadata enhancement in university ETD repositories, arguing for the importance of quality control and the need to improve discoverability through records enhancement.

University libraries often conduct retrospective metadata clean-up projects to improve the quality of the descriptive records and to enhance the discoverability of the information resources.

In terms of consistency, Jensen and Carr (2020) indicate the need for more effective ETD descriptive records and explain some of the retrospective projects aimed at improving the physical and intellectual access to theses and dissertations. Furthermore, Flynn and Ahrberg (2020) discuss different problems regarding ETD metadata management, indicating that unique contextual factors influence workflows, practices and policy development.

The reviewed literature commonly points to the need for ETD subject descriptions of high quality. The discoverability of ETDs is improved through subject analysis to identify geographic topics and subtopics relevant to the description of their subject content. The theoretical foundation of subject analysis is important in the support of quality analysis. This study investigates the effect of theory on the effectiveness of the approaches that subject metadata creators follow in subject analysis aimed at determining the geographic subject metadata of the ETDs.

Current works relevant to the theoretical discussion conducted in this study include that of Hjorland (2021), who highlights the importance of underlying theoretical principles and updated scientific knowledge in the development of systems and processes in knowledge organisation and information retrieval. Issues of theoretical foundations are viewed as being inherent to the practices of subject metadata creation. However, they are not always explicitly discussed in the literature, which specifically focuses on ETDs. The theoretical basis as a foundation for ETD geographic subject metadata creation is investigated in this study.

1.1.1 The role of geographic subject metadata in description of ETDs

The purpose of metadata creation is to allow easy organisation and retrieval of information resources. Geographic subject metadata is useful in finding and bringing together information on a particular geographic area.

The research topics involved in ETDs are usually limited to specific geographic localities or "places". Theses and dissertations may be assigned "place" as their subject metadata. Furthermore, place names may be used as concepts to represent the subjects of ETDs (Zeng, Zumer & Salaba 2010:43). Geographic subject metadata formulated from place or jurisdiction names help to collocate different works on a specific place to facilitate the discovery of related research information contained in the ETDs.

Research on the subject of geographic subject representations can be traced in scholarly literature published over time. Works that discuss subject metadata in general also include the aspect of geographic subject metadata. The existing literature addresses the different contexts around subject metadata creation that are introduced by developments in publication formats and related information processing technology. Such studies include, among others, those of Zavalina (2014), White (2012), Lubas 2011, Chung (2006), Sauperl (1999) and Bake (1974). New research in this area may help to address the inadequacies of old approaches and practices, while filling existing information gaps on subject analysis practices. This study focuses on geographic subject metadata of theses and dissertations in electronic formats in the context of specific South African university libraries – an issue that has not been studied before.

The specific South African setting involved in this study is explained to facilitate the understanding of the discussions.

1.1.2   Overview of ETD subject metadata in South Africa

This section outlines the historical and contextual background on ETDs and the use of subject metadata to describe their content. The academic institutions from which ETDs originate make them available on their institutional repositories. Most ETDs in repositories are openly available, with a potential of satisfying varied information needs (Fox 2021:2). The information content of ETDs is often relevant to specific geographic areas. The emergence of digital formats has introduced a context characterised by an increased user base and potentially unique usage of ETD content to satisfy diverse information needs.

This study investigates ETD geographic subject metadata creation approaches followed by the metadata creators in South African university libraries, with specific reference to the use of subject analysis theories. The focus is mainly on the way in which theory is linked to practice. Tennis (2005) indicates the need to investigate the contemporary practice of subject analysis and how they apply to knowledge organisation in the networked environment. Background information for this study is explored through a detailed literature review on ETDs and subject metadata creation. The United States of America (USA) – in particular University Microfilm International (UMI) – has played a leading role in testing and implementing ETDs. An understanding of subject metadata creation in countries that are at different ETD implementation levels and that follow different management approaches provide a balanced view on subject analysis approaches and practices for metadata creation in different contextual backgrounds.

The investigations in this study are placed in the context of South Africa, with regard to developments in digitisation of theses and dissertations and accompanying metadata creation practices, including geographic subject metadata creation. The development of ETDs in South Africa is facilitated by the adoption of digitisation as a form of publishing information resources. This historical development background is discussed inSection 2.2.4 (Chapter 2). Metadata is used in South African university library repositories to describe various information resources, including the ETDs. The use of metadata in South African institutional repositories is reflected in the results and findings of studies conducted by Bangani (2018) and Matizirofa and Ramalibana (2015), which highlight the dominance of ETD collection in the institutional repositories at South African universities.

The adoption of digitisation of theses and dissertations by the different universities suggests that most ETD metadata is publicly available on the Internet. The adoption of open access policies, as reflected in the study, affirms the intent of the universities to make ETD metadata openly accessible and searchable. The metadata is also aggregated in different cooperative databases, such as the National Electronic Thesis and Dissertation (NETD) and the Networked Digital Library of Theses and Dissertations (NLDTD), which make ETD metadata widely accessible to a variety of uses.

At this stage, very little research has been conducted on ETDs in South African university libraries. As revealed by the studies that have been conducted, which are addressed in the literature reviewed in Chapters 2–4, there is no specific focus on the creation of geographic subject metadata for ETDs in university libraries. The discussions in Section 2.2.4 (Chapter 2) highlight this gap of information in the existing literature. The findings of this study will address these gaps in information.

1.1.3   Conceptualising subject analysis

Different meanings are attached to the process of subject analysis. Definitions of the concepts of *subject, subject metadata* and *subject analysis* are important to understand what the process involves. As opined by Hjorland (1997 & 2017), these concepts are defined in different ways in the Library and Information Science literature, and they are used inconsistently at times. This section outlines how these concepts are defined and used in this study, in order to enhance the conceptualisation of the subject analysis process.

1.1.3.1   The concept of "subject" in the context of ETDs

The organisation of information resources is based on adequate descriptions that are made for the purpose of easy retrieval during information searching. Subjects are used for the purpose of description and to facilitate discoverability of the information resources. The concepts used in this study are interpreted as understood in the field of Library and information Science – in information organisation in particular.

ETDs are digital information resources containing research output from different subject disciplines. ETDs are described by using different bibliographic elements, including the subject terms that aid the discovery of their subject content. The content of ETDs is commonly centred on specific places and these place names can be used as a form of subject description.

The concept "subject" is understood according to the definition given by Hjorland (1997:84) and adopted by the IFLA Working Group on the Functional Requirements for Subject Authority Records (FRSAD) (Zeng, Zumer & Salaba 2010:10), which states that, "The subject of a message is its informative potential". Subject access is facilitated by well-formulated subject descriptions, with the potential for a variety of uses.

1.1.3.2   Subject metadata

In the view of Greenberg (2005:20), there has been confusion in the field of Library and Information Science (LIS) regarding the distinction between metadata creation and cataloguing. The commonality between the two processes is that both are involved with information resource description and they follow similar steps, which begin with subject analysis.

As described by Mayernik (2020), metadata is an evolving topic that continually introduces different meanings attached to the concept. As a way of indicating the differences in meaning, Greenberg (2005:20), Rubin (2010:157) and Hider (2012:4) agree that the distinction can be drawn by commonly applying "metadata" specifically to electronic information. The common definition for *metadata* is "data about data". *Metadata* can also be defined as an aggregation of data elements used in representing an entity's important description features. This definition emphasises the importance of different elements, including "subject", which are necessary for the description of an entity. In this study, *metadata* is considered in terms of its application to the description of ETDs.

It is important to draw a distinction among different types of metadata and their uses. Metadata is commonly categorised into descriptive, administrative and structural metadata (Khoo & Hall 2013:88, Jackson 2011:96, Ma 2006:4). Ma (2006:4) explains the category of "use metadata", which is debatable whether it has anything to do with the information resource itself, as it is used to track usage of the information resource.

Different descriptive metadata elements, including subject, are used to describe information resources. This study focuses on descriptive metadata, which is also referred to as *discovery metadata* (Hider 2012:15).

Subject metadata can take various forms, based on the element used to represent the content of an information resource, which may be a place name. Place names are used to formulate geographic subject metadata.

Metadata plays a significant role in the description of digital information resources to improve their discovery (Husic 2014:1). Subject metadata requires performing subject analysis to represent what an information resource is about effectively. In this study, the theoretical approaches to subject analysis are investigated to contribute to the improvement of the process.

1.1.3.3 Subject analysis

Different definitions and conceptualisations are advanced in literature about subject analysis and what it actually involves. The varied conceptualisations reveal the complexity of the process (Hjorland 1997:39; Chan 2000:15). Human intervention is commonly practiced to conduct effective analysis. The *Online Dictionary of Library and Information Science* (1996) highlights the importance of human intellect to enhance the quality and efficiency of the process. The differences in the interpretation of the concept are important to facilitate the understanding of its meaning and proper use. These meanings are important in terms of informing the processes of resource description in the historical print era and in the current electronic or digital era. Furthermore, scholarly literature presents different perspectives on the topic of subject analysis and subject metadata creation in general.

In Library and Information Science (LIS) literature, subject analysis is presented as a process that facilitates the construction of subject representations during indexing, subject cataloguing and recently in metadata creation. Furthermore, subject analysis is commonly identified as the first stages of analysis of the subjects of information resources, followed by translation into subject representation data in a form of a preferred vocabulary. Hjorland (1997:39), Hjorland (2017:56), Mai (2005:600), Lancaster (2003:9), Albrachtsen (1993:219) and Fourie (2008:112) observe that different terms like *conceptual analysis, aboutness analysis, content analysis* and *subject analysis* are used interchangeably in the LIS field.

However, Hjorland (2017:56; 1997:43) states that clear separation must be made between the stages of conceptual analysis (when subjects are attributed to information resources) and the translation stage (when subject labels are attached to information resources). Similarly, Golub, Soergel, Buchanan, Tudhope, Lykke and Hiom (2016:6) consider the process of creating subject representations to begin with the determination of "aboutness", which is essentially concerned with deciding or conceptualising the main subject or topic of an information resource.

In this study, subject analysis is understood in terms of the stages of conceptual analysis or aboutness analysis, which is conducted by examining the document to determine its subject content and the most important concepts of the subject and, finally to express this analysis by assigning subject terms (translation) (Hjorland 1997:39). The conceptualisation (conceptual analysis or aboutness analysis) is regarded as the basis of quality subject description (Fourie 2008:112). Consistent with the definitions of the afore-mentioned authors, Hjorland (2001:776) observes that subject analysis is meant to identify what the subject is. Hjorland (1997:39) indicates that subject analysis can take the form of an intellectual or automated process that can be conducted by humans or as an automated process. Automation can occur in the form of automatic-assisted or fully automated process.

The empirical investigation in this study focuses on subject analysis that involves human intellectual intervention and how it assists high quality subject description. In addition, the role played by theory in the subject analysis and resource description process is investigated.

1.2   Theoretical foundation of the study

This section outlines the theories and models that support this study. The theoretical framework of the study is based on models and theories that are applicable to subject analysis conducted for the purpose of describing the subject contents of ETDs by creating geographic subject metadata. No theories or models could be traced in the literature that specifically combine subject analysis and the context of geographic subject metadata of electronic theses and dissertations.

However, generally accepted subject analysis theories and models are adopted and outlined to provide a basis suitable for addressing the research problem and questions of this study.

Furthermore, a synthesis of the theories of subject analysis and the view on how ETD geographical subject metadata can serve as boundary objects is proposed, so as to provide a theoretical grounding for this study. The boundary objects theory is used as a foundation to explain a recommended model that can serve as a framework of the way in which the different theoretical conceptions of subject analysis can facilitate the practical use of ETD geographic subject metadata across different contexts.

Different theoretical models are used to represent the different approaches to subject analysis. Highlighting the different models helps to explain the theoretical background of the conceptual model for subject analysis used in this study. Models facilitate the understanding of a problem area within a specific domain (Hadar & Soffer 2006:569). Many examples of theoretical and conceptual models are found in the reviewed literature. Examples of related and widely applied models that address resource description, including the use of "place" as subject, are the Functional Requirements for Bibliographic Records (FRBR) and its associated models and the consolidated FRBR Library Reference Model (LRM). The ongoing discussions in the literature on the bibliographic description models is an indication of the need to address the challenges that continue to emerge with new developments in this field.

The conceptual model of subject analysis is adopted as the basis for the study of subject analysis approaches and how they are associated with theory. The conceptions proposed by the model highlights how different orientations influence priorities and focus areas in subject analysis. Tennis (2005:107) compares different theoretical works that discuss the different conceptions of subject analysis and indicates that they have common and different attributes. The author's conclusion affirms the importance of theory in understanding the background of subject analysis practice and recommends the incorporation of empirical data to build on the subject analysis theory and to integrate it with practice.

Tennis (2005) discusses, among others, the document-oriented and user-oriented subject analysis conceptions or theories and affirms the need for expansion of subject analysis theory. These theories are explored by, among others, Golub et al (2016), Hjorland and Albrachtsen (1995) and Albrachtsen (1993).

An understanding of the different subject analysis orientations will facilitate the adoption of improved approaches to subject analysis that are relevant for describing digital works available on public platforms on the Web. Information resource relevance and use may be wider and may benefit from a broadened approach to subject analysis. ETDs metadata is publicly available on the Web and subject metadata is considered in this context.

Furthermore, the literature review in this study reveals different theoretical conceptions of subject analysis that have unfolded over time. These different views – e.g. those of Hjorland (2021), Albrachtsen (1993 & 2015), Tennis (2005) and Hjorland and Albrachtsen (1995) – help to explain the process of subject analysis from a theoretical viewpoint and to provide a basis for improving the understanding of its theoretical principles. Knowledge of the conceptual models may help metadata creators to understand the theory that informs the process of subject analysis.

An existing model of the conceptions of subject analysis was used to provide a framework for the interpretation of the research findings of this study; particularly for the interpretation of the perceptions about the appropriate focus to be maintained during subject analysis. The adopted model of conceptions of subject analysis presents different theoretical approaches that inform the practice of subject analysis. It outlines the possible approaches for deciding and identifying what to concentrate or focus on during subject analysis. This model is discussed in detail in Chapter 4.

According to the specified model, the process of subject analysis begins with determining what an information resource is about and ends when the subject representations are assigned. Literature published over time, e.g. Hjorland (2021), Albrachtsen (2015), Fourie (2008) and Tennis (2005), discusses the different analysis orientations identified in the model, indicating the different orientations on how to determine the subject content of an information resource.

While this framework serves as a basis for the discussions in this study, all the information gathered from the research results and findings of this study inform the interpretations on which subject analysis focuses.

In addition to the literature reviewed on the theoretical conceptions of subject analysis, this study explores how subject analysis and subject metadata creation are experienced and perceived in practice. The theoretical perspective on subject analysis provides a basis to examine how the process is conducted to create geographic subject metadata for ETDs in the context of South African university libraries. Theory is fundamental to the procedures and processes of subject metadata creation, although its application is not always explicit in practice. The significance of the findings of this study and the conclusions will be discussed in relation to the model to indicate how theoretical knowledge influences the practice of subject analysis.

Furthermore, subject metadata is considered as serving as boundary objects that could facilitate ETD geographic subject content across different contexts by using pooled subject analysis theoretical approaches. Therefore, the study is also inspired by the boundary object theory, advanced by Star and Griesemer (1989). This theory facilitates the discussions on the recommended model that encompasses the different theoretical approaches to subject analysis and geographic subject metadata creation and informs the focus of the interpretations and recommendations on how to manage diversity and cooperation. Subject metadata can be used across different contexts and by a broad information user community. Therefore, metadata is considered as serving as boundary objects, created by different actors and crossing the boundaries of different use interests or purposes. Associated Library and Information Science (LIS) theories and related disciplines are outlined to broaden the understanding of the theoretical background of the study.

This study is placed in the context of LIS. However, other related studies, for example information technology and systems development, do influence the theoretical and practical development of subject analysis. Subject description is an aspect of information organisation and retrieval that is grounded in the LIS discipline; particularly in bibliographic description.

The study investigates how the perspectives and practices of subject metadata creators are shaped by the general historical and current developments in information or knowledge organisation. Furthermore, subject metadata creation is based in the historical context of information resources description processes of subject cataloguing and indexing, of which the purpose is to provide access to subject content (Xie & Matusiak 2016:130; Rubin 2010:157).

1.3 Understanding the theoretical context from associated disciplines

Knowledge organisation theories or knowledge management theories inform the practices of information organisation and retrieval (Zavalina, Kizhakkhethil, Alemneh, Phillips & Tarver 2015; Zeng, Zumer & Salaba 2010:10; Rubin 2010:298). The philosophies associated with knowledge organisation can be linked to perspectives on subject analysis for the creation of geographic subject representation. Knowledge organisation principles, including Otlet's "universalist" principle, which promotes a wide, knowledge-based organisational approach, and Kaiser's "localism" principle, which emphasises a context-based perspective on knowledge organisation, are discussed by Dousa (2010) and Dousa and Ibekwe-Sanjuan (2014). Convergence of these principles and their ideas may help to strike a balance in metadata creation in the current Web context, where resources like ETDs are accessible to the wider public, with a potential variety of uses.

This study is also placed within the background of social informatics, which deals with the description of information resources, the purpose of retrieval and the related technologies influencing them, considering their institutional and cultural context (Rubin 2010:295). The theory explains the sociotechnical context within which people process and use information. Furthermore, it highlights the interaction and social context of information technology. The manner in which information technology is used and its social impact is relevant to the study of developments in information description.

Knowledge organisation, changing user needs and information resource usage patterns are fundamentally technology-driven. Viewed in the same light, the emergence of ETDs was driven by digitisation technology, transforming the interaction of users with the content of these valuable research items. The developments facilitated the availability of the information contained in the theses and dissertations, across social boundaries, enabling cross-domains information use and sharing.

Technology permits metadata processes and functions across social boundaries. Continuous assessment and improvement of ETD descriptions is required to cater for the changing information use landscape, e.g. the increasing multidisciplinarity of research and changing user behaviours. The electronic information and its Web-based potential is grounded in social informatics. The theories within Knowledge Organisation and Informatics, as applied to Information Science, give perspective on the representation of knowledge, in order to improve its use (Zeng, Zumer & Salaba 2010; Rubin 2010:294 & 299). They also provide a contextual basis for metadata creation. These fields, in turn, draw from other social disciplines, including Psychology, Sociology, Computer Science and Communication (Hjorland 2021; Rubin 2010:272 & 299). The theoretical context and the views on related theories point to theory developments that are related and signify advances that justify the need for this study to promote appropriate subject analysis in the era of digital theses and dissertations and the potential of improving the management and accessibility of these resources.

1.4   Problem of the study

The scholarly literature shows evidence of ongoing efforts to study the way in which the representation of geographic place names on information discovery platforms can be improved. Various IFLA reports cited in this study and the research of scholars like Zavalina (2014:87) serve as evidence of such studies conducted over time. Existing geographic subject metadata for electronic thesis and dissertations (ETDs) that is publicly available on various databases on the Web indicate the possibility of different approaches to geographic subject metadata creation. Furthermore, inconsistencies in the approaches adopted to create geographic subject metadata for ETDs is evident.

This study investigates the problem of how the subject metadata creators conduct the analysis of the subject contents of ETDs to determine geographic subject metadata; particularly on how subject analysis theories are applied and their role in subject description. Some ETDs in the various databases clearly reflect geographical focus in their titles and abstracts, but their subject metadata either lacks specificity in displaying the geographical element, or the geographic subject metadata is completely missing.

Despite the availability of common established standards – e.g. the Library of Congress Subject Headings (LCSH) – which reflect when and how geographic headings and subdivisions are to be used, there is no application consistency of the standards on the records of different ETD repositories in South African university libraries and on the metadata hosted on collaborative platforms, such as the NETD and the NDLTD.

Metadata helps to improve the discovery of ETDs. Information resources description is a continuously evolving practice that is shaped by the changing context within which it occurs. This situation necessitates continuous review of the practice for effectiveness within new or different contexts. As part of the description strategies, it is important to decide what metadata will be necessary and also how it should be constructed to maintain quality and integrity. Descriptive subject metadata of all types, including geographic, also need to be carefully implemented. One of the considerations in the creation of metadata should be how to conduct subject analysis, based on relevant theories. The university library environment defines the academically inclined metadata practices. Such contexts may facilitate or obstruct the appropriate application of subject analysis theories. It is necessary to establish how the context affects the application of theories during creation of ETD subject metadata for geographic places.

There are various factors in a particular context that may affect metadata creation. Rubin (2010:135) indicates that a contextual factor like bias influences subject analysis. Subject metadata created to represent the ETD subject content commonly has an academic bias, which may affect its usefulness to the wider public.

As observed by Husic (2014:2 & 8). the practice of depending on authors to supply keywords, which is employed by most libraries, may be a limiting factor in terms of insufficient knowledge of metadata creation, which may confine the metadata usefulness within a specific academic discipline.

In the construction of subject metadata, analysis is a crucial stage, in that it determines the forms and appropriateness of representations (Wolverton, Hoover & Fowler 2011:215; Mai 2005:601). Several factors can be attributed to the quality of subject analysis approaches and practices, but in this study, the focus is on the part played by subject analysis theories.

The application of subject analysis theories is assumed to be intrinsic within the approaches used for metadata creation. Theories are useful in generating ideas and knowledge to guide practice. Careful investigation of metadata creation approaches and practices may help to determine if metadata creators do have knowledge of the theories and whether they are applied in practice.

There are persistent problems in relation to the adequacy of subject representations (Diao & Hernandez 2014:131). These problems necessitate a comprehensive investigation into the process of subject analysis – particularly the application of its theoretical principles, and how it contributes to effective subject metadata creation for ETDs.

The discussed challenges suggest an inadequate approach to subject metadata creation. A shift of focus in metadata creation is necessary to accommodate the current broader information user community introduced by digitisation and the Web. Research must also address the theory gaps to provide support for the implementation of new approaches. Metadata creators can be enabled to adapt and operate efficiently by following suitable theories that address the current trends. The research findings of this study may help to establish how the current situation can be addressed, in order to accommodate the variety of user needs beyond the contexts within which the metadata is created.

This study is motivated by the need to establish how the approaches – mainly how the application of subject analysis theories – are connected to geographic subject metadata creation. A phenomenological approach to the problem is adopted, which allows a deep understanding of the experiences and perceptions of the metadata creators in the context of South African university libraries and the use of subject analysis theories to facilitate geographic subject metadata creation.

Additionally, the role played by individual and institutional factors in subject metadata creation practices is examined, as they are considered to impact on subject metadata creation. Furthermore, it is also considered significant to analyse the existing metadata records and to report objectively on the evident metadata creation approaches and practices as a basis for addressing the research problem.

1.5   Research questions

The research questions provide focus for the research in terms of the areas of interest the research will investigate and attempt to answer (Collis & Hussey 2014:103). The following research questions are derived from the research problem of this study:

1. To what extent and in what ways is geographic subject metadata used for description of ETDs in South African university libraries?
2. What are the current approaches being followed to create geographic subject metadata for ETDs in South African university libraries?
3. How and why do the current analysis approaches being followed during geographic metadata creation for ETDs in South African university libraries affect the process?
4. How can the findings of this study be applied to develop a model for subject analysis and creation of geographic subject metadata for ETDs in South African university libraries?

1.6   Research purpose and objectives

1.6.1   Research purpose

The aim of this study is to investigate the approaches that are being followed during subject analysis and how subject analysis theories facilitate geographic subject metadata creation for ETDs within the South African university libraries. The influence of the theoretical foundation and the factors that create context for the different approaches are analysed. Based on the research findings, a model is recommended for a subject analysis approach that could potentially contribute to meaningful theses and dissertations content representation practices among metadata creators – particularly ETD geographic subject metadata.

Information resources – including ETDs – can be used to satisfy a wide range of information needs. This study addresses the theoretical and practical aspects of the different subject analysis approaches, which are important in terms of helping to create metadata that can improve discoverability of information resources relevant to the different information needs. According to Rubin (2010:271), LIS professionals' perceptions of their roles may need to change to align with developments in information retrieval and use. Metadata creation perceptions and practices also have to be adapted to changes occurring in digital publishing. Understanding the theoretical foundation of subject analysis may provide a common understanding and a framework for developments in ETD geographic subject metadata creation within the context of South African University libraries and the changing perspectives on information and its use.

1.6.2   Research objectives

The main objective of this study is to collect data from metadata creators on the subject analysis approaches; particularly the application of subject analysis theories, adopted during geographic subject metadata creation for ETDs, and to understand how they influence practice. Mouton (2001:122) indicates that the state of existing knowledge of a phenomenon may help to provide a basis to determine the objectives. The objectives of this study are determined, based on the knowledge gathered from the metadata creation and subject analysis literature.

The research objectives are to:

1. Investigate the application of geographic subject representations during the creation of ETD metadata in South African university libraries;
2. Collect data from the practitioners involved in metadata creation in South African university libraries on how they approach the determination of geographic subject metadata for ETDs;
3. Establish from the subject metadata creators in South African university libraries what the implications are of their different analysis approaches for the determination of geographic subject metadata for ETDs; and
4. Establish what type of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries.

1.7 Scope and limitation of the study

This study is limited to the context of metadata creation in South African university libraries. Although there are different forms of subject metadata, this study focuses on geographic subject metadata created for ETDs.

ETDs are the only type of information resource considered in this study. In the university library environment, there is a general consensus towards making theses and dissertations available in electronic format. ETDs are easy to manage and they support the open access availability of information, which is globally endorsed to make information easily and freely available. More research on enhancing access to the information and research results and findings in ETDs is essential.

This study is limited to the context of metadata creators working in university libraries in South Africa. An extensive literature review is conducted to study metadata creation approaches and practices in other countries, thereby providing a basis for the research. This study is aimed at investigating whether the current subject analysis approaches in South African university libraries – particularly the application of subject analysis theories – facilitate the creation of subject metadata that fulfils the purpose of discovery and access in all instances where they are used. The results are anticipated to be applicable to most situations in university libraries with regard to subject metadata creation.

Subject analysis leads to the creation of a specific type of metadata – the subject metadata – which is the specific type of metadata to be investigated in this study. Other types of metadata that do not serve the subject description purpose are excluded from the study. The different types of metadata are explained in Chapter 3.

Various descriptive metadata elements, other than the subject element, are excluded from the analysis of ETD records for the content analysis. In addition, subject metadata that does not serve the geographic description purpose does not form part of this study.

1.8   Significance of the research

The importance of this study lies in its potential contribution to closing the existing knowledge gap in subject analysis theoretical foundation and improvement of current practices. The research findings will aid the development of a model to address the current challenges involved in the creation of geographic subject metadata for ETDs. The findings are aimed at improving the subject analysis process by facilitating the application of subject analysis theories.

In addition, the study is expected to raise awareness among subject metadata creators about the role of subject analysis theories in guiding subject metadata creation for information resources, including ETDs. The practical significance of this research will manifest if the proposed approaches enhance subject analysis practices and ETD geographic subject metadata creation, which, in turn, will significantly improve the discovery of ETD geographic subject content.

Ongoing research in the LIS field is necessary to continue building on its theoretical foundations. Additionally, ongoing studies are necessary to establish ways to develop the theory to keep up with developments affecting LIS practices and to maintain relevance in the different contexts.

This study is considered beneficial in terms of reviewing and redefining the practice of subject description, in order to remain relevant in the current information society, where people depend on information to solve their daily problems.

The societal information needs evolve with developments in technology and publication formats. Information sources that are available on the Web, including theses and dissertations, present an opportunity for the broader user community to benefit from the information in various ways.

There is continued interest in the use of research information, some of which is contained in ETDs, for the development and improvement of human lives. The findings of this study have the potential to contribute to the enhancement of subject descriptions that will make information of geographic relevance contained in ETDs easy to find by the different types of users beyond the academic sector.

Geographic information has the potential to address current societal challenges and improve people's lives. Current areas of interest in the use of geographic information include land redistribution, geographical boundaries, emigration and genealogical studies, which have created interest in the use of geographic information. Therefore, a shift of focus is required in the creation of geographic subject metadata, so as to remain relevant to the current context presented by these social needs, coupled with the advances in technology. Careful description of theses and dissertations will assist with the discovery of the important content of geographic nature.

Yasser (2012:373) affirms the importance and the increased role played by metadata in the management, retrieval and dissemination of information as online collections continue to grow. The foregoing study recommends further research in metadata creation. Husic (2014:1), Buckland (2012:155) and Sauperl (2004:56) indicate the prevalence of literature that discusses other aspects of metadata creation, e.g. alternative metadata schemes, over those focusing on descriptive metadata and subject analysis and the existing information gap on the later aspects.

This study presents new empirical data focused on the importance of descriptive subject metadata and subject analysis practice; particularly in the current ETD context discussed above. Tarver et al (2015) point out that, by 2015, little research had been conducted on subject metadata in digital libraries. Later studies focusing on metadata are evident with more empirical research conducted and published in different platforms, including the *Journal of Metadata*.

Studies that include subject metadata are, among others, those conducted by Terra et al (2021), Ho and Stokes (2019), Steele and Sump-Crethar (2016) and Tarver et al (2015).

This study involves a contribution focusing on one type of metadata – i.e. geographic subject metadata – and it was conducted comparatively across different university libraries in South Africa. The ETD content can be described through this type of metadata to facilitate subject access.

It is important to determine how ETDs are analysed, so as to determine the geographic elements that will serve as subject metadata. The findings about the subject analysis practices will inform the recommendations made on the proposed model for subject analysis and ETD geographic subject metadata creation.

Analysis that is conducted for the purpose of subject metadata creation is based on subject analysis foundational theories and, therefore, the theories have a bearing on the nature of the subject descriptions. Tennis (2005:14) remarks that ambiguities of theoretical understanding of subject analysis theories affect practice. Through its findings, this study attempts to contribute to knowledge of analysis theories and to promote metadata approaches and practices that yield improved ETD geographic subject metadata. Metadata creators can apply the information to produce quality subject representations that will maximise recall, while aiding specificity during information searching. Quality and precision in description is maintained through the use of knowledgeable contributors to information resources description (Earle 2014:10).

This brief overview confirms the general consensus in the literature on the importance of subject access and geographic subject metadata and the importance of subject analysis theory, which is the focus of this study. This study will provide information that will support informed decisions in the future of subject analysis and subject metadata creation.

In addition, the study is aimed at raising awareness among subject metadata creators of the role of subject analysis theories in guiding subject metadata creation for information resources, including ETDs. Uncertainties can be reduced and consistency can be maintained to improve the effectiveness of ETD geographic subject metadata.

1.9   Originality of the research

Previous studies provide a background for ETD management and metadata creation, whereas this study fills the gap of areas of ETD metadata, which has not been researched before. Subject analysis is not a new concept in the LIS field, but the recommendations will be made with relevance to the current ETD environment and geographic subject metadata creation, based on the observed practices and the understanding gathered from the collected primary data.

In order to maintain quality and integrity, it is important to understand who creates and manages the metadata for ETDs. This study is an original contribution that uses different methods to provide in-depth information on the different ways of contributing ETDs subject metadata in South African university libraries. Metadata creators in different university libraries commonly include the theses and dissertations authors, information professionals and other library staff, including library assistants or library technicians. This study focuses on the metadata creators functioning within university library environments.

According to Terra et al (2021:2) and (Husic 2014:2), a common practice in most libraries is for the authors of theses and dissertations to submit subject metadata in the form of keywords, with metadata specialists enhancing their quality and usefulness by effecting enrichment or creating new subject metadata.

The reviewed literature has not yielded any work that is precisely the same or similar to this study, or that directly matches the approach and scope of this study. Related studies include those of Husic (2014), White (2012), Wolverton, Hoover and Fowler (2011), Zavalina (2010), International Federation of Library Associations and Institutions.

Other related studies include those of the Working Group on Metadata for Digital Resources (2009), Zeng, Zumer and Salaba (2010), Chung (2006), Tennis (2005), Shubert (1996) and Bates (1986). These studies address issues on or related to metadata creation and subject analysis. Prior studies focus on various issues related to this study, in that they uniquely cover different metadata-related issues and different geographic areas, while using different research approaches.

Existing research on ETDs and geographic subject metadata issues are commonly focused on areas outside South Africa. For example, Sauperl (2005) conducted a subject cataloguing study among Slovenian librarians, comparing the results to those of a study on subject determination and subject analysis conducted among American librarians.

Zavalina (2010) conducted a study on subject metadata for digital collections, which focuses on the USA and Europe. It investigated subject access in aggregations of digital collections. The study also looks into geographic subject access as part of its investigations (Zavalina 2010:146). The recommendations to extend research on ETDs and how they are handled beyond Europe and the United States of America (USA) was made by Wolverton and Hoover (2004:14) and Zavalina (2010:149). In the study that they conducted, Steele and Sump-Crethar (2016:67) also propose further studies on ETD metadata in other countries.

Evidence of a study that aligns to the recommendation for studies outside the USA is that of Al Salmi (2014), who conducted a study on ETDs in the Arab Gulf States; particularly on factors affecting the adoption and implementation of ETDs. The findings of this study indicate that factors like the availability of resources and administrative and financial support may affect the adoption and implementation of ETDs, either positively or negatively (Al Salmi 2014:235). This is affirmed by Steele and Sump-Crethar (2016:68), who indicate the effect of these factors on the quality of existing metadata. These issues also impact on metadata implementation and quality, as discussed in detail in Chapter 3.

In Africa, a related study conducted by Ezema and Igbo (2016) proposes strategies for the development and adoption of ETDs, which includes development of policies, formation of ETD teams, digitisation of print theses and dissertations, efficient collection procedures, efficient ICT infrastructure, electronic submissions, creating awareness of ETDs and training of staff. These issues indicate the basic nature of research that still needs to be done regarding ETDs.

A survey conducted by Matizirofa and Ramalibana (2015) reveals the state of institutional libraries and ETDs in South Africa. The biggest collections of institutional repositories in South African universities are the ETDs. However, there is no evidence of literature that specifically addresses the issues of ETD subject content access in South Africa. This study focuses on this issue in the South African context.

The originality of this study also lies in its research methodology. The methodologies used in studies on ETDs and metadata vary. Different research conducted over time has led to varied conclusions on ETD-related issues. Wolverton and Hoover (2004) report on research covering wide periods, spanning across several decades, from the 1970s to 2000s, with good size populations and good response rates (i.e. higher than 60%). Most studies use the survey method to collect primary data on theses and dissertations, including ETDs and their management practices (Wolverton & Hoover 2004:1). The results of these empirical works are compared to the findings of this study.

Zavalina (2010) conducted a study by following concurrent mixed method research, employing content analysis, transaction logs, interviews and observation data, which confirms the importance of collection-level subject access. The mentioned study recommended for further studies on issues of subject access. Consistency was maintained in a study conducted by Zavalina (2014). The study, which involved a qualitative content analysis of the metadata of three digital collections, revealed the importance of subject metadata, including the geographic description element.

In contrast, Al Salmi (2014) employed qualitative semi-structured interviews and the quantitative survey by using questionnaires to study the factors affecting the implementation of ETDs. The findings of this study indicated that the availability of resources and administrative and financial support may affect the adoption and implementation of ETDs, either positively or negatively (Al Salmi 2014:235).

Several studies, conducted over an extensive period, which signify the evolution of ETDs and the associated trends in metadata creation, used different research approaches to the one employed in this study. The research design applied in this study – particularly in the study of subject metadata – is considered to be original in its application. The multiple methods used allowed for the collection of new data that was combined in various ways to discover new factors that influence subject metadata creation, which forms part of the original findings of this study.

Theoretical principles provide a foundation for appropriate subject metadata creation and to improve the discoverability of the intellectual content of information resources. Knowledge on the theoretical basis for subject analysis and ETD subject metadata creation in South African University libraries has not been studied before, which was examined in this study.

This study makes a unique contribution that focuses on geographic subject metadata. There are no similar studies conducted comparatively and across different university libraries in South Africa. Furthermore, this study makes an original contribution through the recommendation of a model that serves as a basis for improved approaches to ETD subject analysis in South African university libraries, with the aim of addressing the problem of effective geographic subject metadata creation. The combination of a specific type of subject metadata (geographic) and specific resource type (ETDs) in the South African context is expected to yield new insight into ETD subject metadata creation.

1.10   Research methodology

Research methodology is defined as the overall approach to the research process that encompasses the body or group of research methods suitable to accomplish the purpose of the research (Collis & Hussey 2014:10, Henning, Van Rensburg & Smith 2004:36). The pragmatic research paradigm provides the philosophical framework for this study. Pickard (2013:325) emphasises the relationship between a research paradigm and a research approach.

The reviewed literature indicates how the mixed methods approach combines the elements of the quantitative and qualitative research approaches (Collis & Hussey 2014:5; Leedy & Ormrod 2013:95; Teddlie & Tashakkori 2009:97). The two approaches are combined to address the research questions that could not be sufficiently addressed through only one of the approaches (Creswell 2014:11), which indicates that the choice of research methodologies is based on what is efficient at a particular time.

The mixed methods research, employing both quantitative and qualitative approaches, was considered as ideal for addressing the research problem in depth. The research objectives involved in this study required both qualitative and quantitative approaches to provide comprehensive answers. Therefore, the mixed methods research approach was employed to gain a comprehensive understanding of the complexities of the process of subject analysis, its theoretical basis and its application during the creation of the ETD geographic subject metadata. Furthermore, the pragmatic approach allows the discretionary integration of the qualitative and quantitative approaches in an explanatory sequential mixed method design that would work best for this study. Creswell (2014:224) explains this design as involving the collection of quantitative data in the first phase and using the results to plan for the second, qualitative phase, in order to achieve the practicality of using the interviews findings to explain the quantitative findings further. Further discussions on the reasons for using the mixed methods approach and the explanatory sequential mixed method design used in this study are given in Chapter 5. Pragmatism supports the practical use of the questionnaire survey, interviews and content analysis methods in a single study.

The data for the quantitative part of the study was gathered using a questionnaire, which asked about the perspectives of the ETD subject metadata creators about Subject analysis and subject metadata creation.

A structured questionnaire was used, which was convenient for extensive data collection (Creswell & Plano Clark 2011:177) that was planned for the survey. The advantages of using the instrument were considered, because its design allows easy analysis. As indicated by Collis and Hussey (2014:213), the range of potential answers from which respondents can choose is limited and can be coded in advance.

Semi-structured interviews were used to collect data on the librarians' experiences and perspectives as a follow-up to the findings of the quantitative study. Semi-structures interviews start with pre-set themes to guide the process (Saunders, Lewis & Thornhill 2016:727), thereby allowing the researcher the flexibility to vary the order in which the questions were asked and to ask new questions in the course of the interview, whenever necessary.

Table 1.1 presents a summary of the research design of this study.

Table 1.1: Research design summary

Elements of the research design	Description of the methods
Research approach	Mixed methods: Quantitative + Qualitative
Research design	Explanatory sequential mixed methods
Population and sampling	- Entire population of subject metadata creators in university libraries in South Africa for the questionnaire survey - Purposive sample for the interviews and the ETDs records for content analysis
Data collection methods	- Questionnaire - Interviews - Content analysis
Research instruments	- Questionnaire - Semi-structured interviews

Elements of the research design	Description of the methods
	• Structured observation or structured record reviews
Data analysis	• Quantitative and qualitative analysis • Questionnaire: descriptive statistics by means of SPSS • Semi-structured interviews: NVivo • Analysis of metadata records: Excel analysis to determine frequencies of manually identified characteristics and thematic discussions.
Credibility: reliability and validity checks	Validity and reliability of the quantitative questionnaire and trustworthiness of the qualitative interviews and the content analysis

The quantitative data was analysed deductively, whereas the qualitative part employed inductive analysis to collect in-depth information on the research problem.

Ethical considerations were observed in the course of this study, so as to ensure that the privacy rights of the participants were not infringed upon, and that the collected primary data was kept confidential. The researcher complied with the following ethical considerations: the principles of research ethics for quantitative and qualitative research; the *Policy on research ethics* and ethical clearance of the University of South Africa (UNISA); and permissions and ethical clearance from the universities included in this study. The research methodology is discussed in detail in Chapter 5.

1.11   Definition of terms

The section presents the terminology used in this study and the understanding of their meaning with relevance to their use in the current study. Leedy and Ormrod (2013:43) emphasise the importance of providing meaning in relevance of the context of the research problem and the sub-problems.

1.11.1   Information resource

An *information resource* can be defined as any entity, electronic or otherwise, that is capable of conveying or supporting intelligence or knowledge (Dublin Core Metadata Initiative 2018). The digital era has introduced developments in the formats through which information resources are made available. These changes have also affected the theses and dissertations, which have evolved from print to digital formats that are accessible on the Internet.

1.11.2   Electronic theses and dissertations (ETDs)

ETDs are theses and dissertations – normally Masters dissertations and doctoral theses – published in electronic format, either submitted by authors as "born digital", or converted from print to electronic format, suitable for electronic storage and distribution. Technology has enabled wider access to the ETDs than was possible with the print versions (Wolverton, Hoover & Fowler 2011:202). This has increased their potential for use in a variety of ways.

1.11.3   Metadata

Metadata is defined differently, based on the context in which it is used. Metadata includes data associated with either an information system or an information object for purposes of description, administration, legal requirements, technical functionalities, use and usage and preservation (Dublin Core Metadata Initiative 2018). From an information organisation perspective, *metadata* can be defined as an aggregation of data elements used in representing the entity's important description features. An *entity* refers to an information resource, which, in the context of this study, are ETDs. Metadata elements, including the subject element for schema like the Dublin Core, were designed for better discovery of electronic resources on the Web (Mayernik 2020:697; Dublin Core Metadata Initiative 2018).

1.11.4   Descriptive metadata

Descriptive metadata is used to describe or identify information resources (Ma 2006:4). Metadata of descriptive nature is created to facilitate discovery and access of the information resources.

Hider (2012:5) indicates that descriptive metadata can also be named *discovery metadata*. Soergel (2009:24) regards this type of metadata as serving the purpose of discovering and using information objects.

1.11.5   Subject metadata

Subject metadata can be defined as information concerning what the resource is about and what it is relevant for (Zavalina 2014:78; Soergel 2009:25). This definition is in line with the definition of a *subject*, given by Hjorland (1997:84) and adopted by the IFLA FRSAD workgroup (Zeng, Zumer & Salaba 2010:10), which states that: "The subject of a message is its informative potential". Controlled standards, like the commonly used Library of Congress Subject Headings (LCSH), can be used to assign subject metadata (Husic 2014:1). Subject metadata for various forms of publications, including electronic information resources, can also be derived from natural language. However, discussions are ongoing in favour and against both approaches. Zavalina (2012:141) indicates the role played by subject metadata in promoting subject access to information resources.

1.11.6   Geographic subject metadata

Geographic subject metadata is metadata that is useful in finding and bringing together information on a specific locality. The "places" may be used as concepts to represent the subjects of information resources (Zeng, Zumer & Salaba 2010:6). The place names can be used to formulate geographic subject metadata.

1.11.7   Subject analysis

Hjorland (1997:39) defines subject analysis as follows:

> *Subject analysis is the intellectual or automated process of analysing the subjects of a document (or information resources) and the subsequent expression of this analysis as subject representation data*

The *Online Dictionary of Library and Information Science* (1996) defines *subject analysis* as the "Examination of bibliographic item by a trained subject specialist to determine the most specific subject headings or descriptors that fully describe its contents ...".

1.11.8   Subject analysis theories

Subject analysis theories are information organisation theories that guide and lay a general philosophical foundation for subject analysis.

1.11.9   Subject metadata creators

Subject metadata creators are information practitioners and other participants, who are involved in assigning subject content representations that facilitate the subject description of information resources. The assigned subject representations constitute part of the descriptive information of an information resource that promotes effective discovery and retrieval. According to Hjorland (1997:55) and Moulaison and Dykas (2016), the information professionals provide value-added subject descriptions.

1.11.10   University libraries

In the context of this study, *university libraries* are defined as all the public universities in South Africa that fall within the three categories of traditional universities, comprehensive universities and universities of technology.

1.12   Organisation of chapters

This section outlines the structure of the thesis and gives an overview of the main points discussed in the different chapters of this study. Detailed discussions of the main themes of this study are presented in the different chapters.

Chapter 1

The first chapter introduces the study and discusses the background to this study, in order to provide background knowledge of the research problem in a thematic format.

Chapter 2

Chapter 2 presents a critical analysis of the reviewed literature to provide background knowledge about the history of ETDs and ETD geographic subject access. The chapter is divided into the following sub-themes: historical development of ETDs; ETD development in South Africa; developments affecting geographic subject access for ETDs; the importance of geographic subject access; the applicability of geographic

subject access within different contexts; and geographic subject access in the context of ETDs.

Chapter 3

The third chapter discusses metadata quality and standards. Measures of quality control and the use of standards and schema are explained.

Chapter 4

Chapter 4 presents a detailed discussion of subject analysis theories. Basic knowledge theories and specific theories on subject analysis from the field of Library and Information Science are presented to serve as the foundation for this study. Furthermore, the chapter outlines the model of conceptions of subject analysis that is adopted as a framework for the discussions of the subject analysis orientations.

Chapter 5

Chapter 5 discusses the research design and the chosen research methodology for the study. The pragmatic research paradigm that informs this study is outlined. The general strategy of this study, which includes the planning and overall structure for the procedure, is also presented. The mixed methods approach is explained in detail to clarify how the questionnaires, interviews and the content analysis are employed in this study.

Chapter 6

Chapter 6 presents the results of the data collected by using the questionnaires, the interviews and the content analysis. The types of data collected, which are the responses to questionnaire, interviews and the metadata subject entries on ETDs records are outlined. The quantitative and qualitative data analysis techniques and procedures are explained in detail. A summary of the types of data and how they were collected and analysed is presented.

Chapter 7

Chapter 7 presents the discussions that outline the findings of this study in detail and provides the interpretation of the data. The findings from the questionnaires, the interviews and the content analysis phases are interpreted and discussed. Furthermore, the findings are compared with previous findings from literature for the interpretations.

Chapter 8

Chapter 8 presents the model that is recommended to help enhance geographic subject metadata creation for ETDs, based on the reviewed literature and the findings of this study.

Chapter 9

As the last chapter, Chapter 9 presents the summary of the findings, and the final conclusions and the recommendations.

1.13   Chapter conclusion

The first chapter introduced the current study and discussed the background to the problem of geographic subject metadata creation for ETDs – particularly the application of subject analysis theories. A relationship among the concepts of geographic subject metadata, subject analysis theories, subject metadata creation and ETDs and how they impact on the discovery and access of ETDs were explained. The chapter introduced the argument on the relevance of grounding geographic subject metadata creation on an appropriate theoretical foundation. The chapter also provided a brief context from which the research questions are derived. The research purpose, objectives and questions, scope and limitations of the study and the significance of the research were also addressed.

The next chapter presents a detailed literature review to provide a framework for the study of ETDs and geographical subject metadata.

CHAPTER 2: ETDs AND GEOGRAPHIC SUBJECT ACCESS

2.1 Introduction

This chapter provides a synthesis of the existing literature relevant to the research problem involved in this study, stated in Section 1.2 (Chapter 1): To investigate how information resource content is analysed by subject metadata creators to determine geographic subject metadata for electronic theses and dissertations, particularly how subject analysis theories are applied and the role that they play in subject description. ETDs and geographic subject metadata form part of the key elements of the research problem. A review of the literature on the different viewpoints on these contextual aspects is conducted under the following topics: (i) the development of ETDs; and (ii) ETDs and geographic subject access.

The discussion in this chapter forms the basis for the next two chapters that present the literature review (Chapters 3 and 4), which involves the metadata quality issues and the theories and conceptions of subject analysis. Further background information on the investigation of the research problem is provided, as well as information on the groundwork for contextualising the investigations into the subject analysis approaches, as set out in the following research objectives:

1. To investigate the application of geographic subject representations during the creation of ETD metadata in South African university libraries;
2. To collect data from the practitioners involved in metadata creation in South African university libraries on how they approach the determination of geographic subject metadata for the ETDs in South African university libraries;
3. To establish from the subject metadata creators in South African university libraries what the implications are of their different analysis approaches for the determination of geographic subject metadata for ETDs; and
4. To establish what type of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries.

ETDs and geographic subject metadata are the common fundamentals involved in all these research objectives.

Further background information to the investigation of the problem is provided and groundwork to contextualise the investigations into the subject analysis approaches, as set out in RO1 and RO3. Subject analysis approaches to the ETDs for the purpose of geographic subject metadata creation are further discussed in Chapter 4. An understanding of the development of the ETDs and how subject access to their contents is enabled – particularly in the form of geographic subject access – is important for this study. This understanding facilitates:

- The study of ETDs as a specific type of publication format; and
- The study of geographic subject representations as a specific type of subject form that can assist the discoverability of information contained in ETDs.

The literature on existing studies and their findings on ETDs and subject access is reviewed, in order to obtain insight into the progress of research in this field, and it impacts on this study. A critical review of the literature on ETDs and geographic subject access provides a balanced approach to subject analysis in theory and practice.

2.2   Development of ETDs

As defined in Section 1.4.1 (Chapter 1), ETDs are theses and dissertations in electronic format, either submitted by authors as "born digital" or converted from print to electronic format, suitable for electronic storage and distribution. Wolverton, Hoover and Fowler (2008a:6) and Angadi, Bandi and Koganuramath (2012:886) refer to an *ETD* as a "digital dissertation". There is consensus in the literature in terms of linking ETDs with the general development of digital information resources and the services of digital libraries. ETDs repositories form part of the digital revolution in libraries. In addition, Ubogu (2002:5) and Angadi, Bandi and Koganuramath (2012:887) point to the building of digital libraries of ETDs and that ETDs are facilitated by digital technologies.

This study does not attempt to discuss digital libraries in detail, but a definition is considered necessary for understanding the context of the development of ETDs and how they can be made discoverable through their metadata.

A *digital library* is defined as:

> *An online collection of digital objects, of assured quality, that are created or collected according to internationally managed principles for development and made accessible in a coherent and sustainable manner, supported by services necessary to allow users to retrieve and exploit the resources* (International Federation of Library Associations and Institutions 2013).

As stated by (Yiotis 2008:106), the establishment of an ETDs repository is made easier if a digital library already exists. In consensus with the last part of the foregoing definition on the facilitation of retrieval and exploitation of resources, Panage and Bonde (2016:36) argue that the two important aspects of the content of a digital library are data and metadata that describe the various aspects of data. ETDs subject metadata is meant to fulfil the purpose of retrieval and exploitation of these digital resources. To fulfil this purpose effectively, the intellectual analysis of the subject content of ETDs is required, because they are complex digital resources.

As a special format, ETDs are considered as having complex intellectual content (Beuhler 2013:164). Several authors, including Wolverton, Hoover and Fowler (2008b:44) and Beuhler (2013:164), point out that the new technologically enabled facilities support the inclusion of multimedia and other enhancements in ETDs and they introduce further complexity to ETD bibliographic control issues. As a result, ETDs have unique bibliographic control problems. Such non-traditional formats and features within or accompanying the ETDs contain important information and require careful description. The research problem of this study also involved looking into the way in which subject analysis approaches and metadata practices consider such unique ETD bibliographic control issues. Further views on ETDs description practices are discussed in detail in Chapter 3.

2.2.1   Historical development

ETDs developed from traditional print theses and dissertations. Their emergence can be associated with the general information resources shift from print to electronic formats. According to Cox (2015:20), this general shift from print to electronic information resources in academic libraries was prompted by the introduction of technology and other contextual factors unique to specific environments, e.g. budget reductions. Major periods and events in the global history of ETD development, as set out in the literature that was reviewed and the websites of the contributing organisations, are outlined in Table 2.1.

Table 2.1: Major ETD events

Events	Time period
The first transition from print theses and dissertations to microform	1938
The University Microfilms international (UMI)	
The UMI theses and dissertations became digitally available with Proquest.	1997
Concurrent to the above Proquest move, Virginia Tech and its partner libraries started working on the National Digital Library of theses and dissertations (NLDTD).	1996
Virginia Tech began requiring ETDs.	1997
The Database for African theses and dissertation (DATAD Online)	Initiated in 1998 and launched in 2003
South African National Electronic Theses and Dissertations (NETD)	Development began 2009 and implementation occurred in 2011.

Theses and dissertations were traditionally, and in some instances still are, available in print as a prerequisite for Master's or doctoral graduation. According to Beuhler (2013:164), an ETD is a born digital departure from the written and hard bound version, a scanned and digitised paper thesis or dissertation is also considered an ETD. University Microfilms International (UMI) played a foundational role in the development of the ETDs in the late 1930s (Fox 2021:4; Proquest 2017).

The first transition from print theses and dissertations can be traced to 1938, when UMI made the works of the British Museum available in microfilm format, later also working with dissertations (Proquest 2017). According to Beuhler (2013:167), the path for openness to other types of research was influenced by publication trends in graduate work (theses and dissertations). The UMI collection of theses and dissertations became digitally available with Proquest in 1997 (Beuhler 2013:167). Virginia Tech is also considered a worldwide leader in the development of ETDs for more than 20 years (Virginia Tech 2017).

Fox, McMillan and Eaton (1999:1) and Virginia Tech (2017) indicate that Virginia Tech began requiring ETDs in 1997. Furthermore, Mao (2004:6) and Maurer, McCutcheon and Schwing (2011:281) indicate that a concurrent move to the Proquest open access platform was taken by the Virginia Tech and its partner libraries, which had started working on the National Digital Library of Theses and Dissertations (NDLTD) in 1996. Beuhler (2013:167) mentions that this concerted effort to develop ETDs collections was, however, first taken in 1992. The NLDTD plays a crucial role in making ETDs globally available on open access.

Some libraries already have mandatory policies for electronic submission of theses and dissertations, with print copies no longer accepted; others have only fairly recently transited to electronic copies only, while others still accommodate both formats (Yiotis 2008:106; Cox 2015:21; Angadi, Bandi & Koganuramath 2012:18). The possible future trend is towards increased adoption of the digitised ETDs format, due to the numerous advantages of digital resources. The attributes that differentiate ETDs from the theses and dissertation in print format are discussed by Fox, McMillan and Eaton (1999:4), Wolverton et al (2008a:3), Yiotis (2008:103), Hider (2012:10), Angadi, Bandi and Koganuramath (2012:885), Beuhler (2013:164) and Ezema and Igbo (2016:25).

ETDs have a number of unique attributes, in that they:

- Are technologically advanced medium of expressing ideas;
- Have enhanced features and embedding multimedia within the ETDs is possible;
- Have the capability for wide distribution, which improves global visibility and ranking;

- Offer advanced ways of descriptions: existing digital resources metadata schemes and systems allow use of more metadata fields to describe ETDs comprehensively and consistently;
- Offer improved searching capabilities, including full-text searching: centrally collected metadata allows comprehensive searches from diverse collections;
- Have links in the ETD content and their metadata create greater connectivity to other sources of information;
- Offer enhanced management of ETDs as information resources and their metadata.

These attributes are discussed in Chapter 3, which concentrates on ETDs metadata management in detail.

2.2.2   ETD developments in the United States of America (USA)

Most of the early ETD developments occurred in the United States of America (USA). It is important to understand the development in the USA, because it has had significant influence on the global progress of ETD implementation and management and may have influence on current and future progress of metadata in the South African context. The USA is discussed here to provide insight into the general development of ETDs. Table 2.1 (p. 38) shows the USA involvement in ETDs, from as early as the 1930s. Several surveys – e.g. those conducted by Wolverton and Hoover (2004) and McCutcheon (2007) – were conducted in the USA that reflect the status of ETD development, including the general history and theoretical perspectives, status in individual institutions, management and organisation and technical and system related aspects.

Additionally, Fox et al (2004) trace the national development of ETDs in the USA. They indicate the early development led by Virginia Tech, with a vision of the international community joining the NDLTD project to produce a global multimedia ETD portal (Fox et al 2004:34). The NDLTD has global membership from institutions, consortia and individuals. Membership from the USA on the NDLTD involves 40 institutions and two consortia (Networked Digital Library of Theses and Dissertations 2018), which is the highest number, compared to other countries. Beuhler (2013:172) indicates that the

USA was an early adopter of this unique model, with institutional, regional archiving and the NDLTD being established to archive local and global ETDs. Other countries are at different levels of development in adopting this model. South Africa implemented the NETD national aggregator, which is further discussed in Section 2.2.4.

Collaboration at state level is shown by McCutcheon (2007) in a survey, concluded in 2007, which indicates collaboration in ETDs work by member universities filing ETDs with Ohio's Academic Library Consortium (OhioLINK. Electronic Theses & Dissertations Center).

The report covers issues on processes and general management of ETDs, revealing that the majority of the respondents had mandatory ETDs policies. The report shows early advancement and commitment to ETDs by the surveyed USA libraries. Thirty two (32) libraries are listed as having browsable ETDs by institutions' names on the OhioLINK website (OhioLINK. Electronic Theses & Dissertations Center 2018). There is an indication of growth from the 30 collaborating institutions, earlier indicated by Evans and Schwing (2016) as being affiliated to the Electronic Theses and Dissertations Centre at OhioLINK.

Information on developments in individual institutions in the USA is widely available. The USETDA database hosts a live community driven report, which is continuously updated on the status of ETDs in USA institutions (United States Electronic Theses and Dissertations Association 2018). The information provided covers the status of ETDs and the years in which various processes were implemented. The information of 121 institutions appears in an updatable spreadsheet format, showing the lead that the USA is taking in ETD development.

One of the signs of fast and high level of development of ETDs in the USA is revealed in the report, with most institutions accepting electronic submission only. High numbers of submissions and high annual growths occurring during the early years of ETD implementation are evident (Hall, Hoover & Wolverton 2005:3). The Ohio ETD survey compiled by McCutcheon (2007) supports this rapid growth in ETD implementation in the USA.

Management and organisation of ETDs in the USA is also at a highly advanced stage. The study of Wolverton and Hoover (2004) makes a comparison of survey results between 1970 and 2000 in the USA, revealing how they treated and catalogued theses and dissertations in print and electronic formats, also indicating strengths in the richness of metadata descriptions for ETDs in the USA collections. Most of the findings on these topics are discussed in Chapter 3. Another fairly recent work revealing the status of management and organisation of ETDs in USA institutions is a survey conducted by Middleton, Jean and Gilbertson (2015) at the University of Arkansas, showing the common use of full level descriptions and the use of subject headings to promote online access and visibility of ETDs.

Guidelines on best practices also prevail in the USA. The ETD_MS guide, provided on the NDLTD database, is an example of leadership provided by the USA on the description of ETDs and issues of interoperability (Networked Digital Library of Theses and Dissertations 2018). Positive developments in metadata creation practices are shown in a survey of academic institutional repositories in the USA, with quality control being widely practiced (Steele & Sump-Crethar 2016:67).

However, there are challenges that still need to be addressed. There is indication of a lack of consensus on acceptable standards for ETD description. Disparities on how metadata and quality issues are handled in the different institutions are also revealed in the survey by Steele and Sump-Crethar (2016:67). The quality of metadata is further discussed in Chapter 3.

Moreover, Fox et al (2004), Mao (2004), Hall, Hoover, and Wolverton (2005), Yiotis (2008) and Cox (2015) show progress and high developmental stages in policy development, advanced technology adoption, automation of processes, multimedia ETDs, collaborative models and implementation guidelines initiated in the USA over the years.

2.2.3   ETD developments in Africa

Comparative analysis of ETD development in African regions and South Africa reveal significant information that may help shape developments with regard to metadata creation and adoption of practices that facilitate knowledge sharing on the continent. Knowledge of practices in Africa provides important contextual information for the creation of quality interoperable metadata.

Africa faces unique challenges with the development and implementation of ETDs. In his work published about five years after the Virginia Tech implemented ETDs, Ubogu (2001:25) notes that African research material is still poorly represented on the international bibliographic databases. The project findings show that, in the period of the study, only South Africa and Egypt have centralised access through which the bibliographic information of theses and dissertations produced in their countries could be found Ubogu's (2001:70). Ezema and Igbo (2016:25) confirm that, despite the recorded successes, Africa still faces a challenge of global dissemination of its research output, including the adoption of ETDs, as compared to other countries.

ETDs have an advantage, in that they can be made available on open access and, therefore, address the dissemination problems. The importance of open access is gradually realised among African institutions (Dulle 2010:238; Baro & Etiode (2014:116). However, challenges that involve key problems to ETD implementation as open access material, mainly technological, are experienced in Africa. This is one of the major problems that commonly affect all open access initiatives on the continent, resulting in little progress being made (Dulle 2010:45). Dulle (2010:238) and Kagoro, Khayundi and Oyelana (2016:16) point out to Internet connectivity problems that are still experienced in some African countries, which delay open access efforts. This also affects ETD progress.

In addition, poor exploitation of technology is also shown by Dulle (2010:68) as one of the factors affecting development in scholarly communication in developing countries. However, The Database for African Theses and Dissertations (DATAD) and the African Digital Repository, further discussed below, are examples of efforts in Africa to publish ETDs.

DATAD is a product of the Association of African Universities (Association of African Universities 2017). This is an initiative, started in 1998 and with its infrastructure implemented in 2002, to ensure that research conducted in Africa is made widely visible 2002 (Ubogu 2002:3). DATAD Online was launched in 2003 (Association of African Universities 2017). DATAD emerged from the need to expose research work on Africa and by African scholars (Ubogu 2002:70), which would include research in the form of theses and dissertations. This is more like the Africana collections in different libraries, where works produced on Africa and by Africans are housed. As an online tool, DATAD promises to disseminate the research information in the form of citations and abstracts on a wider scale. In 2017, the only populated collection on the DATAD repository was that of the University of Pretoria, reflecting a total of 11941 items (Association of African Universities 2017). Currently, the only populated ETDs collection is that of Ghana (Association of African Universities 2017).

Another effort to enhance open access to African research materials is that of SABINET implementing the African Digital Repository. The scope of the repository coverage is aimed at harvesting records from institutional repositories in South African and other African institutions, with links to the full text.

At the time of the study, only ETDs of South African institutions were accessible on the platform. Universities that were listed as source institutions include: Stellenbosch University, University of Pretoria, University of Kwazulu-Natal, University of the Witwatersrand, Rhodes University, University of Zululand, Durban University of Technology, University of Limpopo, University of Johannesburg, University of the Western Cape, North West University and Cape Peninsula University of Technology. The harvested metadata are put in a central repository that allows integrated searching.

The exposure of ETDs in this central repository is increased by providing citations and abstracts, with links to full-text stored in the holding institutions' repositories. Wolverton and Hoover (2004:14) and Zavalina (2010:149), recommend for more studies on ETDs to be conducted, focusing on other countries outside the USA.

The development of ETDs in Africa follows common global trends in terms of implementation and development, although not at the same pace as that of the USA (as indicated in Section 2.2.2). South Africa is experiencing common implementation and operational challenges, as other countries in the developing parts of the world. The background information on Africa provides a framework for the investigation of ETD development in South Africa.

2.2.4   ETD development in South Africa

Literature coverage and research on ETDs in South Africa and the creation and management of metadata is observably low. According to Maurer, McCutcheon and Schwing (2011:280), the development of ETDs can be traced to the 1980s. South Africa and other developing countries followed the trends in the developed countries in the implementation of ETDs. Ubogu (2002:5) traces the appearance of the first thesis from Africa on the World Wide Web (WWW) to a South African University, the Rhodes University in 1998, although the initiative had started earlier in that institution in 1996.

Efforts to make theses and dissertations visible, taken by the Union Catalogue of Theses and Dissertations (UCTD) database, currently hosted by SABINET and included in the Current and Completed Research Projects database hosted by the National Research Foundation (NRF), are recognised as major contributions to the organisation and management of theses and dissertations in South Africa.

It is important to share existing knowledge among countries that are well-established in the implementation of ETDs and countries that are still developing in this area. This is important to facilitate the promotion of open access and the global sharing of research information.

Although the significance of ETDs has been widely embraced in South Africa, ETD development and implementation is at its different stages in universities in South Africa, as evident on the websites of the different universities. The institutional repositories of various South African institutions offer full-text access to their resources. Compared with the situation in the USA, South African studies at individual institution level are not sufficiently covered in the literature. Chowdhury and Foo (2012:16) explain the model that existed in South Africa during the initial phases of the introduction of digital resource, which involved the hosting of ETDs by some institutions on behalf of those with no capacity to archive their own digital collections. This model greatly assisted in spreading the implementation of ETDs to some South African universities.

Information on the NETD database (Table 2.2), which is an aggregation of metadata for the theses and dissertations repositories in South African universities, shows an uneven development in the various institutions. Furthermore, the membership presence of the South African ETD repositories on the international cooperative ETD databases, like the Networked Digital Library of Theses and Dissertations (Networked Digital Library of Theses and Dissertations 2018), is observably low. At the time of the study, only five universities showed membership on the NDLTD: Durban University of Technology, Rhodes University, University of Johannesburg, University of Pretoria and University of South Africa. Other factors, like the requirement for membership subscription, may account for the absence of other South African universities. However, links to the NETD, where South African ETDs can be searched, are provided on the NDLTD.

The status of the descriptive metadata creation and the associated standards for ETDs still have to be researched in the South African context. Literature that specifically addresses these issues in South Africa is still lacking and there is no evidence of the study of specific types of subject metadata.

2.3   ETDs and geographic subject access

Subject access, including its geographic form, is important for information resources discovery. Various types of information resources continue to benefit from representation of their content in the form of subject description. However, critical views on subject access in general are common, and usually encompasses negativity towards geographic subject access.

Table 2.2: NETD database collection statistics (as at September 2018)

Collection	Total
1. Cape Peninsula University of Technology	1774
2. Central University of Technology	651
3. Durban University of Technology	2582
4. Nelson Mandela Metropolitan University	4040
5. North-West University	7670
Collection	**Total**
6. Rhodes University	7661
7. Stellenbosch University	12216
8. Tshwane University of Technology	1690
9. UCT Computer Science	97
10. University of Cape Town	17300
11. University of Fort Hare	1171
12. University of Johannesburg	9729
13. University of Kwazulu-Natal	9535
14. University of Limpopo	2081
15. University of Pretoria	12370
16. University of South Africa	15803
17. University of the Free State	1202
18. University of the Western Cape	6809
19. University of the Witwatersrand	9161
20. University of Venda	894

21. University of Zululand	1475
22. Vaal University of Technology	185
23. Walter Sisulu University	84
Total	**126180**

2.3.1   Developments affecting geographic subject access of ETDs

Geographic subject access for various forms of information resources is affected by the general state of development of the subject representation practice. The continued efforts to improve information resource description are evident in the review and development of major international models and guidelines mentioned here, some of which are further discussed in Chapter 3.

Advancements in standards for subject description like the new FAST headings, which are based on the Library of Congress Subject Headings (LCSH), are an example of these efforts. Dean (2004:332) explains the emergence of the FAST headings as being as a result of the re-examination of the way subject access is provided to cater for the fast growing resources on the Web. ETDs and their descriptive metadata are commonly available on the Web and their description practices are affected by these developments. These tools shape the nature of subject access practice in general, including the geographic form.

The FAST headings uphold the LCSH approach of accommodating geographic subject headings. LCSH instructs on the use of geographic subject headings as main topics, where a specific geographic area precedes a topic (e.g. South Africa – History), or as topic subdivisions, where a topic is subdivided by place (e.g. Agriculture – South Africa), to reflect the contents of a work. The LCSH instructs that geographic subdivisions are not mandatory, but are beneficial in situations in which the content of a resource is not focused on a place, but important information on the place may exist. The FAST headings maintain the relevance of geographic subject headings by using a simplified, faceted, application format. The facets in FAST are divided into geographic names, personal names, corporate names, events, titles, time periods, topics and form/genre (FAST Project 2018).

Other developments in information resource description guidelines, including the Functional Requirements for Bibliographic Records (FRBR) model, the Resource Description and Access (RDA) guidelines and the Functional Requirements for Authority Data (FRSAD) model, also promote the significance of subject access to information and the use of "place" as a subject term.

Various other subject access guidelines for specific types of publications, like the Subject Access for National Bibliographies, also address subject access issues, including the levels of subject access. Geographic subject description features in all these guidelines as a recommended approach. A place name is crucial for finding information and, therefore, "place" is identified as one of the major entities that can serve as a subject of an information resource (Riva, Le Boeuf & Zumer 2017:86).

This study is limited to the use of "place" as a means of subject access. It is important to identify the importance attached to geographic subject elements in relation to other types of subject metadata elements. Geographic issues have increasingly become a global concern and warrant intensive investigations – i.e. on how geographic subject metadata can contribute to find relevant information that address these important developments.

2.3.2   Importance of geographic subject access

Subject access has been studied from various perspectives to evaluate its worth in the different eras of information usage and systems. According to Fattahi (1998:212), the choice and form of access points is influenced by their functions. The role of geographic subject access relates to the functions of a subject element in the descriptive metadata of an information resource. According to Dimec, Zumer and Riesthuis (2005:215), these functions of the subject element in the description of a work are closely associated to Cutter's principles of a catalogue, which include the subject element function. The principles include the role played by the subject of a work as: (i) enabling a person to find a book when a subject is known; and (ii) showing what a library has on a given subject.

From the perspective of subject access, retrieval by geographic location is beneficial in finding information about a specific place from the subject content of an information resource. As shown above, the LCSH indicates that a specific jurisdiction or place can be used as a geographic subject description. In support, Pretorius (2005:231) shows that the importance of geographical representation of content is to assist in identification of the locality related to the content. The success of a search on a geographic location commonly depends on matching the right search terms and the information resources about the place. This is commonly performed by specifically searching for the place name or limiting topic searches to specific geographic locations. The geographic subject descriptions are assigned, in order to aid retrieval of information resources of various kinds.

"Place" is also important in making connections in finding documents (Soergel 2009:26). If used in subject descriptions, "place" plays the collocation function by allowing the gathering of information about a particular place. In addition, Bidney and Clair (2014:70) hold the opinion that place is interdisciplinary. Information on specific geographic places may be scattered among different subject fields in ETDs. The value of information linked to various geographic locations, contained in the theses and dissertations, transcends the specific contexts around which the different researches have been conducted.

The value of geographic information can also be far reaching in terms of its worth for a variety of uses in different contexts. Knowledge of the theories underlying geographic subject metadata may help achieve the quality that promotes the cross-contextual discovery and use of information relating to specific geographic locations.

Various studies also confirm the significance of geographic subject access. Zavalina (2010) conducted research that endorsed the importance of subject access. The study reveals the importance of consistent application of different metadata elements, including the geographic coverage. The study also recommends for further investigations into the issues of subject access in general. Consistency is maintained in a study conducted by Zavalina in 2014, on the importance of the subject element of description for digital information resources, including the geographic values, in improving recall and precision.

Comparisons of the results yielded through the use of the various research approaches, indicate a consensus in support of subject access for ETDs, including its geographic form, and reveal the continuous changes in the subject metadata creation practices, which are mainly technology-based.

Debates around the different approaches to metadata and the fields used for the description of various publication formats are continuing in the Library and Information Services (LIS) literature, e.g. Gerolimos (2013), Earle (2014), Husic (2014), Zavalina (2014), Cox (2015) and Yang (2016). However, the types of subject metadata are commonly not treated distinctively, thereby creating a gap in the understanding of how the current developments affect the different levels of subject metadata creation.

Geographic subject metadata is often encompassed with other types of subject representations in the discussions on subject metadata. Zavalina (2014) follows a slightly different approach by briefly indicating the need for full subject metadata coverage, including geographic subject metadata. Through a comparison of metadata in different collections, the study also indicates the lack of geographic information in some of the descriptive metadata records. Gross, Taylor and Joudrey (2015) give an analysis of two decades of work in support of and against continued use of subject headings as opposed to keywords, and conclude that there is major support for subject headings, which confirms the continued importance of subject access.

"Place" continues to be projected in various studies as a necessary geographic subject description form that needs careful consideration. Additionally, Maurer and Shakeri (2016:233) discovered in their study that geographic subject description for ETDs were prominent in the bibliographic records of the ETDs at the Kentucky State University, thereby supporting their observed worth. In addition, the survey conducted by Wolverton, Hoover and Fowler (2011:207) among academic libraries indicates the prevailing use of geographic subject headings. Both positive and critical views on the importance and usage of subject access can be traced in LIS literature.

2.3.2.1   Critical views on geographic subject access

Opinions against subject access, which encompasses the geographic subject access form, are expressed in views on the practice of subject access that are evident in existing literature. Scholars such as Wolverton et al (2008b:44) and Gross, Taylor and Joudrey (2015) point to several works that advocate the discontinuation of subject headings in favour of keywords. After analysing past studies – particularly those on theses and dissertations – Wolverton and Hoover (2004) also established the discontinuation of the use of subject headings by some libraries. This decrease was, however, shown to be a smaller percentage, compared to those using subject headings. Wolverton et al (2008b:46) cite an extensive survey of NDLTD members, which, in consensus with the above findings, also shows a decrease in subject headings usage, compared to other fields and use of keywords. On the other hand, the results of the survey also show a bigger percentage of institutions that do use subject headings (Wolverton et al 2008b:61). Limited use of subject headings is repeatedly shown in the study by Wolverton, Hoover and Fowler (2011:205).

Problems in subject access can also be traced in Larson's study (2011:207), which investigated subject access practice in a large online catalogue and pointed to the general decline in subject headings use. Such trends also imply problems in the use of geographic representations.

In addition, the Bibliographic Services Task Force of the Universities of California (2005:31) is critical of the application of controlled vocabulary for topics due to improved library systems search capabilities. However, it promotes the continued use of controlled vocabulary for names, uniform titles, date and place. Other critical views about subject access, emanating from increased technological developments that facilitate full-text searchability, automated subject representation practices and user contributed metadata, are discussed in Chapter 3. Despite the different practices and criticism, "place" continues to feature as an important subject access form.

As discussed, existing literature confirms the general importance of subject access, including geographic subject metadata, and recommends ways to provide subject access for electronic resources in the broad information provision context.

ETD format presents its own challenges, including subject description practices. Their subject analysis and description approaches are affected by these changes. As a format, ETDs and their special characteristics are introduced in Section 2.3.3.1 and the theoretical conceptions of their subject analysis and the description approaches form part of the discussions presented in Chapter 4.

2.3.2.2 Applicability of geographic subject access within different contexts

An analysis of existing literature demonstrates the different contexts in which geographic subject access has been researched and reveals the broader context in which it is practiced. It further points to the importance of geographic subject access in the different contexts. The literature approaches discussions on geographic subject access from different perspectives, in different eras, and is contextualised to different types of disciplines and special users of geographic information. Bake (1974) studied geographic subject access in the context of assisting genealogical researchers to find information that is focused to their needs from history collections. Geographic content is also dominant in geospatial studies, with focus on access to geographical collections. The geospatial field concentrates on geographic datasets, used with computation techniques, to meet the information needs of expert users of this type of data. Vardakosta and Kapidakis (2013:796) define *geospatial* as mainly used in scientific papers to refer to digital spatial information with a geographic component.

This study maintains the use of the term "geographic" to refer to the place names, as used in the library ETDs contents and their metadata. Such place names are regarded as beneficial in information searches conducted by the experts within specific subject domains. The wider information user community, beyond specialised fields of knowledge, benefits from using place names to search for information.

The information contained in ETDs has potential geographic value that can be beneficial in different contexts and at different times. Environmental, social developments and other contextual factors in specific areas, as indicated in Section 1.1.1 (Chapter 1), have the potential to affect the relevance and value of the information in ETDs in different time periods. These factors introduce potentially new ways of using their content, including the geographical information.

A step towards aligning information to economic, environmental and social development to help people improve their lives, is evident in the stance taken by the Library and Information profession to address how libraries contribute to the United Nations (UN) 2030 agenda (International Federation of Library Associations and Institutions 2016). Another new contextual development is the emergence of ETDs written in indigenous languages, with place names given in those languages. These may present new description challenges, not catered for any existing systems and standards.

These discussions provide a background to understanding how context affects subject access issues, and allows an informed discussion that is focused on ETDs. The geographic subject access approach, using "place" as a description element, plays an important role as a form of subject description for ETDs across various subject disciplines. Understanding the different views and contexts of using "place" as a subject description element and any findings in the reviewed literature on how it is used for ETDs subject description is important for this study.

2.3.3   Geographic subject access in the context of ETDs

Developments related to publication formats affect the nature of subject access practice. As shown in Section 2.2, the development of ETDs is linked to their print predecessors. It is important to establish how the change in theses and dissertations formats from print to electronic has impacted on geographic subject access. As observed by Wolverton et al (2008b:46), the various types of access points that were traditionally assigned to the print versions of theses and dissertations are also commonly used for the description of ETDs. A review of the traditional practices reveal the trends in theses and dissertations description practices, also showing the geographic subject description trends. The progress from the historical print to the current electronic contexts aids in understanding the implications of the changes on the approaches to subject access for the ETDs.

The contemporary digital era shapes the nature of theses and dissertations management. Subject access to theses and dissertations in the electronic environment is currently being studied from different viewpoints, in order to establish its role and guide its practice. This section focuses on geographic subject access as practiced in the ETD context. These issues form the basis for the empirical investigations on geographic subject access for ETDs, as set out in Research Objectives (ROs) 2 and 3:

- RO2: To collect data from the practitioners involved in metadata creation in South African university libraries on how they approach the determination of geographic subject metadata for the ETDs.
- RO3: To establish from the subject metadata creators in South African university libraries what the implications are of their different analysis approaches for the determination of geographic subject metadata for ETDs.

Although the benefits of ETDs are widely accepted (Wolverton et al 2008b:68), there are fewer studies conducted on the description of ETDs (Wolverton et al 2088:66). Therefore, it is important to determine how access to ETD content can be improved.

One way of improving access to ETD content is deciding on the subject relevance, which can be done through geographic representations (Soergel 2009:26, Zeng, Zumer & Salaba 2010:39). Relevant and specific geographic information should be easy to search and retrieve from ETDs. Suitable discovery systems and access points are essential to make ETDs effectively accessible; as is the need to be reviewed on a continuous basis. The empirical data gathered through this study will further assist in this regard, particularly in terms of addressing the theoretical foundation on which the access points in the form of geographic subject metadata for ETDs are created.

This chapter will continue to discuss ETD geographic subject access in the context of the ETDs by considering the following issues: (i) ETDs as a special format; and (ii) the benefits of geographic subject access metadata element versus other subject access metadata elements in ETDs.

2.3.3.1   ETDs as a special format

Geographic subject access and its use is not limited to specific types of publication formats or information resources collections. Bake (1974) conducted a study in a manual cataloguing environment, with specific types of publications, the print archival history materials, focusing on their geographic descriptions. Although conducted several decades ago, the study established the relevance of geographic subject description for archival materials. In a study concerning subject determination, Sauperl (2004:63) recommends for research to investigate whether subject description is influenced by the document types.

The recommendation put in question form is: "Should subject description be conducted differently for different types of publications"? Any type of information resource may contain information related to a specific place or places. Therefore, geographic subject metadata is important for all types of information resources. This study partly answers the question by concentrating on one specific type of information resources, which are ETDs, and a specific type of metadata – i.e. geographic subject metadata.

As special types of digital objects, ETDs have also been found as benefitting from detailed subject descriptions. A survey conducted by Middleton, Jean and Gilbertson (2015:243) indicates the high importance of detailed description, including subject headings, in improving access to theses and dissertations. The authors further support subject access and its benefit to electronic theses and dissertations (Middleton, Jean & Gilbertson 2015:235).

Past research provides evidence that subject access improves discoverability of ETDs, as also revealed in a study conducted by Maurer and Shakeri (2016:236). All these works agree on the importance of subject access for ETDs. A further affirmation of the importance of subject access for ETDs is seen in the efforts of the IFLA Subject Analysis and Access Section, Working Group on Guidelines for Subject Access by National Bibliographic Agencies, to contribute to the continued improvement of methods of providing subject access (International Federation of Library Associations and Institutions. Working Group on Guidelines for Subject Access by National

Bibliographic Agencies 2011). ETDs that are covered in this group's work specifically addresses subject access to web resources.

Various media formats often form part of the ETD content and also need to be described. These heterogeneous content causes description challenges (Alemneh & Philips 2016:2). It is also common to link various types of images in ETDs, including maps and photographs, to specific geographic areas. On the other hand, the new technologies allow finding and linking geographic information from these unique formats, necessitating a revisit of their description practices. Furthermore, the Library of Congress, which gives guidance on the provision of access to images through geographic subdivisions, states that, as new technological methods are developed for linking topic and place names, there may be less need for geographic subdivisions (Library of Congress 2004).

In light of this assertion, it is necessary to establish how technological developments have impacted on geographic subject access for theses and dissertation, with the shift from print to electronic format, and the various multimedia included as part of their content. This study also serves as a basis for the description of future types of publications that may emerge.

2.3.3.2 Geographic versus other subject access metadata elements in ETDs

Hider (2012:7) opines that there are differences in the value attached to different information resources description attributes and, therefore, also in the metadata elements necessary to describe them. Different forms of subject metadata elements can be assigned to ETDs. These, in addition to geographical headings, would include common forms like topics, cooperate names and personal names. The benefits of the geographic subject access metadata element in comparison to other subject access metadata elements in ETDs will determine their continued use.

As indicated in Section 2.3.2, theses and dissertations typically focus on specific geographic locations as their areas of research. They often contain important primary information in the areas where the institutions of their origin are located, or areas in close proximity to the institutions of their origin.

The information contained in ETDs has potential value to address the information needs of the specific geographic areas. It is essential that they should be retrievable through the geographic place names to meet those localised information needs. De Groat (2009:17) and Liebetrau (2010:30) agree that place can be linked to the intellectual content of a document, and that information on "place" is often added as metadata relating to the intellectual access of a resource.

In addition, Soergel (2009:18) supports the idea of allowing different depths of searching of digital items, which suggests the use of different forms and levels of subject descriptions. The geographic element in subject representation is reflected in various vocabulary control standards as the main topic or as subtopic, with subtopics meant to introduce more specificity. This study examines both levels of geographic subject representations. The levels mentioned above allow for the different levels of specificity during searching.

ETD repositories form part of the sources of varying quality available and searchable by the broad information user community on the Internet. ETDs contain information of geographic value that should be discoverable by the information seekers. The challenge is to determine, through proper content analysis, the significance of a place name in the content of information resources. It is important to base ETD geographic subject representation on an understanding of suitable theoretical foundations, so as to enhance their discoverability.

2.4   Chapter conclusion

The reviewed literature reveals a consensus on the great strides in the development of ETDs and the attempts at having them discoverable through efficient metadata, including geographic subject metadata. Although there is evidence of the implementation of ETDs in South Africa, it can be concluded that this is occurring at a lower level than that of developed countries. The literature review reveals that not much research has been conducted on ETD subject metadata creation practices in the South African context.

ETDs can be used to satisfy various information needs and, based on the discussions in this chapter, it can be deduced that ETDs have become more widely accessible to a broader user community than the traditional print theses and dissertations. This situation presents unique description challenges, in that this information has to be made discoverable.

Additionally, subject descriptions are relevant to make ETDs searchable from various information platforms. According to the reviewed literature, geographic subject access remains an important form of subject description for electronic information resources, including ETDs, and requires further study. However, subject access in the form of geographic subject metadata is often not discussed distinctively or in much detail in literature. Furthermore, it can be concluded from the definition of a digital library given in Section 2.2 and followed by discussion on the development of the ETDs in Section 2.2.2, that quality is important in the consideration of ETD metadata.

Chapter 3 explores metadata quality issues in detail and discusses how the effectiveness of metadata can be enhanced through the use of standards.

CHAPTER 3: METADATA AND QUALITY CONTROL

3.1   Introduction

Engaging in this study requires a deep understanding of metadata. This chapter offers in-depth discussions on metadata for ETDs and expatiate on issues of metadata quality, particularly geographic subject metadata, together with associated tools and standards that are used to measure quality. The developments in information organisation and management necessitate a review of the type of metadata required and the quality issues that are surfacing. Quality metadata plays an important role in describing digital resources of various types to enable their discovery and access. Any attempt to evaluate quality can only be affected according to recognised criteria.

Discussion on the quality of metadata are important to address Research Objectives (ROs) 2, 3 and 4:

- RO2: To collect data from the practitioners involved in metadata creation in South African university libraries on how they approach the determination of geographic subject metadata for ETDs;
- RO3: To establish from the subject metadata creators in South African university libraries what the implications are of their different analysis approaches for determination of geographic subject metadata for ETDs; and
- RO4: To establish what type of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries.

These objectives aimed at promoting an understanding of the theories of subject analysis and their translation into practical metadata creation approaches in the new digital era. There is a relationship between the subject analysis theories and the quality of the subject metadata created. Information on metadata quality serves as a basis that facilitates the study on the influence of theories of subject analysis on quality subject metadata creation for ETDs. Greenberg's (2005) work is an example of a conceptual investigative approach on how to transform theoretical understanding into practical application of metadata creation.

Furthermore, RO4 (To establish what kind of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries) can only be reached, based on a clear understanding of what quality metadata is and how it can be measured and improved. The methodology of this research, as discussed in Chapter 5, which involves the analysis of geographic subject metadata on ETD records, also requires this understanding. The quality of the descriptive metadata – particularly geographic subject metadata – is of specific interest in this chapter.

Research results confirm that the use of "place" as a search approach is maintained by the users (Zavalina 2012:159), which confirms the need to maintain quality in the creation of geographic subject metadata to enhance the discovery and access of information resources. Tolosana-Calasanz, Alvarez-Robles, Lacasta, Nogueras-Iso, Moro-Medrano and Zarazaga-Soria (2006:233) regard place names as a problem that needs quality control for proper geographic information resource discovery and access.

Standards are commonly used to control the quality for geographic subject metadata. In spite of the measures taken to control metadata quality, Ochoa (2014:64), Pal (2016:111) and Yang (2016:7) remark on the ambiguities that prevail in the understanding of what quality metadata is and how it is measured. In support, Riley (2017:40) holds the view that there is a redefinition of what "authoritative" or "good" metadata is, due to openness and interconnectedness on the Web. The approaches followed during ETD geographic subject metadata creation are important in defining the required quality and determine how it can be achieved.

3.2   Approaches to ETD metadata creation

An understanding of the metadata background, particularly in the context of ETDs, reveals the aspects that are important in addressing quality maintenance. The discussions in this section are presented according to the following three themes: (i) the structure of metadata; (ii) the levels of description; and (iii) the sources of ETD metadata.

3.2.1 Structure of ETD metadata

The structure of the ETD metadata provides the basis for addressing issues of quality and standards. Soergel (2009:20) and Hider (2012:5) indicate that literature commonly categorise metadata according to the functions that it supports. Ma (2006:4), Jackson (2011:96), Hider (2012:5) and Khoo and Hall (2013:88) categorise metadata into the following three basic types: descriptive, administrative and structural metadata. Various other types – including use metadata, terms and conditions, content rating, provenance and linkage metadata – are mentioned by Greenberg (2005:21) and Mendez and Van Hooland (2014:27–29). In this study, the emphasis was on descriptive subject metadata.

Descriptive metadata is characterised according to different elements that describe the resources. There has been a shift in the definition of descriptive metadata to highlight the importance of changes introduced by technological development, e.g. increased requirement for interoperability. *Interoperability* is defined as the ability of multiple systems with different, hardware and software platforms, data structures and interfaces to exchange data with minimal loss of content and functionality (Riley 2004:2). Riley (2017:6) gives a related definition as the effective exchange of information between systems. Metadata records that do not comply with interoperability standards are a problem to information sharing. The NISO guidelines mention the uses of descriptive metadata as discovery, display and interoperability Riley (2017:7). These changes indicate the need for continuous studies on metadata in different eras, mainly characterised by technological developments. Various guidelines have been developed to assist in the creation of contextually suitable metadata.

The ETD_MS guidelines provided by the Networked Digital Library of Theses and Dissertations (NDLTD), which is a common global ETDs aggregator, guide the use of the Dublin Core elements, in an effort to provide a common metadata standard for sharing ETDs information. The different metadata elements, as presented in Table 3.1 (p. 63) are from the commonly used Dublin Core scheme. They are used as an example of possible metadata elements that can be used to describe an ETDs.

Metadata should provide multiple access points, including a subject, in order to improve content visibility (Tmava & Alemneh 2013:857). Mendez and Van Hooland (2014) categorise the Dublin Core metadata into content, intellectual property and instantiation. The subject metadata element falls into the content category.

Table 3.1: Dublin Core metadata elements set

Metadata element	Description
Contributor	An entity responsible for making a resource available
Coverage	The spatial/temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Creator	An entity primarily responsible for creating a resource
Date	A point or period of tie associated with an event in the life cycle of a resource
Description	An account of a resource
Format	The file format, physical medium or dimensions of the resource
Identifier	An unambiguous reference to a resource within a given context
Language	A language of a resource
Publisher	An entity responsible for making the resource available.
Rights	Information about the rights held in or over a resource
Source	A related resource from which the described resource is derived
Subject	The topic of the resource If a controlled vocabulary is used, the mandatory qualifier should be indicated, e.g. LCHS (the vocabulary used to determine subjects)
Title	The name of the resource
Type	The nature or genre of the resource

This study is centred on subject metadata, which is represented through the subject metadata element. The ETD_MS lists the subject element as mandatory and repeatable for the description of the ETDs (Networked Digital Library of Theses and Dissertations 2018). The subject represents the content of a resource and can take various forms, including the geographic subject form that is being researched in this study.

There may occur a potential overlap in terms of the values for geographic locations that are used in the "subject" and "coverage" metadata elements. This may affect metadata quality, since judgement lies with the metadata creators on the assignment of data values. The geographic subject metadata is one of the key concepts of the problem of this study: "To investigate how information resources contents are analysed by subject metadata creators to determine geographic subject metadata for electronic theses and dissertation, specifically how subject analysis theories are applied and the role that they play in subject description".

Zavalina (2014:78) recommends the inclusion of the subject element for improved recall of information from digital collections and aggregations, which, among others, allows the use of place as a "subject" value. The subject element, particularly the controlled subject metadata, does not usually appear in the descriptive metadata that is meant for minimal level of description only. The different elements are often part of the description records of information resources, depending on the level of description desired. The levels of description are summarised in Table 3.2.

3.2.2   ETD subject metadata description levels

The construct "level of description" is sometimes used interchangeably with the quality of metadata (Zavalina 2014:78). The level of description is a concern for relevant subject metadata creation and quality evaluation. These levels are addressed in the IFLA Subject Access guidelines for National Bibliographies (International Federation of Library Associations and Institutions. Working Group on Guidelines for Subject Access by National Bibliographic Agencies 2011:10). The different levels of description, as applicable to different types of materials – particularly in the online and digital environment – are crucial for metadata creation. The guidelines indicate that indexing level decisions are necessary to describe materials differently and recommend that the levels of significance of information resources should be considered. Hider (2012:78) holds a similar view, in that the level of detail in the description of resources will be different according to the nature of the material, highlighting the importance of considering the nature of the user.

For example, due to its research value, a thesis is considered suitable for the high level of description (International Federation of Library Associations and Institutions. Working Group on Guidelines for Subject Access by National Bibliographic Agencies 2011:38).

In addition, the levels of description determine the detail of information used during metadata creation, in the form of metadata elements and data values. The quality of description can be linked to the amount and type of information used during the description. This situation also applies to subject metadata creation. Frank and Rowe (2004:352) opine that the user needs should drive the decision about the level of descriptions. By opting for the full level description, which from a subject metadata perspective, also accommodates the use of both keywords and the Library of Congress Subject Headings (LCHS), the University of South Florida is an example of a decision taken about the level of description for cataloguing ETDs (Frank & Rowe 2004:347). On the other hand, the current environment, characterised by scarce resources and improved technology, has resulted in the questioning of the need for detailed information resources descriptions.

The perceived level of importance of an information resource factors influence the decision on the level of description to be used. These levels of descriptions are categorised into, high, minimal and basic (no controlled subject access) levels (International Federation of Library Associations and Institutions. Working Group on Guidelines for Subject Access by National Bibliographic Agencies 2011:38). The levels of description are depicted in Table 3.2 (p. 66). High-level descriptions for ETDs will normally include as many elements listed in Table 3.1 (p. 63) as possible, and use both controlled and uncontrolled vocabularies. The kind of vocabulary used is usually determined by the level of description that is anticipated. The sources of description also determine the success of vocabulary use.

Table 3.2: Index levels decision matrix

Recommended indexing level	The value (perceived level of importance) of the resource	Type of resource
Full	High importance	Resource with research value or intended for use as research or reference tool
Minimal	Intermediate importance	Resource with value to the users of the catalogue, but not considered to need full cataloguing
No controlled subject access	Low importance	Resource with ephemeral information, judged to be of little interest to contemporary or future audiences

(Adapted from International Federation of Library Associations and Institutions. Working Group on Guidelines for Subject Access by National Bibliographic Agencies 2011)

3.2.3 Sources of metadata and quality achievement

The existing possibilities for metadata creation result in information resources descriptions of various origins. The different sources often present associated benefits and challenges for the production of quality metadata. Creating ETD metadata locally in institutions owning the ETDs is a common practice in libraries. Metadata is also harvested and shared across ETD repositories. Martin and Mundle (2014:242) express the view that transformation and harvesting may not be necessary in the future data environment. However, in the current situation, where some metadata elements can be published as linked data to enhance discoverability, quality and standardisation are highly important. Metadata that is commonly shared and searchable in aggregated platforms needs to be evaluated for quality and standardisation to enhance discovery of and access to various types of information resources (Hider 2012:8). The correct application of standards requires a good understanding. According to Palavitsinis (2013:47), there is a paradigm shift in terms of who creates metadata and the quality issues that arise as a result of the different metadata origins.

In addition, "provenance", the source of the metadata, is identified in literature as a factor associated with the quality of metadata (Margatipoulos, Margatipoulos, Mavridis & Manitsaris 2008:107; Ochoa 2014:6; David & Thomas 2015:805). The knowledge and expertise of the metadata creator play an important role in the quality of metadata created (Ochoa 2014:68).

The issue of provenance is important for ROs 2 and 3 of this study, which involve making investigations into the approaches followed by metadata creators, who are the source of the metadata. Margatipoulos et al. (2008:107) and Park (2009:219) link provenance to reputation and authority of the metadata creator. The authority of the metadata is normally associated with its source of creation.

Professional contribution to the creation of metadata is regarded as increasing efficiency (Riley 2004:10), although Margatipoulos et al. (2008:107) acknowledges that good source of metadata on its own cannot be proof of quality. Hider (2012:87) in affirmation, states that "Professionalisation of metadata creation goes *some* way toward assuring quality", but also recognises that it is not the only variable that affects quality. As an example of quality issues related to the origin on metadata, Yasser (2012) discusses the errors in existing metadata that cause the metadata quality problems, and states that this is often due to different training levels of the metadata creators.

Table 3.3 (p. 68), adapted from the IFLA metadata for digital resources (International Federation of Library Associations and Institutions Working Group on Metadata for Digital Resources 2009), depicts the typical subject metadata creation scenarios and the origins of metadata.

Table 3.3: Metadata origins

Approach of subject metadata creation	Metadata creator (source of origin)
Full text searchability	Full-text searchable format
Student authors/researcher created	Student authors/researchers
Automatic	Automatic extraction/computerised (facilitated by publishers or access providers)
Semi-automated	Computer technology assisted/software assisted extraction & final human decision
Professional intellectual analysis	Information professionals in various institutions
User contributed metadata (Tagging)	Users

(Adapted from: International Federation of Library Associations and Institutions Working Group on Metadata for Digital Resources 2009)

3.3   Metadata quality evaluation

The quality of metadata is highly significant for the discovery and access of digital information resources. Quality evaluation of metadata ensures its effectiveness and efficiency. This need reflects in metadata quality and evaluation literature reviewed in this study.

3.3.1   Exploring metadata quality background

The discovery of and access to information resources in the digital environment is aided by quality metadata, as was traditionally done through appropriate cataloguing (Riley 2004:1). Despite the importance of quality metadata, it is not adequately addressed in the existing literature. South African literature generally lacks information on metadata quality evaluation. However, there is a noticeable persistence in addressing metadata quality from various perspectives in Library and Information Science literature (Potvin & Thompson 2016; Southwick, Lambert & Southwick 2015; Palavitsinis 2013; Cruikshank, Daniels, Meissner, Nelson, & Shelstad 2005).

Other Library and Information Science literature specifically highlights the importance of quality metadata for digital information resources and suggests criteria and models for quality evaluation (Sandy & Dykas 2016; Alemneh & Phillips 2016; Ochoa 2014; Palavitsinis 2013 & Park 2009).

An observation in the literature shows that there is also a realisation of the need to address quality and standardisation pertaining to specific metadata elements. Potvin and Thompson (2016) specifically investigated the metadata "date" element in ETDs repositories and its effect on interoperability. The study confirms that the different practices in ETDs metadata practices result from the influence of the metadata philosophies, tools and systems. All the description elements listed in Table 3.1 (p. 62) have unique attributes that can be utilised for quality measurement, sometimes evaluated according to different standards. The metadata "subject" element quality issues are addressed in this study, while considering arguments in the literature against the worth of subject metadata, as presented in Section 2.3.2.1 (Chapter 2). Such critical perspectives raise the question as to whether it is necessary to invest resources in the costly practice of creating and maintaining subject metadata of high quality.

ETD metadata evaluation is not sufficiently covered in the literature. Rasuli, Alipour-Hafezi and Solaimani (2016:13) observe that most research on ETDs concentrates on the technical aspects, as compared to other aspects that are important to successful ETD implementation and management. The next section investigates quality management of ETD metadata and the quality of the ETD geographic subject metadata.

3.3.2   ETD geographic subject metadata quality evaluation

The general complexity involved in the evaluation of metadata quality was highlighted in Section 3.1. ETDs are made available through various networked platforms and on the Web and are prone to the stated metadata challenges. The management of theses and dissertations in their traditional print format has always emphasised quality control.

Debates are continuing around the current status and future prospects of ETDs metadata management and control to address the widened accessibility and the broad user community with their varied information needs.

In the results of their study, Alemneh and Phillips (2016:4) affirm the value that quality ETDs subject metadata adds to information discovery and access and caters for the varied and evolving use of ETD information. Alemneh and Phillips (2016) also address quality in relation to multidisciplinarity and interdisciplinarity of ETD information and the different information needs that, as a result, can potentially be satisfied. Alemneh and Phillips (2016:4) confirm the assertion made in Section 1.1 (Chapter 1), in that quality metadata is necessary to cater for the different needs resulting from the changed information landscape resulted in the widened use of ETD information.

Quality issues are commonly addressed, together with developments in information systems and search platforms, which influence the subject metadata creation practices. These developments allow different search approaches, including full-text searchability, for ETD content. Despite these developments, quality metadata is still found to add value to easy discovery and access in this enhanced information search environment (Alemneh & Phillips 2016:16). Libraries that have implemented ETDs are striving to make their collections more visible to the public through effective and efficient metadata records. Subject access is one way of increasing this effectiveness.

Subject descriptions provide bibliographic metadata for information resources. In order to make information resources easily discoverable, subject descriptions need to conform to relevant bibliographic control and quality requirements. Chan (2000:4) argues that the contributions made by bibliographic control and the importance of quality subject access in the networked and Web environment, need to be addressed. Specifically, changes in subject access need to be examined. Critical views on the need for subject access, as a result of all the search capabilities introduced by technology, tend to refute its value and, thereby, any need to evaluate subject metadata quality. Lubas (2009:255) and Laskowski (2016:150) point out that some arguments, mostly arising from the view that full-text searchability, have made subject access obsolete, thereby causing subject access to be undervalued and misleadingly making the issues of metadata quality and standardisation appear less significant.

The creation of ETD subject metadata is commonly based on recognised standards or guidelines. This is also the case with ETD aggregated platforms like the NDLTD, which recommends the ETD_MS guidelines to control how metadata is created. Inconsistencies in the application of these standards and guidelines are one of the reasons for the need to evaluate metadata quality continuously. Chan (2000) observes that resource description and organisation in the networked information resources space are still evolving and issues of how to maintain consistency and quality to achieve efficiency remain a challenge.

Standards control quality for geographic subject metadata. The role of standards in maintaining the quality of ETD metadata, including the geographic subject metadata, is discussed in the next section.

3.3.3   Role of standards in measuring ETD geographic metadata quality

Quality is measured fairly, if there is a set of accepted assessment criteria and standards regulating it. General metadata standards guide the creation of geographic subject metadata. ETD geographic subject metadata creation standards form part of the general standards guiding metadata creation for digital resources. Special guidelines, such as the ETD_MS, show how the standards apply to ETDs. According to Hider (2012:103), metadata standards have been developed for values, elements, formats and transmission. These concepts, which are commonly used in the digital and information retrieval system environments, have relevance to the library environment, due to the use of metadata to describe digital information resources.

Hider (2012:6–7 & 103) discusses these concepts in the context of metadata creation and management and the standards used to control them. Based on the discussions, the following definitions are provided to assist in understanding the standards listed in Table 3.4 (p. 74) and the quality criteria listed in Table 3.5 (p. 81):

- Values: information resource attribute representations recorded in the metadata elements. Geographic subject headings are the types of values that may be used to represent the subject content of information resources, including ETDs.
- Elements: fields for recording various information resources attributes.
- Formats: rules for recording the values, e.g. on the form of arrangements.

- Transmission: protocols on how metadata should be input into a particular system to facilitate transmission.

Standardisation is critical for high quality metadata in digital repositories that publish scholarly material (Rasuli, Alipour-Hafezi, Solaimani 2016:101; Xie & Matusiak 2016:129). Hider (2012:8) and Bothmann (2011:1) hold the view that standards facilitate the sharing of information resources and maintenance of consistency in descriptions. However, a common view is that standards alone do not necessarily guarantee quality.

Historically, quality evaluation and adherence to standards have always been a critical issue for information resources description. Moving into the digital resource era, standards are still recommended to maintain quality metadata. However, Chan (2000:2) states that the suitability of well-established tools used for bibliographic control requires re-examination, in order to suit the nature of information retrieval in the current networked environment. Yang (2016:7), Palavitsimis (2013), Southwick, Lambert and Southwick (2015) and Cruikshank et al (2005) demonstrate the continued discussions, over an extended period, on the evaluation of digital information resource descriptions and the role of standards in quality maintenance. These authors show how digitisation has changed the nature of information resource description and the need for standardisation.

In the literature, standardisation of metadata in the digital space is approached from various perspectives. Various concepts are used for metadata standards, sometimes resulting in confusion in the meanings. Hirwade (2011:18) indicates the common interchangeable use of the words *standards, scheme* and *schema*. The stated author continues to show the distinction between the terms *standards* and *schema* by indicating that metadata standards are defined as "... a set of metadata elements and rules for their use that have been defined for a particular purpose" and metadata schema are "the set of descriptor types available to be applied to information (Hirwade 2011:18). On the other hand, IFLA (International Federation of Library Associations and Institutions. Working Group on Metadata for Digital Resources 2009:1) specifies the concepts of tools (DCMI Online, MARC); rules (e.g. AACR2, RDA, controlled vocabularies) and protocols (OAI-PMH) that are used to create and manage metadata.

These concepts are explained in Table 3.4 (p. 74). The controlled vocabularies and other authority subject term lists are the descriptive metadata standard types used to control the creation of subject metadata.

Bothmann (2011:4) shows that there was a change on focus from descriptive cataloguing standards, which occurred at the turn of the 21st century, due to the rise of automated information systems, and the resulting focus on metadata and schema. Later, interest in descriptive metadata and its standardisation developed, resulting, among others, in the development of standards like the RDA (Bothmann 2011:4). There is an observable increase in the literature covering topics on the creation of metadata and related descriptive standards. A comprehensive perspective on common metadata standards, including the controlled vocabularies, is adopted by Mendez and Van Hooland (2014:15) and Jackson (2011:97–98), who categorise the standards based on their roles as follows: data structure, data content, data value, data format and data exchange.

These standards are presented in detail in Table 3.4 (p. 74). All the presented standards are essential for the management and organisation of ETD, which are designed to be open access materials and are commonly aggregated and searchable from integrated platforms and on the Web, with the assistance of standardised metadata practices. Metadata standards have been developed for values, elements, formats and transmission (Hider 2012:103).

The standards presented in Table 3.4 (p. 74) provide a background understanding of the data value standards (subject headings, authority files and thesauri).

Table 3.4: Common descriptive metadata standards

Descriptive metadata standard type	Common standards examples	Purpose
Data structure standards (schema) composed of data elements represented in the form of a schema	MARC Dublin Core	Provide the metadata structure by defining the fields or elements to be used (e.g. subject element)
Data content standards	AACR2 RDA	Provide guidelines for creating metadata (e.g. guidelines on content structure & rules for input on capitalisation)
Data value standards (Controlled vocabularies, thesauri)	LCSH NAF Getty Thesaurus of Geographic Names	Predefined lists of terms to be used in metadata subject elements
Data format standards	XML (Extensible Markup Language) MARCXML	Bind together elements from a structure standard and allow elements to be arranged hierarchically
Data exchange standards (Exchange protocols)	OAI-PMH Z39.30 protocol	Allow records to be exchanged between institutions/systems (e.g. allow records harvesting)

Data values are of importance to this study and, therefore, this chapter will not attempt to discuss all of the standard types presented in Table 3.4, but concentrate on the standards used in metadata subject elements. These are the controlled vocabularies (data value standards), which are predefined lists of terms that can be used as values to be recorded within the subject element. A brief look into data structures and data content standards, reflected in Table 3.4 will facilitate the discussion on the data value standards (controlled vocabularies) and the related authority lists and indicate how different standards fit into the purpose of metadata quality evaluation.

The metadata standards listed in Table 3.4 show their different types, including schema and the data value standards that apply to the management and organisation of ETDs. Schema, elements sets or formats, often used interchangeably, are a way of facilitating the adherence to standards of access and interoperability (Riley 2017:14). On the other hand, the data value standards are applied in the different metadata elements. Different data value standards, including standards for assigning subject representations, may be applied to the same item that is being described.

The traditional terminologies of information resource description are used in this study to facilitate the understanding of the role of standards in measuring ETD metadata quality. *Schema, content rules* and *controlled vocabularies* are the common terms used for data structure standards, data content standards and data value standards. The use of schema, content rules and controlled vocabularies helps to ensure the quality of metadata (Ma 2006:12). Subject metadata standardisation can be achieved with the use of the controlled vocabularies

Controlled vocabularies – particularly those used for the data values for "subject" element – are discussed in this section. However, it is considered necessary to indicate that, according to a survey conducted by Matizirofa and Ramalibana (2015), with a 100% response rate, counts for 21 open access repositories in South Africa using the Dublin Core schema as a framework for descriptive metadata. The ETD_MS guidelines show how the common Dublin Core metadata element set, including the subject element, can be used to describe ETDs and to facilitate interoperability (Networked Digital Library of Theses and Dissertations 2018).

Potvin and Thompson (2016:99) express the critical role played by standards in enabling or constraining ETD representation in the form of records. ETD records consist of different metadata fields, with values that are populated based on the standards presented in Table 3.4 (p. 74). Riley (2004:3) and Jackson (2011:97) also affirm that the role of standards is to define the content of the metadata elements. This happens by controlling how the values should be represented (e.g. in terms of consistency of the terms used). Data values, which are the content of one metadata element or the information assigned to a metadata element, can be uncontrolled or derived from a controlled vocabulary.

3.3.4 Subject data values and controlled vocabularies

Controlled vocabularies play a critical role in the standardisation of digital information resources subject metadata. They contribute to metadata creation and quality enhancement (Ma 2006:10). The vocabularies are named differently in different contexts. Miller and Perkins (2015:136) indicate that the traditional controlled vocabularies used in libraries, such as name authorities, subject headings and thesaurus are called the "value vocabularies" in the semantic web environment. Additionally, The Dublin Core glossary defines a controlled vocabulary as "A set of consistently used and carefully defined terms" (Dublin Core Metadata Initiative 2018).

This section highlights the use of the controlled vocabularies and the impact on the subject metadata quality.

Subject metadata can be uncontrolled or controlled (Zavalina 2014:77; Greenberg 2005:24). In the context of subject metadata, the standards can be the vocabulary control tools, e.g. Library of Congress Subject Headings (LCSH), Faceted Application of Subject Terminology (FAST) headings, and the Getty Thesaurus of Geographic Names. The Dublin Core, for example, states that values for the "subject" and the "coverage" metadata elements can be derived from a controlled vocabulary (Dublin Core Metadata Initiative 2018). The schema commonly indicates that the value (e.g. subject heading) in the subject element is from a controlled vocabulary (e.g. dc.subject.lcsh). Furthermore, Riley (2017:17) indicates that metadata standardisation can be exercised through standardised control of the actual values used.

Developments concerning the controlled vocabularies affect the ETD metadata practices. In the same way, the changes in vocabulary standards influence how geographic subject metadata is represented. The common vocabulary control tools and other domain specific heading lists specify content rules as to how values within the subject metadata element must be formulated.

The development of Faceted Application of Subject Terminology (FAST) headings from the LCSH introduced a new way of looking at standardisation in the subject description element.

FAST headings have been developed to be simple and suitable for the Web environment and have introduced change to the past practices, while simplifying control subject representations. Chan (2000:5) suggests that greater usage of controlled vocabulary may improve subject access of networked resources.

The discovery of and access to ETDs in the networked environment is facilitated by subject metadata that is commonly created based on these controlled vocabularies, often used together with keywords. The application of controlled vocabularies is optional as recommended by different metadata schemes, which is also the case with the ETDs. Due to the optional use of standards, consistency in their application for quality maintenance is a problem.

Another argument is that the values for metadata elements are normally chosen based on the localised practices to suit specific contexts (Park & Tosaka 2010:109), usually adapted from common globally recommended standards and practices. This will affect the application of standards.

In addition, Zavalina et al (2015:15) indicate that metadata quality maintenance should consider balance among the needs of users, standards, and local environments of the metadata creators. An example of a localised practice in the application of standards based on geographic locality is seen in the work of Robinson, Edmunds and Mattes (2016:192). The work indicates that, at the Penn State (USA), a minimalist approach was adopted, which requires that LCSH to be used only when Pennsylvania or a town in Pennsylvania appears in the title of a work. This practice suggests the importance of indicating geographic relevance by using a controlled vocabulary only where it concerns the local geographic area. However, the use of controlled data values for the "subject" field is commonly recommended (Riley 2004:3, Dublin Core Metadata Initiative 2018).

A common practice is to represent ETD subjects in the form of keywords or concepts from a controlled vocabulary. Lately, user contributed tags are also a possible way of identifying the subject content. The tagging practice, which involves allowing users to contribute description terms to information resources, is an example of uncontrolled subject description system.

Martin and Mundle (2014:244) show that a review of literature published between 2011 and 2012 reflects criticism of the rigidity of controlled vocabularies, although the authors conclude that user contributed tags cannot be a replacement for controlled vocabularies and recommend the use of both to improve resource discovery. The decision to maintain balance between authority control and enriched subject metadata becomes critical with the adoption of user-contributed metadata.

The new transition into linked data as a way of providing access to data also requires the use of quality metadata based on controlled vocabularies (Southwick, Lambert & Southwick 2015:180; Miller & Perkins 2015:136). However, linked data will not be discussed in detail in this study. These changes also necessitate the re-examination of the quality evaluation criteria that are meant to assist interoperability. In the context of this study, it was important to establish how the changes related to standards influence geographic subject metadata.

Standards like the LCSH provide rules for the use of geographic subject headings and geographic subdivisions, now improved by the FAST headings, which allow more flexibility in terms of specificity in using geographic headings. It is common practice to use other standards to create subject metadata. Geographic name standards are used to derive authority forms of different place names. The Getty Thesaurus of Geographic Names is an example of a common standard recommended for use as a data value standard at the point of creating metadata (Zavalina 2010:64). The South African Geographical Names Council is the authority body responsible for standardising geographic names in South Africa (South African Geographical Names Council 2017). Quality descriptions are important for geographic subject metadata that is fit for purpose (Tolosana-Calasanz et al 2006:233).

After investigating the quality of ETD representations and considering the terms from controlled vocabularies and free text, Alemneh and Phillips (2016:1) noted the need to supply adequate and functional subject terms, instead of being mere descriptions of a topic. One of the challenges that emanate from their arguments on quality is accuracy (Alemneh & Phillips 2016:1), which is defined in Section 3.4.4.2. Standards can help to maintain the required functionality in terms of providing different alternative terms and the accurate ways of using them.

The factors used to evaluate the quality of metadata, in terms of their intended functional purpose, will be addressed in the next section.

3.4   Functional evaluation criteria for ETD geographic subject metadata

Metadata must be functional and suitable in providing quality representations that assist the users to retrieve information effectively (Park 2009:214; Alemneh & Phillip 2016:1; Yang 2016:17). Ochoa (2014:65), in agreement, points out that fitness for purpose is the relevant approach to quality measurement.

According to Tolosana-Calasanz et al (2006:233), fitness for purpose also applies to geographic metadata in terms of how well they describe geographical data and if the descriptions are beneficial to their users. This "fitness for purpose" approach may be viewed as involving the evaluation of the functional ability of metadata to help fulfil the user tasks – i.e. to find, identify, select and retrieve information (Ochoa 2014:65). Articulation of efficient criteria that facilitate the fulfilment of these tasks is necessary to guide decisions on the creation of functional metadata.

Different criteria for ensuring quality metadata can be applied. The common quality evaluation criteria, which are consistency, accuracy and completeness, are discussed in Sections 3.4.2.1–3.4.2.3. The adoption of evaluation criteria may differ in different contexts and, therefore, the criteria need to be assessed in terms of their adequacy for evaluating metadata effectiveness within the different contexts. The functional requirements of efficient service, as described by Ma (2006:11), influence the identification of the criteria on which creation of quality metadata is based. In addition, Liebetrau (2010:31) cautions that metadata standards used to facilitate quality must be appropriate for the type of resources and the community for which the metadata is created. ETDs are the type of resources investigated in this study and, therefore, the issues of functional quality – particularly pertaining to geographic subject metadata for ETDs – are emphasised in this study.

Alemneh and Phillips (2016:4) hold the opinion that there are multiple stakeholders in the digital ecosystem whose needs need to be accommodated by creating representations that enhance information use across different knowledge disciplines. In terms of geographic information, functionality assessment should consider the potential to satisfy information needs across many disciplines. In addition, the open access nature of ETDs presents these multiple context issues, which may require their metadata to be evaluated on criteria that accommodate both the localised, shared and networked environment use.

In addition, the different potential purposes, highlighted above, influence policies and approaches to metadata creation and further complicate metadata quality control (Park 2009:219). The recommendation that resources of research value should be considered for high indexing value, as shown in Table 3.2 (p. 65) implies an identified common purpose on which metadata for research information resources is founded and suggests the need for common criteria. The high metadata value status of ETDs, as shown in Table 3.2, suggests the need to adopt stringent metadata creation and quality evaluation criteria. Subject representations of ETDs; particularly those originating from professional sources, should comply with this high value status of metadata creation requirement. The criteria for evaluating metadata quality should consider the need to promote functionality, and to accommodating easy use and flexibility in the described complex ETD context.

Additionally, functional evaluation means the assessment of specific problems that need to be controlled. – As observed by Yasser (2012:373), metadata is only beneficial when it is problem-free. Riley (2004:10), Park (2009:215) and Palavitsimis (2013:17) identify the common problems of metadata quality, which include inconsistencies, incompleteness, nonconformity to controlled vocabulary, redundancy, lack of clarity, and incorrect use of elements. These problems, which are linked to the criteria for quality evaluation identified in this study, affect the quality and functionality of the metadata and require set criteria to control them.

Consistency, accuracy and completeness, further discussed below, are regarded as the most important criteria for assessing metadata quality, which can also be applied to digital resources (Ma 2006:12, Park 2009:214, Zavalina 2014:78). Ochoa (2014:65) labels these as the independent characteristics of the metadata record that affect quality. Table 3.5 lists the basic criteria and their definitions.

Table 3.5: Quality evaluation criteria

Quality criteria	Basic features evaluated	Definition
Consistency	• Metadata elements Consistency in metadata elements used and format in metadata elements • Data values Consistency in the data values used to represent similar concepts and their format	Consistency in the creation of metadata implies the use of the same elements (e.g. subject elements) to describe similar attributes (e.g. subjects); the use of a common structure or format in the elements; and the use of common values to represent or describe similar attributes.
Accuracy	Data values • Relevant: accurate representation of content; • Correct: lack of errors	Using correct data values to represent the attributes (e.g. subject) of an information resource (relevance) and lack of errors in capturing the data values (correctness)
Completeness	• Metadata elements All the required metadata elements used for description • Data values Representing content comprehensively by using all possible values, e.g. subjects	In this study, completeness is understood as the use of all the required metadata elements and comprehensive representation of content by using all important concepts necessary to describe the subject content an information resource

3.4.2.1 Consistency

In this study, consistency involves uniformity in the creation of metadata – particularly geographic subject metadata. Inconsistencies in local practices and across various systems may result in poor information search results (Tmava & Alemneh 2013:857–858).

Park (2009:221) and Zavalina (2014:78) demonstrate that consistency applies at both the structural (data format) level and conceptual (data values or elements) level. In this study, consistency is basically considered on two levels. Firstly, in terms of the elements used, consistency means that the same elements, as prescribed by the scheme applied, are used to describe similar attributes (e.g. subject) and the use of a common structure or format in the subject metadata element (e.g. geographic subject metadata in the format "Country: Province"). Secondly, consistency involves the use of a common value, like a common geographic name, to describe a similar place and captured in a similar format (e.g. "Tshwane", used in place name for old name "Pretoria" and spelled uniformly).

High quality metadata, placed at an incorrect descriptive element, causes recall failure in information searches. Subject metadata has designated elements provided by specific metadata schema or vocabulary. The studies conducted by Zavalina (2010; 2014) demonstrate how subject representations are captured in metadata, sometimes with complementarity experienced in the use of the different elements (e.g. keyword, subject and coverage) to represent the same value. A correct understanding of how the elements are used is important and can be achieved through training. Consistency in the use of the correct metadata elements should be enforced to maintain metadata of quality.

Consistency and quality are ensured by using standards (Hider 2012:81), which can be achieved, for example, in the creation of subject metadata, by enforcing the use of a vocabulary standard (Yasser 2012:374). Furthermore, various standards organisations offer guidelines for quality metadata and stresses consistency in metadata creation. ISO 9000:2015 emphasises the importance of standards in maintaining consistency in processes, in order to achieve required performance. However, the standards are generic and it is the responsibility of every institution to implement them within their contexts. This flexibility leads to varied levels of adherence to the standards and inconsistencies in metadata practices. ETD metadata in a single collection and in shared environments need to be controlled for consistency and to maintain quality that allows interoperability.

Furthermore, consistency is a key quality issue for the interoperability of digital resources, including ETDs. Therefore, interoperability is considered here as being enabled by the sharing of consistent metadata. Consistency in terms of data structure and data values improves access and facilitates interoperability (Tmava & Alemneh 2013:858; Hider 2012:80–81; Dublin Core Metadata Initiative 2018). The NISO principles of good metadata, which endorse quality metadata, list interoperability and authority control as principles for quality metadata as criteria (Riley 2004:10).

Interoperability is defined in Section 3.2.1. Metadata records that do not comply with interoperability standards, inhibit information sharing. Despite this problem, Park and Tosaka (2010:113), who conducted a study on digital resource metadata creation practices, identified the lack of confidence among institutions that the quality of their metadata can meet the interoperability requirements. Ochoa (2014:65) emphasises that interoperability should be a basic characteristic of all information provision systems.

Interoperability may occur at different levels in a repository or resource collection platform. According to Potvin and Thompson (2016:99), interoperability at a collaborative level requires cooperation at technical, content and organisational levels. This argument clarifies that interoperability involves structure (how to share data across different systems through coding schema, e.g. MARC, Dublin Core and the various standards for metadata elements of content presentation) and the values used to represent content.

Adherence to schema facilitates consistent encoding and improves discoverability of information resources (Tmava & Alemneh 2013:858). However, different schema are used to cater for different needs and audiences (Riley 2004:11), resulting in inconsistencies in metadata practices across different systems. The data structure standards (schema) address the ability of the chosen scheme to allow sharing of metadata across different platforms on which information can be searched. All other types of standards in Table 3.4 (p. 74) are useful in metadata creation and in enabling consistency and interoperability.

Value standards are the basis for the interoperability of subject metadata, which are investigated in this study. In a case where records from various collections are integrated and a common vocabulary is used, consistency of subject representations is expected. The standardised use for values in the subject elements ensures consistent description of information resources content and enhances information discovery and interoperability in networked environments (Liebetrau 2010:31). On the other hand, consistency may be a problem, if records based on different data value standards, are to be merged or shared. Furthermore, Hider (2012:175) warns about the difficulty to map vocabularies due to their extensive and complicated nature.

Interoperability is crucial for sharing information from ETDs and metadata quality guidelines have been compiled in an effort to facilitate it. The Electronic Theses and Dissertations Metadata Standard (ETD-MS), which an interoperability metadata standard, provides guidelines for interoperable ETD metadata. These guidelines are meant to facilitate quality metadata creation and global sharing of ETD information.

3.4.2.2 Accuracy

Accuracy is a key aspect that promotes discovery and exchange of digital information. In the context of this study, accuracy is defined as using correct data values to represent the subject content of an information resource (relevance) and lack of errors in capturing them (correctness) in the data values (Park 2009:2020; Hider 2012:79; Ochoa 2014:67). Based on the description of accuracy given by Park (2009:220), subject metadata can be understood as the accurate description of the subject content of a resource. Accuracy is highlighted in this description.

Relevance and correctness of the metadata values are crucial for accurate ETD descriptions. In this study, accuracy specifically deals with the relevance and correctness of geographic subject metadata. Relevance can be judged in terms of matching the attributes of the content described and the data values (geographic subject terms) (Stvilia, Gasser, Twidale & Smith (as cited in Zavalina 2014:78). Correctness can be judged in terms of the absence inaccuracies, like typographical errors (Hider 2012:79), or misspelling. It also involves the use of correct terminology, as prescribed by the vocabulary standard used. Existing literature reveals studies that show concerns about accuracy.

Yasser (2012) conducted a study on metadata errors in digital libraries and repositories and the effect of metadata creation knowledge on accurate metadata creation and confirmed problems related to metadata accuracy. On the other hand, the influence of the perception of the metadata creator on the interpretation of the content of an information resource cannot be ruled out and, therefore, there is always a subjective judgement involved (Ochoa 2014:67). This creates problems in decisions on relevance and populating the subject element with accurate values.

3.4.2.3   Completeness

Completeness is another criterion for the evaluation of metadata quality (Tmava & Alemneh 2013:858). In this study, completeness is understood as the use of all the required metadata elements and comprehensive representation of content by using all important concepts necessary to describe the subject content of an information resource (Zavalina 2014:18). – No essential element for supporting discovery of an information resource should be left out on the item metadata (Yasser 2012:374). The omission of a required element, for example the geographic subject element, makes the descriptive record incomplete. Full adherence to standards and vocabularies assist in the creation of comprehensive metadata. The NDLTD recommends the ETD Metadata Standard (ETD_MS) for maximal exposure of ETDs, by ensuring the creation of complete metadata. Another example of a standard that helps to achieve completeness is the LCHS. The LCSH instruction on the use of topics and geographic subdivisions is a common illustration of how vocabularies support the creation of complete subject metadata.

Despite the presence of quality evaluation criteria that have been discussed, various other issues affect metadata quality. In an attempt to achieve completeness, the problem of redundancy, identified by (Zavalina 2014:80), should be guarded against, e.g. when creating geographic subject matter. Sometimes, problems with geographic subject metadata may relate to other issues, such as the existence of both the "subject" element and the "coverage" element, which both address the geographic attributes of an information resource. Theories and conceptions of subject access in general may provide a basis for understanding geographic subject metadata and how to assess its quality.

3.5   Chapter conclusion

The literature reviewed in this chapter leads to a conclusion that metadata quality is important for information resource discovery and access. The creation and use of metadata evolves with the developments in information resources management. Quality issues are also affected by these changes and different criteria are used to evaluate metadata in the different contexts need to be continuously observed.

The way in which issues of ETD metadata quality are relevant in the digital and networked context was also discussed, as well as the criteria that can be used to measure the quality. There are several criteria for measuring metadata quality , with accuracy, consistency and completeness being regarded as key evaluation criteria.

Additionally, it was also concluded that it was important to investigate how the underlying theoretical viewpoints influenced the creation of quality metadata. The study of metadata quality is interrelated to all metadata development issues, including investigations into the current shifts in theories providing context for metadata creation. Metadata quality and its relationship with the existing theoretical subject analysis principles should be analysed. An underlying theoretical framework facilitates the interpretation of the necessary quality requirements in the different ETD contexts, including the subject description area.

CHAPTER 4: THEORIES AND CONCEPTIONS OF SUBJECT ANALYSIS

4.1   Introduction

This chapter discusses the theoretical and conceptual foundations underlying subject analysis. The purpose of the literature review is to provide a theoretical framework to help understand the process of subject analysis. It also provides a basis for the empirical study of the different metadata creators' subject analysis approaches. The explicit or implicit subject analysis conceptions held by the metadata creators during ETD geographic subject metadata creation are interpreted with the help of the information gathered through the extensive literature review on subject analysis and its theories.

The theory discussions relate to the second research objective (RO2): To collect data from the practitioners involved in metadata creation in South African university libraries on how they approach the determination of geographic subject metadata for ETDs.

The association between the different theoretical and conceptual positions and the approaches to conducting subject analysis is further discussed in Sections 4.2–4.5. Additionally, the theories and conceptions are considered in the context of how subject analysis facilitates the creation of geographic subject metadata for ETDs. The background information from the reviewed literature, together with the research findings of the empirical study help to establish the subject analysis approaches considered most suitable in this context.

Based on this information, a recommendation is made in Chapter 8 for a model or an approach that is considered appropriate to improve subject analysis, as stated in Research Objective (RO4): To establish what type of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries.

The context of ETDs and geographic subject metadata is furthermore considered as fundamental in determining the relevance of the theoretical model or approach that is recommended in this study.

The discussions in this chapter are categorised according to influence of the basic theories and the different historical views on LIS practices considered as a suitable foundation for subject analysis, followed by discussions on a specific theoretical-conceptual model that represents the different approaches to subject analysis. The selected model serves as a representation of how subject analysis is conducted and helps to explain the principles that are commonly followed in the subject analysis process. The model is discussed from different perspectives presented in LIS literature that cover topics on the conceptualisation of a "subject" and subject analysis discussed by, amongst others, Soergel (1985, 1995), Albrachtsen (1993), Fourie (2008, 2010), Hjorland (1998, 2017). The importance of establishing a theoretical base for the practice of subject analysis is further explained below.

4.2 The importance of subject analysis theoretical basis for ETD geographic subject metadata creation

The importance of the subject analysis theoretical framework is firstly discussed by considering the need for theory re-examination in the ETD context and their geographic subject metadata. Hjorland (1992:195) emphasises the importance of the whole context in which subject description takes place. The broad context for the theoretical discussions in this chapter is derived from the ETD background, as discussed in Chapter 2.

It is evident that the development of ETDs has introduced changes in terms of accessibility of research information and the required descriptive metadata to enhance access. The ETD developmental factors identified in Chapter 2 include: specific aspects of ETD digital format; the use of both intellectual and automated/automatic assisted subject analysis and the emergence of other technology enabled subject representation practices, such as user-contributed tags. As a result of these changes, the possibility of a shift in the theory and practice of subject analysis practices needs to be investigated.

The importance of a theoretical basis for the changes related to ETD metadata and subject analysis is pointed out by Potvin and Thompson (2016:102), who present examples of shifts in ETD metadata and related theory issues. The first change is the move from traditional, simplified cataloguing theories and approaches for print materials to a new approach to ETD metadata, treating them as surrogates to information resources. This argument suggests that the metadata should be the best representation possible for ETDs. The second change concerns the shift towards managing electronic and networked objects for long-term access and curation. Quality and interoperable metadata is important are make these resources discoverable. Furthermore, Alemneh and Phillips (2016:4) discuss the quality issues of ETD representations and suggest a need to revisit the traditional approaches on which information resource descriptions were based. It is necessary to re-examine the application of traditional analysis theories and whether they remain relevant for ETDs, or if they need to be interpreted differently to suit the description requirements of these digital resources.

Furthermore, the importance of the theoretical perspectives underlying the subject analysis stage – particularly for geographic metadata creation, or the use of "place" in subject description – is an additional factor that provides context for the discussions in this chapter. Subject analysis that is based on appropriate theoretical foundations facilitates suitable geographic subject description. Literature on the theoretical foundations applicable to geographic subject metadata and its analysis is lacking. The theoretical foundations applicable to subject analysis for geographic subject metadata creation are considered as being embedded in the general subject analysis theory literature. These contextual factors add a different dimension to the conceptions and approaches to subject analysis, while further complicating the process. Theory investigations can help to resolve these complexities.

The role of theory is to help address the complex nature of subject analysis that is partly resulting from the discussed contextual factors. Hjorland (1992:179) indicates that there is no fixed procedure to guide subject analysis – mainly because of the complex nature of the subject analysis process, as pointed out by Fourie (2008:113) and (International Federation of Library Associations and Institutions. Working Group on the Functional Requirements for Subject Authority Records 2010:10).

This assertion was lately confirmed by Howarth and Olson (2016:3), who argue that the identification of a subject is slippery. The subject analysis theory base is considered in the light of these common problems. Therefore, understanding the different theories and conceptions facilitates practical subject analysis and helps to resolve the complexities in relation to subject analysis; particularly for the creation of ETD geographic subject metadata.

In relation to the themes for discussion identified in Section 4.1, the importance of a theory for subject analysis is aligned to the need for theory to provide a basis for LIS practices in general. Bates (1986, 2005) emphasises the significance of the underlying theories of LIS in research and practice. Fuller, Hjorland, Ibekwe-Sanjuan, Ma et al (2013:1) support the importance of theory as a basis for LIS practices, while Hjorland (1998) specifically points out that theory is the basis for an explanation of the functions of different search elements and descriptors. The function of subject descriptors and their analysis is performed for the descriptions in the current context, founded on theory. It is important to base the discussions in this chapter on the role of a theoretical framework in helping to understand the functions of subject metadata and the basis of subject analysis. Therefore, it is considered important to reflect on the question asked by Howarth and Olson (2016:3) as to whether the traditional subject analysis theoretical principles in relation to the representation of an information resource subject are still relevant in the current subject access context.

This question can be addressed by re-examining the traditional theories, their epistemologies and their application in the current subject access context. Furthermore, it is also important to understand how the epistemological positions and ideologies held by major contributors to the history of LIS influenced their understanding of LIS practices, including subject analysis. According to Hjorland (2013:175), epistemological theories influence the way in which subject representation is practiced. Therefore, it is important to establish how the different epistemological positions can foster an improved understanding of subject analysis. In addition, it is important to understand how these theoretical positions relate to the conceptions of "subject" and subject analysis.

In addition, the importance of theory is viewed in relation to how it assists the identification of an appropriate subject analysis model, as indicated in Section 4.1. Hadar and Soffer (2006:569) argue that there is no correct model in any specific discipline, but preference depends on the theoretical viewpoint that is adopted. The discussions in this chapter follow the ideas raised here on the importance of theory in understanding the subject analysis model. Figure 4.1 illustrates the overlapping relationships of theories and conceptions that provide a theoretical basis for the conceptual model to the subject analysis approaches.
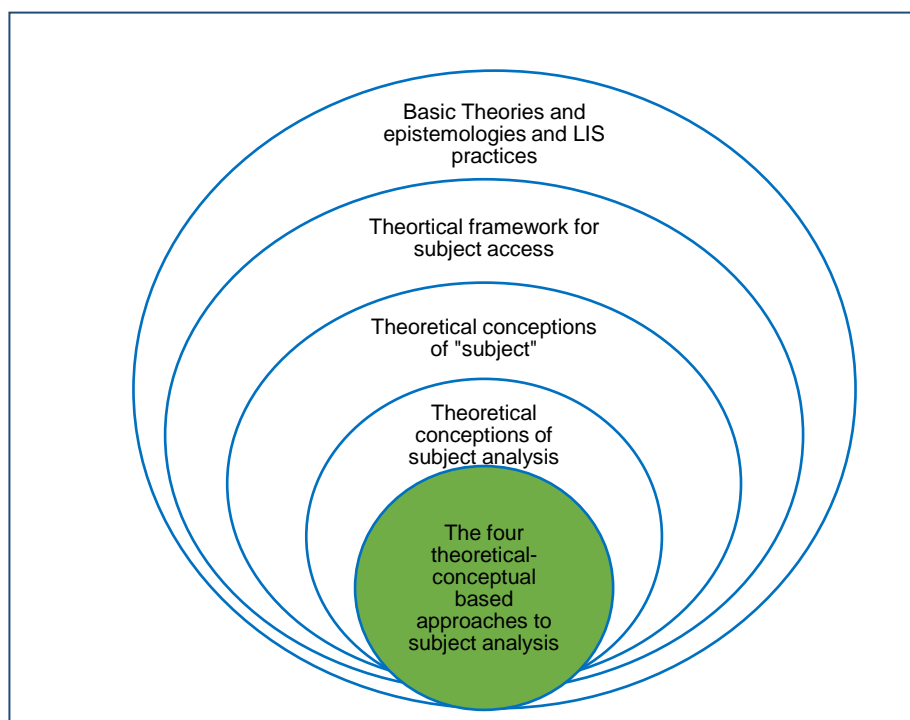
Figure 4.1: Theoretical basis for the subject analysis conceptual model

The first theory aspect to be discussed in the next section is the basic theories and epistemologies that provide a general theoretical framework for LIS practices. These basic theories provide metadata creators with a theoretical and conceptual basis for subject analysis that is conducted for the creation of ETD geographic subject metadata.

4.2.1 Basic theories and epistemologies and LIS practices

The basic theoretical positions commonly form the basis for all theory discussions. As a basis for the theoretical discussions in this study, it is important to consider the relationship drawn in LIS literature between the epistemological positions and LIS practices, including subject analysis. Additionally, it is necessary to understand the basic theoretical positions, in order to understand the theoretical and conceptual foundations of subject analysis. A brief analysis of theories is important to understand the basis of the conceptual model used as a reference for positioning subject analysis theoretically.

Mai's (1999:287) view of epistemological positions forming the basis for various conceptions of indexing, is considered applicable across all subject representation methods, including subject metadata creation. In addition, Hjorland (2002; 2013:17) and Dutta and Dutta (2013:82) support the need for theory and a specific epistemological position on which to base information resources representation approaches. In general, an understanding of the epistemologies is considered as being fundamental to solving LIS problems and is considered to form an interdisciplinary foundation for theories in the LIS field (Hjorland 2002:268).

Furthermore, the basic epistemological positions are considered important to classify the conceptions of "subject" and subject analysis, as discussed in Sections 4.2.3 and 4.3. Hjorland (1998:619) indicates that epistemological positions impact on theories of subject analysis. The subject analysis conceptions and approaches discussed in Section 4.4.2 are influenced by these epistemologies. On the other hand, Dutta and Dutta (2013:82) acknowledge that there is little epistemological knowledge in LIS. This situation exists despite the evident impact of epistemological positions on the LIS practices.

Over the years, Hjorland (2013; 2009; 2004; 2002; 1992) has contributed extensively to the explanation of the basic theoretical positions, arguing their importance and indicating their relevance to LIS practices.

Tennis (2005:20) regards Hjorland's contribution as foundational work for the study of LIS theory; in particular that for subject analysis and the various ways in which it is conceptualised. These epistemological positions are discussed in more detail in Chapter 5.

Hjorland (1992:172; 2013:173) presents the following four common epistemological positions: empiricism, rationalism, historicism and pragmatism:

- *Empiricism*
  Knowledge is primarily based on experience and gained through observation and induction (Hjorland 2004:134). Empiricism is regarded as similar to the present day understanding of positivism, in placing less emphasis on theory due to the argument that it introduces subjectivity (Hjorland 2009:1523; 2004:135).
- *Rationalism*
  Knowledge is gained independently of experience, but based on conceptual clarity and evidence (Hjorland 2004:134 & 135).
- *Historicism*
  Knowledge of history determines understanding of issues. Social context and historical understanding and an individuals' pre-understanding shape knowledge (Hjorland 2009:1525). Historicism fosters the idea of attributing significance to space and time, such as historical period, geographical place and local culture.
- *Pragmatism*
  Focuses on practical or logical response meaning and truth, based on observable practical consequences. From an LIS perspective, Hjorland (1992:179) holds the opinion that subject data has a pragmatic function, and the assignment of subject representations must be done in anticipation of the users' needs that are related to a specific problem for which information is required.

Hjorland (2009:1520) regards pragmatism as the most relevant and closely related approach to study the theory of concepts in Information Science and knowledge organisation. The understanding of the concepts from a pragmatic position is considered helpful in situating the discussions in the unique context discussed in the next paragraph. Furthermore, Hjorland (2013:174) considers empiricism and rationalism as lacking in considering the social and historical nature of information.

However, Fuller et al (2013:2) maintain that the boundaries between e different epistemological theories are not clearly defined. They further indicate that the involvement of technology and the social aspect introduced a complementarity in epistemologies that characterises Information Science and necessitates investigating the possible reconciliation of the different epistemologies. This position may have an effect on how the influence of theories and conceptions are argued in the LIS theories field.

Furthermore, pragmatism is adopted as the theoretical position in this study to provide a basis for understanding subject access and to assist with the adoption of suitable conceptions for subject analysis. In this study, the understanding of subject access, specifically subject analysis is based on the circumstances and practical application, which involve the creation of ETD geographic subject metadata and in relation to anticipation of the potential information user needs to be satisfied by ETDs. This interpretation is pragmatic in approach (Mai 2010). It is important to understand how theory influences subject access and its different approaches in general, as a basis for subject analysis discussions.

4.2.2   Theoretical framework for subject access

Fuller et al (2013:1) observe that there is an indication in LIS literature that, in the past more attention was paid to the practical aspects of LIS practices and less on the theoretical aspects. The dominance of the practical guideline approach compromised the understanding of the theoretical aspects of various practices, including subject analysis. Fuller et al (2013:2) call for attention to a theory basis in the LIS field. Literature that emphasises the importance of a theoretical framework can also be traced to broader LIS theory literature used in this chapter.

Subject analysis theory is often identified from literature that broadly discusses theories on subject access practices like indexing, subject cataloguing and classification. Fuller et al (2013:1) constituted a panel for the ASIST 2013, where emphasis was put on the significance of the underlying theories of LIS in its research and practices.

These authors consider the theory of subject access in its various forms, while highlighting the argument that the basis for all scientific approaches can be categorised under epistemological positions discussed in Section 4.2.1. The theory of subject access is linked to these epistemologies. On the other hand, the theory of subject analysis is embedded in that of subject access and, therefore, also being influenced by the different epistemologies.

Furthermore, Howarth and Olson (2016) and Zavalina (2012), among others, confirm the need to re-examine the theoretical basis for subject access and to compare how change affects metadata practices in the new information organisation contexts, as opposed to the traditional metadata. In relation to this argument, Mai (2010:628) observes that developments in information organisation, mostly driven by technology, present the need for reconceptualisation. In agreement, Bates (1986; 2005), Tennis (2005) and Hjorland (2017) state that a proper conceptual or theoretical basis contributes to improve access to the increasing information resources by aiding that the subject representations provided are relevant to the users' needs.

A few examples of conceptual approaches are learned from subject access literature. Tennis (2005) compares different conceptions of subject analysis and identifies "aboutness'" as a top level attribute of conceptions of subject analysis identified among the compared theoretical works. Additionally, a significant contribution to conceptions of "subject" and subject analysis is made by Hjorland in his different works. Hjorland (2009) argues that there is a connection between the different theoretical positions and conceptions of "subject", further arguing for the social meaning of concepts. Fairly recently, Hjorland (2017) has continued to argue for the importance of the review of concepts in LIS and how theoretical positions help to understand the concept of "subject" during subject analysis.

The discussed theoretical positions provide a basis for different ways in which the concept of "subject" is understood (Hjorland 1992), while, in turn, influencing how subject analysis is approached.

4.2.3   Theoretical conceptions of "subject"

Epistemology shapes the understanding of concepts in a specific discipline, and the "… different conception of subject can therefore be classified into epistemological positions" (Hjorland 1992:172). The epistemological positions discussed in Section 4.2.1 are the foundation of the theory of the determination of a subject (Hjorland 1992:172; 2002:173). Hjorland (1992:181) finds pragmatism a fitting approach to understanding the concept of "subject". This view is maintained despite the limitations of pragmatic conception of a "subject", which is aligned to how a subject may be interpreted from a user's perspective. One of the limitations is that a pragmatic approach has an impact on the priority of assigned subjects and the implied assignment of many subjects (Hjorland 1992:181). Based on the understanding of the different theoretical positions discussed in Section 4.2, it can be concluded that the understanding of "subject" is affected by any epistemological position advanced as an ideal theory.

Hjorland (2017:62), who points out the different conceptions and the association between the concept of "subject" and the process of subject representation, states that:

> *Any approach to subject representation is connected to a certain understanding of a "subject", which is often implicit.*

Based on this view, subject analysis that is conducted for creating subject representations requires a specific understanding of "subject" and how it is determined. Based on the common use of a subject to represent information resources content and searching, Dutta and Dutta (2013:79) and Hjorland (2017:55) consider "subject" as a fundamental concept in the LIS field. Therefore, its meaning is significant in this study.

The understanding of what a "subject" is, has alternatively been interpreted as the determination of "aboutness" of an information resource. Beghtol (1986:84) and IFLA FRSAR (International Federation of Library Associations and Institutions (IFLA).

The Working Group on the Functional Requirements for Subject Authority Records 2010:10) indicates the need for a strong definition of "aboutness" and its theoretical foundations in relation to various LIS practices. In addition, different opinions are raised in the literature about the practice of determining the subject of an information resource and its changing meaning. The continued discussions on this topic are necessary for a practical meaning of "subject" in the context of subject analysis. Howarth and Olson (2016:7) express the need to rethink the principles of subject determination, in view of the contemporary developments, such as social tagging, and how they influence subject access. Social tagging, which is the practice of users participating in the allocation of subject representations, was discussed in more detail in Section 3.4.1 (Chapter 3). An examination of the major LIS theoretical arguments that shape the meaning of "subject" is important to give perspective to how the concept of "subject" can be conceptualised.

The different theoretical and the epistemological positions are considered from an LIS historical perspective to provide a basis for specific LIS theoretical positions. These LIS theories influence the understanding of "subject" and of subject analysis. A brief analysis of these historical LIS perspectives shows the development of the meaning of "subject", which determines how subject analysis is understood and practiced. Hjorland (1992; 2017:56–58) gives a comparative analysis of major historical views on "subject", held by LIS theoreticians. Dutta and Dutta (2013:82) give a further analysis of these LIS historical positions and confirm that they shape the understanding of "subject". The analysis helps to derive a suitable meaning of the concept of "subject" in the LIS field that is specifically relevant to enable access to specific subjects covered in information resources. Different reflections on the meaning of "subject" were made in different times by theoreticians in the LIS field. Examples of the historical LIS theoreticians and their arguments on the meaning of "subject" are outlined in Table 4.1 (p. 98).

Table 4.1: Historical LIS theoretical positions on "subject"

Theoretical position	Views on "subject"
Cutter	The meaning of a subject is determined by the social context and subjects are shaped in social processes. Hjorland (2017a:57) considers Cutter's views on the social nature of subject as more appropriate. However, the social nature is not clearly defined by Cutter (Hjorland 2017a:57). The relations between the understanding of the concept of subject and Cutter's functions of the subject element is presented in Section 2.3.3 (Chapter 2).
Ranganathan	Different definitions of a subject: • An assumed term (Hjorland 1992:177; Dutta & Dutta 2013:80). A subject is understood as an idea that cannot be concretely identified through the content of a document. Hjorland (2017:57) is of the opinion that this definition aligns to the content-oriented view (discussed in Section 4.4.2.2). • An organised body of ideas whose intentions are likely to fall coherently within the field of interest and comfortably within the intellectual competence and the field of inevitable specialisation of a normal person. Ranganathan is criticised, because he is considered as defining the concepts "subject" in a way that suits his Colon Classification system and is not suitable for general scientific use (Hjorland 2017a:57). The organised and logical nature of subject is emphasised, but the definition in itself offers no guidance for subject analysis (Hjorland 2017a:57).
Wilson	Wilson concludes that the determination of a subject can never be definite (Dutta & Dutta 2013:80). Hjorland (1992:176) regards this view as unsuitable to define the concept of "subject" for use in the LIS context.

According to (Hjorland 1992:177), Ranganathan and Wilson are more inclined to an idealistic conception of the meaning of "subject", which does not match the idea of subjects in people's minds. Additionally, this approach is considered lacking in terms of the pragmatic aspects of a subject that recognise its potential to inform the users of information resources and that do not guarantee the correct subject analysis (Hjorland 1992:176–179). Therefore, the views of Ranganathan and Wilson are considered inconsistent with the definition of the subject given in this study, which embraces the information potential of a subject of an information resource.

On the other hand, Cutter's approach is regarded as being more pragmatic. Dimec, Zumer and Riesthuis (2005:215) support the relation of subject access to Cutter's position on the importance of a subject element in finding information resources. Cutter maintains that subjects can be determined, but the stability of subjects and their meanings can be maintained in specific social contexts (Dutta & Dutta 2013:80). In addition, Hjorland (2017:56) traces the chronological history of the theory of the conceptions of a "subject" and finds a more significant link to the understanding of Cutter's theoretical views on the principles of subject cataloguing. The principles are, among others, the importance of assisting a person to identify a book if the subject is known, and to show what the library has on a given subject (Cutter 1904:12). Section 2.3.2 (Chapter 2) also discusses the influence of Cutter's principles on the understanding of the concept "subject". In addition, Hjorland (2017:57) holds the opinion that Cutter's viewpoints are more meaningful by recognising the social element that influence how the meaning of "subject" is determined. Different approaches adopted for LIS practices, including the subject analysis approaches discussed in Sections 4.3.2.1–4.3.2.4 show an association with the discussed historical views.

Furthermore, the understanding of the philosophical meaning of the concept "subject" facilitates the theoretical approach to the study of subject analysis and serves as a basis for the empirical investigations conducted in this study. Hjorland (1997:40) holds the view that the theory of subject analysis needs the theory of a subject. The theory of subject analysis is expressed in the literature as being embedded in the historical development of the meaning of "subject". A good conception of "subject" enables meaningful subject analysis. Furthermore, Tennis's (2005) study endorses the importance of the theoretical basis for subject analysis and compares different theoretical conceptions. The study reveals the common and different attributes of the conceptions that improve the understanding of the contemporary practice of subject analysis. Such studies confirm the evolving complexities of the conceptions of subject analysis.

Albrachtsen (1993:219) poses the question as to whether there are different conceptions of a subject and hence of subject analysis. In addressing this question, different conceptions of subject are highlighted in this section, while the model for conceptualising subject analysis is discussed in the next section.

The definition of "subject" adopted in this study is that "The subject of a message is its informative potential" (Zeng, Zumer & Salaba 2010:10). This can be further explained according to the understanding of Zavalina (2014:78) and Soergel (2009:25), in that the subject is what the resource is about and what it is relevant for. The conceptual model, discussed in detail in Section 4.3.2, represents and distinguishes the different conceptions of subject analysis and serves as a framework to identify the different principles that guide the process

4.3   A theoretical-conceptual approach to subject analysis

There is no evidence of literature that specifically deals with theories and conceptions of subject analysis for geographical subject metadata. However, existing literature on subject analysis does provide a basis for the general subject analysis theoretical framework. Tennis (2005:16) argues that literature discussions on the theories of subject representation reveal differences in the views and approaches towards subject analysis, but argues that more needs to be learned about the role of theory in the contemporary practice of subject analysis. However, contributions in the LIS literature like those of Cutter, Wilson and Ranganathan discussed in Section 4.2.3, including recent contributions of authors like Hjorland (2021 & 2017), indicate continuing efforts to translate theory into effective subject analysis practice. These range from deliberations on the general epistemologies, discussed in Section 4.2.3, to conceptions that shape specific theoretical subject analysis approaches.

Subject analysis theories are information organisation theories that guide and lay a general philosophical foundation for subject analysis. These theories, which serve as a foundation for the principles of subject analysis and subject metadata creation are mostly advanced in cataloguing and indexing literature.

Section 4.1 clarifies the association between indexing and subject cataloguing or subject metadata creation and their common basis, which is the practice of subject analysis.

On the other hand, it is also noted that there is confusion in the understanding of the subject analysis theories and theories about the use of tools like indexes (Tennis 2005:14). Tennis (2005:14) also notes how this ambiguity in the understanding of theoretical aspects and the importance of subject analysis negatively affect the detailed discussions on the subject analysis process.

A chosen theoretical approach has an effect on all stages of subject metadata creation. Subject analysis is the crucial stage in subject metadata creation where these theories and conceptions have effect. Hjorland (1998:606) and Dutta and Dutta (2010:81) observe that the epistemological positions have an impact on the development of theories of subject analysis. However, it is acknowledged that epistemology does not give the final answers to subject analysis, but provide knowledge of the merits and weaknesses of different solutions (Hjorland 1998:613). Therefore, specific subject analysis theories need to be understood. LIS literature presents the understanding of subject analysis from different theoretical perspectives. Tennis (2005:3) argues that subject analysis can only be understood through its different conceptions.

In addition, Guimaraes (2017) demonstrate the importance of theories and conceptions as they relate to the practices of providing access to information. Continuing investigations on subject analysis or determination of subjects form part of the necessary general re-examination of subject access. As shown in 4.2.3, theories of subject analysis are associated with general subject access theories. They are also linked to theories about the concept of "subject" (Hjorland 1997:40). Therefore, conceptions of subject analysis are based on the understanding of subject access the interpretation of "subject". As the meaning of subject access and "subject" is analysed in different contexts, it is also necessary to revisit the different theoretical and conceptual models of subject analysis. This re-examination will help to determine the relevance of subject analysis theories and conceptions in the contemporary context, which includes the digitisation and ETD era.

4.3.1 Theoretical-conceptual approaches to subject analysis

This section discusses the theoretical and conceptual aspects of subject analysis in relation to ETD geographic subject metadata creation. The discussions of the theoretical aspects are conducted according to a conceptual model. The wide-ranging conceptions of subject analysis are not elaborated on, but a common established conceptual model on which subject analysis approaches are based is identified in LIS literature. Scholars like Hjorland (2017; 2013), Zavalina (2012), Chung, Miksa & Hastings (2010), Soergel (2009; 1985), Chung (2006), Chung and Hastings (2006); Fourie (2005); Tennis (2005), Mai (2005) and Albrachtsen (1993) elaborate on different conceptual positions that may be used as a theoretical basis for subject analysis. The different conceptions in respect to subject analysis can be defined as perceptions, viewpoints or approaches about subject analysis (Chung 2006:22; Chung & Hastings 2006:6; Mai 2005; Soergel 1985; Fidel 1994; Albrachtsen 1993).

An existing and generally accepted conceptual model arising from the contributions made by the different authors listed above is used to show different types of emphasis on conceptions of subject analysis. The views presented are from different forms of subject representation practices, including indexing, cataloguing and metadata creation.

Albrachtsen (1993:220) names the conceptual model approach as the "model of conceptions" of subject analysis. This conceptual model conveys the fundamental principles of subject analysis. In addition, the results of the empirical study reported on in the next chapters on the current approaches followed during subject analysis, are analysed in relation to these conceptions of subject analysis. These conceptions are considered as being explicit or implicit in the process of subject analysis. Albrachtsen (1993:220) names the three different conceptions or viewpoints of subject analysis as: (i) simplistic conception; (ii) content-oriented conception; and (iii) requirement-oriented conception. The different contributions made in the literature are discussed to help understand this conceptual model.

The different conceptions on "subject" and subject analysis may derive from any of these positions. There is often an observed interrelatedness among the different conceptions in the theoretical discussions and in the practice of subject representation (Fidel 1994:572). The theoretical principles drawn from these conceptual positions provide a theoretical framework for subject analysis.

Mai (2005:599) regards analysis in the context of subject representation as focused on t either he document or on the user. In addition, Hjorland (2017:55) identifies the most important distinction of theoretical positions on subject representation that are related to the concept ''subject'' as:

- Document-oriented conception (maintains that subjects are inherent within the documents)
- Request-oriented or user oriented conception (maintains that subjects are attributed to documents in order to facilitate their use)

These conceptions are considered helpful to guide subject analysis. Other common names or categorisations are often used for the conceptions of subject analysis, some of which are used interchangeably or in addition to the ones listed above. The terms *entity-oriented, document-oriented, user-oriented, need-oriented, problem-oriented, domain-oriented* and *socio-cognitive view* are often used to identify the different conceptions of a "subject", subject analysis and subject representation practices (Fourie 2008:117 & 2005:26; Chung & Hastings 2006:6; Soergel 1985:236).

In addition to the three conceptions named by Albrachtsen (1993:220), the socio-cognitive conception or view is also discussed in this section, in order to include a common perspective that is based on domains and their influence on subject analysis approaches. The focus on domains is viewed in the current LIS literature as the ideal theoretical position for subject representation (Campos & Gomez 2017:179; Hjorland 2017a:60). This position is aligned to the socio-cognitive or domain-oriented subject analysis approach discussed in Section 4.4.2.4.

According to Guimaraes (2017:90), Hjorland leads the support of the socio-cognitive conception in the LIS context.

On the other hand, Chung (2006:25) presents what he calls the "convergent perspective" to simplify the interrelatedness of the different conceptions of subject analysis and categorises them into three: content-oriented, document-oriented and domain-oriented conceptions. In this study, a four conceptions model – consisting of the simplistic-oriented, content-oriented, request-oriented and socio-cognitive approaches – is used as a theoretical basis for the study of the subject analysis approaches used by subject metadata creators. Table 4.2 (p. indicates the interconnectedness of the different approaches, based on the different conceptions, the type of information considered during subject analysis and the method used to identify the concepts to be used for subject representation. The applicability and relevance of these theoretical viewpoints in different contexts need to be investigated and confirmed.

4.3.2   Four theoretical-conceptual based approaches to subject analysis

Based on the discussions above, the four approaches to subject analysis, which are the simplistic approach, content-oriented approach, request-oriented approach and socio-cognitive approach, were drawn from the identified theoretical conceptions. The different conceptions of subject analysis lead to different approaches discussed in this section. The approaches show different levels of analysis.

Table 4.2 (p. 105) represents the four approaches to subject analysis.

Table 4.2: Approaches to subject analysis and their interconnections

Approaches to subject analysis	Type of subject information considered during analysis	Method of subject representations construction
Simplistic approach	Explicit information contained in an information resource	Subject representations extracted from the information resource
Content-oriented approach	Explicit and implicit information contained in an information resource	Assigned subject representations
Request-oriented approach	Explicit and implicit information contained in an information resource	Assigned subject representations that consider users' information needs and the potential requests for information
Socio-cognitive approach	Explicit and implicit information contained in an information resource	Assigned subject which takes into account the information needs within specific domains

(Adapted from: Albrachtsen 1993:220)

4.3.2.1   The simplistic approach

According to the simplistic approach to subject analysis, concepts used to represent the subject content are derived from the document itself (Albrachtsen 1993:220). Fourie (2008:117) equates the simplistic approach to the entity-oriented approach, because the content representation terms are directly selected from the entity. The subject analysis is conducted in a manner limited by this perspective of selecting terms from the document, and the meaning of the content of the information resource is not considered. Matanji (2012:167) proposes that this approach may be relevant for representing new concepts or in areas of high specialisation. The simplistic approach is considered to have limitations in terms of efficiency in subject analysis.

4.3.2.2   The content-oriented approach

The content-oriented approach to subject analysis involves the representation of subjects in information resources, either explicitly by using the terminology as used in the document, or implicitly as interpreted by a human being reading the document (Albrachtsen 1993:220). The content-oriented subject analysis approach bases description on the attributes of an information resource only (Hjorland 1997:76; 1992:180), or on dominant elements in a document (Chung & Hastings 2006:6).

Due to this characteristic, (Hjorland 1997:77) classifies the content-oriented conception as related to a rationalistic view of knowledge. Rationalism, as discussed in Section 4.2.1, places significance on reason rather that experience.

Fourie (2008:118) states that, similar to the simplistic approach, the content-oriented approach does not consider the users' needs. As a result, these subject analysis approaches do not cater for the use of representations that reflect the potential uses of the information contained in an information resource (Albrachtsen 1993:223).

The content-oriented approach is also criticised, because it considers an item as an isolated entity outside its context of the document collection to which it belongs (Albrachtsen 1993:221). Therefore, it may be difficult to reveal the interrelatedness of information resources with this approach (Hjorland 1992:180). Subject analysis that is based on this approach will provide subject representations that focus on describing the information resources as an individual entity only.

On the other hand, (Fourie 2002:87, 2008:117) argues that both the content-oriented and the simplistic approaches can be compared with the entity-oriented or document-oriented approach. The argument is based on the fact that the entity-oriented approach is focused solely on the entity being analysed and described as accurately as possible (Soergel 1985:227), which is a characteristic of the two approaches. However, the content-oriented approach, in contrast to the simplistic approach, is deemed to cater for the interpretation of the content of an information resource (Fourie 2005:24). Furthermore, Chung and Hastings (2006:6) hold the opinion that the content-oriented approach to subject analysis lies in the boundary between keyword-based subject indexing and subject term assignment-based indexing. This versatility is outlined in Table 4.1 (p. 98). Content-oriented subject analysis results in the assignment of objective terms, as opposed to merely extracting keywords (Chung, Miksa & Hastings 2010:690; Chung & Hastings 2006:6; Albrachtsen 1993:220).

Chung and Hastings (2006:7) and Chung, Miksa and Hastings (2010:690) outline further differences between the content-oriented and document-oriented conception based approaches in the context of automatic subject term assignment, which are considered relevant to intellectual subject analysis.

Both the content-oriented and document-oriented conception based approaches inform subject analysis that is based on the analyses of the actual document content. However, the latter requires the identification of the author's intentions, as they are considered indexable subject matter (Chung, Miksa & Hastings 2010:690). Therefore, the document-oriented approach allows the representation of the implicit information in an information resource. Mai (2005:600) introduces the variant of the document-oriented approach, which is called the document-centred conception.

(Mai 2005:600) indicates that the variant of the document-oriented approach is the domain-centred approach. The commonality of the document-oriented and the document-centred approaches, as depicted by Mai (2005:600), is the lack of focus on context during subject matter determination. Their difference lies in the fact that the document-oriented approach does not consider user needs during selection of concepts, whereas the users' needs and questions are kept in mind with the document-centred approach (Mai 2005:600).

Automatic subject analysis is presented differently in the literature, in relation to the different subject analysis approaches. Automatic indexing is categorised as being document-oriented in approach, with the focus being on the stored text in an information resource (Fidel 1994:575; Anderson & Perez-Carballo 2001:233). On the other hand, Hjorland (2017:59) investigated the influence of automation on subject representation and regards automated analysis as not being purely document-oriented. This is due to its approach of word frequency count that does not sufficiently discriminate the concepts according to their importance.

Further limitations of automated subject analysis are based on the idea that automated subject analysis may not guarantee identification of terms that represent major topics in an information resource. The dependence of automated subject analysis on word frequency count only, may fail to fairly assign value to important concepts within an information resource. Based on this argument, automated subject analysis is considered as an example of a simplistic approach to analysis. In addition to the stated simplistic conception inefficiency, automated subject analysis is criticised for its lack of good theoretical foundation and poor basis for practice (Albrachtsen 1993:221; Chung, Miksa & Hastings 2010:688).

4.3.2.3 The request-oriented approach

According to Soergel (1985:230), the request-oriented approach focuses on the anticipated queries and the subject analysis is guided by that approach. The request-oriented approach, which is also regarded as a user-centred or problem-oriented approach (Fidel 1994:72; Soergel 2009:28), operates based on the potential queries and how relevant information to match these queries is retrieved from information resources (Soergel 1985:65, Golub et al 2016:6). According to the request-oriented approach, the concepts used to represent the information resource content should accommodate the user's search approaches (Fourie 2005:24; Soergel 2009:28; Golub et al 2016:6).

The common argument in the literature is that this approach entails a detailed study and analysis of the needs or requirements of the potential users, in order to obtain knowledge about potential queries (Matanji 2012:21; Chung 2006:25; Soergel 1985:54). Both real and anticipated user needs play a role in the requirement-oriented subject determination (Hjorland 1997:77). Hjorland (1992:180) considers this as a pragmatic approach that considers the user as having a specific need for information or a problem, which has a potential to be addressed through a subject search. Therefore, according to the request-oriented approach, subject analysis should be pragmatic in its approach and be focused on satisfying these specific user needs.

On the other hand, the request-oriented approach may be wrongly confused with the document-centred approach mentioned by Mai (2005:600), because they both accommodate user needs. Mai (2005:600) points to the differences between the document-centred approach and the user-oriented or request-oriented approach to subject analysis. The former does not necessarily require knowledge of the user's needs: it merely requires keeping the users' needs in mind during subject determination, whereas the latter does (Mai 2005:600). Furthermore, request-oriented subject analysis may involve prior assessment of the user needs.

Another characteristic of the request-oriented approach is that it accommodates continuous change, to a certain degree, by catering for change in the users' information needs during the allocation of subject representations.

This happens by considering potential future information needs, as pointed out. The challenge posed by this approach is the difficulty to predict all future information user needs (Albrachtsen 1993:223). Due to the unlimited possible uses of an information resource content, the aim of subject analysis should be to identify the most important potential in relation to user needs (Hjorland 2017a:62). Another shortcoming of the request-oriented approach is that it can be highly subjective (Hjorland 2017a:59). The ultimate decision to determine the representation lies with the metadata creator, who may also be subjective, due to being influenced by situational and other factors.

From a related perspective, Albrachtsen (1993:222) indicates that the requirement-oriented analysis boards between the request-oriented analysis and the sociological-epistemological analysis. According to Albrachtsen (1993:222), the requirement-oriented analysis requires prediction of the potential use of the information resource content and social epistemology introduces the element of user domains to subject analysis. Furthermore, Albrachtsen (1993:222) associates the requirements-oriented approach with both approaches and claims that this position gives it an advantage, in that it supports broad and open transfer of knowledge. This can be considered a way of aligning to the domain-oriented approach.

Furthermore, Hjorland (2017:447), Chung (2006:28), Fourie (2002:83), Mai (2005:600) and Soergel (1985:232) concur that the request-oriented or user-oriented analysis allows focus on interests of specific groups or is domain-oriented. Therefore, the request-oriented approach aligns to a socio-cognitive approach to subject analysis (Fourie 2008:118). The socio-cognitive approach is also named as the *domain-analytic approach* (Hoffman 2009:31; Hjorland 2002:257).

4.3.2.4   Socio-cognitive approach

Bates (2005) opines that people form social or intellectual domains. There is growing recognition that, in order to understand information seeking, the domains should be understood. Domains are accommodated within a subject analysis socio-cognitive approach. A socio-cognitive approach requires a document to be analysed with the purpose of predicting its potential for serving a particular group of users (Albrachtsen (1993:222).

According to socio-cognitivism, the focus on the user as part of a domain should be maintained at all stages of subject representation, including subject analysis (Hoffman 2009:31). Additionally, Hjorland and Albrachtsen (1995:400) hold the opinion that domain-analysis has always been an implicit theoretical basis for past and contemporary approaches to Information Science. Additionally, Ranganathan is considered a predecessor of domain-analysis through his categorisation of knowledge into facets or specific categories, which was the basis for classification according to subject disciplines (Hjorland & Albrachtsen 1995:403).

Jansen, Orland, Hartel and Huvila (2016) further illustrate that the socio-cognitive approach introduces the social aspect to the relationship between people and information. It expands on the request-oriented approach by incorporating the element of looking at users from discourse communities or domains to which they belong. Fourie (2008:118) defines *discourse communities* as the knowledge producing, knowledge sharing and knowledge consuming communities. The domain scope determines the boundaries within which subject analysis should be limited. Based on this conception, the approach to subject analysis should focus on the creation of metadata that accommodates the user needs within their own domains. In order to achieve this objective, the analysis of a document to determine its subject matter should be preceded by the analysis of context (Mai 2005:609). Information on the context is essential in determining the users' information needs. Such an approach is considered to improve the translation stage by ensuring that terminologies used in a specific domain are incorporated in the subject representation (Matanji 2012:167).

In addition, Hjorland (1992:180) holds the opinion that this approach may be highly suitable for specialised information services that serve the needs of specific target groups. Similarly, Albrachtsen (1993:223) and (Fourie 2008:118) support socio-cognitivism, on the notion that the social space or contextual frameworks within which subject analysis and subject metadata creation operate is of utmost importance. This view incorporates the surrounding information of an information resource, thereby allowing consideration of the user's possible future needs within specific domains (Hjorland 1997:41; Chung 2006:28). The user is not considered as an individual and the subject analysis is conducted based on the potential information needs of a group.

However, Albrachtsen (1993:223) states that, although the domain-oriented approach serves target groups better, it has a weakness of preventing generalisation beyond individual domains or it poorly supports interdisciplinarity. Another problem resulting from this approach is that the immediate requirements of a particular group may be highly regarded at the expense of future users (Albrachtsen 1993:222), which may lead to meeting the needs of a wide variety of users becoming a challenge (Hoffman 2009:27). The importance of familiarity with a community in facilitation of subject description (Buckland 2012:156) becomes a problem when the scope of users and information needs is broadened, which is a significant issue in this study.

Another limitation to the domain-oriented approach is identified by Szostak (2014), who advances a different argument from the claim that access to information should be entirely domain-oriented. When making decision about the appropriate approach to any study on information resource description should consider these different theoretical views that are held in the LIS field. Szostak (2014:160) argues in favour of interdisciplinarity and social diversity in information access, as opposed to the narrow domain-oriented approach to subject representation and access.

In this study, the different views are considered as a basis to establish whether the domain-oriented approach to subject analysis is efficient to determine the geographic subject representations, or whether a different approach to subject analysis is necessary. This is in view of the interdisciplinarity of "place" and its use as a form of subject representation. The investigation also considers the potential of the geographic information contained in ETDs to satisfy diverse information needs across different domains and to the wider public to whom ETDs are available on the Web.

A re-examination of the different views in LIS literature is necessary to identify suitable theoretical approaches on which to model subject analysis, or a convergence of some – particularly for ETD geographic subject metadata creation. A solution may be found by re-examining the different conceptions and approaches to subject analysis and finding ways of drawing from the strengths of each.

4.4 A synthesis of views on subject analysis approaches

The synthesis of subject analysis approaches is discussed in terms of the views expressed in the literature on the convergence of epistemological positions that leads to merged approaches, integration of conceptions and emerging and broadened interpretations of the different approaches, specifically on the aspect of domains. Mai (1999) states that there is commonly a combination of the various conceptions in the practice of indexing. This is considered applicable to the approaches followed during metadata creation; particularly in the subject analysis stage. Different opinions on the ideal subject representation and subject analysis approaches and their categorisation are outlined in literature, some showing the convergence of various conceptions and approaches.

According to Hjorland (2017:55) and Fidel (1994:572), the important theoretical positions to subject representation are the document-oriented and or user-oriented views. The arguments presented in Section 4.2.2.2 show that the simplistic and the content-oriented subject analysis approaches are regarded as being document-oriented. On the other hand, the requirement-oriented and the domain-oriented approaches are regarded as being the user-oriented. It is considered important in this study that a clear distinction is between the different subject analysis approaches, in order to identify the specific theoretical conceptual approach re followed during subject analysis.

Hjorland's (1997:40) whose view on the framework of subject analysis is that subject analysis should be need-oriented and not document-oriented also argues (Hjorland 1997:40) that domain-orientation, which is a need-oriented in nature, should be the main approach to subject representation. However, Hjorland (1997:41) and Guimaraes (2017:96) agree that other factors connected with the concrete facts of an information resource (e.g. time period, persons and places) have an effect on how subject analysis is conceptualised and practiced. Due to the possible existence of the given factors, which are important to represent an information resource subject content, the synthesis of conceptions with both a pragmatic (aligned to user-oriented conceptions) and positivistic (aligned to document-oriented conceptions) approach may occur and, therefore, influence subject analysis.

Subject representation has to cater for these aspects, which may not be restricted to the specific domains. Furthermore, Hjorland (1997:41; 1992:195) argues that, in practice, several subject descriptions can be assigned to an information resource. As a result, subject analysis should be based on theoretical conceptions that accommodate the various information resource properties.

In addition, Mai (2010:627) presents an argument that indicates the possible need for synthesised subject analysis conceptions that will serve as a basis for information resource description practices that are relevant to the contemporary context. Mai (2010:627) argues that there is a movement away from domains, caused by social interaction applications that challenge the way information is made accessible.

This situation is applicable to ETD repositories, which have made research information widely accessible to people with varying information needs and in a space where such social interaction applications exist. The theoretical conceptions on which subject analysis in the digital context is based, need to be realigned to these changes. Subject analysis conceptions that support both the domains view and users outside the domains becomes necessary in this case.

Furthermore, Albrachtsen (2015:557) argues for a broadened and flexible interpretation of domains and indicates that domains are not fixed, but can be redefined within different contexts. In line with this belief, Albrachtsen (2015:557) gives the modern definition of a domains analysis as "a method of inclusion, heterogeneity, wholeness".

In support of the flexibility in the meaning of domains, Hjorland (2017:451) indicates that there are different theoretical understandings of domains in different times and across broader society. These arguments are an indication of continuing exploration of better ways of representing information, which impacts on subject analysis for the creation of geographic subject metadata.

4.5   Chapter conclusion

This chapter discussed the role of theory in the subject analysis process. The conclusion that can be drawn from the reviewed literature is that there is a need for theoretical research to establish how theoretical perspectives and their understanding and application impact on subject analysis. It is also revealed that human or intellectual analysis theories play an important role in facilitating subject searching. LIS literature also recommends that automated systems designs can benefit from the understanding of intellectual analysis theory. Furthermore, the reviewed literature revealed an epistemological mix occurring because of the integration of traditional human subject analysis practice and the integration of automation.

The model of conceptions is presented as a theoretical framework for understanding and guiding subject analysis practices. According to this model, subject analysis can be guided by the principles of simplistic-orientation, content-orientation, requirement-orientation and socio-cognitive-orientation.

The literature indicates that preference in the adoption of the principles may vary according to context. However, there is a noticeable inclination towards principles that align with the user-centred approach.

The literature reviewed in this chapter also helped to establish how the traditional subject analysis conceptual approaches could apply in the ETD environment. It is also shown that the theoretical principles or conceptions can form the basis for subject analysis of various formats of information resources. In addition, it is concluded that more investigations are necessary to determine how the conceptual models can contribute to further enhancement of geographic subject metadata.

Chapter 5 discusses the research design and the chosen research methodology selected for and employed in this study.

CHAPTER 5: RESEARCH METHODOLOGY

5.1   Introduction

This chapter discusses the research methodology employed in this study. *Research methodology* is defined as the overall approach to the research process that encompasses the body or group of research methods suitable to accomplish the purpose of a research (Collis & Hussey 2014:10; Henning, Van Rensburg & Smith 2004:36). The aim of this study was to investigate the approaches followed during subject analysis and to examine how subject analysis theories are applied to facilitate geographic subject metadata creation for ETDs. The study investigated how information resource content is analysed by subject metadata creators to determine geographic subject metadata. Specific focus is on the role of subject analysis theories in the geographic subject description of electronic theses and dissertations.

Literature was reviewed to provide context to the different phases of the study – i.e. the questionnaire survey, interviews and content analysis. Different aspects of the research approach and justifications for their appropriateness to address the research problem are discussed in this chapter. The discussions of the purpose and objectives of this study are presented in Section 5.2. The pragmatic research paradigm, within which this study was developed, is discussed in Section 5.3. Pickard (2013:325) emphasises the relationship between a research paradigm and approach. In Section 5.4, the rationale for adopting a mixed method approach is explained and justified, based on the pragmatic paradigm.

The research design is discussed in detail in Section 5.6. Aspects of the research design, which include the methods, data collection and analysis, a framework for prioritising datasets and the approach to the data integration, reliability and validity and ethical considerations are discussed in this section. The research phases are informed by the design followed involved in a study (Creswell & Plano Clark 2018:52, Leedy & Ormrod 2013:260; Brannen & Halcomb 2009:68), which, in the case of this study, is the explanatory sequential mixed method design, with two phases and an additional third phase. The term "phase" is used in this study to refer to the particular stage at which a different research method was introduced to the design.

The research methods are discussed in detail in Section 5.7. Discussions on how the integration during the different phases and interpretation was conducted are presented in Section 5.8. The ethical considerations for this study are explained in Section 5.9. The summary of the research methodology is presented in Table 5.1 (p. 117). The research methods and the design were considered suitable to achieve the research purpose and objectives of this study.

5.2   Research purpose and objectives

The purpose of this study is to investigate the approaches that are followed during subject analysis and how subject analysis theories are applied to facilitate geographic subject creation for ETDs. The methodology and the research design for this study were informed by the outlined purpose and objectives. The use of appropriate approaches improve the prospects of conducting a successful study and to maintain the credibility of the research process and of the results.

The research objectives of this study are to:

1. To investigate the application of the subject geographic representations during the creation of ETD metadata in South African university libraries;
2. To collect data from the practitioners involved in metadata creation in South African university libraries on how they approach the determination of geographic subject metadata for ETDs;
3. To establish from the subject metadata creators in South African university libraries what the implications are of their different analysis approaches for the determination of geographic subject metadata for ETDs; and
4. To establish what type of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries.

Table 5.1 (p. 117) summarises the research objectives that informed the adopted research methodology of this study. These objectives were the foundation to the literature review that formed part of the preliminary information of this study. The information from the literature review is integrated in the discussions and the recommendations made for the approaches to subject analysis.

In addition, literature was reviewed to provide context to the different phases of the study, which consist of a questionnaire survey, interviews and content analysis. Furthermore, the research objectives provided the foundation for the interpretation of the data and the final recommendations.

Table 5.1: Summary of the research methodology for this study

Research paradigm: Pragmatism				
Research objectives	Research questions	Research approach	Research methods	Population
1. To investigate the application of the subject geographic representations during the creation of ETD metadata in South African university libraries	To what extent and in what ways is geographic subject metadata used for the description of ETDs in South African university libraries?	Mixed methods	Quantitative questionnaire; Qualitative interviews; quantitative/qualitative content analysis	Metadata creators in South African university libraries ETD records in institutional repositories
2. To collect data from the practitioners involved with metadata creation in South African university libraries on how they approach the determination of geographic subject metadata for ETDs.	What are the current approaches being followed to create geographic subject metadata for ETDs in South African university libraries?	Mixed methods	Quantitative questionnaire; Qualitative interviews	Metadata creators in South African university libraries ETD records on the institutional repositories
3. To establish from the subject metadata	How and why do the current analysis	Mixed methods	Quantitative questionnaire; qualitative interviews	Metadata creators in South African

Research paradigm: Pragmatism				
Research objectives	**Research questions**	**Research approach**	**Research methods**	**Population**
creators in South African university libraries what the implications are of their different analysis approaches for the determination of geographic subject metadata for ETDs.	approaches being followed during geographic metadata creation for ETDs affect the process?			university libraries ETD records on the institutional repositories
4. To establish what type of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries.	How can the findings of this study be applied to develop a model for subject analysis and creation of geographic subject metadata for ETDs in South African university libraries?	Mixed methods	Quantitative questionnaire; qualitative interviews; quantitative/qualitative content analysis	Metadata creators in university South African libraries ETD records on the institutional repositories

This mixed methods study was guided by the pragmatist research paradigm, which is discussed in the next section. The important relationship between the research paradigm and the methodology of this study is outlined.

5.3 Research paradigm

A *paradigm* can be defined as people's views about the world or reality and how knowledge develops (Saunders, Lewis & Thornhill 2016:132; Collis & Hussey 2014:343). Henning, Van Rensburg and Smith (2004:12) point out that research cannot be conducted in a theoretical vacuum. A paradigm influences the way in which data about a phenomenon should be gathered, analysed and used. Consequently, it guides the choice of research questions and methodology, research approach, data collection methods and instruments (Saunders, Lewis & Thornhill 2016:132, Collis & Hussey 2014:343, Creswell 2014:6).

Different paradigms are presented in the literature. Collis and Hussey (2014:49) discuss the continuum of paradigms, with positivism and interpretivism as the two extremes and other paradigms positioned in between them. The authors highlight that the paradigms changed over time, with new types emerging to improve on the limitations of the earlier ones.

Furthermore, Shannon-Baker (2016:321) argues that paradigms are intentionally used to guide research and are, therefore, applied on the researcher's discretion. This study is based on the pragmatist paradigm, which argues that the research question or problem and the research objectives are the most important determinants of the research approach.

Table 5.2 (p. 120) presents the two extremities or core paradigms mentioned above and the pragmatist research paradigm. Pragmatism was adopted as the philosophical framework for this study. The ontologies, epistemologies and axiologies of the different paradigms are presented, as discussed by, among others, Biesta (2010), Shannon-Baker (2016:322), Saunders, Lewis and Thornhill (2016:137), Creswell 2014:11, Collis and Hussey (2014:43–56), Teddlie and Tashakkori (2009:88) and Henning, Van Rensburg and Smith (2004:2).

Table 5.2: Research paradigms

	Positivism	**Interpretivism**	**Pragmatism**
Ontology (Nature of reality, being and truth)	There is one true reality. It is believed that reality is concrete	Reality is a projection of human imagination and, therefore, it is constructed. It is believed that there are multiple realities	There is no absolute truth or reality. Knowledge draws from the different worldviews wherever practicable
Epistemology (Assumptions about reality) Knower/ known relationship (researcher participant/relationship)	Knower and known are independent The researcher is an objective analyst, independent from the researched.	Knower and known are interactive, inseparable Subjective interpretations by the researcher are a key contribution. Different interpretations of a phenomena occurs	Intervention plays an important part in the way knowledge is obtained. Both objective and subjective points of view. The researcher decides on an epistemological position to endorse (Biesta 2010:10)
Axiology (Role of researcher values in an inquiry)	Inquiry is value free	Inquiry is value bound	Values play an important role in interpreting results.
Methodological assumptions (The process of research)	Quantitative Highly structured methodology	Qualitative	Both quantitative and qualitative. The researcher use the chosen methods, based on what works at the time, to address the research problem

(Adapted from: Collis & Hussey 2014:46)

In the next paragraphs, detailed discussions are given to explain the aspects outlined in Table 5.2 and to demonstrate how pragmatism influenced the design of this study. The nature of knowledge or ontology in pragmatism is shaped by the understanding that there is no absolute unity in what truth is and knowledge allows drawing from the different viewpoints or worldviews wherever practicable.

Therefore, the choice of research methodologies is based on what works at the time (Creswell 2014:11). In this way, it was decided to employ different methods in this study, so as to draw perspectives to address the research problem.

The pragmatist ontology influenced the direction the research followed regarding the methods used, data collection, sampling, data analysis and interpretation of the results. Knowledge about the problem of this study was draws from the different viewpoints through data gathered from the different methods and interpreted quantitatively and qualitatively for a comprehensive understanding.

Consistent with the views of Venkatesh, Brown and Sullivan (2014:448) on the pragmatist paradigm, inductive and deductive meaning were derived simultaneously in this study. Meaning was deduced from data obtained from the quantitative questionnaire, induced from the qualitative interviews data and the quantitative/qualitative content analysis. Content analysis data was subjected to descriptive analysis and interpretative meaning-making for the purpose of supplementing the qualitative data follow-up explanations. This approach provided a comprehensive understanding of how the subject analysis theoretical principles apply to the geographic subject metadata creation for ETDs. The pragmatist epistemology stance adopted in this study is that the objective and subjective points of view are reconciled in terms of their practical consequences in specific contexts (Saunders, Lewis & Thornhill 2016:143). In line with the pragmatist epistemology, the researcher's views played a significant role in the interpretation of the data relevant to the research problem involved in the study.

The researcher decides which epistemological position to endorse (Biesta 2010:10). In this study, the researcher objectively analysed the quantitative data from the questionnaire survey and interpreted them. In addition, the participants' experiences and perceptions were analysed qualitatively to deduce meanings to explain the quantitative findings. The epistemological ideas reflected in Table 5.2 (p. 120) agree with the research approach used in the study, where flexibility was maintained by using different methods to gain the practical understanding from varied perspectives on the practice of subject analysis and how its theoretical basis has an effect on the creation of ETD geographic subject metadata.

The research problem of this study played a significant role in deciding on the research design that would enable it to be addressed in depth. Conclusions were drawn from both the objective and subjective standpoints to address the research problem. This is consistent with the views of Saunders, Lewis and Thornhill (2016:143), who opine that, based on the pragmatic beliefs, the practical effects of ideas and knowledge are supported in terms of how they enable actions to be carried out successfully.

The axiology of pragmatism is that values play an important role in interpreting results and drawing conclusions about a study (Teddlie & Tashakkori 2009:90). Consistent with the views of the authors on the pragmatic axiology, the units of analysis and variables included in all three phases of this study were considered suitable to yield information to address the research problem comprehensively. Practicality dominates the reasoning for the research process in this study.

The methodological assumptions of pragmatism are that both quantitative and qualitative methods are used, depending on what methods answer the question best. The choice of research methodologies is based on what works at the time (Creswell 2014:11). Mixed methods research, which uses both quantitative and qualitative approaches, was considered ideal for the in depth investigation of the research problem. Furthermore, the pragmatic worldview allowed the discretionary integration of the qualitative and quantitative approaches in a design that would work best for this study. Pragmatism supported the practical use of the chosen explanatory sequential design with and added method to support this study.

Pickard (2013:5) acknowledges the prevalent use of pragmatism as a guiding paradigm for research in the LIS field. Pragmatism provided the basis of all the procedures of this study, conducted within the LIS context, because it supported the use of the research approach that was considered suitable to answer the research questions of this study.

5.4 Research approach

A research approach is the plans and procedures for research that outline the decisions from the philosophical assumptions, design, the specific methods and their details (Creswell 2014:247). Three common research approaches used to address the research questions are evident in the social and behavioural sciences (Teddlie & Tashakkori 2009:7). The two basic categories are the quantitative and qualitative research approaches. The mixed methods approach emerged as an alternative to the two basic forms (Leedy & Ormrod 2013:95, Teddlie & Tashakkori 2009:22). The different research approaches are summarised in Table 5.3.

Table 5.3: Core characteristics of the quantitative, qualitative and mixed methods research approaches

	Quantitative	**Qualitative**	**Mixed Methods**
Purpose	Explain and predict, confirm and validate, test theory	Describe and explain, explore and interpret, build theory	Purpose is to seek better understanding of complex situations (Leedy & Ormrod 2013:96). Combination of different approaches, confirmatory and exploratory, is driven by the research question.
Process	Known variables Established guidelines Predetermined methods	Unknown variables Flexible guidelines Emergent methods	Variables are identified, based on the assumption that they can support the understanding of the research problem of a study.
Data collection (Data collection methods)	Numeric data Representative, large sample Standardised instrument	Textual or image-based data Informative, small sample Loosely structured or non-standardised observations and interviews	Use of quantitative and qualitative methods (mixed methods)

	Quantitative	Qualitative	Mixed Methods
Data analysis and deriving meanings	Statistical analysis Objectivity Deductive reasoning	Search for themes and categories Subjectivity Inductive reasoning	Both quantitative and qualitative analysis. Integration of statistical, thematic and conversion Deductive and inductive meanings
Reporting of findings	Numbers Statistically aggregated data	Words Narratives and individual quotes	Integration of quantitative and qualitative data Numeric plus narrative

A mixed methods approach was employed in this study, which resulted from the use of quantitative and qualitative approaches in a single study. Although different in characteristics, these approaches were used together to address the different research questions appropriately and comprehensively.

Different authors, including Tashakkori and Teddlie (2003:x; xi) and Creswell (2014:4), argue that the mixed methods approach originated in the late 20th century, when it was recognised as a research approach, based on the pragmatic way of using the strengths of both the quantitative and the qualitative approaches. The reviewed literature indicated how the mixed methods approach combines the elements of the quantitative and qualitative approaches (Collis & Hussey 2014:5; Leedy & Ormrod 2013:95; Teddlie & Tashakkori 2009:97). The two approaches are combined to address research questions that could not be sufficiently addressed through only one of the approaches.

The pragmatist paradigm discussed in Section 5.3 supports the use of the broadest techniques to answer research questions (Tashakkori & Teddlie 2003:678). The design discussion explains how the mixing of the three methods, the quantitative questionnaire, qualitative interviews and quantitative/qualitative content analysis was conducted in this study. By employing the strengths of the quantitative and qualitative approaches in a single study, it was possible to address the research problem adequately – more so than would have been possible by using either of the approaches by themselves.

Although the elements of the quantitative and qualitative approaches differ, they can be combined in a mixed methods study (Leedy & Ormrod 2013:98), as was done in this study. Understanding the elements of the different approaches outlined in Table 5.2 (p. 123) assisted in the decision on how to approach the design of this mixed methods study. The integration between the quantitative and qualitative methods occurs at various points. In addition, the interpretation draws from the integration of the different analysis results to address the research problem. Data from the different methods connects to explain or build on each other. The research design is discussed in detail in Section 5.5.

The next section explains the reasons for deciding on a mixed methods approach to address the research problem of this study.

5.5   Rationale for using the mixed methods approach

The rationale for using mixed methods in this study was based on the nature of the research objectives. The research objectives in this study, stated in Section 5.2, required both qualitative and quantitative approaches to provide comprehensive answers. Subject analysis is a complex process; particularly when introducing its theoretical aspects. The use of both approaches provided for the research objectives to be addressed comprehensively. The next two paragraphs discuss the justification for the need to use the different methods collectively and how they assist to address specific objectives of this study.

The justifications for combining quantitative and qualitative research approaches in mixed methods research have been discussed by various authors, including Onwuegbuzie and Combs (2010:34), Venkatesh, Brown and Sullivan (2016:44), Saunders, Lewis and Thornhill (2016:173), Leedy and Ormrod (2013:259), Teddlie and Tashakkori (2009:161) and Greene, Caracelli and Graham (1989:259).

The different purposes outlined in the reviewed literature include several that were found relevant to the design of this study:

- *Complementarity*

  Quantitative and qualitative methods were used in a complementary way in this study, which provided the benefit of complementing the weakness on one method with the strengths of the other. The complementary purpose was realised, in that the quantitative results were further explained by the results of the qualitative analysis (Greene, Caracelli & Graham 1989:258). Additionally, the quantitative and qualitative approaches allowed the different research questions to be addressed sufficiently. The quantitative data showed the perceptions ranking and levels of use of the different approaches, while the qualitative data explained how the perceptions and experiences influenced the understanding and practices of subject analysis and subject metadata creation. Furthermore, the content analysis helped to supplement and support the explanations of the quantitative research findings, thereby fulfilling the complementarity purpose.

- *Completeness*

  Another reason for using mixed methods in this study was considered as enabling the production of a more complete or in-depth knowledge of the application of subject analysis theories by the ETDs metadata creators. Completeness focuses on the full address of the problem and its sub-problems (Leedy & Ormrod 2013:259). The common themes used in the different phases to categorise the different aspects investigated in this study guided the questions used for the questionnaire and interview questions. The approach yielded more complete information on similar aspects examined in this study.

  Mixed methods research is considered useful to produce valid and balanced evidence from the more complete findings, thereby resulting in credible results (Saunders, Lewis & Thornhill 2016:173; Venkatesh, Brown & Sullivan 2016:442). A complete investigation on the theoretical aspects of subject analysis and geographic subject metadata creation was facilitated by using the different approaches.

- *Development*

  For the purpose of development, the results of one method are used to develop or inform the other method. The development purpose in this study was realised through the use of the quantitative results to refine the questions for the interviews and to decide on the sample to be used for the interviews. Additionally, the results of the quantitative phase of the design influenced the interpretation of the content analysis data. Greene, Caracelli and Graham (1989:267) indicate that the results of one method can be used to inform the analysis of the other to assess similar phenomena.

- *Expansion*

  The use of mixed methods can achieve the purpose of extending the range and breadth of enquiry to address the different components of the study. The content analysis was used to gather additional information by extending on the explanatory design with a supplementary phase to the explanations of the quantitative findings.

- *Diversity*

  The coverage and depth of the different methods used in a mixed methods study allows varied insights (Venkatesh, Brown & Sullivan 2016:442; Greene, Caracelli & Graham 1989:257). The results of this study encompassed the diversity of mixed methods purposes, in that different methods – i.e. the quantitative questionnaire survey, qualitative interviews and quantitative/qualitative content analysis – were used to examine the subject metadata creation approaches.

It was important to collect varied information by using the different methods in this study, so as to assist balanced arguments about subject analysis and geographic subject metadata creation. The results of the survey questionnaire gave a broad overview of the different metadata creators' perspectives on subject analysis and geographic subject metadata creation, whereas the interviews alluded the perspectives and experiences in depth. The content analysis findings provided evidence from practice. The discussions in the sections reveal how these varied purposes are realised in this study through the use of the mixed methods approach.

In addition, the use of mixed methods was adopted because of its strengths and relevance to address LIS practice problems, as confirmed by Ngulube, Mokwatlo and Ndwandwe (2009). The focus of this study was the investigation of subject metadata creation in university libraries. According to White (2012:79), there is no single best method for researching metadata.

The field of metadata creation and use requires application of multiple methods to understand its complex nature. Historically, both quantitative and quantitative methods were used individually or were mixed to conduct research on metadata. The nature of this investigation involved the complexity, breadth and scope that required the use of a mixed methods approach.

5.6   Research design

Research design can be defined as the whole or general strategy of the research (Saunders, Lewis & Thornhill 2016:163; Leedy & Ormrod 2013:74). The design of this study incorporated the qualitative or quantitative approaches, as opposed to using either alone. The design was based on pragmatism as the underlying research paradigm that informed the mixing of three methods in a manner suitable to address the research problem.

This study aimed at investigating subject analysis theories and their application to geographic subject metadata, from the context of the practitioner's perspectives and the existing records, by using the explanatory sequential mixed design involving the questionnaire survey, interviews with an additional content analysis phase. Creswell and Plano Clark (2018:102) explain the approaches that move beyond the core mixed methods designs. A content analysis phase was introduced to the core explanatory sequential mixed method design to provide additional information to support the explanations. The method helped to establish the actual application or effect of the practices from the existing products in the form of ETD record subject metadata. The ETD records were examined in the third phase of this study.

The approach was found suitable for this study, because, as highlighted by Leedy and Ormrod (2013:259), quantitative and qualitative approaches can be combined to address a complex research problem, as was found necessary to address the complex process of subject analysis creation and its theoretical approaches. The complexity of the process is pointed out by Hjorland (1997:39) and Chan (2000:15), among others.

Three core mixed methods designs are discussed in literature. Creswell and Plano Clark (2018:59) name the core designs as the explanatory sequential design, the exploratory sequential design, and the convergent parallel design. Creswell (2014:242 & 243) explains the designs as follows:

- Convergent parallel design involves simultaneous collection of quantitative and qualitative data and the separate analysis and comparison of the results to establish confirmation.
- Exploratory sequential mixed method design involves collecting qualitative data and following it up with quantitative data collection and analysis.
- Explanatory sequential design involves collecting the quantitative data and analysing it, followed by data collection and analysis of the quantitative phase.

Furthermore, the literature indicates that some mixed methods designs go beyond the core or basic designs mentioned above (Creswell & Plano Clark 2018:102; Creswell 2014:16; Teddlie & Tashakkori 2009:139 & 163). The design of this study involved the use of an explanatory sequential mixed method design, by sequentially involving quantitative, qualitative and an additional quantitative/qualitative method, resulting in a study with three phases. This design is considered an advance on the common two phases design types explained in the foregoing paragraph. The survey was followed by the interviews and the content analysis to elaborate on its findings (Saunders, Lewis & Thornhill 2016:171). The following paragraphs explain how the design, integration and reporting occurred.

The chosen design was guided by the insights discussed in the foregoing paragraph to combine core mixed methods designs with additional methods. Johnson and Onwuegbuzie (2004:20) indicate two possible typologies of mixed methods designs: firstly, based on the decision whether to follow one prominent paradigm or not; and,

secondly, whether to conduct different phases concurrently or sequentially. The design of this study followed the sequential approach, based on the pragmatist paradigm. This research design required conducting the different parts of the study in sequence (Creswell & Plano-Clark 2018:80), with the quantitative findings guiding how the qualitative follow-up was to be conducted to provide detailed explanations, as explained in the following paragraphs.

The research design was conceptualised at the initial stage of the study to use all three methods – i.e. the questionnaire survey, the interviews and the content analysis. Table 5.4 presents a summary of the research design involved in this mixed methods study.

Table 5.4: Research design summary

Elements of the research design	Description of the methods
Research approach	Mixed methods: quantitative + qualitative
Population and sampling	• Entire population of subject metadata creators in South African university libraries for the questionnaire survey. No sampling was performed. Instead, all metadata creators involved in university libraries with ETD metadata on the NETD database were invited to participate in the questionnaire survey. • Purposive sample for the interviews and the purposely identified ETD records for content analysis
Data collection methods	• Questionnaire • Interviews • Content analysis
Research instruments	• Questionnaire • Semi-structured interviews • Structured observation or structured record reviews
Data analysis	• Quantitative analysis, questionnaire (instrument: SPSS for descriptive statistics) • Semi-structured interviews (instrument: NVivo) • Analysis of metadata records (instrument: Excel analysis) to determine frequencies of manually identified characteristics and thematic discussions)
Credibility: reliability and validity checks	Validity and reliability of the quantitative questionnaire and the trustworthiness of the qualitative interviews and the content analysis
Ethical considerations	• Research ethics for quantitative and qualitative research

Elements of the research design	Description of the methods
	• UNISA ethical clearance
	• Ethical clearance from the universities
	• LIASA Professional Code of Ethics and Conduct

The reviewed literature outline variations to the explanatory sequential mixed method designs and different integration types used in different studies. Fetters et al (2007) used a related design in which a core sequential explanatory mixed methods design was extended with an additional phase. Different studies used a design with three phases, integrating an additional method, such as content analysis, which is normally not used as a stand-alone. Morse (2010:483) indicates that a single study can use mixed methods with a supplemental method that cannot be interpreted or utilised alone. In his study, Worrall (2014) used a related sequential three-phase design for a study in the LIS field. However, the phases were carried out in direct sequence in an explanatory sequential, multiphase design. Creswell and Plano Clark (2018:61) remark that the *core design* used in the multiple stages of procedures should be the focus of the design, as it indicates the primary intent of mixing methods.

An explanatory sequential mixed method design was found suitable as a *core design* to conduct this study, with the qualitative findings clarifying the quantitative findings. Creswell (2014:224) explains this design as involving the collection of quantitative data in the first phase and using the results to plan for the second, qualitative phase, which was done in this study, in order to achieve the practicality of using the interviews findings to explain the quantitative findings. Based on the timing, the sequence followed for mixed methods can primarily be categorised into concurrent or sequential (Venkatesh, Brown & Sullivan 2016:445). However, some designs can combine both concurrent and sequential elements in one study or use multiple core designs in one study (Creswell & Plano Clark 2011:66; 2018:61). The process and the timelines that were followed in this study are illustrated in Figure 5.1 (p. 133).

The quantitative questionnaire survey was conducted in the *first phase* of this study. Data was collected according to the planned phases of this study, as shown in Figure 5.1 (p. 133).

In the *first phase* of the study, the questionnaire was sent to the participants at the university libraries that granted permission to conduct the study. Separate analysis was conducted for the data collected through the questionnaire. The quantitative results obtained through the questionnaire informed the way in which the interviews were conducted.

In the *second phase*, aspects that needed further clarity were identified after the quantitative data were analysed and clarified from the experiences and perspectives shared during the qualitative interview phase. Based on the quantitative results, ETD metadata creators were identified at the participating libraries for the interviews. Data collected from the interviews was also analysed separately as a follow-up to the first, quantitative part of the study.

Figure 5.1 (p. 133) illustrates a schematic representation of the design adapted from Creswell's explanatory sequential mixed methods and multiphase mixed methods designs (Creswell 2014:220 & 221). The diagram was guided by the explanatory sequential mixed methods and multiphase designs of Creswell and Plano Clark (2018:85), Creswell (2014:221) and Fetters et al (2007). Creswell and Plano Clark (2018:61) observe that the focus of attention should be on the core design in the multiple stages of procedures.
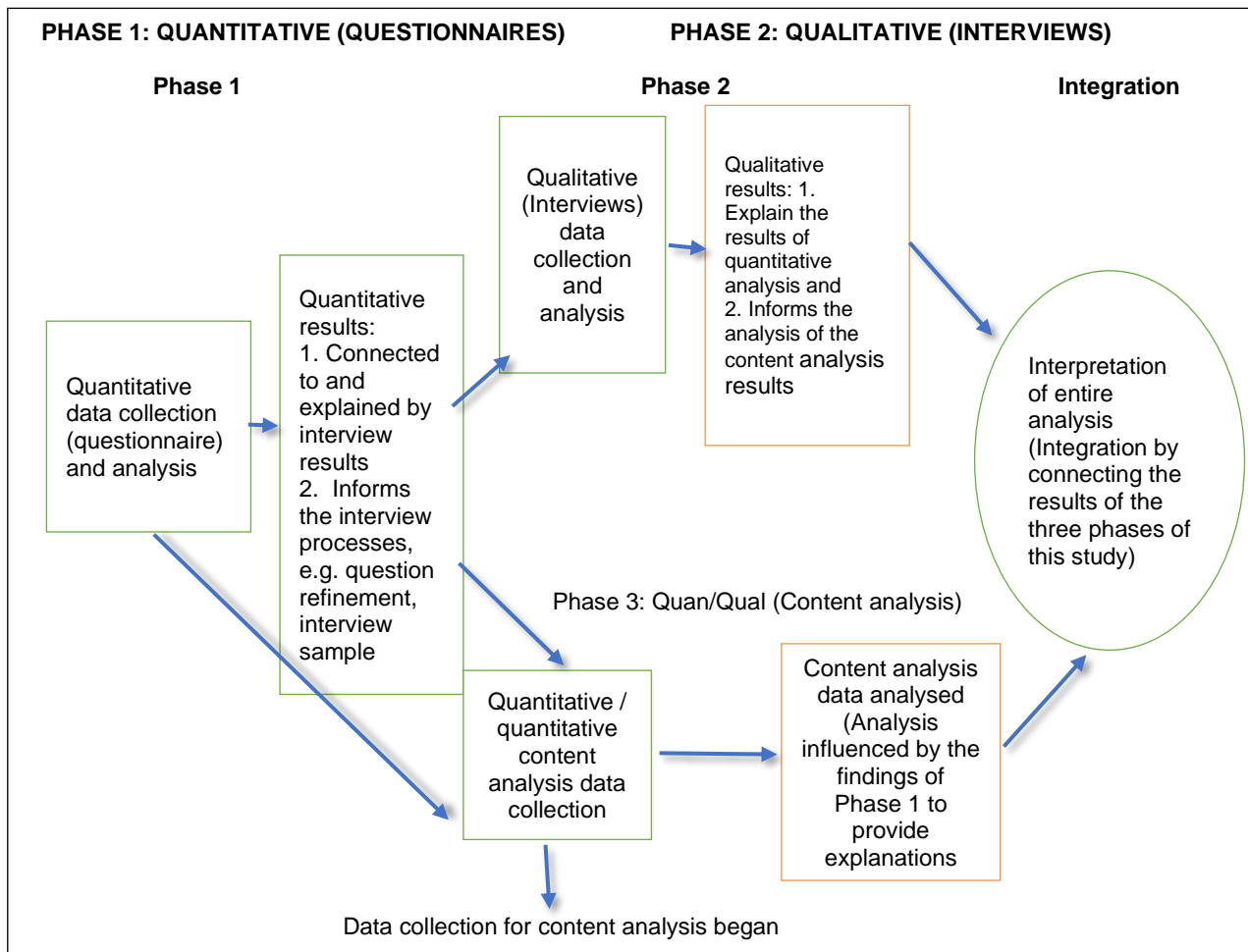
Figure 5.1: Research design

A *third phase* was included in the diagram to show how the content analysis was introduced and integrated in a planned sequence. Fetters et al (2007) followed a design with similar characteristics. The timing of the phases resulted in the third phase taking the form of an additional quantitative/qualitative content analysis phase, so as to enhance the study and to enable the observation of patterns and trends. The introduction of the third method suited the research purpose of this study, which was to obtain wider perspectives and a broad understanding of the process of subject analysis, as practised by metadata creators, supplementing the qualitative explanations to provide comprehensive answers for the research questions.

The content analysis was considered in the design planning stage to collect additional information from the actual ETD descriptive records, introducing a third phase to the design for in-depth understanding of the research problem.

While waiting for the questionnaire responses, the researcher began with the data collection from the ETD records for the third phase of the study. As a result, data collection for the content analysis was based on the pragmatic flexibility of designs and method choices and commenced while waiting for the results of the questionnaires. Based on the intense procedures learned from literature and the researcher's experience of metadata analysis, it was considered practicable in the planned sequential study to introduce content analysis data collection as a third phase and to integrate the findings at the interpretation stage.

Creswell and Plano Clark (2018:61; 2011:104) support this type of combination design that can, for example, use a primary sequential mixed method and combine it with an additional phase, in order to improve on the research process. In view of this, it was considered that the manner in which the content analysis was phased in would not interfere with the core sequence of quantitative and qualitative approaches. The data was collected from existing records on the National Electronic Theses and dissertations (NETD) database and ETD repositories.

The following paragraphs provide an overview of the separate analysis of the data collected in the different phases of this study and the integration of the different stages and during the interpretation and presentation of the findings.

The findings of the quantitative phase influenced the decisions on the final analysis of the content analysis data. The method was used with the aim of enhancing the study and addressing issues that needed clarification. The understanding provided by the content analysis results enhanced the qualitative explanations, validating issues that addressed Research Objective (RO1) on the extent and the nature of usage of the geographic subject metadata.

The introduction of this phase resulted in an analysis stage that was considered as a sequential mixed analysis (Teddlie & Tashakkori 2009:276 & 344). The analysis of the results of the interviews was influenced by the intended use of clarifying the quantitative questionnaire results. In addition, the content analysis phase of data analysis was influenced by the aim of supporting the qualitative explanations for the questionnaire findings to answer the research questions in depth.

Furthermore, the results of the last phase were integrated throughout the interpretation of the overall results.

Integration occurred in various ways and at different stages in the design of this study. In this study, integration was understood as explained by Guest (2014), O'Cathain, Murphy and Nicholl, (2010) and Moran-Ellis, Alexander, Cronin, Dickinson, Fielding, Sleney and Thomas (2006:46), in that it can occur from conceptualisation until the interpretation of the research findings. The three phases were *integrated* as required for a mixed methods design. The sequential introduction of the interviews and the content analysis phases as follow-up was in accordance with the insights from Creswell and Plano Clark (2018:82) about the variants of the explanatory sequential mixed method design. Using more than two phases in an explanatory sequential design introduced a follow-up element, where the interviews and the content analysis further explained the quantitative survey results.

Consistent with the views of Venkatesh, Brown and Sullivan (2016:438), it is understood that a multimethod design is not always mixed methods and, therefore, the phases of this study were *integrated* by connecting the qualitative interviews and the quantitative/qualitative content analysis to build on the quantitative questionnaire phase. The quantitative findings informed the way in which the qualitative phase was conducted. The first quantitative findings guided the choice of the interview participants to explain the aspects from the questionnaire findings that needed further clarity. In addition, they influenced the types of questions asked to seek clarity.

Furthermore, *integration* with the content analysis occurred through influence of the analysis of the results of this phase. This *integration* of the content analysis was consistent with the views of Greene, Caracelli and Graham (1989:267), who opine that, because of the sequential timing of the methods, the results of one method may influence analysis for the subsequent methods of a study. The content analysis data was analysed to obtain supplementary information to address the research problem.

Overall, all the findings from the questionnaire, interviews and content analysis were *integrated* in the interpretation stage. The conclusions of this study drew from the integrated findings of all methods used to fulfil the varied purposes of conducting a

mixed methods study, as discussed in Section 5.5. Interpretation shows how the qualitative interviews provided additional insights into the quantitative results (Creswell & Plano Clark 2018:238) and how the content analysis findings provided clarity on the actual position in practice.

The choice of this research design was informed by the pragmatic view of embracing multiple perspectives to address a research problem. Hall and Howard (2008:248) argue that there is no expectation of a single truth in a mixed methods approach. Therefore, the findings of all three methods of this study contributed important information from multiple approaches to address the common problem of subject analysis conducted for the purpose of creating ETD geographic subject metadata. The findings were integrated at the interpretation stage. The design used for this study followed the mixed methods principles, which meant that the methods must build on one another. Appropriate combination of the methods and procedures was maintained and the findings were successfully integrated.

Further details on the research procedures form part of the discussion of the individual methods. In summary, the collection of primary data was followed by separate analysis and reporting of the results of different perspectives drawn from quantitative and qualitative data. This was followed by the interpretation of all analyses (of three methods) of this study, with the aim of explaining the survey results by means of the interviews and the content analysis findings. The overall structure of the methods followed is discussed in Section 5.7.

5.7   Research methods

Research methods can be defined as the techniques used for collecting and analysing data (Collis & Hussey 2014:55). The form of data collection and analysis for the quantitative questionnaire and qualitative interviews and quantitative/qualitative content analysis are discussed under each method used. The individual research methods are discussed in Sections 5.7.1–5.7.3.

5.7.1 Quantitative questionnaire

The questionnaire assisted in the collection of primary data that gave a broad overview about the ETD geographic subject metadata and subject analysis approaches. The questionnaire survey was meant to address the first three research objectives of the study.

Questionnaires have a limitation in terms of yielding data that does not offer explanations for the findings, due to the closed nature of questions. Detailed explanations and confirmations from additional data obtained by using different methods in this study helped to address this problem. A common limitation experienced with questionnaire studies is questionnaire fatigue (Collis & Hussey 2014:207). There is increased reluctance for participation, due to the increase in questionnaire research conducted for various purposes. This was minimised by keeping the questionnaire shorter and designing questions that are simple to respond to. Several follow-up questions were posed to encourage responses. Item non-response bias, which commonly leads to incomplete results for some questions, occurred less in this study.

5.7.1.1 Data types

Data collected through the questionnaire involved seeking responses to questions categorised under the following four sections: (i) biographical and background information; (ii) metadata procedures; (iii) subject analysis basis; and (iv) ETDs and geographic subject metadata. The data primarily involved the responses to the structured questions. The questionnaire used multiple choice questions, checklists and five-point Likert scale measures (Saunders, Lewis & Thornhill 2016:459).

The data collected through the questionnaire was of categorical or nominal type (measured by using non-numerical codes), and ordinal (measured in counts to identify order or rank) (Collis & Hussey 2014:343). The responses indicated numbers, frequencies, agreement and likelihood about the different content of the questions. The responses yielded data that addressed the research problem on subject analysis and the creation of geographic subject metadata in the form of responses to the questionnaire.

5.7.1.2   Data collection instrument

Questions for the quantitative study were derived from the research objectives and developed based on the identified subject analysis theories, as discussed in Section 4.3.2 (Chapter 4). The data was collected by using a questionnaire (Appendix A), which included categorical variables such as nominal type (measuring non-numerical codes) and ordinal (measured in counts to identify order or rank).

A structured questionnaire was used, which was convenient for extensive data collection (Creswell & Plano-Clark 2011:177) that was planned for the survey. The advantages of using the instrument were considered, because its design allowed easy analysis and, as indicated by (Collis & Hussey 2014:213), the range of potential answers from which respondents could choose was limited and could be coded in advance. The research questions, the reviewed subject literature – e.g. the work of Wolverton, Hoover and Fowler (2011) – and the experience of the researcher in ETD subject metadata creation informed the design of the research questions that were relevant to the research purpose. In addition, the researcher used the guidelines on constructing a questionnaire that were obtained from reviewed literature on research methodology.

The questions were grouped into categories of demographics and background information, subject analysis basis, metadata procedures and ETDs and geographic subject metadata. The questions were aimed at collecting data in relation to the objectives of this study. The mixed types of questions facilitated the evaluation and quantification of the participants' responses, while open-ended questions sought clarifications on the reason for some of the responses and ratings. The intent of colleting this type of data was to obtain an in-depth understanding of metadata practices and the approaches followed by the metadata creators. The presence of the critical elements of geographic subject metadata and their frequencies were determined from the data.

The questionnaire was pilot tested prior to the final version being sent out. Saunders, Lewis and Thornhill (2016:723) indicate the purpose of a pilot test as refining the questionnaire, so as to avoid problems with answering the questions and recording

the data. Pilot testing of the questionnaire was conducted over a period of three (3) weeks. A pilot questionnaire was made available through Google Forms and a link was sent out to metadata creators at five (5) different universities. Two (2) university libraries that did not participate in the final study were included in the pilot study, in order to obtain varied perspectives about the instrument. The piloting exercise assisted in identifying and addressing the shortcomings of the tool, e.g. shortening some questions and correcting the scoring matrix. The pilot phase assisted in the measurement of the validity and reliability of the questionnaire by testing if the questions would help to collect data that was an accurate reflection of what the study intended to establish and if the rating scales could measure the respondents' views consistently in different times (Collis & Hussey 2014). Further discussions on the questionnaire validity and reliability are given in Section 5.7.1.5.

Instructions were sent via email, with guidelines on the aspects that the needed to consider as they responded to the questions. The guidelines presented by Saunders, Lewis and Thornhill (2016:449), which include layout, clarity of questions and instructions and the time it takes to respond, were used. The problems resulting from the answers and the suggested changes were incorporated in the revised questionnaire. Suggestions included the shortening of some questions, where the previous formulation was too long and sometimes difficult to follow. A few questions were also changed from the multiple-choice format to linear-scale questions.

5.7.1.3 Reliability and validity of the quantitative questionnaire

Internal and external validity are important measures for determining whether quantitative research measures what it is supposed to (Koonin 2014:257). Internal validity focuses on whether the research method will answer the research questions. Errors in the questionnaire design and results should be controlled (Leedy & Ormrod 2013:191; Koonin 2014:257). External validity focuses on the ability to generalise the findings from specific samples to a larger population. In this study, the population of the survey questionnaire was not sampled, but approach as a whole by inviting all South African University libraries metadata creators to participate in the study.

The inclusion of numerous libraries in the identified sector presented the possibility of generalising the research results of this study within the South African university library context. Mouton (2001), Leedy and Ormrod (2013:218), Saunders, Lewis and Thornhill (2016:450) and Koonin (2014:257) state that there are different approaches, stated below, for testing the validity of the data gathered through the questionnaires.

Although quantitative studies are easy to replicate, they have a weakness in terms of meaningfulness of the data to explain the research problem, as was the case in this study. In this study, internal validity involved a test on whether the questionnaire survey results correctly answered the research questions. In order to achieve this, the questions were carefully constructed to ensure that they allowed to measure what they were intended to. In addition, proper questionnaire construction was followed regarding the suitability of the questions asked, which contributed to the improved face validity. Internal validity was also enhanced by using the evidence gathered through the other methods used in this study to cross check for the authenticity of the results and to explain the questionnaire data. (Creswell 2014:160) recommends other, traditional forms of establishing validity in quantitative research discussed in the next paragraph.

Content validity was assessed in this study, which involves determining if the items reflected the content they were supposed to (Creswell 2014:160). The questionnaire items were repeatedly read through by the researcher in relation the research objectives to establish if they provided adequate coverage of the research questions. This validity test is called *content validity* (Leedy & Ormrod 2013:89; Saunders, Lewis & Thornhill 2016:450; Creswell 2014:160). The literature reviewed on the topics of subject analysis and ETD metadata creation helped to decide on the items to include in the questionnaire, as well as the decision on "adequate coverage" (Saunders, Lewis & Thornhill (2016:450) to gain appropriate information from the survey.

In addition, pilot testing results, as discussed in Section 5.7.2.2, helped to correct the common questionnaire errors, including condensing some questions and correcting the scoring matrix.

Pilot testing is considered important to establish content validity of scores and improve questions, format and scales (Mouton 2001:153; Creswell 2014:161). Questions regarded as essential were retained and the unnecessary ones were eliminated after the pilot testing.

Construct validity refers to the extent to which the measuring instrument, the questions and the scores used are useful to measure the presence of hypothetical constructs, which are essentially characteristics that are not directly observable (Saunders, Lewis & Thornhill 2016:450; Creswell 2014:160). There are underlying constructs, e.g. the attitudes of the respondents that have the potential to influence on their responses. However, due to the general difficulty of validation of such hypothetical constructs (Collis & Hussey 2014:53), other techniques were used in this study. Therefore, the Cronbach's alpha coefficient test was used to test the reliability, as explained in the next paragraphs. The interviews were used as an additional instrument to test the unobservable attitudes for subject analysis and ETD subject metadata creation.

Validity measurement, which was considered useful for this study, was employed. The specific situation may determine the importance of a validity measurement instrument (Leedy & Ormrod 2013:89); hence different techniques such as *face validity* were relied on in this study. In this study, face validity was ensured by applying the principles mentioned by Koonin (2014:256), Collis & Hussey (2014:53) and Leedy and Ormrod (2013:89) of ensuring that on the surface it would look that the measures used represent what they were supposed to represent, even to the participants of this study. The researcher ensured that the readability and general outline of the questionnaire makes it a well-designed instrument (Koonin 2014:256).

Concepts such as the approaches, procedures and experiences of the participants were important to this study. Therefore, it was important to ensure that the questions in the questionnaire were suitably constructed to represent these concepts. Suitable definitions and measures of variables were used in the questionnaire to ensure that the questions were well understood and that suitable answer options were available to enhance the validity of the instrument. The use of differently structured questions to investigate the same issue provided additional validation, enhanced the credibility of the findings, and prevented any bias in the data.

In the reporting stage of this study, the number of the survey questionnaire non-participants and non-response cases were specified. This practice was done to indicate any possible non-response bias (Leedy & Ormrod 2013:218), which results from the significance of the difference between respondents and non-respondents in the study. Cronbach's alpha coefficient test reliability was used, as explained in the next paragraphs.

Reliability is linked to the consistency in the research findings of the research (Koonin 2014:254). The accuracy and precision of the measuring instrument and consistency in the results when the study is repeated (Collis & Hussey 2014:343) are important criteria for confirming reliability in a quantitative study. Several questions, in different forms, were included in the questionnaire to test consistent responses on the same aspect. Leedy and Ormrod (2013:100) suggest that the degree to which the participants consistently respond to multiple items created to assess a single characteristic can be mathematically determined.

The assistance of a statistician was sought to test the suitability of the rating scales used. This study used the Cronbach's alpha coefficient test, which is commonly used for rating scales. Cronbach's alpha is used to measure the degree to which the items in the instrument are related or the consistency of responses across a set of questions designed together to measure a particular concept (Saunders 2016:714). For most research purposes, an alpha of 0.70 or higher is considered satisfactory (Saunders 2016:714).

Table 5.5 (p. 143) reveals that the Cronbach's alphas for the five research variables were between 0.529 and 0.785. Therefore, the measurement items used in this study were reliable, since the overall Cronbach's alpha coefficient was higher than the acceptable threshold value of 0.7.

Table 5.5: Reliability analysis (n = 29)

	Cronbach's Alpha	N of Items
Subject analysis	0.529	6
Categories of Metadata to be suitable for the discovery of the subject content of ETDs	0.690	5
The creation of the geographic subject metadata for ETDs	0.785	4
Geographic subject metadata for ETDs	0.621	5
ETD geographic subject metadata	0.536	12
Overall	0.752	32

As indicated by the findings given in Table 5.5, the alpha value of Cronbach for each item was within 0.529 and 0.785 and, as suggested by Saunders (2016:714), an overall value of 0.752 is considered reasonable and as a sound indication of the internal consistency of the survey instrument (Wells & Wollack 2003).

5.7.1.4   Population and sampling

Teddlie and Yu (2007:78) demonstrate that sampling in mixed methods research follows a similar pattern to the types of mixing of qualitative and quantitative aspects throughout the research process. The population of the questionnaire survey was the metadata creators, who were responsible for creation of ETD subject metadata. These metadata creators were based in South African university libraries. The libraries have established ETD repositories distinguishable on the NETD database. The NETD database reflects collections of ETD metadata from universities libraries. The metadata creators contributed to the creation of the ETD subject metadata.

The units of analysis for the questionnaire survey were the individual metadata creators, who participated in the study. Metadata creators responsible for the creation of ETDs subject metadata at the university libraries in South Africa are a small population. The total number of metadata creators in all the libraries was not readily available and no sampling was conducted. Additionally, not all libraries had established institutional repositories at the time of data collection. The metadata of 22 libraries was accessible on the NETD database, as indicated in Table 2.2 (p. 46). Collections 9 and 10 given in Table 2.2 belonged to the same institution and were, therefore, combined into one.

As previously indicated, the metadata records on the NETD database link to the institutions hosting the original record (Webley, Chipeperekwa & Suleman 2011). The NETD list was the first reference source for the universities where the metadata creators were found. However, libraries that did not grant permission to conduct the study were excluded. All metadata creators, who belonged to the libraries with ETD metadata on the NETD database, were invited to participate in the questionnaire survey.

The participants of the study were considered as having the knowledge and experience in metadata creation and, therefore, suitable to give informed responses and multiple perspectives about the practice. Bias in the choice of participants was not a problem, as all metadata creators in the identified population had a chance to participate in the study. It is considered acceptable to select and collect data from the entire population, if it is relatively small (Collis & Hussey 2014:167).

5.7.1.4   Data analysis techniques

Data from the quantitative questionnaire survey was analysed descriptively. Descriptive statistics is used to organise, summarise and analyse data to render it more comprehensible (Punch 2005:124; Leedy & Ormrod 2013:277). The frequency occurrence (Collis & Hussey 2014:167) of the identified characteristics was presented in graphs, charts and tables. Responses to a few open questions were transformed into quantitative scales (data transformation), and analysed with the quantitative data, thereby allowing data integration.

Tables were used to present continuous or categorical data types and charts were used to present the single and categorical types of data. Data interpretation in the quantitative part was based on deducing meanings from the collected data by observing the scores of the different variables. The attitudes and behaviours, reflected in the questionnaire responses, were analysed and displayed to answer the research questions of this study. As pointed out by Punch (2005:153), this kind of analysis is possible for quantitative studies.

The aim of this study was to investigate the approaches that are being followed during subject analysis and how subject analysis theories are applied to facilitate geographic subject metadata creation for ETDs. The analysis helped to interpret the responses with relevance to the research aim, objectives and questions reflected in Table 5.1 (p. 117).

5.7.2   Qualitative interviews

Based on the findings of the questionnaire survey, interviews were conducted to gain in-depth information about geographic subject metadata approaches. The findings of the questionnaire survey were clarified in the interviews. This is in line with the nature of the explanatory sequential mixed methods approach where the qualitative phase yields further explanations to the findings of the quantitative phase (Creswell 2014:220, Leedy & Ormrod 2013:260; Brannen & Halcomb 2009:68). This sequence was adopted to conduct this study.

According to Leedy and Ormrod (2013:259), a phenomenological element may be incorporated in a mixed method study to investigate peoples' subjective "realities" perspectives. In this study, an interpretative phenomenological position is chosen as a theoretical basis for the qualitative interviews. Interpretation, in this study, aims to infer how subject analysis is understood based on the data collected from metadata creators who are experienced in the practice. Additionally, Mayoh and Onwuegbuzie (2015:94) indicate that interpretative phenomenological forms of enquiry look for extensive meaning beyond the mere descriptions. This form of enquiry allowed the investigation of the different perspectives held by the metadata creators and the influential contextual factors to metadata creation.

Interviews help to gain insight that compensates for the lack of depth in data gathered through the content analysis and the questionnaire and therefore contribute to collection of more complete information. In addition to completeness being achieved, the purposes of complementarity, focus and diversity, identified in Section 5.4, were also accomplished by using interviews as a second method for this mixed methods study.

5.7.2.1   Data types

The participants, who were the metadata creators at university libraries in South Africa, involved with creation of geographic subject metadata for the ETDs, were interviewed to gather different kinds of data. The interview participants were purposely selected from the South African university libraries used for the questionnaire survey. Fourteen (14) university libraries' names were identifiable from the returned questionnaires. The purposive selection strategy is discussed in Section 5.7.3.3.

Collected data included their experiences, views and other contextual factors. These types of data are examples of suitable data for a qualitative study, as pointed out by Henning, Van Rensburg and Smith (2004:6). Patterns were observed from the participants' responses. There was also an attempt to unearth the reasons for metadata practices as participants gave their responses. The subjective elements (Henning, Van Rensburg & Smith 2004:52) of the participants were managed by asking different carefully constructed probing questions to confirm the explanations.

5.7.2.2   Data collection instruments

Semi-structured questions for the qualitative study were derived from the objectives of this study and based on the subject analysis theories and conceptualisations identified from literature, which are discussed in Chapter 3. The choice of the data collection instrument for the qualitative part of this study is influenced by epistemological assumptions of conducting a qualitative study, which means that the researcher tries to get as close as possible to the participants being studied. Conventionally, interviews are thought of as an interviewer talking to an interviewee (Burns 2010). This commonly involves person-to-person interaction. On the contrary, as indicated by Burns (2010), this study used emergent media technologies as alternative forms of interviewing techniques including skype, where found necessary. Brannen and Halcomb (2009:68) and Guest, Namey and Mitchell (2013:49) confirm that interviews can be conducted by telephone, e-mail, Internet chat groups and other technological innovations, improving on the common traditional ways.

The interview questions were piloted using the face-to-face sessions to ascertain the clarity of the questions and the suitability of using both face-to-face and telephone approaches for the same set of questions.

Due to problems of easy access to participants for the pilot test, only one participant was used for the interviews. Subsequently, face-to-face interviews were used together with telephone and e-mail interviews, where suitable to the participants, to gather information from metadata creators, based on their individual views about the subject analysis practice. Telephone interviews were used where it was considered by the metadata creators to be suitable to give personal experiences with some level of confidentiality guaranteed and to engage without distractions. Proximity, time limits and the costs involved were some of the reason for resorting to telephone interviews, where it was deemed feasible.

Telephone interviews have a number of advantages and disadvantages. The advantages are in terms of proximity, costs, timing and space convenience. Warren and Karner (as cited in Guest, Namey & Mitchell 2013:49) mention the remote interviewing has advantages of anonymity and willingness to share any information that is considered sensitive. Telephone interviews have been used as an alternative method, with a consideration that metadata creators often work in shared office space environments, where it is not usually suitable to express personal experiences. Guest, Namey and Mitchell (2013:6) caution that a public or group setting is not a suitable setting to share personal experiences.

In addition, limitations may be experienced in terms of not having the opportunity to interpret nonverbal cues (tone of voice, gestures, etc.) and the possibility of cross talking occurring between the interviewer posing the questions and the participant responding to the questions, while the interviewer is typing/capturing the participant's responses (Guest, Namey & Mitchell 2013:51). Semi-structured interviews were based on a pre-established set of questions, with a flexibility to control their order as determined by how the interview progresses. This approach allows consistent presentation and comparability across interviews, thereby offering more reliability of data (Henning, Van Rensburg & Smith 2004:54; Guest, Namey & Mitchell 2013:21). An opportunity to explore further on the structured questions was provided and open-ended questions were included for the interviews.

5.7.2.3   Population and sampling

The population used for the qualitative interviews consisted of metadata creators in university libraries in South Africa, identified based on the findings of the quantitative study to have potential to share more information on subject metadata creation practices. This was a purposive selection of participants who had the potential to share in-depth information and reveal diverse perspectives on geographic subject metadata creation.

According to Leedy and Ormrod (2013:152), data sources for the qualitative interview can be chosen through selecting only individuals or objects that will yield the most important information about the topic under investigation, the information-rich participants. This method was used in this study to ensure that the interviews allowed deep probing of the information-rich participants. The units of analysis in the interviews were individuals from the sample of metadata creators from the universities that formed part of the content analysis and the questionnaire survey.

The interview participants for this study were identified based on the nature of the information about the respondents, obtained from questionnaire responses, which also indicated the different positions and experience levels of the metadata creators. Enquiries were made with the different participating libraries based on the information supplied as responses to the questionnaire. Based on this information, seven (7) participants were purposely chosen from six (6) university libraries that formed part of the population used for the survey. The selection strategy is further discussed in Section 6.3.1 (Chapter 6). The interview participants were also identified by consulting the websites of different South African university libraries, where the university name was given in the questionnaire survey and the staff and their responsibilities were listed. Consent by the participants to be interviewed was used to determine the final list of participants (Appendix D1–D10). The participants were limited in terms of the criteria explained in the above paragraphs.

Potential access barriers were considered. Unavailability and willingness of metadata creators or their institutions to participate were some of the potential obstacles considered in this study.

5.7.2.4   Data analysis techniques

As recommended by Saunders, Lewis and Thornhill 2016:550 and Bezuidenhout and Cronje 2014:232, transcription was used to prepare data gathered during the qualitative part of the research. The verbal responses to the interview questions were recorded during the interviews by using a laptop voice recorder and a cell phone recorder and later transcribed into a word processed document format. Notes were captured in handwritten form and on a laptop during the interviews as supplementary data. These were integrated with the verbal responses during the transcription. Data was reduced into manageable categories or coded. This process involved the analysis of the data to discard irrelevant data and to reorganise significant data into related categories (Collis & Hussey 2014:157; Bezuidenhout & Cronje 2014:232).

The interviews data collection method used categories into which the questions were organised. The notes were organised by being input into Microsoft Word tables according to these categories of the interview questions and further according to categorised responses. The research objectives of the study assisted in providing predetermined categories or codes (Appendix A). These were refined to obtain sub-codes after the data was collected. The verbatim recorded interview responses were played back with an audio tape recorder to be reproduced as a written account, in Microsoft Word format.

The NVivo12 software assisted in organising and exploiting qualitative data of this study. Data was imported into the Nvivo12 software and stored according to the different transcribed files. The files were examined and codes were assigned to the data according to the aspects that the study sought to identify. Discussions according to identified themes followed to provide detailed explanations of the data obtained from the participant's responses. According to Braun and Clarke (2006:78), thematic analysis is a form of analytic approach that facilitates the identification of themes or patterns within data, without being confined to the prescriptions of any theoretical approach. Themes similar to the questionnaire topics were drawn from the data on the subject analysis basis and metadata creation practices.

The NVivo12 software was preferred based on its suitability, sophisticated capabilities and speed of data manipulation. Training in the use of the software helped the researcher to understand its application. A statistician provided guidance in application of the software.

Intense interrogation of the interview data, as endorsed by Henning, Van Rensburg and Smith (2004:52), was undertaken. Henning, Van Rensburg and Smith (2004:105) state that the process of inductive making of meaning is highly interpretive. Furthermore, Bezuidenhout and Cronje (2014:234) mention that inductive analysis approach involves the use of raw data during analysis to develop themes without a preconceived conceptual framework.

However, the categories of questions, as described above, guided the analysis conducted in this study. Conclusions about the subject analysis approaches followed by the ETD subject metadata creators in South African university libraries were drawn from the in-depth interview responses to help explain the results of the quantitative analysis. The analysis conducted meet the rationale, as explained in Section 5.4, of complementarity, completeness, focus and diversity of the findings.

Qualitative research is not easily replicable and the reliability and validity of its findings was balanced by use of the quantitative findings of this study.

5.7.2.5   Trustworthiness of the interviews

With their detailed descriptions (Leedy & Ormrod 2013:104) that provided a deeper understanding of subject analysis practices , the qualitative findings accounted mostly for the validity of the findings of this study, which that can be read by others to draw their own inferences. Koonin (2014:258) indicates that there is increasing use of the term "trustworthiness" to represent the validity and reliability of qualitative studies. There are important criteria (credibility, transferability, dependability and confirmability) that were considered in this study, which accounted for trustworthiness. Different authors – including Teddlie and Tashakkori (2009:296), Pickard (2013:21), Koonin (2014:258) and Collis and Hussey (2014:172) – support these criteria for trustworthiness (credible, transferable, confirmable and dependable).

Credibility is defined as conducting research in such a manner that the subject of enquiry is correctly identified and described (Collis & Hussey 2014:26). In this study, the research problem was clearly articulated and the research questions were formulated in line with the research objectives. Additionally, the researcher conducted observations consistent with the recommendations of Teddlie and Tashakkori (2009:296) and (Collis & Hussey 2014:26) and engaged in persistent observation of the subject under investigation to obtain an in-depth understanding of the subject under study, so as to improve the credibility of the research findings. In addition, the researcher used a combination of several methods – i.e. the questionnaire, interviews and content analysis – to collect data from different sources. The lack of credibility was controlled by keeping alert throughout the research process and to control the researcher's subjectivity during the data interpretation. This helped to interpret the interview data with no intent to modify the participants' descriptions, thereby increasing the possibility of the research findings being believable to the participants (Koonin 2014:258).

Transferability is explained as being concerned with the applicability of the findings to a sufficiently similar situation to permit generalisation (Collis & Hussey 2014:172). Generalisation is not the main goal in this qualitative inclined phase. It is not expected that the exact same results would be found when the interview method used in this study is applied in other settings. The purposive sampling of interviewees, who were knowledgeable in the subject under investigation was intended to obtain in-depth explanations to clarify the findings of the quantitative part of the study (see Section 5.7.2.3). Thick descriptions, which are detailed descriptions of the context and implementation of the study (Leedy & Ormrod 104), were used to enhance readers' understanding of the research process of this study, so that the concepts and designs can be applied in other environments (Collis & Hussey 2014:54; Koonin 2014:258).

Dependability refers to the extent to which the process of the inquiry is dependable, in that the data collection instrument can yield consistent results across different contexts (Teddlie & Tashakkori 2009:296). The process of implementing the three phases followed the criteria of well a documented research process and thorough execution of the processes in a systematic order that can allow testing of the consistency of the data (Collis & Hussey 2014:172).

As a result, it may be possible to repeat the methods used in this study, but, due to the qualitative nature of this phase of the study, it may be difficult to replicate the research findings. In addition, the quality of the integration between the data collection methods, data analysis and the interpretation of the data was repetitively verified to ensure that the process was dependable (Teddlie & Tashakkori 2009:296). This was also tested through the assessment of the c supervisor, who examined the research process.

Teddlie and Tashakkori (2009:296) defines *confirmability* as the extent to which the product of an inquiry is confirmable, including whether the results are grounded in data, whether the inferences are logical and whether there is enquirer bias. In this study, detailed descriptions in the reporting of the findings helped to show how the results were linked to the raw data (Collis & Hussey 2014:172) and that they were not the researcher's subjective views. This in-depth information given during reporting provided a basis that helps to confirm the interview findings.

As indicated in Section 5.7.2.2, the interview questions used in this study were piloted and the questions were considered useful for the purpose of measuring different attitudes towards subject analysis and geographic subject metadata creation for ETDs.

The next section discuss the way in which content analysis was employed in this study.

5.7.3   Content analysis

Leedy and Ormrod (2013:148) defines *content analysis* as:

> *A detailed and systematic examination of contents of a particular body of material for the purpose of identifying patterns, themes or biases*

Another definition of content analysis is given by Krippendorff (2013:24):

> *A research technique for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use.*

The two definitions reflect important aspects of data content collection and analysis, in that they indicate the purpose of content analysis – i.e. to identify patterns and themes and to make inferences from the data. All these purposes are aimed at drawing meaning from the textual data.

The existing literature offers different arguments about the nature of content analysis. According to Collis and Hussey (2014:166); Neuman (2014:371), the quantitative aspect of content analysis involves systematic counting and recording procedures to produce a numerical description of the studied content. The results must be expressed numerically (Du Plooy-Cilliers & Cronje 2014:169). However, content analysis that is conducted from a qualitative approach goes beyond statistical measures and seeks meaning in the records (Neuman 2014:372). From a qualitative position, meaning must be sought outside the body of the texts themselves, so as to address a specific purpose or to be relative to a specific context (Krippendorff 2013:29).

In this study, content analysis is understood on the basis of the argument offered by Krippendorff (2013:22 & 88), who rejects the dichotomy of the quantitative and qualitative content analysis, stating that both aspects coexist, and both start with the observation of text, which is qualitative in its nature. Leedy and Ormrod (2013:149) agree by indicating that, for data analysis in any content analysis study, the approach is both quantitative and qualitative, as they all involve tabulating frequencies of important characteristics of the analysed content. Due to the nature of an explanatory sequential design in this study, the emphasis on the findings of the content analysis was the support of the qualitative findings in the explanations of the quantitative findings. Although data was first analysed quantitatively, emphasis was on the subjective (qualitative) interpretation of the data.

The discussion of common advantages and disadvantages of content analysis helps to give further context on way in which the method was used in this study. Saunders, Lewis and Thornhill (2016:612), Collis and Hussey (2014:168) and Krippendorff (2013:45) identify several advantages and disadvantages of content analysis. A common advantage is that it is an unobtrusive technique to reveal hidden realities about a particular phenomenon (Collis & Hussey 2014:167; Leedy & Ormrod 2013:102).

As a result, it is unlikely for the researcher's influence on the subjects under investigation to occur. Content analysis can handle large qualitative sets of data by reducing them to manageable units of analysis. ETD repositories house large collections and, therefore, the extracted data to be used in this study had to be categorised and summarised.

Additionally, content analysis is a context-sensitive approach, and inferences are drawn from the data situated in the context of their origin. Therefore, data in this study aligned to the context of the libraries and the purpose for which they were created. Another advantage is that permanent records can be analysed several times and procedures can be set out clearly. The readily available data used in this study allowed for the validation of accuracy. Furthermore, content analysis allows the study of trends, shifts and patterns over time, which was possible in this study.

There are several common disadvantages of content analysis that were experienced in this study. Limited availability of data or documentary sources may be a challenge. Full metadata was not available from some library repositories and, therefore, they had to be excluded for the content analysis. As explained, only records with accessible full metadata were useful for observation in terms of the research problem. Additionally, the process can be time-consuming and tedious, with large amounts of data to work with. Processing data for content analysis took longer, due to manual processing. The databases were unstable at times, which delayed the collection of data. Furthermore, content analysis requires a consistent approach and demands much concentration. Organising and cleaning the data manually to make it ready for analysis was a lengthy process.

Carefully planned steps on conducting the content analysis were documented and followed, which assisted in improving trustworthiness of the process and the findings. Table 5.6 (p. 155) presents a summary of the process, with more details discussed under the different subsections.

Table 5.6: Content analysis process summary

Criteria	Process steps
Identify a source to harvest metadata of ETDs from South African universities (ETD metadata records in a common format to facilitate easy analysis)	The NETD database was identified as the source for data harvesting. (Only universities listed on the NETD database at the time of harvesting data were included for the content analysis).
Only ETD records were to be harvested for analysis from the NETD database	A record harvesting procedure was performed through a programmer to derive ETD titles from the NETD database. (Permission was obtained from the NETD database administrator to harvest titles from the platform for observation).
The record fields with the required information were the: i.) Title fields containing geographic names	A list of records with titles *that contained geographic names* was harvested from the metadata records on the NETD database. (The metadata records on the ETD database consist of different fields like titles, author, subjects and other bibliographic descriptive fields).
ii.) Subject fields	The subject fields were explored to study the subject types used for description of the ETDs.
The set timeline was to draw a list of ETD metadata records created between 2014 and 2018 only.	Only records created between 2014 and 2018 were included in the harvested list of records to be explored.
A second step of the analysis involved the verification of the types of controlled or uncontrolled subjects used.	Full view of records showed all *subject types* (controlled or uncontrolled) used. The Metadata records on the NETD database hyperlinks to the institutional repositories to facilitate full record view, which shows all subjects used (controlled and uncontrolled).
The description schema (e.g. Dublin Core) subject description elements (e.g. dc:subject) were examined to establish the type of subject metadata used.	Subject elements and their values (type of subject metadata, controlled or uncontrolled) were analysed.

Content analysis follows a systematic approach. The basic components of content analysis include sampling, determining the unit of analysis, coding and analysis (Saunders, Lewis & Thornhill 2016:611). The steps involved in this process steps are briefly outlined to provide further context to the content analysis.

The specific procedures and findings of the content analysis phase can be referenced to the context of the NETD database and the five libraries' ETD repositories included in this study. The NETD database was the primary source of identifying the universities listed on their database. An initial list of the twenty-three (23) South African university libraries collections was compiled from the NETD database. However, further analysis yielded 22 libraries, since two collections on the list belonged to the same university. The NETD database contains ETD metadata from university libraries falling within the categories of traditional, comprehensive and Universities of Technologies. Only ETD repositories listed on the NETD at the time of extracting data were considered for inclusion in the study. The reason for using the NETD database as a primary source was to have a source where metadata from the different universities could be extracted at once in a common format.

The services of a programmer were used to conduct automatic extraction of the records from the NETD database for records with a creation date between 2014 and 2018. Only institutions that granted permission were included in the study. Metadata records were observed for titles with place names. In addition, ETD repositories were consulted to observe the subject metadata of the identified records. This second step was necessary, because the metadata on the NETD database does not show the full nature use of subject metadata, but the record do link to the original library repositories where they were harvested to show full subject metadata.

The next sections provide more explanations on the specific components of the content analysis and the procedures followed in this study.

5.7.3.1 Data for content analysis

The unit of analysis in the content analysis was the texts representing the geographic subject metadata on ETD records. The unit of analysis was the selected words or phrases used in a particular context (Mcnamara 2018:3). Additional units were introduced to make sense of the data. Other types of subject metadata and the schema descriptive element used (e.g. dc:subject) were also examined.

The characteristics to look out for were primarily categorised as the occurrence and frequencies of use of geographic subject metadata in the ETD metadata records and the nature of use of the geographic subject metadata used on the identified records, so as to determine whether they were controlled (e.g. LCSH) or uncontrolled terms. The units of analysis were assigned into categories to make them analysable. This categorisation is a form of coding for content analysis (Mcnamara 2018:4; Krippendorff 2013:381). Mcnamara (2018:4) explains that coding the unit categories will commonly indicate whether representations are coded as positive or negative. The characteristics were coded and first analysed descriptively, with the ultimate purpose of drawing meanings to fulfil the explanatory purpose.

The sources of data for the content analysis consisted of ETDs records, the subject description elements of the records (schema description elements) and the values (kind of subject metadata) they contain. Content analysis was conducted on the final selection of five (5) universities' ETDs repositories, with relevant data to address the research problem. Five libraries were used in the study, because of common problems identified in the literature and by Saunders, Lewis and Thornhill (2016:612) and Collis and Hussey (2014:169) that relate to access to sources of data, and compatible and unusable data, which is common to content analysis studies. Collis and Hussey (2014:169) give an example that "perhaps you want to analyse data for the past five years, but find that one quarter data is not available". The repositories used for the record observations were suitable to provide access to the data that matched the identified criteria of fully accessible metadata, including subject metadata, for the period 2014–2018.

These different library ETD repositories allowed searching in various ways, including collections, issue dates, titles, authors, keywords and subjects. Metadata records created between 2014 and 2018 identified through issue or production dates, were examined through intensive manual analysis. Concentration was placed on observing whether the geographic subject metadata was captured on the ETD records; how they were assigned and which schema description elements (e.g. Dublin Core descriptive elements) were used to capture the geographic subject metadata. It was also examined whether controlled (standardised) or uncontrolled (free text keywords) subject metadata was allocated.

The data was analysed and the findings integrated with those of the questionnaire and interviews for interpretation (as shown in Figure 5.1) and further discussed in Section 5.8.

5.7.3.2   Data collection instruments

The instrument content analysis was a coding scheme from which the checklist or coding form was developed. The coding scheme is a set of rules or instructions on how to observe record content from text (Neuman 2014:373). The initial coding form was developed with the guidance of a statistician. A pilot study was conducted by using records from an initial list harvested by the programmer. Observations were conducted on the existing ETD records hosted on the NETD database and the libraries' ETD repositories to pilot the data collection tool. Computerised searches were conducted by using the search functions of the individual ETDs repositories. The facet limit of "date issued" was used to retrieve ETDs issued within the specified period of 2014 to 2018 only. The coding form facilitated the determination of the characteristics of the geographic subject metadata.

Based on the pilot study experiences, the coding scheme was revised and finalised. A few changes were made to the coding scheme. Additional units of analysis like other types of subjects and description elements, were included to help make sense of the data. After the final coding rules were finalised, an improved coding form was used. The reviewed literature (Krippendorff 2013:127; Neuman 2014:374; Fink 1995:9) explains that the coding scheme contains detailed written rules that coders can apply reliably in content analysis. The final coding form was used to collect data and to record compliance to the identified criteria, observing the developed coding guidelines. The coding scheme and the descriptive information is attached in Appendix B.

Krippendorff (2013:127), Neuman (2014:374) and Fink (1995:9) indicate that measurement for content analysis uses structured observations or structured records review. This involves systematic, careful observation of the texts, guided by explicitly setting rules or criteria on how to categorise the observations. The records were observed according to the rules set out for use in this study, as discussed in Appendix B. The researcher used a coding form or checklist to collect the data and the rules guided how to record data under different categories.

The resultant categorisation from the observations constituted the basic form of measurement (Krippendorff 2013:88). For the instrument to be applied, human observation was necessary as planned in this study.

In this study, the observations of the researcher (coder or observer) involved systematic analysis or careful observation of the content of ETD records, based on the set coding rules and procedures on how to categorise and classify observed data. The researcher's manual examination of the ETD records helped to identify the values in the form of geographic subject representations, used within the subject description elements. The values in the context of this study, which were the types of assigned subject metadata, are discussed in detail in Section 3.3.4 (Chapter 3). These are the subject headings assigned to represent the subject content of ETDs. Specific focus is placed on identifying the values in the form of geographic subject metadata.

The researcher (coder) transformed the units of analysis into analysable data by coding. According to Krippendorff (2013:126), coding can also be regarded as a form of recording criteria. The categorisation of data into manageable forms was conducted by producing summaries of Excel lists before further analysis was conducted. Collis and Hussey (2014:167) recommend the use of a coding frame, which lists the coding units and the analysis for each unit in a structured form. A similar format was used in this study. The procedure summary, presented in Table 5.6 (p. 155), and the coding scheme (Appendix B) outline how the recording of data was executed and state the inclusion and exclusion criteria. To maintain consistency, the coding rule involved primarily looking out for geographic place names, with other forms of subject metadata examined to provide context.

The assigned geographic subject metadata was examined to ascertain their nature of use. The observations were also based on the criteria contained in the coding form (Appendix B). Observations were focused on the type of description elements used (free- text keyword or controlled element) to identify geographic subject metadata type. Validation was conducted against the commonly used description schema, vocabulary and authority standards. The common standards for the bibliographic description elements sets and vocabularies include the Dublin Core and Metadata Object Description Schema (MODS).

The Library of Congress Subject Headings (LCSH), the FAST headings, the South African Geographic Council place names and Getty's Thesaurus of Geographic Names that are commonly used to control the vocabulary and authority forms of subject headings.

The population and sampling, the analysis of the content analysis and the trustworthiness of the content analysis method are discussed in next sections.

5.7.3.3 Population and sampling

The population for the content analysis was the metadata records in the ETD repositories of South African university libraries that were hosted on the NETD database. The libraries with metadata on the NETD were university libraries. The NETD database provides harvested university libraries records that are accessible from one common platform. These university libraries with ETD metadata records that were harvested to the NETD database provided the initial list of all the libraries and their ETDs records. The reason for using records on the NETD as the initial source was that the common platform allowed harvesting of different universities' ETD records that are accessible from one platform and in a common format. The metadata records on the NETD database linked to the libraries hosting the original record (Webley, Chipeperekwa & Suleman 2011). The NETD database administrator granted authorisation to harvest the metadata on the database.

The records selected for the content analysis phase of this study were identified from six institutions that formed part of those that participated in the survey and the interviews. The population was appropriate, based on the focus of the study being ETD geographic subject metadata in South African university libraries. The criteria and the steps followed to select the records that were analysed are explained in the next paragraphs.

Collis and Hussey (2014:167) and Krippendorff (2013:365) indicate that, if there is a large population of relevant texts of data and it is not manageable and practicable to be analysed within a reasonable time, a sample must be selected. Purposive or relevance sampling (Krippendorff 2013:120) was used to obtain records suitable to address the research questions.

Furthermore, Teddlie and Yu (2007:82) observe that, sometimes complex sampling procedures are used due to the complexity of the phenomenon under study, in order to find uniform cases purposely to answer the research questions. Multiple steps were taken between the NETD database and the libraries' repositories to identify cases purposely and to compile the final list of records to be analysed in this study. Varied factors, like library permissions and accessibility of full records, contributed to the complex selection procedures. Initially, all records on the NETD database created by the participating libraries between the years 2014 and 2018 were harvested. Selection was done to identify only those records that matched the coding criteria shown in Appendix B.

Following the analysis of the list according to the identified criteria, a final list was produced. Searches for the identified titles were conducted from the libraries' IR repositories for further analysis. The NETD database was the first source to identify the different university libraries' ETD metadata records in a common format and the second source for examining the full records were the universities libraries' repositories.

The criteria used to determine the records to be harvested from the institutional records can be summarised as follows:

- To conduct a preliminary test, searches were conducted for ETDs published within a five years' period, between 2014 and 2018. Computerised searches were conducted by using the search functions of the ETD repositories. Filtering was done by using the "date issued" facet to retrieve only records for ETDs issued within the specified period of 2014 to 2018.

- Automatic extraction of the records from the NETD database through the services of an identified programmer for records harvested between 2014 and 2018 was conducted in April 2019. Only institutions that granted permission to conduct the research were retained for further analysis.

- To examine the harvested records manually and to select only those titles representing research centred on specific geographic localities by identifying any place names reflected in the titles. A study conducted by Terra et al (2021:2) found that important keywords in the titles could be used as indexing descriptors. Operating from the same understanding, the geographic place names in the titles were considered as important elements to be used for subject descriptions.

- To clean the records, removing the duplicates and titles not in English language. The language criteria was important for the correct identification of place names.

The final lists of records that qualified according to the indicated criteria consisted of 4870 titles. The big data approach of examining everything in the databases was avoided, as it would not be practicable for this study. All suitable records were purposely selected in the outlined manner from the large data sets of records harvested from the NETD database. The hyperlinks from the NETD database and titles were used to conduct searches for the records to be analysed from the institutional repositories of the libraries involved. Libraries that keep their full metadata in the background were excluded.

A manual analysis was conducted from Excel lists to select only those records containing geographic localities by identifying any place names reflected in the titles. A sample refining step was to use existing sampling *(N 5000 / S 357)* guidelines from the literature (Krejcie & Morgan 1970:608). According to this formula, a sample of 1785 is representative for a total of 4870 initially identified records for 2014–2018.

The second step was conducted to improve representativeness of records created over time. To be more systematic and to improve sample representativeness, the records were systematically selected by including every second record on the individual libraries' lists.

However, for one library's list with more records, 2169 in total, every fifth records was selected and the sample size of 434 records met the criteria (*N2200 / S327*) set by Krejcie and Morgan (1970:608).

Collis and Hussey (2014:199) endorse the sample size guidelines. Generalisation is not the main goal in this qualitative inclined phase, but transferability of the findings is intended to apply across similar settings (Collis & Hussey 2014:45).

The records analysed in this study were representative of the variety of records created in the ETDs repositories of the different South African universities. It was considered that the data that was based on these criteria would provide information on the extent (frequency at which they are assigned) and the nature of the usage, including how much variability (if there was no consistency) in the practices of assigning geographic subject metadata in the different years. The findings were not meant to be generalised, but the sample was representative of all the three university category types and the results could be applicable to the contexts of the different libraries.

5.7.3.4   Data analysis techniques

Content analysis follows a sequential approach, including analysis of the data (Saunders, Lewis & Thornhill 2016:611). In line with the views of O'Cathain et al (2010), different data integration techniques were considered in this study, to improve the credibility of the process of integration of data during the analysis stage. An integrated and sequenced analysis was conducted for the data collected by means of content analysis. Analysis was influenced by the follow-up nature to help explain the questionnaire findings. The findings of the content analysis were used, together with the interview findings, to help explain the quantitative analysis findings. The analysis was influenced by the questionnaire results and the purpose to supplement the interview explanations. The introduction of the data analysis for the content analysis phase resulted in an iterative sequential mixed data analysis (O'Cathain et al 2010), defined by Teddlie and Tashakkori (2009:277) and as the analysis of data from a sequential study that has more than two phases. The analysis of the different phases built upon one another.

A mixed quantitative and qualitative analysis was conducted by using frequencies and percentages and thematic discussion based on the interpretations and the meanings drawn from the data.

According to Onwuegbuzie and Combs (2010:15), qualitative analysis can be conducted through counts of a particular element. Descriptive statistics, which can be used as needed to answer the research question (Leedy & Ormrod 2013:150), was used to analyse the content analysis data. The units (geographic subject metadata, other subjects, controlled and uncontrolled descriptive elements/fields), identified from the ETD records were recorded and analysed by using the Excel spreadsheet. The Excel software data filtering function helped to clean and manage the data. Tables and graphs were inserted through the Excel analysis functions to display the data summaries.

Tabulating frequencies of each characteristic found in the examined material or texts is crucial in any content analysis, whether quantitative or qualitative (Leedy & Ormrod 2013:149; Krippendorff 2013:88). Frequencies of occurrences of geographic subject metadata, together with additional units of other subjects, were tabulated and presented through graphs. Frequencies in the use of geographic subject metadata helped to confirm the extent of usage and to establish whether there was dominance in the use of specific categories of subject metadata fields, or any significant differences in usage with relevance to the objectives of this study.

Furthermore, the quantitative results were used to identify patterns and relationships to address the research questions. The discussions were aimed at supporting the interview findings in line with the explanatory design purpose. A qualitative reflection allowed inductive analysis of the data categories with reference to their context. Providing descriptions from the summaries of collected data helped to explain trends and patterns of use across the different libraries.

As indicated in the literature, content analysis can be used in combination with other qualitative techniques to support the explanatory purpose, where data is analysed for that purpose (Saunders, Lewis & Thornhill 2016:612). Qualitative analysis used themes derived from the quantified data to describe the findings. The researcher used themes related to those used for interviews analysis, as the findings were meant to supplement or validate the qualitative explanations. The themes enabled descriptions that yielded supplemental information on the observed of patterns and trends.

The techniques used to analyse the content analysis data were meant to identify the extent of use of geographic subject metadata and to obtain information on the nature of geographic subject metadata to address the research problem.

Furthermore, the categories on the use of controlled terms, as compared to uncontrolled terms, were used to find meaning about the vocabulary control trends. Additionally, the information from the descriptive summaries was used to align the different data categories (unit categories) to the interview themes, so as to explain the use of controlled and uncontrolled terms. Moreover, the content analysis data was analysed to infer the subject metadata creators' approaches to subject analysis from the meanings interpreted from of the descriptive summaries. Krippendorff (2013:55) affirms that inferences can be drawn on the patterns of behaviours and decision-making of the actors from content analysis data. The relationships with prior findings, from Phases 1 and 2, on the metadata creators' perspectives and practices were verified. Existing literature indicates that analysis is completed by describing the findings and reporting the trends and patterns to provide answers for the research questions (Saunders, Lewis & Thornhill 2016:610; Krippendorff 2013:84).

In conclusion, the data analysis techniques used in this phase of the study were applied with the understanding that the quantitative results did not give insight into the underlying reasons for the patterns and trends observed in the data. Additionally, the understanding of the metadata creators on subject analysis could not be revealed through the quantitative analysis technique alone. However, including qualitative analysis made it possible to interpret patterns and trends (Krippendorff 2013:53 & 55) and the underlying metadata creators' practices.

During interpretation, the results of the content analysis were combined with those of the questionnaire and the interviews, in order to provide a comprehensive understanding of subject analysis approaches and geographic metadata creation. This type of analysis and integration, facilitated by the mixed methods approach, strengthened the integrity of the research findings. The research problem of this study is addressed through the interpretation of the aggregated findings of the different analysis phases of the study.

5.7.3.5 Trustworthiness of the content analysis

The quality measures were relevant for testing the rigour of the entire content analysis process, based on its quantitative and qualitative qualities employed in this study. Trustworthiness or qualitative validity (Creswell & Plano Clark 2018:217) was a critical evaluation criterion because of the qualitative emphasis of the content analysis. Trustworthiness (credible, confirmable and dependable) is the key criterion mentioned by, among others, Teddlie and Tashakkori (2009:26) and Krippendorff (2013:89). The criterion was used to test the quality and rigour and the results of the content analysis. The context for the evaluation was provided by the emphasis on qualitative follow-up explanations.

Credibility: According to Leedy and Ormrod (2013:262), trustworthiness involves the extent to which the research findings are compelling. To maintain credibility (similar to internal validity), the researcher ensured the proper identification of appropriate sources (the records); proper capturing of the collected data; rechecking data for verification; and controlling subjectivity, guided by the explanations from Saunders, Lewis and Thornhill (2016:206) and Neuman (2014:218) and guided by the research objectives.

Confirmability involves the degree to which the process steps are described fully, so that it can be confirmed that the results are derived from the collected data (Collis & Husey 2014:172). The study provided full descriptions of the process to allow validation. As indicated in the earlier discussions under Section 5.7.3, the content analysis process was clearly detailed to allow verification. Furthermore, globally acceptable subject metadata creation standards were readily available to compare the schema description elements use and the geographic subject metadata used in the ETDs records and assist interpretations of data.

Dependability, which was one of the criteria found to be relevant for this study, was used to measure quality and rigour of the content analysis. *Dependability* is defined by Collis and Hussey (2014:172) as the systematic, rigorous and comprehensive documentation of the research process. Document analysis processes are systematic and detailed and are well documented to improve trustworthiness of the procedures.

A pilot study was conducted to test the reliability of the content analysis by testing the feasibility of the procedures. The first pilot test was done in February 2019, with a script written by a programmer. ETD records created between 2014 and 2018 were harvested from the NETD database and the records were examined for suitability and before a re-harvesting was done for records created within the identified period. The method was refined before the actual investigation was conducted.

Validation procedures were employed to enhance the acceptance of the findings of the content analysis. Krippendorff (2013:329) points out that validation procedures can help to justify a content analysis research. Semantic validity Krippendorff (2013:334) was ensured by means of accurate descriptions of concepts for the different aspects of content analysis. These descriptions allowed correct interpretations of their use in the context of the research problem. Furthermore, functional validity was ensured by conducting content analysis according to the correct steps – as explained in Section 5.7.1 – in order to accomplish what it was meant to confirm (i.e. the frequency of use of geographic subject metadata). Functionality was based on the understanding of the acceptable meanings and practices in the subject metadata context.

The next section shows how the collected primary data was integrated and reported.

5.8   Data integration and reporting

The data interpretation accommodated the integration of the findings from the complete pool of data collected in the course of the study. Integration in a sequential explanatory design involves planning how to collect data, conducting separate analyses and connecting them to the quantitative results to help plan for the qualitative phase (Creswell & Plano Clark 2018:80). This section explains how integration also occurs during data interpretation and in combining the findings from the different methods. The qualitative phase helps to explain the quantitative findings.

Integration at different stages involved the following three phases building on one another: (i) quantitative results informing the sample choice and question refinement for the interviews; (ii) influence on the content data analysis by the findings from the preceding phases; and (iii) integration at the interpretation stage.

The integration of the findings into a coherent whole, as fulfilled in this study aligns with the requirements for a mixed methods study (Creswell & Plano Clark 2018:219 & 2020; Teddlie & Tashakkori 2009:324). The integrated interpretation involved the use of combined findings that allowed conclusions to be drawn based on the overall results. Tashakkori and Teddlie (2003:667) highlight the importance of the researcher's values during interpretation. The values of the pragmatic worldview influenced the interpretation of the results.

The reporting of the findings included the results from the questionnaire, the qualitative interviews and the content analysis. Separate analysis was conducted on the data, after each phase of the study. The reports from the results of the different data collections were integrated during interpretation, with qualitative findings helping to explain the quantitative findings, to draw the final findings of this study. Furthermore, the literature reviewed in Chapters 2–4 was used to assist the interpretation and reporting of the results. Morgan (2014:41) states that there needs to be a plan for integrating the results of the different methods. Integration of the findings from the quantitative questionnaire and the qualitative interview and the quantitative/qualitative content analysis occurred at the interpretation stage of this study, as shown in Figure 5.1. Integration at this stage of research is supported by the literature (Creswell & Plano Clark 2011:84; Bryman 2006:101). The reporting reflected the integrated findings.

As argued by O'Cathain et al (2010), in a mixed method study, it is important for the integration to yield knowledge that reflects comprehensive findings that are presented as parts from independent studies. The findings of the survey, interviews and the content analysis are presented as a whole in Chapter 7, thereby presenting a comprehensive report of the findings and an in-depth understanding of the theoretical approaches to subject analysis and ETD geographic subject metadata creation.

5.9 Ethical considerations

The ethical considerations for research, as outlined by Leedy and Ormrod (2013:104) and Punch (2005:227), were found relevant to and were observed in this study. The research guarded against the harming of participants, coerced participation, privacy invasion, dishonesty, unprofessional conduct, lack of research integrity and quality. No physical or psychological harm was inflicted to the participants. This was avoided by pilot testing the questionnaire and interview questions, as pointed out earlier. The participants of the pilot test were requested identify any questions with which they felt uncomfortable and that needed to be avoided.

Voluntary participation was attained by circulating a consent form (Attachment D) prior to conducting the study. The informed consent sought from the participants was to ensure that they willingly participated in the study. No incentives were unlawfully administered.

Anonymity, confidentiality and privacy clauses were included in the questionnaire and explained prior to the interviews. The identities of the participants and the institutions were protected. Permission was sought from the ethics committees of the various universities where it was required. Requests were also made to the management of the libraries in which participants were based, in order for them to be officially cleared and permitted to participate in the study. Ethical clearance was sought from and granted by the UNISA Research Ethics Committee.

The researcher guarded against data falsification by ensuring that the findings were reported without any distortions to suit the research. The data that was provided by the participants was acknowledged as such to avoid plagiarism. Research ethics and acceptable conduct were observed to guard against unacceptable practices.

5.10 Chapter conclusion

This chapter discussed the research methodology employed in the study. The pragmatist research paradigm informed the study, while a mixed methods research design was found to be relevant and, therefore, adopted to the study.

An explanatory sequential mixed method design allowed conducting the quantitative and qualitative study in a suitable way to achieve the purpose of complementarity, completeness, focus and diversity. A questionnaire, semi-structured interviews and coding scheme for the structured observations were employed to collect the primary data.

Integration in a sequential explanatory design involves connecting the results from the quantitative to help plan for the qualitative phase (Creswell & Plano Clark 2018:80). The qualitative phase helps to explain the quantitative findings. In this study, a design with an additional quantitative/qualitative phase was used. The collected data was analysed in sequence: questionnaire data, followed by the interview data and lastly the content analysis data. The questionnaire findings were explained through the interview and the content analysis findings. Findings from the content analysis provided supplementary information for the interview explanations. The combined results were integrated for interpretation and to draw comprehensive conclusions. This was followed by a credibility check of the findings by evaluating the trustworthiness, validity and reliability. An explanation of the ethical issues observed was given and all limitations clarified.

The chapter concluded by indicating how the results were synthesised in light of the research questions. The validity and reliability discussed under each of the three methods used in this mixed methods study were maintained. The use of mixed methods requires establishing the validity of scores from the quantitative measures and discussions of the trustworthiness of the qualitative findings (Creswell 2014:225). The varied methods used in this study also accounted for the credibility of the findings. For explanatory sequential designs, issues of validity include connecting the initial quantitative results with the qualitative follow-up (Creswell & Plano Clark 2018:52). The methods discussed in this study were suitable for the different types of quantitative and qualitative data that was collected.

Chapter 6 presents the data analysis and reports on the research findings.

CHAPTER 6: DATA ANALYSIS AND RESEARCH FINDINGS

6.1   Introduction

The study investigated subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations. The study was conducted to address the following research objectives, which were the foundation of the analyses of the data collected through the three methods:

1. To investigate the application of geographic subject representations during the creation of ETD metadata in South African university libraries;
2. To collect data from the practitioners involved in metadata creation in South African university libraries on how they approach the determination of geographic subject metadata for ETDs;
3. To establish from the subject metadata creators in South African university libraries what the implications are of their different analysis approaches for the determination of geographic subject metadata for ETDs; and
4. To establish what type of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries.

The data analysis is aligned to the order followed in the explanatory sequential mixed method design that was employed in this study. Creswell (2014:224) explains this design as involving the collection and presentation of quantitative data in the first phase of the study, followed by the qualitative phase. The study was conducted in sequential phases by using a questionnaire survey, followed by the interviews and the content analysis. This chapter presents and analyses the primary data collected through the questionnaire survey, interviews and content analyses phases of this study. The presentation of the findings are presented is based on the three methods used in this study.

A questionnaire survey was conducted among the South African university libraries, followed by the interviews. Content analysis was used as a third method for this study. An overview of the types of data and the data collection methods is presented as an introduction to the analysis.

The research design, approach and the quantitative and qualitative data analysis techniques, and the procedures of the mixed methods research approach are outlined in detail in Chapter 5.

6.2   Questionnaire survey quantitative analysis

Following the data collection, separate analyses and reporting of the results drawn from the questionnaire responses were conducted. Descriptive statics were used to analyse the data. The frequency distributions (Collis & Hussey 2014:235) of the identified characteristics were presented in graphs, charts and tables. The quantitative data analysis techniques used to analyse the questionnaire survey results are outlined in detail in Chapter (Research methodology). The survey was conducted among ETDs metadata creators at South African university libraries in. The libraries are presented in Table 5.1 by using codes to protect their confidentiality. The data was collected on subject analysis and the creation of geographic subject metadata in the form of responses to the questionnaires (Annexure A) that was distributed online.

Responses were received from the university libraries that granted permission to conduct research among their ETD metadata creators. The libraries that responded had identifiable ETD repositories on their library websites and on the National Electronic Theses and Dissertation (NETD) database. It was predetermined that the names of the participating libraries would not be disclosed, as the aim was to gain general understanding of on subject analysis and ETD geographic subject metadata creation. Naming the libraries by the respondents was voluntary and, in order to protect their confidentiality, codes are used to identify and categorise the libraries. This was maintained throughout the presentation of the findings from the survey.

The respondents were requested to answer individual questions. The questionnaire primarily consisted of closed-ended questions, with a few open-ended, follow-up questions (Appendix A). The data was analysed with the assistance of a statistician, by using Version 25 of the Statistical Package for the Social Sciences (SPSS) to determine frequencies of manually identified characteristics.

The descriptive statistics are presented in the form of tables and graphs that indicate the scoring patterns for the different variables. The analysis of the findings will be presented in the sequence that was followed in the questionnaire.

Item non-response error (Bickman & Rog 1998:477) was low in this study, in that failure of the respondents to answer individual questions was very low. The responses to the questions posed in the questionnaire are presented in the next sections.

6.2.1   Demographic and background information

This section reflects the general demographic information and other job aspects that gives important background data about the surveyed libraries, the respondents and their roles. The researcher concentrated on several aspects: participating libraries, the number of ETD metadata creators, positions of participants and experience in ETDs metadata creation.

6.2.1.1   Participating libraries

Permission to conduct the study was requested through the various ethical clearance procedures, as advised by the university libraries. The libraries of the different universities that granted the permissions were approached for final permission to proceed with the study. A total number of 29 questionnaires were returned. They were completed appropriately according to the aim of data collection. Table 6.1 identifies the university libraries where the data was collected and the responses from each institution. During the time of conducting this survey – October 2018 and May 2019 – 22 South African university libraries were identifiable on their library websites and on the NETD database as institutions with established ETD repositories.

The library management of the different university libraries were requested to refer the invitation to the ETDs metadata creators in their libraries. Respondents were asked to state the name of their libraries. Out of the total of 29 responses received, 24 responses (82.76%) indicated the names of the libraries where the data was collected and five (5) respondents (17.24%), were coded as a category of anonymous libraries, where the library names were not disclosed.

Table 6.1 (p. 174) outlines the university libraries at which data was collected.

Table 6.1: University libraries where data was collected

Library	Number of responses
A	1
B	1
C	5
D	1
E	3
F	1
G	1
H	2
I	3
J	1
K	1
L	1
M	1
N	2
O (Anonymous)	5
Total	**29**

6.2.1.2   Number of ETD subject metadata creators at various libraries

The following survey question was posed to the respondents:

How many metadata creators in your institution are involved with creation of subject metadata for the ETDs? (Please indicate the number).

This question was asked, in order to get insight into the number of ETDs subject metadata creators in each participating library who were involved in subject metadata creation for ETDs. There is no readily available statistics on the numbers of ETD subject metadata creators in South African university libraries. The researcher also aimed at establishing if ETD metadata creators involved in the general metadata creation roles were also involved in subject metadata creation.

Table 6.2 presents the responses given on the numbers of ETD subject metadata creators in the participating libraries. Not all the responses could be ascribed to specific libraries, as some of the responses were anonymous. Varying numbers of metadata creators responded at each participating library.

No specified categories of responses were given to choose from and, therefore, the respondents had to state the numbers of subject metadata creators. All 29 (100%) respondents answered this question, indicating the numbers in their libraries. The lowest number of metadata creators in an individual library was only one individual. Five (17.24%) respondents indicated that there was only one ETD subject metadata creator in their libraries. The next lowest number indicated by three (10.34%) respondents was of two subject metadata creators. The responses showed that prevalent cases are four (4) or more ETD subject metadata creators in a library.

Table 6.2: Number of ETD subject metadata creators

Number of respondents	No. of ETD subject metadata creators in the participating libraries
5	1
3	2
3	3
6	4
5	5
1	6
2	8
1	11
1	12
1	13
1	15

The results also revealed that there was prevalence of the use of a pool method. When added together, (62.07%) responses indicated four (4) and more staff members responsible for ETD metadata creation. The different staffing practices for information resources description are a common situation in the different university libraries, at times resulting from the size of the library.

6.2.1.3 Description of positions at the different libraries

The following survey question was asked to the respondents:

How would you describe your position at the institution?

The researcher aimed at establishing the participants' job position and their levels, whether professional or nonprofessional, in order to understand the job status of the respondents. All the participants responded by indicating their job positions as summarised in Table 6.3.

Table 6.3: Description of positions at the different libraries

Position	Number of respondents	Percentage (%)
Professional Librarian	28	96.55
Cataloguing and Metadata Services Librarian	1	3.45
Total	29	100

The majority of respondents, 28 (95.55%), referred to their positions as Professional Librarians. It was not clear if there were different job titles used for the Professional Librarian positions. Only one (1) respondent (3.45%) chose the category of "other" and indicated a different position, with the title of Cataloguing and Metadata Services Librarian.

6.2.1.4 Experience in ETD metadata creation

The following survey question is involved in this section:

How long have you been involved with metadata creation, specifically for the ETDs?

It is generally understood that the general experience in metadata creation shapes the understanding of the various activities involved in the process. The findings illustrated shown in Figure 6.1 (p. 177).
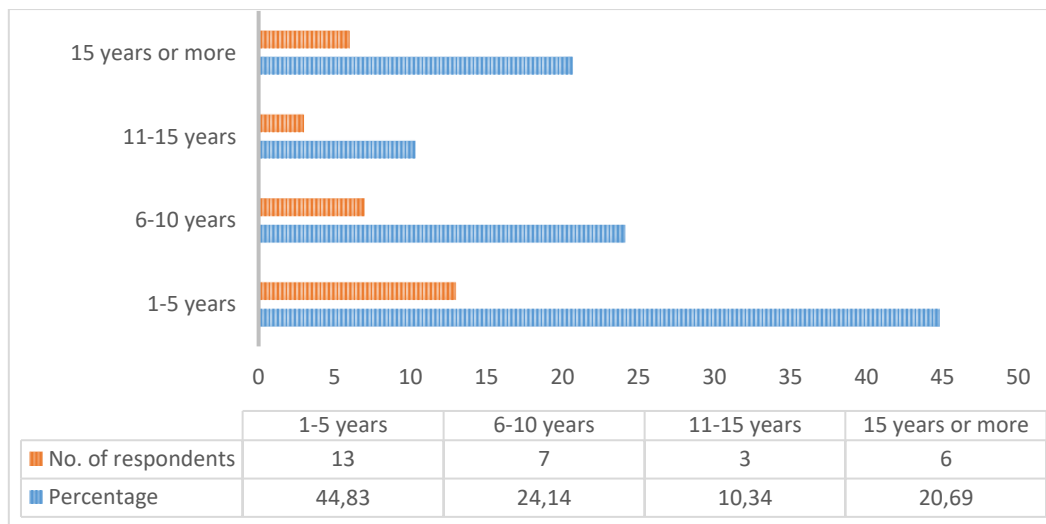
Figure 6.1: Period involved in ETD metadata creation

A moderate number of 13 participating metadata creators (44.83%) had been involved in ETD metadata creation for a period of 1–5 years. This was followed by seven respondents (24.14%), who were in the 6–10 years category. The 11–15 years category showed only three (3) (10.34%), respondents.

Six (6) respondents (20.69%) were well experienced, with more than 15 years of involvement in ETD metadata creation. However, from the collected data, it could not be confirmed with certainty if the 13 (44.83%) respondents, who had been involved with ETD metadata creation for less than six years (1–5 years) were newly qualified professionals or experienced professionals who assumed new responsibilities. Significant differences were observed among the respondents' number of years involved in ETD metadata creation.

6.2.1.5   Experience in subject description

In order to establish the extent of the respondents' experience in subject description, the following survey question was asked:

Altogether, how many years of experience do you have in the subject description of information resources (i.e. assigning subject headings or subject metadata)?

This question was important, since ETD subject metadata creation is commonly performed by staff who have historically been involved in metadata creation of other types of information resources in the different libraries.

The responses captured in Table 6.4 showed the overall experience of the respondents in subject metadata creation.

Table 6.4: Experience in subject description

Number of years	Number of responses	Percentage (%)
1–5 years	5	17.24
6–10 years	8	27.59
11–15 years	6	20.69
15 years or more	10	34.48
Total	29	100

The data in Table 6.4 indicates that five (5) respondents (17.24%) had been involved in subject description for a period between one and five years; eight respondents (8) (27.59%) between six and ten years; six (6) respondents (20.69%) between eleven and fifteen years; and 10 respondents (34.48%) for more than 15 years. Overall, 16 respondent (55.17%) fell in the category of more than 11 years, compared to the 13 (44.83%) who had been involved in subject description 10 years and more. A significant number of the respondents had extensive experience in general subject headings or subject metadata creation.

6.2.1.6   Metadata creation role

In order to gain an understanding of the different roles of the respondents in metadata creation, the following survey question on metadata creation roles was posed to respondents:

Indicate the option that best describes your role in metadata creation.

The responses describing the different roles are outlined in Table 6.5 (p. 179).

Table 6.5: Respondents' roles in metadata creation

Role types	Total numbers	Percentage (%)
Metadata Creator	22	75.86
Metadata Editor	4	13.79
Metadata Quality Controller	1	3.45
Other	2	6.90
Total	29	100

As reflected by the responses captured in Table 6.5, 22 respondents (75.86%), concentrated on the general metadata creation role. Out of the 29 metadata creators, four (4) respondents (13.79%) indicated that they were involved in metadata editing. Only one (3.45%) respondent amongst the different categories could specifically be identified as a metadata quality controller. There were two (2) responses (6.90%) responses in the "other" category: one (1) respondent performed metadata creation, editing and quality control and the other respondent in this category indicated the role of "manager", with no further specifics as to what the role entailed.

6.2.2   Metadata procedures

The researcher asked the questions on metadata procedures, with the aim of determining the extent and nature of geographic subject metadata use and subject analysis approaches and to establish how the theoretical background informed the practices.

6.2.2.1   Responsibility for assigning subject metadata for ETDs

The respondents were asked to indicate who contributed or assigned subject metadata for ETDs in their libraries. The question sought to establish who were mainly fulfilled the responsibility for ETD subject metadata creation.

Nineteen (19) respondents (65.52%) indicated that library metadata creators were responsible for assigning ETD subject metadata. According to the results in Table 6.6 (p. 180), other categories, like authors, seemed to participate less in this role.

Table 6.6: Responsibility for assigning subject metadata for ETDs

Responses	Frequency	Percentage %
Theses and dissertations authors	8	27.59
Library metadata creators	19	65.52
Contributed by authors and updated by library staff	1	3.45
Cataloguing and Metadata Services Librarians	1	3.45
Total	29	100

The second highest rated response to the question on who contribute or assign subject metadata for ETDs was that of authors at eight (8) respondents (27.59 %).

Because it is considered as part of metadata creation library staff, the category of Cataloguing and Metadata Services Librarians was included in the list of choices to accommodate all possible workflows and staffing practices. Only one respondent selected this category. The low rating for the option on "contributed by authors and updated by library staff" may suggest that librarians performed less updating to author-supplied metadata.

6.2.2.2   Terms assigned to represent ETD subjects

Table 6.7 shows the responses to the question:

Indicate the terms that are used at your institution to represent the ETD subjects.

Table 6.7: Terms used to represent ETD subjects

Terms used	Frequency of responses	Percentage %
Keywords from text	13	44.83
Terms from natural language	1	3.54
Controlled subject headings	15	51.72
Total	29	100

An option was given to select more than one response to this question. The results indicated varying ways through which metadata creators described ETDs subjects and used different terms to represent the ETD subject contents in the different libraries.

Controlled subject headings was selected by 15 (51.72%) respondents as being the most common way used to assign ETD subjects; 13 (44.83%) respondents used keywords from the text and only one respondent (3.45%) the use of terms from natural language to assign ETD subject metadata. Although the option was given to indicate any other types of terms used, none were identified.

6.2.2.3 Suitable subject metadata for the discovery of ETD content

Varied responses were given to the question:

How often do you find the listed categories of Metadata to be suitable for the discovery of the subject content of ETDs?

Table 6.8 summarises the responses.

Table 6.8: Suitable categories of metadata for the discovery of the ETDs subject content

Items Nos.	Statements	Always suitable	Sometimes suitable	Not sure	Rarely suitable	Never suitable	Mean	Std. dev.
Q18.1	Keywords contributed by theses and dissertation authors	11 (37.98%)	14 (48.28%)	0%	4 (13.79%)	0%	1.90	.976
Q18.2	Subject metadata contributed by library staff	21 (72.41%)	6 (20.69%)	1 (3.45%)	1 (3.45%)	1 (3.45%)	1.45	.948
Q18.3	Subject metadata created through automated library processes	2 (6.90%)	7 (24.14%)	15 (51.72%)	5 (17.24%)	0%	2.79	.819
Q18.4	Subject metadata created in the form of user contributed tags	2 (6.90%)	5(17.24%)	14 (48.28%)	4 (13.79%)	4 (13.79%)	3.10	1.081
Q18.5	Keywords in the full-text of information resources	14 (48.28%)	10 (34.48%)	4 (13.79%)	1 (3.45%)	0%	1.72	.841

Figure 6.2 presents a graphical representation of the percentages of suitable categories of metadata, as displayed in Table 6.8 (p. 181).

	Q18.1	Q18.2	Q18.3	Q18.4	Q18.5
■ Always suitable	37,93%	72,41%	6,90%	6,90%	48,28%
■ Sometime suitable	48,28%	20,69%	24,14%	17,24%	34,48%
■ Not sure	0,00%	0,00%	51,72%	48,28%	13,79%
■ Rarely suitable	13,79%	3,45%	17,24%	13,79%	3,45%
■ Never suitable	0,00%	3,45%	0,00%	13,79%	0,00%
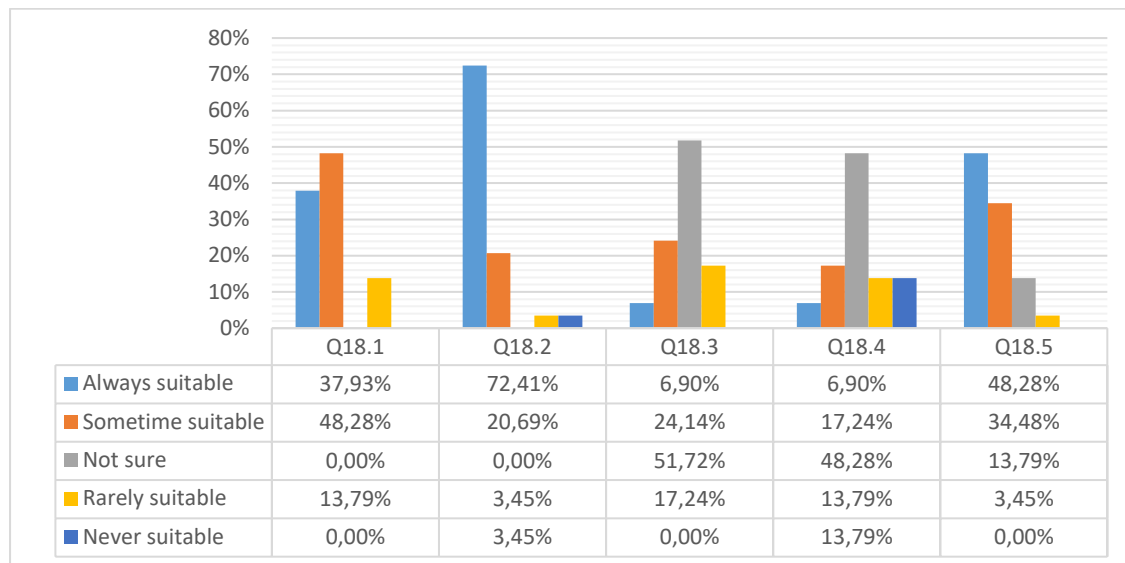
Figure 6.2: Suitable metadata categories

Twelve (12) respondents (72.41%) identified subject metadata contributed by library staff as "Always suitable"; and six (6) respondents (20.69%) identified it as "Sometimes suitable". Furthermore, Table 6.8 (p. 181) shows the mean and standard deviation for this approach as ($\bar{x}$ = 1.45, s = 0.948). The standard deviation for the mean was low, compared to that of the other items, which suggested that there were more definitive, positive answers on the adequacy this approach.

The second highest rating for "Always suitable" was for the "Keywords in the full-text of information resources", followed by the "Keywords contributed by theses and dissertation authors". However, when the categories of "Always suitable" 11 (37.93%) and "Sometimes suitable" 14 (48.28%) were added, the "Keywords contributed by theses and dissertation authors" was rated suitable by 25 (86.21%) respondents. The deviation from the mean ($\bar{x}$ = 1.72, s = 0.841) was the second lowest, suggesting less disagreement in the responses on the preference for this approach.

The suitability rankings for the keywords contributed by theses and dissertation authors for the discovery of the subject content of ETDs indicated positive perspectives.

However, it was noted that a small number of four (4) respondents (13.79%) considered such metadata as "Rarely suitable". No uncertainty was expressed about the suitability of the authors keywords and the keywords assigned by the librarians, but views were clearly expressed as either suitable or not.

There was smaller range of answers ($\bar{x}$ = 2.79, s = 0.819) for "Subject metadata created through automated library processes", but a noticeable uncertainty, with 15 (51,72%) expressing uncertainty about the suitability of the approach.

There is noticeable uncertainty about the suitable nature of subject metadata in the form of user contributed tags. As indicated, 14 (48.28%) respondents showed that they were not sure. The most negative ratings were observed for the user-contributed tags, rated "Rarely suitable" by four (4) respondents (13.79%) "Never suitable" by four (4) (13.79%). Five (5) respondents (17.24%) rated the metadata created through automated library processes as "Rarely suitable", while 15 (51.72%) respondents were uncertain about the suitability of this approach. Observably, the highest standard deviation from the mean ($\bar{x}$ = 3.10, s = 1.081) suggested a wide range of responses to the question and more uncertainty.

Keywords contributed by theses and dissertation authors (Q18.1) were overall considered as suitable and were rated "Always suitable" by 11 (37.98%) and "Sometimes suitable" by 14 (48.28%) respondents, followed by Keywords in the full-text of information resources (Q18.5) at 14 (48.28%) "Always suitable" and 10 (34.48%) "Sometimes suitable". The two types of approaches were observed as being more preferred, as respondents indicated more ratings in the affirmative in terms of suitability rankings.

6.2.2.4   Activities for the creation of the ETDs geographic subject metadata

The metadata creators responded in different ways to the question:

How often do you engage in the following activities during the creation of the geographic subject metadata for ETDs?

Table 6.9 indicates how often the various types of activities were performed by the respondents.

Table 6.9: Activities for the creation of the geographic subject metadata for ETDs

No.	Activities	Always	Often	Sometimes	Rarely	Never	Mean	Std. dev.
Q 19.1	Creating completely new geographic subject metadata	3 (10.34%)	13 (44.83%)	5 (17.24%)	5 (17.24%)	3 (10.34%)	2.72	1.192
Q 19.2	Adding geographic terms as subdivisions to the existing subject headings	7 (24.14%)	13 (44.83%)	1 (3.45%)	6 (20.69%)	2 (6.90%)	2.41	1.268
Q 19.3	Using the keywords derived from the information resource to create additional geographic subject headings	6 (20.69%)	12 (41.4%)	5 (17.24%)	4 (13.79%)	2 (6.90%)	2.45	1.183
Q 19.4	Creating new geographic subject metadata or subdivisions using geographic terms not derived from the ETD itself	3 (10.34%)	6 (20.69%)	5 (17.24%)	9 (31.03%)	6 (20.69%)	3.31	1.312

From the presented data, it was evident that geographic terms were, in most cases, always added as subdivisions to the existing subject headings. This occurred always at 24.14% and 13 (44.83%) often, compared to always using the keywords derived from the information resource to create additional geographic subject headings selected by 20.69% and always creating completely new geographic subject metadata indicating 10.34% rating.

More dispersed ($x̄$ = 2.41, s = 1.268) responses were observed regarding how often the activity of adding geographic terms as subdivisions to the existing subject headings was performed. It was observed that different ways were followed to create geographic subject metadata. Many of the metadata creators chose more than one of the given options. Consistent with the results in Q81.1 to Q18.9, it was also noted that using keywords from the texts was indicated as a commonly used method, with la smaller range, ($x̄$ = 2.45, s = 1.183) of responses.

The other two dominant activities were creating completely new geographic subject metadata (Q19.1), or adding geographic terms as subdivisions to the existing subject headings (Q91.2). Although the two activities were indicated as not necessarily always carried out, as rated "Always" by three (3) respondents (10.34%) and seven (7) respondents (24.14%) respectively, both activities were commonly rated by 13 (44.83%) respondents as "Often" used. This was a general indication of more performance of these activities.

It was indicated 13 (44.83) respondents – that creating new geographic subject metadata or subdivisions by using geographic terms not derived from the ETD itself was often practiced. However, it appeared to be the least practiced, compared to all other forms of geographic subject metadata creation methods.

6.2.2.5   Maximum number of subject headings

The responses reflected in Table 6.10 show varied practices performed by different libraries regarding the maximum number of subject headings that were allowed to be assigned to individual ETD metadata records.

Table 6.10: Maximum number of subject headings

		Frequency	**Percentage**
1–5		11	37.93%
5 or more		3	10.34%
No limit set		15	51.72%

A notable number 15 (51.72%) of the respondents indicated that there was no limit set by their libraries to the number of subject representations that could be assigned. These responses indicating "no limit set" can relate to the findings in Section 6.2.2.1, which showed that policies for subject analysis were non-existent in a considerable number of libraries and uncertainty was observed in some cases.

The second highest number of respondents, 11 (37.93%), indicated that they were limited to assign 1–5 subject headings. The least group of three (3) respondents (10.34%) respondents were allowed to assign five (5) subject headings or more.

6.2.2.6 Finding relevant geographic terms to be used as subject metadata

Metadata creators find relevant geographic terms to be used as subject metadata from different parts of the information resource. This is shown in the responses to the question:

Where do you usually find the relevant geographic terms to be used as subject metadata?

Figure 6.3 presents the frequency and the percentages of responses.

	Title of a work	Abstract	Contents of a work
Frequency	24	3	2
Percentage	82,76	10,34	3,45
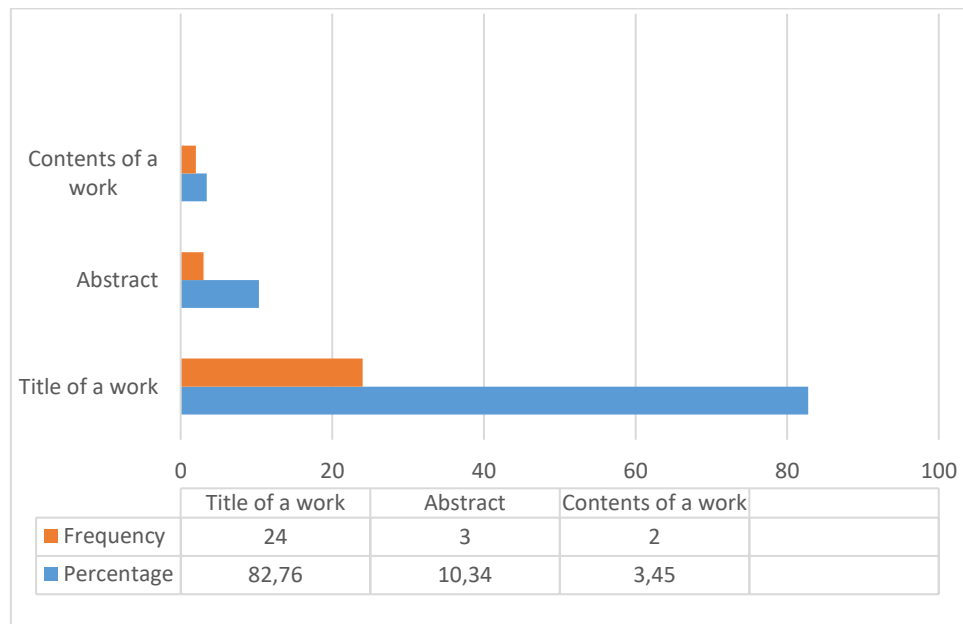
Figure 6.3: Finding relevant geographic terms

Figure 6.3 shows that 24 (82.76 %) respondents usually found the relevant geographic terms to be used as subject metadata in the title. The second highest rating was three (3) respondents (10.34%), who found subject terms from the abstract. The lowest rating was two (2) respondents (6.90%), who indicated that they found terms in the content of the work. As titles do not always reflect the subject content accurately, it could be assumed from the results that, when that was the case, the metadata creators also consulted the abstract and, in a very few cases, they checked the ETD content.

6.2.3   Subject analysis basis

This section presents the findings about the effect of current approaches on ETD subject metadata creation; particularly how the metadata creators comprehended the impact of the basis used for subject analysis. The information assisted in determining the impact of the approaches on ETD geographic subject metadata creation.

6.2.3.1   Existence of policies

A question was asked as to whether the participating libraries had any policies or guidelines that specifically instructed how to conduct subject analysis during metadata creation. The question sought to establish whether the different libraries in which the respondents were employed had any specific policies or guidelines that instruct how to conduct subject analysis for subject metadata creation.

The results are captured in Table 6.11.

Table 6.11: Existence of subject analysis policies

Response	Frequency	Percentage (%)
Yes	17	58.62
No	8	27.59
Not sure	4	13.79
Total	29	100

The results showed that most respondents,17 (58.62%), replied "yes" to the question. Another eight (8) respondents (27.59%) pointed to the absence of policies to guide subject analysis practices in their libraries. Four (4) respondents (13.79%) answered that they were not sure about the existence of policies to guide their subject analysis practices.

6.2.3.2 Policy enforcement

The question asked to the respondents was: *Indicate the extent to which you feel subject analysis policy should be enforced.*

The results are displayed in Figure 6.4.

	Not enforced	Enforced	Sometimes enforced	Always enforced
■ Frequency	3	2	1	23
■ Percentage	10,34	6,9	3,45	79,31
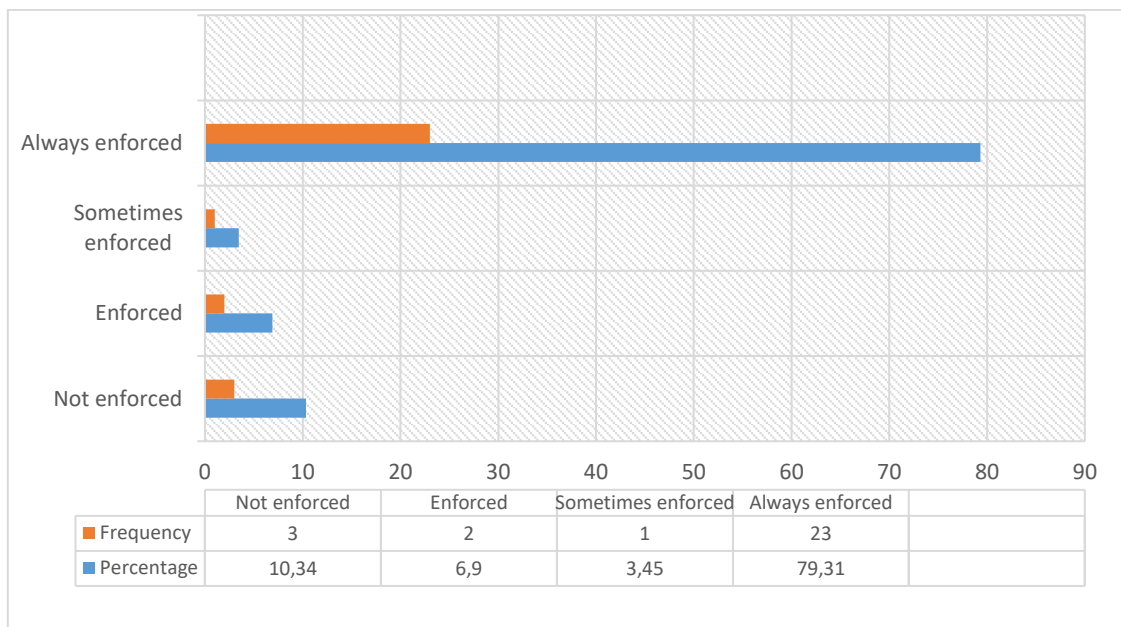
Figure 6.4: Policy enforcement

As indicated by Figure 6.4, there was a high number of affirmative responses – 23 (79.31%) – regarding the need for policy to guide ETD subject analysis. Although the majority (represented in the "always enforced", "sometimes enforced" and "enforced" categories) of the respondents agreed that policy should somewhat be enforced, an insignificant number of three (3) respondents (10.35%) did not agree that policies on subject analysis should be enforced.

6.2.3.3 Important aspects when determining what a resource is about

The respondents were asked to indicate the level of importance that they attached to several listed aspects when determining what a particular information resources was about. They were requested to identify the statements that best reflected their approaches and to rate them in terms of the perceived importance. The results are captured in Table 6.12.

Table 6.12: Importance of specific traits when conducting subject analysis

Item No.	Statements	Very Important	Fairly Important	Important	Slightly Important	Not Important
Q9.1	Only identifying the keywords within an information resource	13 (44.83%)	6 (20.69%)	8 (27.59%)	1 (3.45%)	1 (3.45%)
Q9.2	Analysis of the entire content (All key topics covered within an information resource)	16 (55.17%)	10 (34.48%)	3 (10.34%)	0%	0%
Q9.3	Determining the author's intention for producing the work	14 (48.28%)	7 (24.14%)	3 (10.34%)	3 (10.34%)	2 (6.90%)
Q9.4	Focusing on all the potential needs to be satisfied by the subject content of the ETD	20 (68.97%)	3 (10.34%)	5 (17.24%)	1 (3.45%)	0%
Q9.5	Only focusing on satisfying the information needs of a specific subject area	8 (27.59%)	5 (17.24%)	12 (41.38%)	4 (13.79%)	0%
Q9.6	Satisfying user needs across different disciplines	19 (65.52%)	6 (20.69%)	1 (3.45%)	3 (10.34%)	0%

Figure 6.5 displays the responses to Questions 9.1–9.6, indicating the percentages.

	Q9.1	Q9.2	Q9.3	Q9.4	Q9.5	Q9.6
■ Very Important	44,83%	55,17%	48,28%	68,97%	27,59%	65,52%
■ Fairly Important	20,69%	34,48%	24,14%	10,34%	17,24%	20,69%
■ Important	27,59%	10,34%	10,34%	17,24%	41,38%	3,45%
■ Slightly Important	3,45%	0,00%	10,34%	3,45%	13,79%	10,34%
■ Not Important	3,45%	0,00%	6,90%	0,00%	0,00%	0,00%
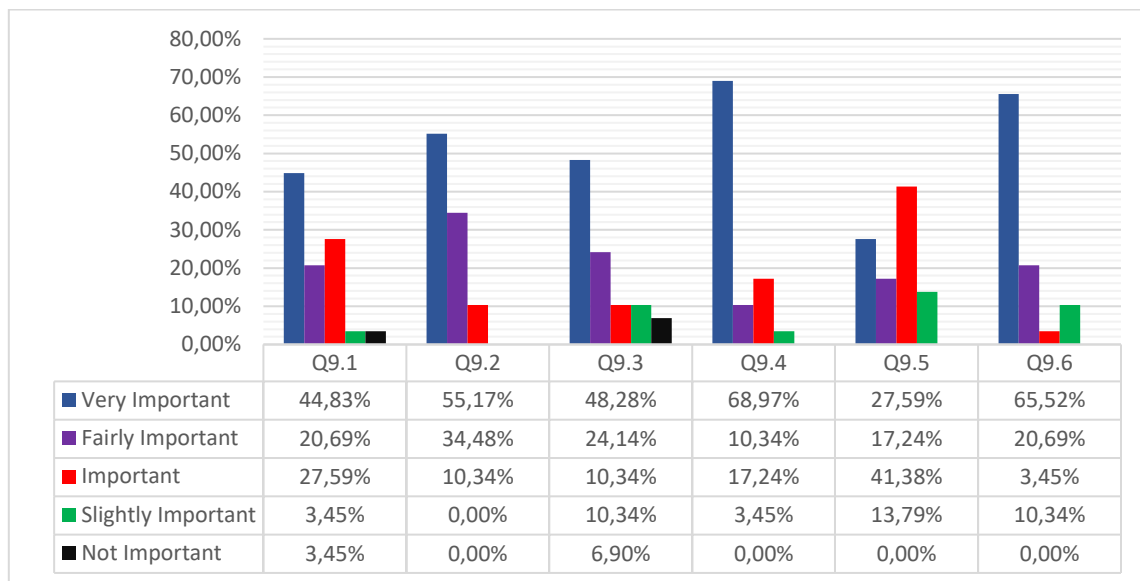
Figure 6.5: Specific traits when conducting subject analysis

The overall positive ratings (i.e. Very important; Fairly important; and Important) were the highest for the trait of analysis of the entire content (All key topics covered within an information resource), Q9.2. A total of 16 (55.17%) respondents rated the approach as "Very important". The "Important" rating was selected by ten (34.48%) respondents.

However, the highest ratings in the affirmative "very important" were given to traits in Questions 9.4 and 9.6, as reflected in Table 6.12 (p. 189). The highest rating of "Very important", at 68.97%, was given to the need to focus on all the potential needs to be satisfied by the subject content (Q9.4). The other three categories of responses to this same trait indicated moderate ratings for "Fairly important": three (3) (10.35%); and "Important": five (5) (17.24%). Only one (3.45%) respondent gave a rating of "Slightly important" for this aspect. It was also evident that none (0%) of the respondents completely disagreed with the importance of this approach.

The second mostly rated in the affirmative "Very important" was the aspect of satisfying user needs across different disciplines (Q9.6). Nineteen (19) respondents (65.52%) gave this rating. There were no respondents (0%), who rated this aspect as "Not important".

The indication of the high regard held for analysing the entire content of an information resource was consistent with the views expressed in the preceding question (Q9.1), where the respondents gave low ratings for the use of "keywords only".

Furthermore, the results of this survey indicated an inclination towards focusing on a broader approach to subject analysis in terms of focus across different subject disciplines and on a wide range of users.

The authors' intentions were rated as "Very important" by 14 (48.28%) respondents, compared to importance attached to keywords, where 13 (44.83%) respondents selected the "Very important" rating. The responses to the importance to focus on "All the potential needs" (Q9.4) were compared to the ratings of the importance of "Satisfying the needs of a specific subject area" (Q9.5) in terms of positive and responses. The comparison showed fewer negative ratings by one (1) (3.4) for the former trait, compared to four (4) (13.79) negative ratings for the latter.

6.2.3.4 Familiarity with the theoretical principles for subject analysis

This section presents the findings from the data collected in response to the question:

How familiar are you with the theoretical principles that provide the basis for subject analysis?

Table 6.13 indicates the varied responses to this question.

Table 6.13: Familiarity with the theoretical principles for subject analysis

Responses	Frequencies	Percentages %
Very familiar	16	55.17
Fairly familiar	3	10.34
Familiar	7	24.14
Slightly familiar	3	10.34
Total	29	100

Sixteen (16) respondents (55.17%) answered that they were familiar with subject analysis theoretical principles. However, different levels of familiarity were observed among the respondents.

The results indicated that three (3) respondents (10.34%) were "fairly familiar"; seven (7) (24.14%) were "familiar"; and three (3) (10.3%) rated themselves as being "slightly familiar" with subject analysis theoretical principles. A small number of three (3) (10.34%) respondents responded that they were unfamiliar with subject analysis theoretical principles.

6.2.3.5   Library and Information Science (LIS) training for subject analysis

This section presents the analysis of the responses gathered for the question: *In your Library and Information Science studies, did you receive any training in the theoretical principles for subject analysis?*

Figure 6.6 illustrates the results of Question 11.
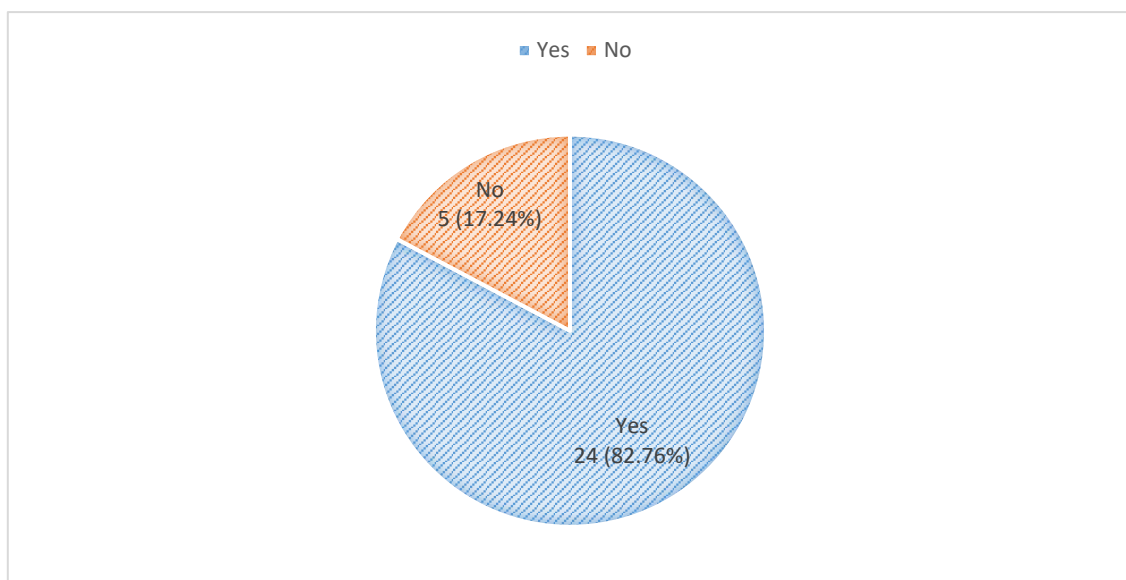
Figure 6.6: Library and Information Science training

A significant number of 24 (82.76%) respondents indicated that they received training on these principles during their Library and Information Science training. It is noticeable that this was in line with a higher number – 16 (52.17%) – of respondents who indicated that they were familiar with the theoretical principles of subject analysis, as shown in responses in Section 6.2.2.4. Five (5) respondents (17.24%) indicated that they did not receive training on subject analysis foundational principles in their LIS studies.

6.2.3.6 Importance of knowledge from LIS Schools training

As a follow-up, the respondents were asked to respond to the question, *If you answered "Yes" to Question 11, how do you rate the importance of the theoretical knowledge that you received?*

The results are displayed in Figure 6.7.

	Very important	Fairly important	Important	Slightly important	Not important
■ Frequency	21	3	3	1	1
■ Percentage	72,41	10,34	10,34	3,45	3,45
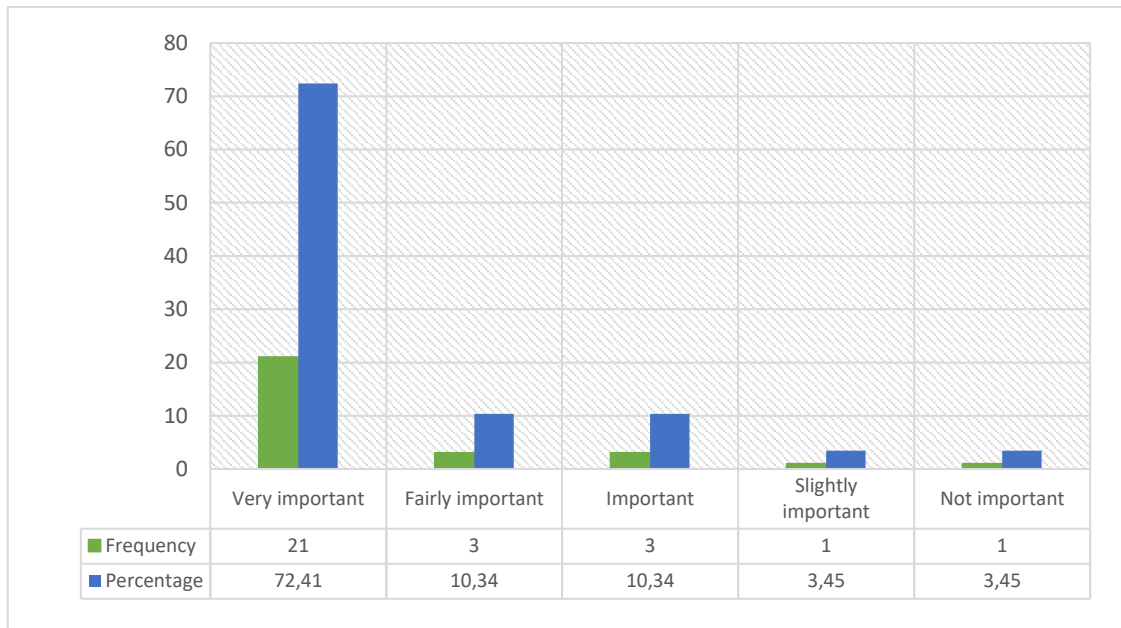
Figure 6.7: Importance of the theoretical knowledge on subject analysis

Most respondents, 21 (72.41%), answered positively, in that they rated the theoretical knowledge of subject analysis as "Very important". The three (3) respondents (10.34%), who rated this aspect at "Fairly important", where equal in number to those who chose the rating of "Important". Only one respondent selected the "Slightly important" ranking. One respondent regarded the theoretical knowledge of subject analysis as "Not important".

6.2.3.7 On-the-job training on the theoretical principles of subject analysis

The respondents were asked to indicate if they received training on the theoretical principles of subject analysis in the form of on-the-job training. Figure 6.8 (p. 194) reflects that 16 (55.17%) respondents did receive training on the theoretical principles of subject analysis in the form of on-the-job intervention.
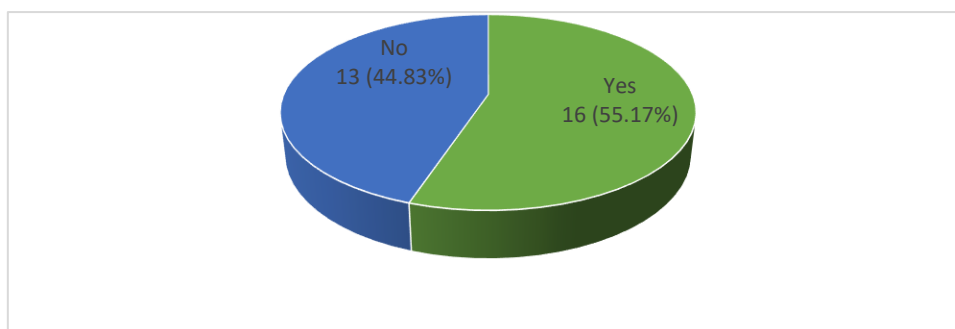
Figure 6.8: On the job training on the theoretical principles of subject analysis

On-the-job training is normally offered to orientate and train the metadata creators and to make them aware of the policies, rules and procedures of the specific libraries. Thirteen (13) respondents (44.83 %) indicated that they did not receive any such training through on-the-job interventions.

6.2.3.8 Importance of on-the-job training on promoting the theory knowledge for subject analysis

As a follow-up to the responses given in Section 6.2.2.7, the respondents were asked to rank the importance of on-the-job training as a way of acquiring the theoretical knowledge for subject analysis, if they had responded positively to the question whether they received on-the-job training. Table 6.14 presents the rankings.

Table 6.14: Importance of on-the-job training on the theoretical knowledge of subject analysis

Responses	Frequency	Percentage
Very important	12	41.38
Fairly important	3	10.34
Important	1	3.45
Slightly important	1	3.45
Not important	1	3.45
Not indicated	11	37.93
Total	29	100

Less than half of the respondents – 12 (41.38%) –indicated that on-the-job training for subject analysis theoretical knowledge was very important. This was followed by three (3) (10.34%) respondents, who considered on-the-job training on the theoretical

knowledge of subject analysis as fairly important and one (1) respondent (3.45%) regarded it as important. A small number of low rankings were also assigned to this aspect, with one (1) respondent (3.45%) rating it as slightly important and one respondent (3.45%) ranking it as not important.

6.2.3.9 Other ways of learning the subject analysis approaches

Table 6.15 presents the analysis of the responses to the question asked to the respondents, if there was any other ways through which they learned the different approaches for subject analysis – i.e. other than in the ways stated in the previous questions (Section 6.2.2.5 (formal LIS training) and Section 6.2.2.7 (on-the-job training).

Different responses were given to this question, with an indication of respondents pursuing other ways of learning and acquiring knowledge of the theoretical basis of subject analysis. The most common learning methods stated by the respondents included personal research, experience and workshops. These appeared to be the common ways of metadata creators acquiring knowledge about subject analysis. These results are captured in Table 6. 15.

Table 6.15: Other ways use to learn about the different approaches for subject analysis

Responses	Frequency
Not stated	15
By familiarising myself with the tool (Library of Congress Subject Headings)	1
Determining subject heading, begins with understanding the main classes on a broader scale, e.g. decide if you working on philosophy, religion or social science, then scale down to the specific content by checking key areas on the thesis like the abstract. Often thesis have keys words, and this work as a good starting point to further create subject headings.	1
Experience	1
It is not always clear to analyse a subject from the title and abstract. Always search the authors, fields of specialization for further understanding of what related topics he/she mostly focuses on.	1
LIASA IGBIS workshops to gain practical knowledge	1
Looking at similar records in our database and on OCLC	1
NO other way	1

Responses	Frequency
On the job self-taught because of the various groups you have to take into account that will work with the data, i.e. cataloguers, users/recipients and software programmes	1
Online documents: Introduction to LCSH & LC Subject Headings Manual (SHM)	1
Reading the abstract	1
Relationship between the class number and subject topics/analysis	1
Through being the member of Interest Group for Bibliographic Standards	1
Through consortium meetings	1
Through own research.	1

These responses suggested that, other than the formal education route and structured on-the-job training, there were other ways that metadata creators pursued to acquire knowledge on how to perform subject analysis for ETDs.

6.2.3.10   Other contributory factors to effective subject analysis

Varied responses were given to the question:

Which other factors are significantly contributing to effective subject analysis for the creation of geographic subject metadata for ETDs?

The results are shown in Table 6.16.

Table 6.16: Additional factors in subject analysis

Additional subject analysis factors
Abstract area
AUTHORITY CONTROL
Awareness of the name changes taking place around South Africa and beyond
Experience
General knowledge. e.g.: metadata creator must know there is more than one Middelburg in South Africa
Name changes of places and institutions
No sure
None
Subject knowledge, subject analysis training
The title as well as the abstract are usually the starting point to indicate what the work entails.
The user's in finding the relevant information source
three authors choice of subject keywords
To keep the standards, uniformity and consistency of subject analysis.

Additional subject analysis factors
To provide quality bibliographic record
Understanding the authors intention of producing the work and reading further than just a title, introduction, abstract etc.
USMARC
You have covered the factors

Important information was gathered from the additional information given by the respondents, as reflected in Table 6.16 (p. 196), where they stated various factors that they considered as contributing significantly to effective subject analysis for the creation of geographic subject metadata for ETDs.

The abstract was considered important during subject analysis, while the author's intention and keywords were also stated as important aspects to consider during subject analysis for ETDs. The identified aspects in this section related to the responses in Section 6.2.2.3. Together they showed dominant aspects identified by the respondents concerning ETD geographic subject metadata creation to be, among others, the use of standards, considering the users' needs and the authors' intentions.

Furthermore, authority control was also considered as important. Authority control related aspects involves the issue of understanding the area in which the place name belongs, as one respondent clearly stated, there is a possibility of similar place names being used in different areas. The name changes mentioned by one respondent also aligned with authority control.

The identification of training and experience as important factors related to the expressed efforts to acquire knowledge and skills for subject analysis, shown in Section 6.2.2.9.

6.2.3.11 Other subject analysis and geographic subject metadata challenges

The respondents were given an opportunity to state any other challenges that they commonly experience with subject analysis and the creation of geographic subject metadata for ETDs. The respondents stated 13 different challenges, as shown in Table 6.17 (p. 198). Metadata creators experienced various challenges that may differ from library to library, as observed in these findings.

Table 6.17: Other challenges

Challenges
Authors use keywords, and librarian guided by those keyword then create subject heading using LCSH.
Challenge is the delay of including new geographic places in the authorized the subject headings thesaurus.
Controlled lists such as the LCSH do not have the best terms for South African content. Uncontrolled vocabulary might have solved the problem, but scatter the resources that can be grouped using controlled vocabulary.
Exclusion of other geographic name in the geographic subject metadata
Frequency of the assignment of author-supplied keyword strings that are not relevant in the ETDs are misleading the newly qualified Cataloguing and Metadata Librarians
Lack of knowledge in a specific subject area
LCSH does not really cater for African countries
Metadata librarians are not subject specialists and difficulty arises with disciplines like engineering, medicine, health sciences, etc.
None
Not finding the exact subject on the LCSH
Not so clear subject content and the authors unclear objectives
Specificity in representing geographic areas
There needs to be uniformity about the way the metadata is entered.
Training not good adequate

No predefined categories were set for this category of questions in the questionnaire. However, it was observed that four of the challenges stated by the respondents related to the inadequacies of controlled vocabularies to guide geographic subject description. The need for timely updating of standards was indicated in the statement involving a delay in including new geographic places in the authorised subject headings thesaurus. Comprehensiveness in terms of coverage of the standards was shown as important by the statement that controlled lists, such as the LCSH, did not have the best terms for South African content. Additionally, this was shown by the comments about the exclusion of other geographic names in the geographic subject metadata and the view that the LCSH did not sufficiently cater for African countries.

Three responses mentioned that the author supplied keywords issues, including irrelevance and unclear subject content representation. One respondent addressed process issues, stating that the authors only supplied keywords, whereas the professional metadata creators assigned subjects from a controlled vocabulary. In a related response about author-supplied keywords, one respondent expressed the view that the author-supplied keywords were misleading the newly qualified librarians.

The problems of sometimes experiencing unsuitable author supplied keywords was also mentioned. The other problem that could be read in the responses addressed training, as one response mentioned "training not good or adequate".

The challenges faced by ETD metadata creators regarding comprehension of the specialised subject areas were raised in two responses. The lack of uniformity in metadata creation was pointed out. Adequate training was specifically pointed out as necessary for subject analysis and the creation of geographic subject metadata for ETDs.

Based on the research results, the conclusion on the data analysis and the findings of the questionnaire survey phase of this study is that it was observed that the university libraries in South Africa participated in creating metadata; particularly geographic subject metadata for ETDs. The number of metadata creators in the different libraries varied and the metadata creation roles played are different. In addition, the status on the use of policies and their non-existence in some cases was notable. The analysis and findings also covered issues of familiarity with the theoretical principles for subject analysis and views on training. Although different responses were recorded for how to conduct subject analysis and perspectives on the importance of geographic subject metadata for ETDs, the response were mostly positive. Furthermore, different views were held about the focus that should be maintained when creating subject metadata for ETDs. Further probing through the interviews was considered necessary to acquire more information on metadata creators' perspectives on ETD geographic subject metadata and the theoretical principles guiding their creation.

6.2.4   Current approaches for ETDs and geographic subject metadata

Geographic subject representations are an additional form of describing the subject content of information resources. The participants expressed different views on how current metadata creation approaches had influenced the creation of geographic subject metadata. The current practices, including varying practices regarding the application of policies, are presented in the next sections. A theoretical understanding of how to perform analysis for the determination of geographic subject metadata was fundamental to the problem of this study.

6.2.4.1 Policies and specific guidance on geographic subject metadata

A question was asked to the respondents as to whether the policies at their institution gave specific guidance on geographic subject metadata creation. The results are captured in Figure 6.9.

As illustrated in Figure 6.9, 12 (58.62%) responded "Yes" to this question, while 17 (41.48%) responded "No". When compared to those in Section 6.2.2.1, the responses showed that, while policies for subject analysis generally did exist in most libraries (58.62%), they provided less guidance (41.48%) in terms of specific focus on geographic subject metadata creation.
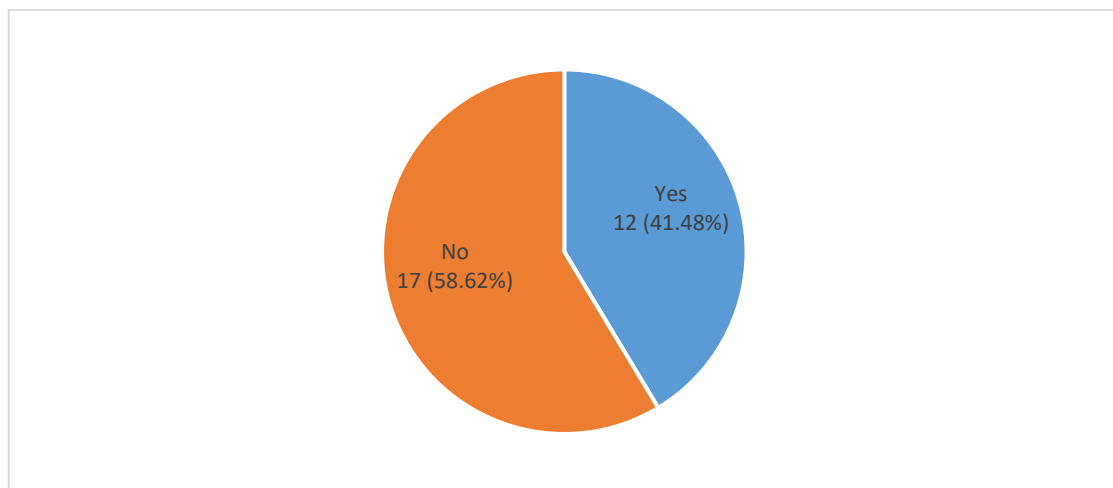
Figure 6.9: Policies on geographic subject metadata

6.2.4.2 Standards used to control assigned geographic subject metadata for ETDs

Subject metadata can be assigned by using controlled headings or other approaches like natural language terms. The respondents were asked to indicate the standards that they used to control geographic subject metadata for ETDs.

The standards and the choices are reflected in Table 6.18 (p. 201).

Table 6.18: Standards used to control assigned geographic subject metadata for ETDs

Response choices	Frequency	Percentage
Library of Congress Subject Headings (LCSH)	24	82.76
Library of Congress Subject Headings (LCSH) and NLM Medical Subject Headings (MeSH)	2	6.90
Library of Congress Subject Headings (LCSH) and FAST headings	1	3.45
Library of Congress Subject Headings (LCSH) and Geographic names authority lists	2	6.90
Total	29	100

The responses presented in Table 6.18 indicate that the majority of the respondents – 24 (82.76%) – used the Library of Congress Subject Headings (LCSH) as a standard to control or assign geographic subject metadata for ETDs. Two (2) (6.90%) respondents used both LCSH and the Medical Subject Headings (MSH). This could be libraries housing medical collections or medical libraries, as is commonly the case. Geographic name authority lists were used in addition to LCSH by tow (2) (6.90%) respondents. LCSH and FAST headings were e used together by only one (3.45%) respondent, as shown by the results.

6.2.4.3   Metadata description elements used

The respondents were asked the following question:

Which metadata description elements do you use to assign geographic subject metadata for ETDs? (Please specify the name of the standard and the elements that are used for your ETDs in the space provided below, e.g. Dublin Core subject, (dc:subject).

The findings indicated that metadata creators used different metadata description elements to assign geographic subject metadata for ETDs. As shown in Table 6.19 (p. 202), it was also evident that there were different metadata schema used by the different libraries that participated in this study. No predefined response categories were given for this question.

Table 6.19: Metadata description elements

Dublin Core Subject using MODS schema
Descriptive metadata
Dublin Core Subject
Dublin Core Subject
dc.subject; dc.coverage.spatial
Dublin Core (dc.subject)
dc.subject and dc.subject (LGSH)
dc.subject
Dublin Core Subject (dc.subject)
Dublin Core
Subject
dc.subject - Dublin Core
dc.subject.lcsh
ContentPro
dc.subject
Dublin Core subject
MODS

Dublin Core was shown to be the most used description schema. The commonly used description element was the uncontrolled subject element (dc:subject). Among the responses given for this question, there was only two respondents, who explicitly indicated that they used the Library of Congress Subject Headings (LCSH). One respondent mentioned the use of the Dublin Core, Coverage: spatial (dc.coverage.spatial). Some of the responses mentioned the schema, but did not specify which of its elements they used to assign geographic subject metadata to ETDs. It is common for the description scheme and its descriptive elements to exert a possible influence on the type of metadata assigned.

6.2.4.4 Adequacy of geographic subject metadata

The respondents were asked to respond to the question:

How would you rate the current level of adequacy of geographic subject metadata attached to your institution's ETDs records?

Table 6.20 (p. 203) outlines the responses to this question.

Table 6.20: Adequacy of geographic subject metadata

Rating	Frequency	Percentage
Highly adequate	12	41.38
Fairly adequate	9	31.03
Moderately adequate	8	27.59

The respondents' ratings for the level of adequacy of geographic subject metadata attached to their library's ETD records were recorded as highly adequate by 12 (41.38%) respondents' fairly adequate by nine (9) (31.03%) respondents; and moderately adequate by eight (8) (27.59%) respondents. There was no indication of views pointing to total inadequacy. It was observed that the respondents' perspectives indicated that confidence levels about the current adequacy levels of the geographic subject descriptions for ETDs in their collections was mainly positive.

6.2.4.5 Influence of understanding of the theory of subject analysis

An analysis of the responses given to the question, *In your opinion, to what extent does an understanding of the theory of subject analysis influence geographic subject metadata creation?* is shown in Figure 6.10.

	Major influence	Fair influence	Moderate influence
■ Frequency	14	11	4
■ Percentage	48,28	37,93	13,79

■ Frequency   ■ Percentage
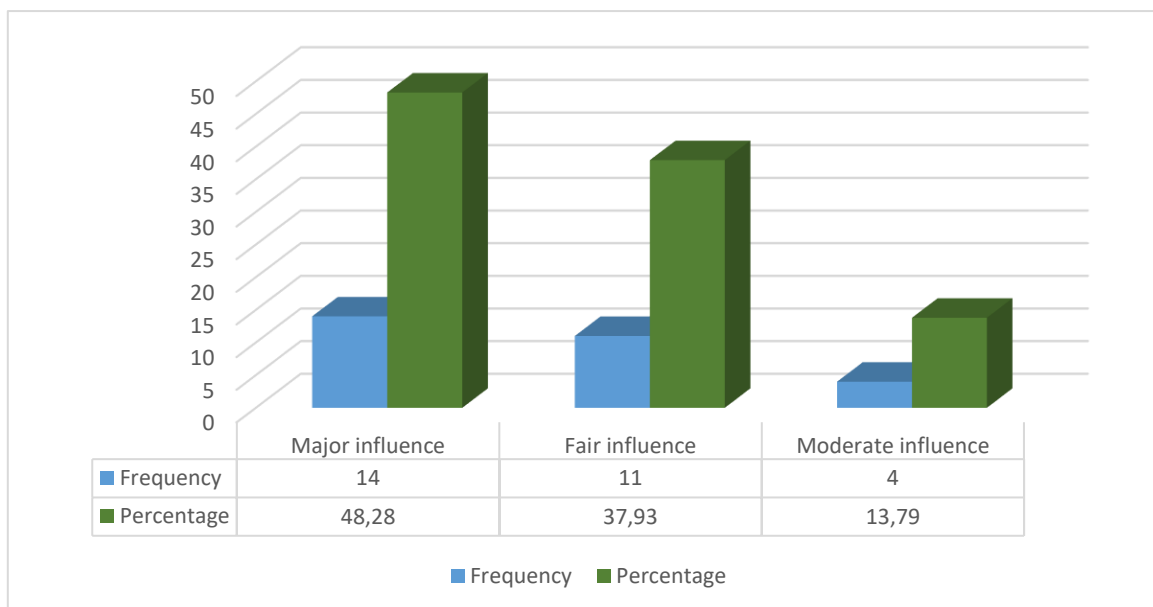
Figure 6.10: Influence of understanding of the theory of subject analysis

These responses indicated that none of the respondents expressed entirely negative views about the need to understand the theory of subject analysis as a basis for geographic subject metadata creation. A noteworthy number of 14 (48.28%) respondents believed that the theory of subject analysis had a major influence on geographic subject metadata creation. In addition, 11 (37.93%) respondents rated the influence of theory of subject analysis as being fair; and four (4) (13.79%) respondents identified the extent of the influence as being moderate.

6.2.4.6   Important aspects when assigning ETD geographic subject metadata

The researcher aimed at establishing what the metadata creators regarded as important aspects to consider when assigning geographic subject metadata for ETDs and how would they rate their importance.

Table 6.21 presents the results of this question.

Table 6.21: Important aspects when assigning geographic subject metadata

Item no.	Statements	Very Important	Fairly Important	Important	Slightly Important	Not Important
Q27.1	The author's intentions for producing the work	19 (65.52%)	6 (20.69%)	2 (6.90%)	2 (6.90%)	0%
Q27.2	The overall content of the ETDs (topics coverage)	20 (68.97%)	7 (24.14%)	2 (6.90%)	0%	0%
Q27.3	The information needs of all potential users of the ETDs	15 (51.72%)	7 (24.14%)	6 (20.69%)	1 (3.45%)	0%
Q27.4	The information needs of users within a specific subject field	20 (68.97%)	5 (17.24%)	4 (13.79%)	0%	0%
Q27.5	Researching the past information users' requests	8 (27.59%)	7 (24.14%)	9 (31.03%)	5 (17.24%)	0%

The two aspects that were equally ranked as "highly important" by 20 (68.97%) respondents were (Q27.2) the overall content of the ETDs (topics coverage) and (Q27.4) the information needs of the users within a specific subject field. When considering the affirmative responses (very important, fairly important and important) for these two aspects, the overall positive rating for both was significant. None of the two aspects was rated low by any of the respondents, in the ranks of "slightly important" or "not important". The author's intention (Q 27.1) followed with the highest positive responses at 19 (65.52%).

Researching the information users' request patterns received negative rankings: only five (5) (17.24%) respondents ranked it as "Slightly important. The respondents did not view researching the information users' needs as the most crucial aspect when assigning geographic subject metadata for ETDs. However, there was no complete negative responses for any of the listed aspects.

6.2.4.7 Other perspectives on geographic subject metadata analysis and creation

The participants expressed different views when they were asked to indicate their level of agreement or disagreement to several statements concerning ETD geographic subject metadata. The questions sought to learn additional perspectives that could assist in addressing the research questions of this study. Table 6.22 presents the responses to this question.

Table 6.22: Views on ETD geographic subject metadata

Items. No	Statements	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
Q29.1	Geographic subject metadata is important for the description of the subject content of ETDs.	18 (62.07%)	11(37.93%)	0%	0%	0%
Q29.2	Library staff involved with metadata creation assign the most suitable geographic subject metadata for	6 (20.69) %	12 (41.38%)	0%	0%	0%

Items. No	Statements	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
	the discovery of the subject content of ETDs.					
Q29.3	The authors of ETDs supply the most suitable geographic subject representations for their subject content.	12 (41.38%)	8 (27.59%)	5 (17.24%)	4 (13.79%)	0%
Q29.4	It is important to consider the public information users during the creation of geographic subject metadata for ETDs.	15 (51.72%)	7 (24.14%)	7 (24.14%)	0%	0%
Q29.5	Standardised approaches are essential in the creation of geographic subject metadata for the ETDs to facilitate the sharing of metadata.	18 (62.07%)	8 (27.59%)	3 (10.34%)	0%	0%
Q29.6	Guidelines on how to conduct subject analysis to determine the geographic subject metadata are essential.	18 (62.07%)	11 (37.93%)	0%	0%	0%
Q29.7	Training on how to conduct subject analysis is essential.	19 (65.52%)	10 (34.48%)	0%	0%	0%
Q29.8	Knowledge of subject analysis theories is important to facilitate professional subject metadata creation.	19 (65.52%)	9 (31.03%)	1 (3.45%)	0%	0%

Items. No	Statements	Strongly agree	Agree	Not sure	Disagree	Strongly Disagree
Q29.9	Technology has made subject analysis for the creation of geographic subject metadata unnecessary.	0%	3 (10.34%)	7 (24.14%)	19 (65.52%)	0%
Q29.10	The ETDs subject metadata creation approaches that are currently followed in my library accommodate the geographic information needs of the broad public.	13 (44.83%)	14 (48.28%)	2 (6.90%)	0%	0%
Q29.11	Subject analysis should only focus on satisfying the information needs within a specific subject field.	4 (13.79%)	5 (17.24%)	9 (31.03%)	10 (34.48%)	0%
Q29.12	Changes in place names affect ETDs discoverability.	9 (31.03%)	8 (27.59%)	5 (17.24%)	7 (24.14%)	0%

All the respondents (100%) agreed that geographic subject metadata was important, although to varying extents, with 18 (62.07%) respondents perceiving it as highly important and 11 (37.9%) respondents as important. Furthermore, there was consistency among the respondents in considering the library staff involved in metadata creation to be assigning the most suitable geographic subject metadata, with 16 (55.17%) respondents strongly agreeing and 12 (41.38%) respondents agreeing to this view.

The responses to the question (Q29.3) whether the authors of ETDs supply the most suitable geographic subject representations showed reasonable agreement that this was the case, with 12 (41.38%) respondents strongly agreeing and eight (8) (27.59%) agreeing. However, an element of uncertainty could be traced, with five (5) respondents (17.24%) indicating that they were not certain. Consistent with the responses to Q18.1, bone (0%) of the respondents held a completely negative view about the suitability of the author-supplied geographic subject representations.

The responses to the question (Q29.4) on whether it is important to consider the broad public information users during the creation of geographic subject metadata for ETDs were mostly positive. This approach was rated by 15 (51.72%) respondents strongly agreeing, and seven (7) respondents (24.14%) agreeing. An element of uncertainty also prevailed, with seven (7) respondents (24.14%) not being sure about this approach.

A related question (Q29.10) asked whether the current approaches followed in the respondents' libraries accommodated the geographic information needs of the broad public. The results showed that 13 (44.83%) respondents strongly agreed; 14 (48.28%) agreed; and seven (7) (24.14%) respondents were not sure. It was also observed that the majority of the respondents were positive – 18 (62.07%) respondents strongly agreed and eight (8) respondents (27.59%) agreed that standardised approaches were essential in the creation of geographic subject metadata for ETDs to facilitate the sharing of metadata.

Among the ranked statements, more negative ratings were assigned to questions Q29.9 and Q29.11, which asked if technology had made subject analysis for the creation of geographic subject metadata unnecessary and if subject analysis should focus on satisfying the information needs within a specific subject field only. Disagreement with the statements was expressed by 19 (65.52%) and 10 (34.48%) respondents respectively.

6.3   Qualitative interview data analysis

Interviews were conducted to collect detailed and in-depth information to explain the quantitative findings about subject analysis approaches and their effect on geographic subject metadata for ETDs. The analyses and findings are presented in line with the sequential nature of the study, where the questionnaire survey findings were analysed first. The perspectives and experiences described by the metadata creators helped to clarify and broaden the understanding of how subject analysis facilitated the creation of ETDs geographic subject metadata.

The interview phase of this sequential explanatory study was conducted from a phenomenological position to investigate the metadata creators' perspectives, their practical experiences and the influential contextual factors to metadata creation. According to Mayoh and Onwuegbuzie (2015:94), a phenomenological approach helps to investigate perceptions and beliefs and how they inform practice. The phenomenological approach also informed other steps followed in this study, including the determination of the sample and interpretations to explain the quantitative findings. A theoretical approach informs the decisions on how sampling is conducted (Lopez & Willis 2014:730). The approach influenced how a sample of seven metadata creators was considered suitable this for study, as described in Section 6.3.1.

Thematic analysis was used for the qualitative interview phase of this study. According to Braun and Clarke (2006:78), thematic analysis is an analytic approach that facilitates the identification of themes or patterns in data without being confined to the prescriptions of any theoretical approach. The use of thematic analysis was deemed appropriate for the analysis of the collected qualitative data and to help address the research questions.

The thematic analysis procedures used in this study were guided – with adaptations – by the steps outlined by Braun and Clarke (2020:4). The researcher used the NVivo 12 software and, where necessary, complemented the automatic assisted procedures with manual analysis to build further relationships of the meanings identified from the data. Alhojailan (2012:39) supports the approach of using manual analysis to complement the analysis tools, where it may necessary.

The following procedures were followed in this study:

1. Data familiarisation
2. Systematic data coding
3. Generating initial themes from coded and collated data
4. Developing and reviewing themes
5. Refining, defining and naming themes
6. Writing the report

The procedures are discussed in detail in Chapter 5 (Research methodology).

Furthermore, the qualitative phase of this study was based on the understanding of qualitative research being a form of interpretative inquiry. The researcher used personal judgement or subjective interpretations to induce meaning from the data, in order to understand the subjective perspectives of the participants' responses. The researcher's prior knowledge of metadata creation assisted in constructing meaning from the perspectives in the analysis process.

6.3.1   Participating libraries

Permission to conduct the study was requested through the various ethical clearance procedures, as advised by the university libraries. A total of seven (7) metadata creators participated in the interviews, with two coming from the same library, as they considered it ideal, due to their workflow. The seven participants were based in libraries that met the following criteria:

- All libraries having institutional repositories (IRs) with metadata created for ETDs; and
- Libraries having granted permission to conduct the study and the availability of experienced metadata creators to respond to questions on subject metadata creation; particularly geographic subject metadata.

The libraries that participated in the interviews fell within the three categories of universities in South Africa – the traditional universities, comprehensive universities and the universities of technology.

The inclusion of the three categories was meant to assist to use a sample that would be varied and representative of all the three sectors. Among the libraries that granted the permissions to conduct the study, six (6) were approached with a request to conduct the interviews. These were purposively identified as libraries with staff who are experienced in the creation of ETD metadata and who were easily reachable in terms of proximity for the personal interviews and additionally as a result of accessibility and the interviewees' availability to participate in the telephone interviews. Table 6.23 indicates the university categories where the interviews were conducted and the codes used to identify the participants.

Table 6.23: University libraries where interviews were conducted

University	Category	Participants' codes
1	Traditional University	Participant 1TR
2	Traditional University	Participant 2TR
3	Comprehensive University	Participant 3aCU Participant 3bCU
4	Comprehensive University	Participant 4CU
5	University of Technology	Participant 5UT
6	University of Technology	Participant 6UT

The approached ETD metadata creators were considered to be involved and experienced in ETD metadata creation and with a potential to share in-depth information to substantiate the information collected during the questionnaire survey phase of this study.

Five (5) participants were interviewed in a face-to-face conversations and two (2) interviews were conducted telephonically. The interviews were recorded on a laptop recording software and notes were written as supplemental information. The interviews were conducted between December 2019 and March 2020. The period during which the interviews were conducted resulted from the availability of the participants during specific periods.

6.3.2   Background information

Brief demographic questions were asked during the interviews. The data collected from the questionnaire responses reflected the general background information on the participants and their libraries.

Questions were asked on the following aspects:

- *Position description and metadata creation role*

  All the participants indicated that they were professional librarians. The positions they occupy, which involved ETDs metadata creation, were labelled differently in the six libraries where they are working. The description of the positions and roles include Quality Controller, Metadata Specialist, Cataloguer: Institutional Repository, and Cataloguer: ETDs and Cataloguing. Furthermore, the roles of the participants in ETD metadata creation differed in terms of workflow and procedures followed in their libraries. The common practice was that all processes from item ingestion to final metadata output being done by an individual metadata creator.

- *Experience in the subject description role*

  All the participants mentioned that they were metadata creators with more than five years' experience in the practice of ETD subject metadata creation and with a background in cataloguing of information resources in the university libraries sector.

6.3.3   Interview data analysis hierarchical exploration summary

This section presents the data coding summary of the interviews conducted with the research participants. The discussions in this section are presented in line with the questions asked in the interviews, which were categorised according to a designed guideline (Appendix A). Participants' files were coded for analysis and relationships were drawn across the different interviews' data. The summary of the findings was made available for all interview data by means of NVivo 12 software. The data yielded information that was coded and analysed qualitatively. The qualitative results were summarised through the NVivo 12 software. Although similar categories of questions were asked to all participants, the types of responses to the questions were different across the different participants.

A hierarchy chart in Figure 6.11 (p. 214) indicates the coding and the response categories. The sizes of the squares show the aggregate number of times that the reflected information units were coded.

Table 6.24 presents the summary of the codes and sub-codes assigned to the interview responses. The codes were used to categorise the responses and to facilitate the analysis of the data. The main codes are the ETD geographic subject metadata, metadata procedures and experiences, subject analysis experiences and perspectives and vocabulary control.

Table 6.24: NVivo coding summary (codes and sub-codes)

Codes	Sub-codes
ETDs and Geographic Subject Metadata	Geographic subject metadata current experiences
	Knowledge acquisition and recommendations
Metadata procedures experiences	Usage extent and procedures
Subject analysis (SA), experiences and perspectives	SA importance
	SA improved knowledge necessary
	SA theoretical basis
	SA analysis orientation
Vocabulary control	Mixed approach
	Need for authority control
	Use controlled vocabulary
	Vocabulary control important

The rates at which different aspects were identified in the participants' responses are reflected in Figure 6.11. The size of the square reflect the number of coding references. Some themes are main themes and other subthemes (Saunders, Lewis & Thornhill 2016:585). This is a form of representing the main ideas from the participants' narratives. Although the frequency of occurrence of ideas was not used as the foundation for determining the findings, it is highlighted, because it reflects the coding results as per the exercise conducted through the NVivo 12 tool. The codes were matched with the participants' responses to perform the analysis. The coding is displayed in Figure 6.11 (p. 214).
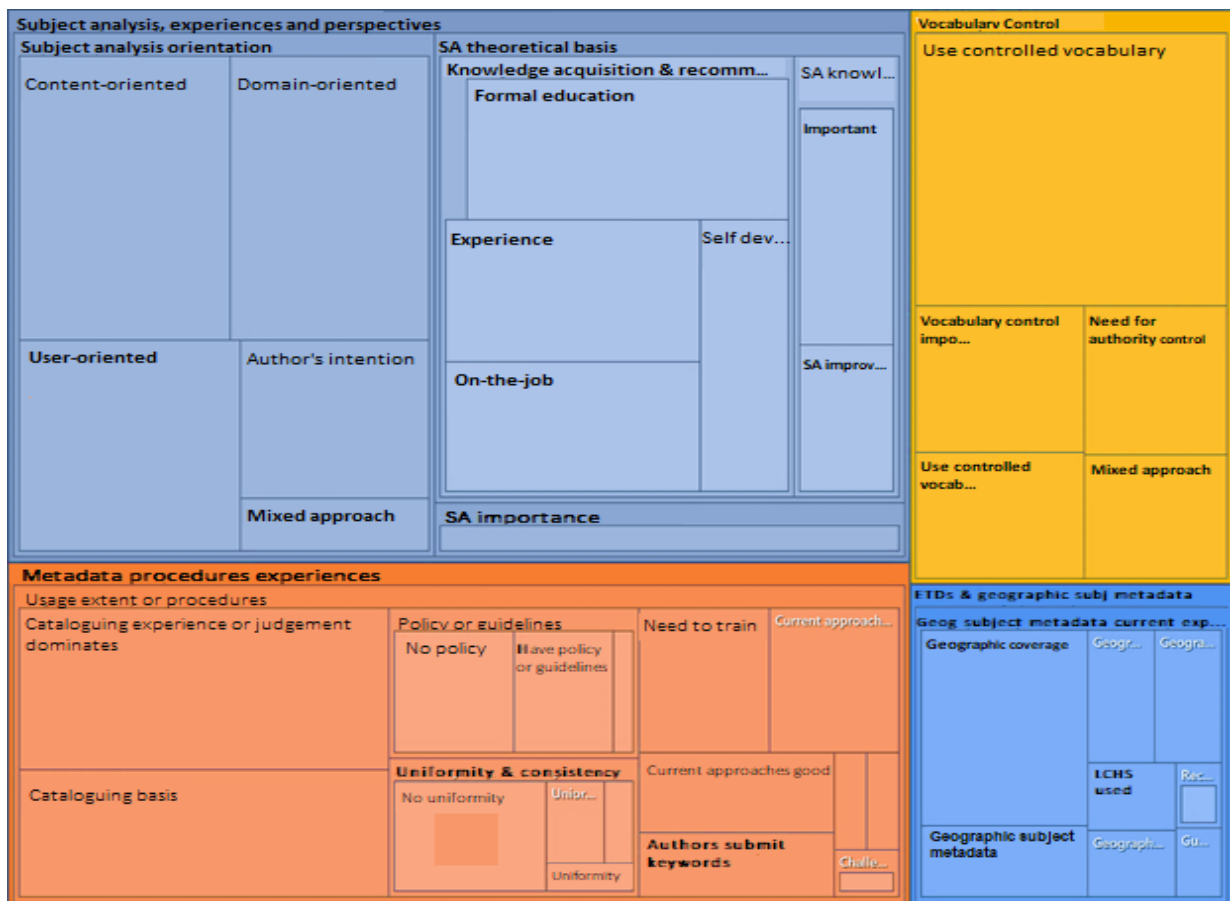
Figure 6.11: Items coding summary: hierarchical display

6.3.4   Metadata creation practices: experiences and perspectives

The first part of the interviews involved an investigation into the general experiences and perspectives with metadata creation – particularly ETD geographic subject metadata. The questions were meant to gather information on the experiences regarding the strategies used for subject metadata creation and e to investigate the perceptions of the processes and procedures of creating geographic subject metadata. The data revealed that geographic subject metadata was used to describe the ETDs. It was also observed that the traditional cataloguing practices played a role in influencing subject analysis and metadata creating experiences and perspectives. Inferences about the application of subject analysis theories were drawn from the different experiences and perspectives, in order to explain the practices identified in the quantitative findings.

In response to the question, *Explain how you experience the process of the utilisation of geographic subject metadata creation for the ETDs*, the metadata creators described different experiences and gave different views on the creation of subject metadata and the procedures that they were following. The data was analysed in relation to the research problem of investigating the application of subject analysis theories in the creation of geographic subject metadata for ETDs.

6.3.4.1 Role descriptions

The common role description by all participants was that cataloguing was part of their responsibilities and that metadata creators also performed other cataloguing duties for information resources in other formats. The cataloguing experiences emerged as the dominant basis of influence to the general approaches to ETD geographic subject metadata creation. This was common across all the participating libraries and it was the same finding as that in the questionnaire results. The general finding was that there appeared to be a profound link between the traditional cataloguing subject metadata creation experiences and the subject metadata creation in the digital space – particularly the ETD geographic subject metadata creation. The finding was consistent with views presented in literature – e.g. those of Xie and Matusiak (2016:130) and Rubin (2010:157) – in that subject metadata creation has links with cataloguing. This association, which continues in the digital information resources description space, suggests a continued influence of cataloguing theories on ETDs subject analysis and geographic subject metadata creation.

When asked to describe their roles in terms of ETDs and subject metadata creation, Participant 3aCU indicated who was responsible for creating subject metadata for the ETDs by remarking that:

Participant 3aCU: *It is cataloguers,* and a further comment by the participant on the same question was that *… going back to the cataloguers, what I know is that they are all qualified librarians. They are trained in librarianship; their experience ranges between five and ten plus, most of them have been doing cataloguing for a long time.*

More references to cataloguing and cataloguers were evident from most role descriptions. Cataloguers and cataloguing were mentioned in several responses, e.g. in the following six (6) responses:

Participant 4CU	*All of us we are five cataloguers and all of us are cataloguing the ETDs and as such they assign the subject headings*
Participant 5UT	*We are four cataloguers and all the four cataloguers have been doing electronic dissertations.*
Participant 6UT	In response to the question on the participant's experience with the process of metadata creation, Participant 6UT responded: *Because I am a cataloguer, what we do is we catalogue our theses on OCLC then we import it down to our local library management system.*

Similar to the previous responses, the link to cataloguing was evident in spite of not being explicitly stated in other responses. One of the participants did not indicate the association with cataloguing in the description of the metadata creation role, but responded when the question was asked about the need for knowledge about subject analysis theories.

Participant 1TR remarked that*:*
It is true as a cataloguer, especially as a metadata creator you need to have the knowledge about these different areas.

It was also found as a confirmation to the questionnaire findings that experience in cataloguing seemed to have an influence on metadata creation for ETDs. Two of the participants recommended the involvement of cataloguers to facilitate consistency of practice.

Participant 4CU	Recommended: *Having someone as a cataloguer mentoring the person who is going to be working on the IR*
Participant 6UT	*I would suggest that it must be a cataloguer who oversees it or did it at least.*

Further probing into the role descriptions revealed additional findings about subject metadata experiences and perceptions that reflected the extent and nature of use of geographic subject metadata as a form of description for ETDs. The general ETD subject metadata creation procedures and experiences described, varied.

6.3.4.2 Process and procedures for metadata creation

As observed in the foregoing discussions, all six libraries endorsed cataloguing experience and its basis for judgements made dominate the experiences of ETDs subject metadata creation. Furthermore, there were indications that geographic subject metadata was used for the description of ETDs and there were different accounts given on the experiences and perspectives regarding their application. When outlining their experiences, metadata creators stated different views that were interpreted as an indication of usage and clarified the different approaches regarding the process of geographic subject metadata creation.

Additionally, the following question was asked to the participants: *Please explain how you experience the process subject metadata creation for the ETDs*. Follow-up questions and further probing questions were asked to seek further clarity to the responses. The explanations given included the following:

Participant 1TR	*… with the geographic metadata creation, what we normally do is … the Library of Congress Subject Headings leading us as we do metadata creation … what we do is to try and make it more searchable for our users as possible.*
Participant 2TR	When posed with a probing question to ascertain the situation specifically with regard to geographic subject metadata use replied: *Yes, and (we consider) the geographic area.*
Participant 3aCU	Participant 3aCU explained the use of geographic description and an indication of the procedures as follows: *You will also check the scope, how far is the scope of the ETD, just to guide yourself … to start checking your geographic coverage. Like which geographic area is he covering.*

Participant 4CU	*I think it is definitely necessary that you specify which geographic area the research was done… I think it is absolutely necessary because that research has been in a specific area, so it might differ in another area … So I think it is very necessary.*
Participant 5UT	*…I must admit that myself although I find that my ETD has a specific ….(the) study concentrated around an area, specific, then I… assign that. ….. that's how we have been taught that to analyse your subject, add place name because most of the theses you will find that in the title … they would indicate where the research was done.*

Additionally, the varied procedures followed in the different libraries were described as follows:

Participant 4CU:	*Say your ETD subject is experiential learning, then you read through and you see which geographical (is) about … then you assign your subject headings according to that.*
Participant 5UT:	*We focus on the fields…we know which fields to populate and one of the fields is subject headings, so you will always fill in the subject headings*
Participant 6UT	Participant 6UT explained the workflow: *… we catalogue our theses on OCLC and then from there when we have to update it onto our institutional repository, we do copy and paste of what we already created. Then we edit it.* When probed further about the process followed with specific regard to geographic subject metadata, the participant (6UT) said: *Because I am used to working with Library of Congress Subject Headings, we follow that standard and rules, that's how we create ... it starts … with your country and you go smaller (areas).*

Agreement was noted on the use of geographic subject metadata for the description of ETDs. Several additional issues around the procedures emerged from the explanations, including how linking with cooperative bibliographic description spaces is practiced, the use of standards, and observing the principle of specificity in assigning subject metadata.

6.3.4.3　Consistency and uniformity of local practices

Further probing revealed that, despite geographic subject metadata being regarded as important for content description of ETDs, the participants raised problems of the lack of consistency and uniformity. Inconsistencies in general metadata creation practices may suggest the same to happen with the creation of geographic subject metadata for ETDs and the performance of subject analysis.

Participants 3aCU and 3bCU from the same library were not explicit in confirming their perspectives on the status of their current practices in terms of consistency and uniformity regarding subject metadata creation in general and including the geographic subject metadata creation practices. However, the need for consistency of practices was affirmed by their responses when they were asked if they had a policy that guides their processes of metadata creation:

Participant 3aCU:	*… as cataloguers now and then we will sit together and discuss guidelines how to deal with issues. And we agree … that If you are going to do an ETD and input the authors, this is how you are going to input the supervisors. You are going to follow the same format as we catalogue, because we want to maintain uniformity …*
Participant 3bCU	Participant 3bCU responding about consistent practices said: *If we are going to do cooperation it will be good to have a policy so that other new Librarians who start will (understand) it very well.*
Participant 2TR	Participant 2TR indicated the lack of consistency by stating**:** *… because currently we don't even have a supervisor, we are just deciding on our own… as long as it is accessible*
Participant 5UT	Participant 5UT, highlighting the possibility that there may not be consistency in the subject analysis practices said: *As I say, it is cataloguers' judgement. Unfortunately … due to our workload …, we never peer review each other's theses, you would sometimes peer review others' original cataloguing but not the theses and dissertations.*

Consistency is an important quality standard in metadata creation. In response to questions examining the level of consistency of practices, the majority of the participants shared similar views, identifying consistency as a challenge in the different libraries.

6.3.4.4 Perceptions about policies and guidelines

Responses about policy and guideline issues were picked up from responses for different questions asked during the interviews. In respect to the perspectives held about the aspects of consistency and uniformity in Section 6.3.4.3, some of the participants professed the need for policy or guidelines to direct the metadata creation practices as important. Policies may be useful to explain any inconsistencies for different practice. However, perceptions were gathered concerning the lack of appropriate policies for subject analysis approaches and practices in assigning subject metadata. Three participants pointed out the absence of policy or guidelines.

Despite the lack of policies and guidelines, participants view them as important to guide the decisions and processes as observed in the following responses:

> Participant 1TR: … *for now I think we can come up with a new way saying if you feel you can add keywords do it as a metadata creator, … there is no policy or a restriction from the institution to say you cannot add, … metadata creators for now are allowed based on their own judgement … whether there are new terminology that could be added as keyword.*

When further probing was done, the participants' offered the following comments with regard to guidelines:

Participant 1TR	*…. Because more so we did not have a guiding document that this is the (thing) we have to consider. I think it will depend on the level of understanding from the metadata creation.*
Participant 2TR	Participant 2TR confirmed and responded by saying: *No (policy), In my former institution there was a policy on adding and here you can add as many. Yes, policy is very important.* *Each and every individual will have their own procedure on how they do the analysis, so definitely you will need some guidelines on procedure to guide you as a new ….*
Participant 3bCU	Participant 3bCU indicated that their library did not have a policy that guided ETD metadata creation. This was confirmed by the second participant (3a) from the same library, with emphasis on the need for the guidelines for inexperienced staff.

Participant 3bCU	This view was expressed by
	In agreement, when asked if we need policy to maintain consistency in subject metadata creation commented as follows:
	If we are going to do cooperation, it will be good to have policy so that other new librarians who start will (understand)l.

One participant, who commented on policy and general knowledge to assist in improving the practice of subject analysis, raised a different perspective about policy. The comment was understood as indicating that the library did not have a policy, while the participant did regard policy as not freestanding concerning issues of performance competency.

Participant 5UTsaid: *That's why I say at this moment … that's just general knowledge. You as a cataloguer know how to create a subject.*

In the absence of policy or guidelines, it seemed that the practices adopted by new metadata creators would depend on the person who trained them. This was realised when a question was asked about the existence of a policy, and one participant responding on the situation in their library by responding as follows:

Participant 2TR: *Nothing, I was only told by my colleagues*

Further policy issues emerged in the responses to questions on subject analysis. Issues of policy emerged again when specific descriptions were given about subject analysis basis. The absence of policies appeared to have an impact on the current practices in geographic subject metadata creation. The responses were understood to imply that uniformity and consistency would be compromised in the absence of a guiding policy. When participants were asked if there was guidance about focus areas, one participant responded as follows:

Participant 2TR for example commented: *We don't have a policy, so it is the cataloguer's judgement*

The responses that implicitly reflected the absence of policy or guidelines and their importance were also observed. The foregoing responses seemed to suggest that an individual's discretion was an influential factor for subject analysis and assignment of geographic subject metadata. In spite of this current situation, there was an observed desire to maintain consistency, as expressed by the participants.

6.4.4.5 Perspectives on the need to improve current practices

Questions on two different aspects of the need for improvement received responses reflecting similar participant views. The need to improve on geographic subject metadata creation practices was highlighted several times by the participants. A question was asked as to whether the participants were satisfied, or if there was room for improvement concerning the approaches used by the different libraries. Two (2) participants commented on ways that implied the perceived need for improvement:

Participant 1TR	*It is true that at some stage you feel that you can add, but because we might not have the knowhow of particular disciplines keywords and terminology … what we are doing is to transcribe only what the author has sent … But there are times when you feel you can add because you feel some of the keyword should have been added but they are not there.*
Participant 5UT	*There is always room for improvement. It is not always easy because we don't have a Head of Cataloguing who can tell the others what to do. I can just go back with suggestions when we have a meeting as a group …We can … make suggestions, and observe …*

With further probing, it was asked if there was something that the participant could regarded as important and if it was necessary to introduce changes. The following responses were captured:

Participant 1TR	*I think we can come up with, but for now there is no policy …*
Participant 5UT	*Even after training you find quite a number of the cataloguers they are willing to do personal name authorities but they shy away from the geographic ………You would rather than omit the geographic heading because you don't know how to do it and you are not too sure and (how) to show the link.*

With further probing, two participants – 3aCUand 3bCU from the same library – indicated the need to improve, with specific reference to geographic subject metadata and geographic coordinates and linked data. The researcher's observation was that these were the views expressed by only two participants from the same library.

On the contrary, in response to probing on recommendations that could be proposed for the improvement of subject analysis approaches to enhance the geographic subject metadata for ETDs, one participant expressed satisfaction and did not recommend any need for improvement at the library. An emphasis on the use of standards in the library was observed:

> Participant 1TR said:
>
> *Answering for my institution I would say so far the method that we have been using … having to adopt the international standard. The set standard … is our vocabulary control procedure. I don't think there is anything so far as long as the information that we worked on serve the purpose of (access).*

Furthermore, the researcher identified statements in which the participants mentioned the possibility of benchmarking their processes.

6.4.4.6 Subject metadata creation challenges

Most participants raised a challenge pertaining to less experienced metadata creators' understanding. The need to train less experienced metadata creators in the practice of subject analysis and for assigning subject metadata was expressed by five participants.

For example, on two occasions, Participant 1TR emphasised the need to transfer knowledge from the experienced metadata creators to the less experienced ones. A further observation was that of Participant 2TR, who addressed the need for training new staff and furthermore indicated his/her challenge was the lack of skills for training others. Participant 5UT expressed an opposing view by saying, *I think we are doing enough in guiding (new staff) ...* This suggested there could be libraries where it was believed that reasonable training was being offered.

The findings drawn from the data presented in Sections 6.4.3.1–6.4.3.6, regarding the general experiences with geographic subject metadata creation, revealed varied perceptions about the efficiency of current practices and the need for improvement.

6.3.5 Subject analysis basis

The perspectives on the importance of the theoretical basis in terms of informed decisions and practices in subject analysis and assigning of geographic subject metadata were important for addressing the research questions of this study. The understanding was critical to facilitate the enhancement of ETD geographic subject metadata creation. The interviews sought to gather in-depth views of the participants on these aspects. Questions were asked on the awareness and level of understanding of the importance of subject analysis; and its impact on effective subject metadata creation; important aspects when determining what a resource is about; how subject analysis knowledge was acquired; and how different forms of training facilitated the understanding of subject analysis theoretical basis.

6.3.5.1 Perspectives on the importance of subject analysis

A general agreement was observed among the interview participants on the importance of subject analysis to facilitate geographic subject metadata creation. The participants also indicated that they recognised the importance of the theories of subject analysis and the basis that they provided. However, participants indicated the need to improve this knowledge. In response to the question that examined the perspectives on whether it was considered necessary to have knowledge about theories, the remarks were as follows:

Participant 3aCU	*It is necessary, even the general knowledge ….. is very important,*
Participant 4CU	*Yes. Absolutely.*
Participant 5UT	When asked if it was necessary for metadata creators to have this knowledge, Participant 5UT emphatically responded as follows: *I agree. Definitely, I do agree that we should take this back* (to our libraries).

The implicit meaning derived from these responses was that knowledge of the theoretical foundations to assist in subject analysis was valued by participants.

The responses also indicated how the acquired knowledge of subject analysis theories was understood as impacting on the subject analysis processes.

Participant 6UT, in affirmation, gave a detailed comment:
Yes, it is important and it is good to have a good general knowledge background, for instance … Some of our students … always do their studies and research based on the specific cities or specific towns that they live in or village. … it is important to point that out.

An indication of perceived challenges regarding the knowledge possessed by novice metadata creators was identified in the following comments:

Participant 2TR commented: … *the problem is the people we have employed don't know cataloguing at all …*
Additionally, as a response to further probing on the importance of formal training to make people to get the background information, Participant 2TR confirmed the above assertion and commented:
... the new librarians, it is difficult for them to understand …

6.3.5.2 Important aspects when determining what a resource is about

According to Tarver et al (2015:31), during subject analysis, an information resource is examined from various viewpoints to assign subject metadata. Prioritisation of focus during subject analysis is an important consideration when determining what a resource is about. The metadata creators should be aware of the effect of their prioritisation decisions on the effectiveness of the geographic subject metadata produced. Responses to the question on the priority of focus during the analysis were coded into categories, which revealed different perspectives and approaches.

It was observed that all the participants indicated the different focus that they maintained during subject analysis and the underlying motivations. The participants were asked what could be prioritised when conducting subject analysis, followed up by probing questions on whether metadata creators maintained a common focus and did that consistently.

The analysis revealed that metadata creators did hold different perspectives on subject analysis and its basis. Responses from all participants suggested that there was diversity in terms of what was prioritised when determining the subject, with an indication of where more focus was placed, e.g. focus on the resources' content was expressed as follows:

Participant 1TR	Indicated focus on content: *There are times where because we are not knowledgeable in all areas, what we consider is to look at the relevance of the content.* Furthermore, indicating the most preferred choice of concentration, the participant said: … *I think most of the research output will lead you through the summary … you go through the summary and you are more convinced that the subject you identified is tailored to the content then I think that's where it will take the lead.*
Participant 2TR	Referred to focus being placed on content and on several occasions emphasised the use of the abstract, e.g.: ... *we basically focus on the abstract, the summary of what the thesis is all about.* … *looking at the content is preferred.*
Participant 4CU	Indicated that concentration is put on the content: … *you read through the first page of the title, then you go through the abstract, and go through all the chapters and then you …. see which geographical it is about and then you assign your subject headings according to that, and say your ETDs subject is experiential learning, and then you read through and see … this is in Gauteng …* Affirmed the above position: *How we go about. You start with the title and you go through the abstract and then you go through the content and then you go through the whole thesis and you assign the subject heading. We normally try to stay as close as possible to the Library of Congress subject headings, but sometimes … they (authors) are very specific about what they write about and then you also assign those subject headings as well.*
Participant 5UT	In support of focus on the content, said: *It depends on the content I don't want to stray off the content. If that content is only for that specific thing…*

A different opinion mentioned was that participants focused on subject areas or subject disciplines addressed by the ETD content:

Participant 1TR	*What we consider is to look at the relevance of the content, and discipline into which the author falls in particular. If someone is writing from the social aspects, then it will dictate or will give us direction as to how … we create a geographic subject metadata ...*
	In addition, Participant 1TR explicitly stated:
	Subject specification, because I think most of the research output will lead you through the summary … and you are more convinced that the subject you identified is tailored to the content then I think that's where it will take the lead.
Participant 4CU	Gave an example and stated:
	Say your ETD's subject is experiential learning, and then your read through and you see which geographical it is about … then you assign your subject headings according to that.

An alternative approach was stated as follows by three participants, in line with the concentration placed on the user:

Participant 1TR	*By clearly defining our specific subject and document keywords, it helps us …to be able to reach the most potential user, so that we have our work usable and searchable and discoverable to the users.*
Participant 3bCU	*In a university library, I think that you should focus on the subject but also in context with the needs of the users.*
Participant 4CU	Commenting on the subject analysis theories, said:
	Yes, absolutely, I would say, in the first place it is must be user-oriented, but you must also know what the author is talking about in order to assign those subject headings.
	Additionally, on indicating the priorities when assigning better geographic subject metadata, Participant 4CU replied as follows:
	Definitely the users

There were several instances in which the participants mentioned more than one focal point, suggesting a diversified focus approach. The responses suggested a combination of priorities in terms of how the subject analysis was performed. An example of the combination of approaches (user and author) was mentioned as follows:

Participant 4CU: *Ye. Absolutely. I would say, in the first place it must be user oriented, but you must also know what the author is talking about in order to assign those subject headings.*

The foregoing findings contributed significantly to explanation of the practises based on the existing theoretical models in literature. The question the participants' choice, if they were given a chance to decide on the focus to maintain, was followed up with questions on whether the participants thought there was consistency and uniformity in approach among the metadata creators in their libraries. The perceptions of most participants showed that consistency on prioritised focus for specific aspects during subject analysis was not certain. An example is the comment indicating there was no consistency in the practices:

> Participant 1TR: … *Because … we did not have a guiding document … I think it will depend on the level of understanding from the metadata creation.*

6.3.5.3   Knowledge acquisition strategies

The participants were asked for their views on ways of improving knowledge about the theoretical basis of subject analysis. The questions aimed at investigating how improving this knowledge would impact on the performance of subject analysis and subject metadata creation. Despite the metadata creators affirming that they were knowledgeable of the theories of subject analysis, some comments suggested that they mostly considered what was prescribed by the encoding schemes to assign the subject metadata. Their cataloguing knowledge background also played an important role to inform their practices.

Furthermore, the responses indicated different ways in which knowledge was acquired and updated (kept current). When the participants were asked how they acquired the knowledge about the basis for subject analysis, responses from some of the participants were that their training exposed them to such knowledge. In addition, the responses identified different methods through which knowledge was acquired. The participants explained that knowledge was acquired through formal Library and Information Science education, on-the-job training, through experience and other forms of interventions.

Experience featured most in the responses given by all participants about how knowledge was acquired, for example:

> Participant 1TR said: *It started with formal training, but as you go to the workplace, you develop even further.*

Another indication of the views held by participants on the importance of the role of experience was observed when a question was asked on how the knowledge about subject analysis was acquired. The response given by three participants pointed to experience:

> Participant 1T
>
> *I think having to work with bibliographical utilities in the section as cataloguers, helped a lot when we to come to the issue of metadata creation. We move along with that; it gives you a basic … So it goes a long way to the metadata creation to have that knowledge.*

Another participant expressed a similar view by emphasising the levels of experience in describing his/her role in ETDs metadata creation:

> Participant 3aCU
>
> *…. that was a long time ago after gaining a lot of experience with regard to digitisation and cataloguing … now I have been working for the institutional repository for the last ten years ….*

Further affirmation of the important role of experience in expanding knowledge levels was given when a participant was asked if there could be something could be done at their library to facilitate improvement of creating geographic subject metadata for ETDs.

When asked about how knowledge for theoretical basis was acquired, Participant 4CU affirmed the perceptions held about the role of experience. After further probing, the participant identified experience again as a significant factor, as some of the knowledge was received during formal training. The following response was recorded:

> Participant 4CU
>
> *I think at this stage we are very fortunate that we have got a wonderful group of people together and I think we are all on par with, they know what they have to do and how to go about, that works very well for us.*

	…. so it definitely comes with experience, what you have learned while studying, you could apply but over the years you definitely gain some experience, and yes, that's also very important.
	Definitely yes; it gives you …. some background and … you know what to look out for... I would say definitely your academic.

Formal education was the identified as the second, most prominent way of acquiring knowledge of the subject analysis theoretical basis. For example, the following responses were expressed:

Participant 3bCU	*During formal education we got very basic thing about subject headings, assigning subject headings and so on.* In addition the participant said:
	… I already had this basic thing in my mind, because I was doing it already when I was still a student.
	Furthermore, in responding on the role of formal training, further affirmation was given
Participant 4CU	*Definitely yes; it gives you some background and … you know what to look out for. But with technology and all that over the years, yes, your perspective also change but I would say definitely your academic background and your experience over the years (counts).*
Participant 5UT	*I did my Library degree … It was in the curriculum how to apply subject matter and geographic headings; I mean that was part of cataloguing.*
	I studied the theory when I did my degree.

There was a number of references made to knowledge gained **on-the-job** through various means. Experiences in the form of informal in-house meetings were described as follows:

Participant 3aCU	Expressed the following view on in-house meetings:
	It is just that as cataloguers now and then we will sit together and discuss like guidelines ...
Participant 5UT	*We do talk about having meetings between cataloguers where we share knowledge … specifically to do headings. Again, to alert people on how to do headings. The correctness of headings. We do have meetings as cataloguers where we help each other or talk about things.*

In addition, there was noticeable reference to **self-development** as a method of acquiring knowledge for subject analysis and for metadata creation. The participants indicated other ways of gaining knowledge of the basis of subject analysis through workshops, conferences, webinars, etc., e.g. when questions were asked about the specific training interventions, the following responses were captured:

Participant 4CU	*We also had some workshops on that, we had a couple of workshops ... just to sort of look at more or less assigning the Library of Congress Subject Headings. And so that also helped a lot.*
Participant 5UT	Expressed a similar view by saying: *The general rule is cataloguing training, but working with other cataloguers, having the workshops … you get alert … networking and … colleagues.*
Participant 6UT	Recommended the following: *As I said practice, joining webinars, read up on it, keep abreast on the changes on the different rules, what is new, what is old and how to replace it, keeping abreast of all the trends.*
Participant 1TR	Shared related views about self-development: *… it is important to keep on attending these formal or informal trainings, get to network and know what other institutions are doing and also familiarise yourself with new trends especially with these new areas of metadata creation as a metadata creator.*

Learning from the responses, it could be assumed that there was a general effort to keep abreast of developments in subject analysis and subject metadata creation.

Furthermore, the participants indicated other training methods that they considered valuable and the different challenges associated with them. They expressed that experienced metadata creators could play an important role in mentoring and coaching the inexperienced colleagues.

However, the availability of experienced trainers, who were able to mentor and coach others, was pointed out to be a challenge, for example:

Participant 1TR	Recommending how knowledge about the theoretical background can be acquired, said: *… I think there are those that (I) will consider knowledgeable. I think it will be important to transfer this information (timeously).*
Participant 3bCU	*I think it is also the internships … to get an opportunity to learn and to realise the importance of the development of those skills.*

Implied in the foregoing responses, was an indication of the need for experienced metadata creators to transfer knowledge. The participants highlighted an additional method that could assist in disseminating knowledge.

The participants highlighted the need for guidelines to assist in gaining the subject analysis theoretical knowledge. The following responses were given:

Participant 2TR	*We are creating a document, where we are creating … rules, procedure manual, for when you are uploading etc. I think if we can have a … guide, that must include examples (for future use).* The comment is understood to be an indication of the need to document procedures for current and future use.
Participant 3aCU	*Each and every individual will have their own procedure on how they do the analysis, so definitely you will need some guidelines on procedure to guide you as a new person, (with time) somehow you develop your own way of just working…*
Participant 6UT	Implying the use of guidelines, Participant 6UT expressed the following view: *The more you work …, the more you learn ... More practical, as I work through it that helps. Working on examples*.

Additionally, Participant 6UT affirmed the need for guidelines when recommending other ways of facilitating the gaining of theoretical knowledge about subject analysis by recommending practical examples to repeat exercises.

The participants' perceptions on how training enhanced the understanding of the foundational knowledge of subject analysis were positive. There was a general consensus among them about the importance of the training received on subject analysis theories.

The views can be summarised by the comment made by Participant 4CU, who, when probed about subject analysis theory training method, indicated the value of subject analysis theory and remarked as follows:

Participant 4CU

Definitely yes, it gives you some background and … you know what to look out for

---

The perspectives shared in this section were important for assessing the need for improvement of theoretical understanding and what measures needed to be adopted to facilitate it. The next section looks into the perspectives that explained the dominance of controlled vocabulary observed in the quantitative findings.

6.3.6   Perspectives on the role of controlled vocabulary

The explanations on the use of controlled vocabulary and the recommendations to adhere to the standards to create geographic subject metadata reflected the participants' preferences for standardisation of ETD geographic subject metadata practices. It was regarded as important to report on these perspectives about the use of standards, as the participants were observed as prominently indicating their use in their questionnaire responses.

Saunders, Lewis and Thornhill (2003:89) indicate that one of the aims of following inductive approach to analysis may be to examine how participants cope with the challenges or the problems they experience. The data revealed that the participants' responses mentioned vocabulary control as one of the ways in which they approached subject metadata creation.

The responses about the general experiences across all the participants indicated the use of controlled subject metadata. The importance of vocabulary control was also explicitly pointed out when responding to different questions during the interviews, with extensive reference to the use of the Library of Congress Subject Headings (LCSH). In response to a probing question on the role played by the participant in ETD metadata creation, the answers indicated the use of a vocabulary control standard:

Participant 3aCU	*I will still use the same format that I use when I catalogue. I will start with surname, initials or as it appears on the ETDs, and you go through your keywords that are there because most of the theses come with the keywords. Guided by those keywords you are able to create your subject headings using the Library of Congress Subject Headings.*
	As a follow-up response, Participant 3aCU said: *DSpace allows you to put the keywords as they are. But I think … we also agreed that we should also input the authorised subject headings using the Library of Congress.*
Participant 4CU	*… We normally try to stay as close as possible to the Library of Congress subject headings, but sometimes … the (authors) are very specific about what they are talking about and then you also assign those subject headings as well.*
Participant 6UT	*If everyone is following the Library of Congress Subject Headings, then it should not be an issue. And I think that is the international rule that we are all following. So it should be fine.*

Responding to the question on the approaches that were currently being followed, their impact on the creation of geographic subject metadata, and suggestion on possible different approaches, another participant commented in favour of the LCHS.

All interviewees confirmed the use controlled vocabulary for the creation of subject metadata. In contrast to the use of controlled vocabulary, the use of keywords also emerged from some of the responses. This observation was an indication of the possibility of a diversified approach of employing both controlled and uncontrolled vocabulary

6.3.7   Other experiences, perspectives and recommendations

The interviews were concluded by asking additional questions on the theoretical and practical aspects of subject analysis; assigning subject metadata; and their effect on the geographic subject metadata for ETDs. The questions were asked to probe for more information and to fill gaps identified from the earlier responses to specific questions asked during the interviews. In-depth meaning of the different aspects that were considered useful for addressing the research questions were derived from these responses to further probing.

The interviews were ended, once the researcher was convinced that saturation had been reached. Responses that pointed out information relevant for this study were selected for reporting. The main findings from the responses are presented below.

The need to maintain specificity in geographic subject metadata allocation was expressed. Learning from the responses by one of the participants, it was observed that it would assist to be as specific as possible in the creation of geographic subject metadata for ETDs.

Participant 3bCU said:

You can also go to repositories' own interface, which is not a very specific search engine, but it will be a little bit tricky or you should know the system to narrow down your results because you might end up with a lot of irrelevant things.

Some participants also emphasised the importance of using other standards. Perspectives on standard that guide the formulation of geographic subject metadata, like the Geonet, were observed, for example:

Participant 3aCU remarked that:

… what is interesting with the technologies …, with the change of street names, with the towns ... Those ... incorporated in the Geonet. So, even if they are not on OCLC, you know that … whoever is going to create an authority form of this name. You will still go to Geonet.

Some participants raised considerations about standardisation for the purpose of consistency among university libraries. When asked what approaches could be followed to make sure that people had theoretical understanding of subject analysis, Participant 1TR mentioned the need to close the gaps between institutional practices. Benchmarking was also suggested as a means of facilitating the closing of gaps amongst institutional practices.

Participant 1TR remarked as follows:

We are trying to explore more as to how other institutions are doing this, but so far we are all conversant with the way we are doing it. So there hasn't been any development, we are trying to benchmark with other institutions as to how they are doing it.

Different perspectives expressed highlighted challenges and proposed solutions. Some solutions were in the form of suggestions to improve the ways of conducting subject analysis. The systems used for hosting the ETDs and for creating the subject metadata for ETDs have been pointed out as having a shortcoming of not accommodating the authority files.

A recommendation was made as to adopting authority control. In response to a question on what to incorporate into the current procedures. The recommendation was given as follows:

> Participant 3bCU:
>
> *I think there is one gap…… In a system like DSpace, we don't have an authority file. And authority is very important it plays a very important role in the description of subjects. Though we have the Library of Congress subject headings, you can assign them and you can put the on the database but there is no authority file in the system, that is definitely a gap and it is a challenge.*

Some participants commented on workloads and time for quality control.

> Participant 5UT remarked that:
>
> *Of course. … it is cataloguers' judgement. Unfortunately … due to our workload and … we never peer review each other's theses, You would sometimes peer review others original cataloguing but not the theses and dissertations.*

6.3.8 Current approaches for ETD geographic subject metadata

The respondents expressed different views on how current metadata creation approaches impacted on the creation of geographic subject metadata. Theoretical knowledge of subject analysis and its influence on the current approaches being followed for creating of geographic subject metadata for ETDs was investigated in all phases of this study. The participants were in consensus that geographic subject metadata was affected by the current practices, which were experienced differently, as presented in the next section.

Different experiences were observed from the participants' comments. Experiences regarding the perceived levels of complexity of geographic subject metadata creation were described. From the shared experiences and perspectives, observations were made about the understanding regarding the impact of the current practices of subject analysis, particularly on geographic subject metadata creation. Understanding how the effect was perceived should help to understand the intervention needed to promote the knowledge of the theories of subject analysis.

Two references were made to the fact that geographic subject metadata was difficult to assign, compared to other types of subject metadata.

Participant 5UT:
The geographic headings is a little bit technical. In addition, the participant said: *Even after training you find quite a number of the cataloguers they are willing to do personal name authorities but (not) the geographic because the geographic is next level of subject headings. It has got more work, it is more difficult, you have to research more.*

A different perspective that was expressed by some participants was that subject metadata creation was performed without numerous challenges. This was observed from several references that affirmed good experience with metadata creation:

Participant 6UT	*… because I am a cataloguer, it has been fine, it is easy.*
Participant 3bCU	Sharing an additional perspective, 3bCU was of the view that subject metadata creation was easy, the following experience was shared: *At this stage, our cataloguers are already experienced, they are doing it automatically, they know.*

Responses to the same question about experiences with metadata creation showed that views about complexities of subject analysis were based the current local practices, or they were judged according to individual standards.

Participant 1TR commented:
… what we have been doing so far, it has been a success, it might not be that higher level that other institutions are at but then is sufficient that when information created is searchable information can be retrieved. I think so far so good for our database.

Additional views were obtained in relation to the perceptions that geographic subject metadata creation was a complex process. Some participants believed that novice metadata creators in some libraries needed more knowledge and guidance to perform subject analysis and to create subject metadata:

Participant 2TR	*… I can give you example, the new librarians, it is difficult for them to understand why I put this and sometimes it is difficult for me to explain to them why.*
Participant 3bCU	*If we are going to do cooperation, it will be good to have policy so that … new librarians who start will get it very well.*
Participant 4CU	In response to further probing for suggestions on how to guide new staff, commented: *… (I) ask (new staff) how they would go about doing it and then guide them in the right direction and just show them what to look out for and … better analysing ….*

6.3.9   Concluding remarks on the qualitative data analysis

Conclusions drawn from the interviews data analysis were that the experiences and perspectives clarified the quantitative findings by affirming that the university libraries in South Africa participated in creating ETD metadata – particularly geographic subject metadata. The results of this qualitative phase of the study provided insight into the current practices of subject analysis and geographic subject metadata creation for ETDs.

Additionally, issues regarding familiarity with the theoretical principles for subject analysis were explained. Furthermore, how training for subject analysis was offered was pointed out. It further observed that different views were held on the focus that should be maintained when creating subject metadata for ETDs. In addition, views were raised about the use of standards for vocabulary control among all the libraries. The response also covered the issues of metadata creation policies or guidelines.

Varied perspectives were identified regarding theories of subject analysis and their effect on ETD geographic subject metadata. Additionally, several recommendations were made to fill the knowledge gaps.

The general observation was that participants shared different perceptions about the process of subject analysis and the geographic subject metadata. Moreover, the results suggested that the levels of experience shaped the different understandings on how knowledge on the theories of subject analysis affected practice.

6.4   Data analysis and findings of the content analysis

This section presents the results of the analysis of the data collected by means of the content analysis method. The collected data was recorded and reported by means of Excel analysis functions to present the quantitative results. The data was analysed descriptively, summarised and reported in tables and graphs.

Patterns and trends of subject representations and the nature of the use of ETD geographic subject metadata were observed. Patterns and trends were studied to find information to support the interview explanations, in line with the explanatory design purpose. Additionally, a qualitative analysis was conducted with themes that emerged from the categories of the quantitatively summarised data used for discussions of the latent or underlying meanings. The themes drew from the findings of previous phases of the study, which the content analysis was intended to supplement. Qualitative data analysis can use deductive data analysis (Teddlie & Tashakkori 2009:252).

Similar content analysis studies, conducted previously on information resources descriptive records were used as a reference for successful research of a nature related to this study. The studies involved exploring metadata developments over time. Such studies include those of Terra et al (2021), who studied theses records in a digital repository, created between 2001 and 2019, and Preminger et al (2020), who conducted a related study on public library metadata landscape, studying metadata records created between 2018 and 2019. A study on editing library metadata records, analysed records created between 2013 and 2015 (Zavalina et al 2016).

Additionally, Marc (2016) conducted a study on assessing metadata quality for the period between 2012 and 2014. Another related study that was conducted on library metadata change and how this can be related to metadata quality change in the library catalogues, analysed metadata records for October 2009 and December 2012 and January to April 2014 and between 2014 and 2016 (Zavalina et al 2015).

The records in Examples 1–2 show excerpts from records on the NETD database and on the institutional repositories (IRs). These examples show how geographic subject metadata is used to describe ETDs in the records of one of the studied libraries. The tiles, which contain place names and the assigned subject metadata, including the geographic subjects, are displayed.

Example 1: Excerpt of a record in the NETD database

Title	Mandarin as foreign language in higher education in China and South Africa
Author	Wang, Yuhua (Educator)
Subject	Mandarin as foreign language
Subject	Comparative study
Subject	Provision
Subject	China
Subject	South Africa
Subject	Chinese language -- Study and teaching (Higher) -- China -- Foreign speakers
Subject	Chinese language -- Study and teaching (Higher) -- South Africa -- Foreign speakers
dc.type	Thesis

The NETD records link back to the repositories where they were harvested. An example of a full record from the ETDs repository is shown in Example 2.

Example 2: Excerpt of a record from ETDs IR

Title	Mandarin as foreign language in higher education in China and South Africa
dc.subject	Mandarin as foreign language
dc.subject	Comparative study
dc.subject	Provision
dc.subject	China
dc.subject	South Africa
dc.subject	First language dominant
dc.subject	Higher education
dc.subject.lcsh	Chinese language -- Study and teaching (Higher) -- China -- Foreign speakers
dc.subject.lcsh	Chinese language -- Study and teaching (Higher) -- South Africa -- Foreign speakers
dc.title	Mandarin as foreign language in higher education in China and South Africa
dc.type	Thesis

Chapter 5 discussed the full analysis procedures. The records that were examined were taken from the final list drawn from the first list of records harvested from the NETD database.

Content analysis follows a systematic approach, including sampling, determining the unit of analysis, coding and analysis (Saunders, Lewis & Thornhill 2016:611). Initial deductive steps involved the categorisation of units, according to the coding scheme, into positive (comply) and negative (not comply) values to indicate frequencies of occurrence and to indicate and compliance to identified types of use (Appendix B).

Subsequently, the counts were converted to numerical values. Tabulating frequencies of each characteristic found in the examined material was crucial in a study involving content analysis (Leedy & Ormrod 2013:149). Descriptive statistics were used to analyse the data. Tables and graphs are used to display the results, displaying frequency counts and percentages.

The next phase of the analysis involved qualitative reflection on the quantified data and using themes to do the description of the findings regarding the use and nature of use of geographic subject metadata, describing patterns and trends observed. Qualitative data analysis can use deductive analysis (Teddlie & Tashakkori 2009:252). Themes were derived to support the interview findings in line with the explanatory design purpose. Deductive analysis in a qualitative approach is driven by the researchers' analytical interest (Braun & Clarke 2006:84). Emphasis remains on the explanatory qualitative aspects of the findings.

The quantitative analysis yielded categories that could be analysed qualitatively. Krippendorff (2013:29) states that meaning must be sought outside the physicality of the texts themselves to address a specific purpose or relative to a specific context. The next sections report on the findings of the data analysis.

6.4.1   Extent and nature of usage of subject metadata for ETDs description

Table 6.25 captures the results of the data analysis for the practices in subject metadata creation. The results indicated how subjects were used to describe the content of the ETDs between the years 2014 and 2018. Subjects form part of the core elements for the description of information resources, including the digital formats.

Table 6.25: Subject metadata usage

Dates	Number of titles with place names examined	Number of Records with subjects metadata	Number of records without subjects metadata
2014	555	406 (73.15%)	149 (26.84%)
2015	505	378 (74.85%)	127 (25.14%)
2016	235	226 (96.17%)	9 (3.82%)
2017	281	277 (98.57%)	4 (1.42%)
2018	1576	1287 (81.66%)	289 (18.33%)

The results in Table 6.25 reflect how subjects were used to describe the content of the ETD records on the NETD database, created between the years 2014 and 2018. Consistently, between 2014 and 2018, most titles with place names were consistently assigned with subject metadata in the period 2014–2018. The percentages increased from 73.15 % in 2014 to 98.57% in 2017. An insignificant drop occurred in 2018, where the assigned subjects' percentage was 81.66%. On the other hand, the number of titles with place names, but not assigned with subjects, varied in the different years. It was more (18.33%) in 2018, compared to the preceding two years, 2016 with 3.82% and 2017 with 1.42%. The absence of subject metadata implied the nonexistence of geographic subject metadata too.

6.4.2   Current approaches for ETDs and geographic subject metadata

Subject metadata can take various forms, for example, topics, personal names, geographic place names. The basic unit of analysis in this study was the geographic subject metadata.

Table 6.26 (p. 243) shows how geographic subject metadata was used to describe the contents of the ETDs.

Table 6.26: Geographic subject metadata usage

Dates	Records assigned with subject metadata	Records with geographic subject metadata	Records without geographic subject metadata
2014	406	134 (33.0%)	272 (66.99%)
2015	378	177 (47.82%)	201 (53.17%)
2016	226	175 (77.43%)	51 (22.56%)
2017	277	208 (75.09%)	69 (24.90%)
2018	1287	748 (58.11%)	539 (41.88%)

The results of the data analysis on ETD records assigned with subject metadata, records assigned with geographic subject metadata and records without geographic subject metadata results is given in Table 6.26. The analysis revealed inconsistency in the use of geographic subject metadata to represent ETD subject content over the five-year period. There was also an indication of less use of geographic subject metadata, compared to the general subjects assigned, as seen in Table 6.25 (p. 242). The highest number of 272 (66.99%) reflected cases where geographic subject metadata was not assigned to describe the ETD records in the year 2014. However, more records were generally assigned with other types of subjects.

However, an increase in assigned geographic subject metadata was observed in the years 2016 and 2017, when geographic subject metadata assigned were 175 (77.43%) and 208 (75.09%), respectively. Differently, a decrease was experienced in the year 2018, when only 41.88% of records were assigned with geographic subject metadata. The increase in subjects assigned in the year 2018 was not consistent with the allocation of geographic subject metadata in that year. The findings suggested that the presence on subject metadata did not imply the application of geographic subject metadata.

Figure 6.12 (p. 244) represents the distribution, in percentages, of the subjects assigned, compared to the geographic subject metadata created for the examined records.
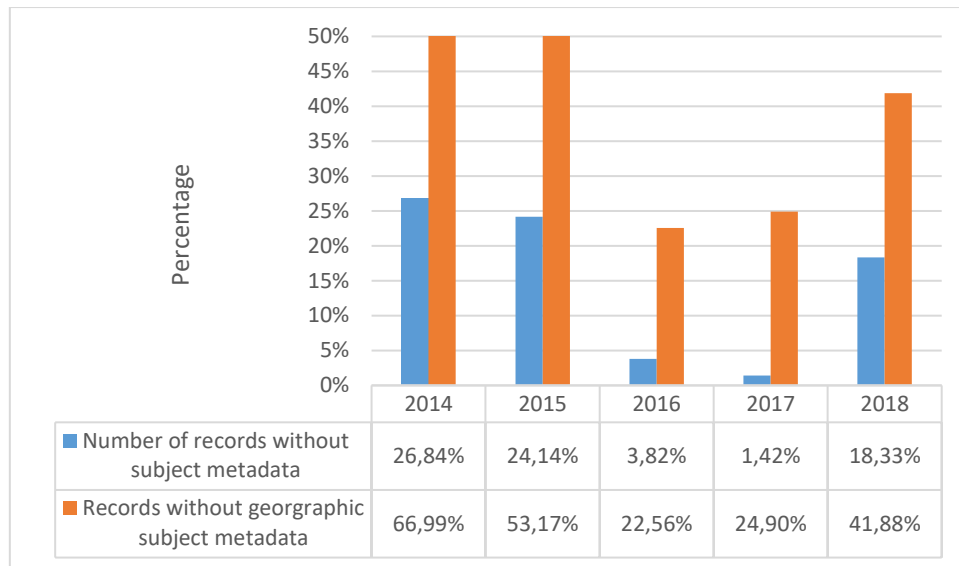
	2014	2015	2016	2017	2018
■ Number of records without subject metadata	26,84%	24,14%	3,82%	1,42%	18,33%
■ Records without georgraphic subject metadata	66,99%	53,17%	22,56%	24,90%	41,88%

Figure 6.12: Percentages of subject metadata usage on ETD records

The figure shows a decrease in the number of records without subjects, from 26.84% in the year 2014 to 1.42% in the year 2017. A slight progression of subject use was noticed in the year 2018, when the percentage was more at 18.33%. Consistent with the decreasing trend in general subject usage, a decrease occurred in the number of records not assigned with geographic subject metadata – from 66.99% in 2014 to 22.56% in 2016. However, inconsistencies in trends in geographic subject metadata use could be seen, with fluctuations displayed in the percentages: 24.90% in the year 2017, compared to 2018, during which 41.88% records were not assigned geographic subject metadata. Although most records created between 2014 and 2018 were assigned subjects, geographic subjects were not assigned in most cases where they were required.

6.4.3   Role of controlled vocabulary

An exploration of the nature of the vocabulary used for subject metadata was enabled by examining the full ETD records on the library' IRs. In a qualitative content analysis approach, trends can be identified from the identified categories of the analysed text (Krippendorff 2013:1). The metadata records on the NETD database link to the institutions hosting the original record (Webley, Chipeperekwa & Suleman 2011). As such, full metadata was examined from records on the IRs to observe the trends.
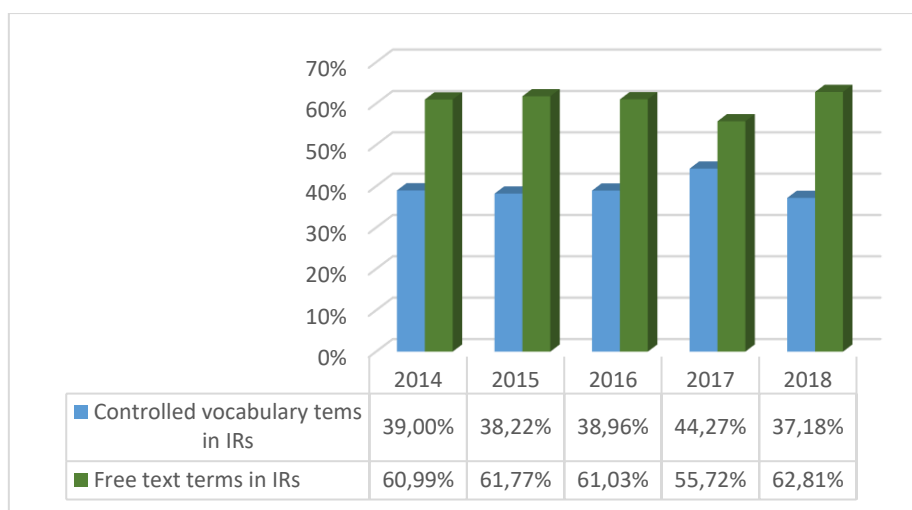
Globally acceptable vocabulary standards and subject metadata creation schema are readily available to compare the geographic subject metadata use in ETD records and assist the interpretation of data. The nature of the vocabulary was observed against a common standard, the LCSH. The commonly used Dublin Core metadata schema (DC) and the Metadata Objects Description Schema (MODS) recommend the use of controlled vocabulary for subject descriptions.

The questionnaire findings indicated that most libraries participating in this study used the Dublin Core description schema and the LCSH. This was confirmed by the explanation of the processes during the interviews. Table 6.27 shows the total subjects assigned for ETDs and their different forms, whether they were derived from standard vocabularies or assigned as free-text subjects. The percentages of the different subject metadata forms showed that, between the years 2014 to 2018 there was more use of uncontrolled subject metadata, compared the controlled types.

Table 6.27: Subject metadata elements in IRs

	Examined subjects in IRs	**Controlled vocabulary terms in IRs**	**Uncontrolled vocabulary terms in IRs**
2014	2182	851 (39.00%)	1331 (60.99%)
2015	3197	1222 (38.22%)	1975 (61.77%)
2016	1668	650 (38.96%)	1018 ((61.03%)
2017	1992	882 (44.27%)	1110 (55.72%)
2018	1721	640 (37.18%)	1081 (62.81%)

Figure 6.13 (p. 246) displays the percentages of controlled and uncontrolled subject terms used to describe the contents of ETDs.

	2014	2015	2016	2017	2018
■ Controlled vocabulary tems in IRs	39,00%	38,22%	38,96%	44,27%	37,18%
■ Free text terms in IRs	60,99%	61,77%	61,03%	55,72%	62,81%

Figure 6.13: Subject metadata elements in IRs

Between the years 2014 to 2018, both controlled and uncontrolled terms were used to represent the subject content of the ETDs by different university libraries in South Africa. More use of uncontrolled terms was evident, with percentages of 60.99% in the year 2014 and 62.81% in the year 2018. Percentages of free-text elements used in 2015 and 2016 were: 61.77% and 61.03% respectively, showing that free-text or uncontrolled terms were used the most in those years. The 55.72% of free-text terms used in 2017 was the lowest in the five-year period, which suggest that there was more use of controlled vocabulary in 2017, compared to other years.

Compared to fee-text terms, controlled vocabulary was used with 39.0% less in the year 2015. The situation improved in 2017, when 44.27% controlled terms were assigned. There was an observed decrease again in the year 2018, when only 37.18% of the assigned subject terms were in controlled form. The inconsistency in the use of controlled vocabulary for ETDs subject metadata is evident from the presented results.

Overall, the data analysis of the content analysis supported the explanations that subjects have been used to describe ETDs in South African libraries between 2014 and 2018 and that geographic subject metadata was assigned to the records. There was fewer instances of subjects not being assigned, compared to records assigned with subject metadata. However, records without geographic subject metadata were more, compared to those assigned with geographic descriptions, which was contrary to the questionnaire and the interviews findings.

It is commonly expected that titles with place names should be assigned with geographic subject metadata. Consistency with the application of schema rules and controlled vocabularies varied during the period covered in this study. Comparatively, free-text or uncontrolled subject terms were the more used than the terms from controlled vocabularies.

6.4.4   Subject analysis basis

Trends and patterns may lead to interesting content analysis observations (Krippendorff 2013:59). One of the inferred observations from the content analysis data has been how they reveal the decision making patterns of the subject metadata creators. Krippendorff (2013:55) affirms that content analysis data can be applied to draw inferences from the actors' patterns of behaviours and decision-making.

The continued use of geographic subject metadata suggested confirmation of the perspectives from the questionnaire and the interview findings that geographic subject metadata was perceived as a valuable practice for ETD subject description. The use of controlled vocabulary for ETD geographic subject metadata indicated its relevance for use with the digital formats like ETDs. However, the ETD subjects were described more in free-text terms. This was consistent with the findings of a recent study conducted by Terra et al (2021), which confirms more use of keywords for ETD subject descriptions. However, the authors support the enhancement of subject metadata creation practices by using controlled vocabulary.

The identified nature of assigned subject metadata showed that the theoretical understanding of subject analysis was less inclined towards the use of vocabulary control standards. This was despite the indication of the findings of the other phases of this study in which the metadata creators considered this practice as important. Further studies may explain this contradiction and learn more about factors that limit the use of controlled vocabulary, despite the expressed preference.

General observations during the studied period showed the usage of subjects, particularly geographic subject metadata. However, trends showed the continuous prevalent use of free-text subject elements and geographic subject metadata assigned in uncontrolled forms. Inconsistencies of practices among institutions need to be explored to understand the impact of contextual factors. Evaluation of accuracy and other quality measures can be conducted in future metadata studies by using the content analysis method.

6.5   Chapter conclusion

The data from the questionnaire, interviews and content analysis phases produced important information that highlighted the existing practices in and perceptions on subject analysis and ETD geographic subject metadata. The questionnaire results were analysed quantitatively by using descriptive statistics. The questionnaire results of metadata procedures indicated a significant use of geographic subject metadata. The findings pointed out that professional library staff were chiefly responsible for metadata creation. However, the limited use of controlled vocabulary use was noted, contrary to the positive views held about its value. The majority of perceptions on the adequacy of the existing geographic subject metadata were negative. Subject analysis approaches aligned more to focus being placed on the content of the ETDs, followed by focus on specific subject areas. Insights into the effects of the subject analysis theories knowledge were drawn from the shared perspectives.

Thematic analysis was used to conduct a quantitative analysis of the interview data. The results, which indicated general support for the questionnaire findings, offered more clarity to the questionnaire findings. Probing on the approaches used revealed practices that are aligned to the theoretical conceptions discussed in literature. A key clarification was given to the general indication of high preference for keywords, in addition the use of controlled vocabulary. It was found that a hybrid approach of using both approaches was suitable to offset the weakness of one with the other. Contextual factors were also pointed out to limit adequate description practices. The explanation of the processes followed confirmed ETD content and subject domains as the main focal points for subject analysis. However, the metadata creators could not explicitly communicate the impact of the theories.

The data from the two instruments also revealed more perceptions in favour of controlled vocabulary. However, notable inconsistency on this aspect was revealed between the data from the first two phases and the content analysis. The latter revealed more use of other forms of subject headings and the use keywords, compared to more positive perspectives shown regarding the use of geographic subject metadata and controlled vocabulary.

The next chapter will discuss the proposed model for theory and practice for subject analysis and geographic subject metadata creation.

CHAPTER 7: INTERPRETATION AND DISCUSSION OF THE FINDINGS

7.1 Introduction

This chapter presents the analysis and interpretation of the findings of this study. As previously indicated, the findings of the three phases involved in the study were integrated in the interpretation phase. The rationale for the mixed methods study and the background information of the participants provided important information that served as the basis for the interpretations.

The findings were interpreted in line with the aim of this study, which was to investigate the approaches that are being followed during subject analysis and how subject analysis theories are applied to facilitate geographic subject creation for ETDs. Furthermore, the findings were examined according to the rationale of completeness, complementarity, expansion and diversity, discussed in detail in Chapter 5. Contradictory evidence was also examined. The varied lessons learned from the findings addressed the research problem – i.e. involving subject analysis, ETDs geographic subject metadata creation and, in particular, the effect of subject analysis theoretical principles – from different perspectives.

The important contribution of this study is that it provides new insight into the effectiveness of subject analysis and the relevance of its theories for subject metadata creation in the ETD context. It is important to describe the scientific information contained in ETDs appropriately and relevantly to facilitate discovery, considering the different information provisions contexts and the changing information needs over time. Conclusions are drawn from the integrated quantitative and qualitative findings from the three phases of this study to address the research objectives.

The research objectives were translated into the following research questions:

1. To what extent and in what ways is geographic subject metadata used for description of ETDs in South African university libraries?
2. What are the current approaches being followed to create geographic subject metadata for ETDs in South African university libraries?
3. How and why do the current analysis approaches being followed during geographic metadata creation for ETDs in South African university libraries affect the process?
4. How can the findings of this study be applied to develop a model for subject analysis and creation of geographic subject metadata for ETDs in South African university libraries?

The interpretations of the findings were organised in the following three key topics that addressed the research questions: (i) the nature and extent of the use of geographic subject metadata for ETDs; (ii) current approaches to ETD geographic subject metadata creation; and (iii) perspectives on the significance of current analysis approaches on the geographic subject metadata creation process. Based on the findings, a model is presented and recommended to promote effective subject analysis and ETD geographic subject metadata creation.

The existing subject analysis theories helped to explain the findings. The conclusions and recommendations were made at the end of each section. Concluding summaries of the interpretations and the recommendations are presented at the end of the chapter. The background information and the rationale for using the mixed methods approach are discussed to provide a basis for the interpretation of the findings.

7.2   Rationale for mixed methods study

The reviewed literature recognised subject analysis as a complex process often studied from the practitioners' perspectives. In contrast, this study aimed at investigating this process from both the theoretical and practical perspective. Additionally, studying the actual ETD records provided information on the application of theoretical principles and outputs in the form of the ETD geographic subject metadata.

The mixed methods approach was found suitable to address the research in depth. – As highlighted by Leedy and Ormrod (2013:259), quantitative and qualitative approaches can be combined to address a complex research problem. The mixed methods approach provided the advantage of gaining from the strength of the quantitative and qualitative methods to reach a complete understanding of the problem under investigation. Quantitative and qualitative approaches were combined in this mixed methods study to provide an in-depth understanding of the practice of subject analysis, as applied to ETD geographic subject metadata creation. It would have not been possible to achieve that by relying on one approach. Understanding the topic in a broader sense helped to identify the possible ways of improving the process of subject analysis and the geographic subject metadata for ETDs.

Furthermore, the methodology was considered relevant because of the varied nature of the research questions, which could be answered more meaningfully through the use of both quantitative and qualitative methods. The research questions of this study involved enquiries on varied aspects regarding procedures that are performed; how they are performed; and the effect of the practices of the subject metadata creators. Addressing the research questions properly required the use of different approaches – quantitative and qualitative.

The questionnaire survey phase of the study produced the initial findings that gave a broad outlook on the research problem. Several issues that needed clarity or additional information to clear ambiguities were identified and addressed from the qualitative findings. The interpretations of the findings of the content analysis was partly influenced by the findings of both the questionnaire survey and the interviews. Moreover, the findings of the content provided additional information on the extent and nature of usage of geographic subject metadata by revealing patterns and trends from the existing ETD records. Additionally, the findings confirmed the findings of the questionnaire and the interview phases, so as to provide a better understanding of the effect of the subject analysis approaches.

The study was conducted in three phases. The core quantitative and qualitative phases were carried out in an explanatory sequential design, with an additional quantitative and qualitative content analysis phase to supplement the qualitative explanations. The analysis of the data from the content analysis phase was influenced by the results of the questionnaires and the interviews to gather more complete insights into the study problem. The findings of the three methods were integrated for interpretation. The discussions in Chapters 1–6 provided a detailed theoretical and contextual background for the discussion of the findings of the study.

7.3   Framework for the interpretations

The pragmatist research paradigm rendered the overarching theoretical framework for the interpretations. Furthermore, the literature review revealed different theoretical conceptions of subject analysis that unfolded over time. Different views from literature – e.g. those of Hjorland (2021, 2017), Albrachtsen (2015, 1993), Tennis (2005) and Hjorland & Albrachtsen (1995) – helped to explain the process of subject analysis from a theoretical viewpoint and to provide a basis for improving the understanding of its theoretical principles. Knowledge of the conceptual models may help metadata creators to understand the theory informing the process of subject analysis.

An existing model of conceptions of subject analysis was used to provide a framework for the interpretation of the research findings – particularly for the interpretation of the perceptions of the appropriate focus to be maintained during subject analysis. This model is discussed in detail in Chapter 4. The adopted model of conceptions of subject analysis presented different theoretical approaches that informed the practice of subject analysis. It outlined the possible approaches for deciding on and identifying what to concentrate or focus on during subject analysis.

According to the mentioned model, the process of subject analysis begins with determining what an information resource deals with and ends when the subject representations are assigned.

Literature published over time, e.g. the work of Hjorland (2021), Albrachtsen (2015), Fourie (2008) and Tennis (2005), discusses the model and its advances, indicating different orientations on how to determine the subject content of an information resource. While this framework served as the key guide for the discussions in this study, all the information gathered informed the interpretations.

In addition to the theoretical conceptions of subject analysis, this study explored how subject analysis and subject metadata creation are experienced and perceived in practice. The theoretical perspectives on subject analysis provided a basis to examine how the process is conducted to create geographic subject metadata for ETDs in the context of South African university libraries. Theory is fundamental to the procedures and processes of subject metadata creation, but its application is not always explicit in practice. The significance of the findings of this study and the conclusions will be discussed in relation to the model of the influence of theory knowledge on the practice of subject analysis.

Furthermore, subject metadata is viewed as a boundary object that could facilitate ETD geographic subject content to be used across different contexts. This view informed the focus of the interpretations.

7.4   Background information

The background information and other characteristics of the study participants indicated that it was mostly professional librarians who were responsible for ETD subject metadata creation. Most participants had a cataloguing background, which was consistent with the literature (Xie & Matusiak 2016:130; Rubin 2010:157; Greenberg 2005:20) that identifies the association of metadata creation and cataloguing. Representation of participants in the questionnaire survey and the interviews was fair in terms of the university types and the levels of experience in ETD subject metadata creation. The questionnaire responses revealed that almost all the respondents specified their positions as professional librarians. All the interview participants further confirmed this position.

The experience of the metadata creators who participated in this study varied, with most being in the beginning and intermediate phases of experience levels in ETD metadata creation roles. Less than one third of experienced metadata creators with over fifteen years' experience were involved in ETD subject metadata creation. The questionnaire responses revealed that the respondents with a lengthy experience (15 years and longer) in general subject metadata creation were more than a third of the total respondents. This suggested that there was a reasonable number of respondents who, in addition to the specific role of ETD metadata creation, brought into their new role an extensive experience in general subject metadata creation.

The questionnaire and the interviews provided information that aided the understanding of the background of the metadata creators responsible for the creation of ETD subject metadata. The interviews provided more information as to when and where the metadata creators studied the theories of subject analysis and their different experiences. The researcher's assumption was that the different backgrounds had an influence on the kind of information that the metadata creators shared about ETD subject metadata creation. The questionnaire survey findings provided quantitative data on the characteristics of metadata creators in South African university libraries and also linked them to their roles. Theory knowledge will assist the novice staff with the foundation to build their subject metadata creation skills. There is a need for the experienced staff to compile properly outlined procedures and guidelines to help to link the theory and practice for ETD geographic subject metadata creation processes and procedures.

In addition, the findings revealed that, despite systems such as DSpace being used among the participating libraries and provided the functions that facilitate different roles for creators, editors and archiving, most interview participants confirmed that all these activities were performed by a single person. This finding explained the indication from the quantitative findings where only one respondent selected the metadata creation quality control role option.

7.5   Use of geographic subject metadata for ETDs

The investigation into metadata procedures was conducted to determine the extent and nature of the current usage of geographic subject metadata for description of ETDs in South African university library ETD repositories. The findings mainly addressed the first research question, which sought to investigate how the geographic subject representations were applied for the creation of the ETD metadata. The question was answered by interpreting the questionnaire findings, the qualitative findings from the interviews phase and the quantitative findings from the content analysis phase.

7.5.1   Extent of use of geographic subject metadata for ETDs

The questionnaire findings revealed that subjects, including the geographic subject type, were used to provide descriptions for ETD content. An indication of participation by more than two thirds of library metadata creators, almost a third of the authors and other participants in assigning subject metadata was an indication of an activity in creation of metadata of subject type. The qualitative findings of the interviews consistently provided explanations from participants regarding the activities being performed during ETD subject metadata creation, which confirmed the use of geographic subject metadata. Furthermore, the information helped to explain the metadata creators' decision-making process for the different activities.

The most common approach to subject metadata creation was shown as creating geographic subdivisions for existing subject metadata by using terms derived from the ETD content. This was shown to be the function that was performed the most. Almost a quarter of the questionnaire respondents indicated that they were always adding geographic terms as subdivisions to the existing subject headings. In addition, almost half of the respondents often used this approach. The dominance of this activity suggested that much of geographic subject metadata was created as an enrichment to existing subject metadata. This was an unexpected finding, because the professional subject metadata creators were dominantly involved in subject metadata creation and, therefore, it was expected that they participated more in creating original subject metadata.

The expectation was also partly based on the finding that other functions, like quality control, that are commonly assigned to professional metadata creators, were found to be less prevalent.

In contrast, the quantitative findings showed that new, main geographic subject metadata was also created by using terms from different sources. One of the common origins of these terms is the resource itself, where metadata creators derive subject representations.

Almost half of the participants used this approach often and sometimes never. In addition, other sources outside the resource itself were seen as being used by an insignificant number to derive subject description terms. However, this approach appeared to be less common compared to the former. This finding suggests that, comparing the use of terms from the resource itself to usage of other sources, a stronger focus was on the ETD content when creating geographic subject metadata. The findings were considered as additional evidence of the use of geographic subject metadata.

The qualitative findings provided clarity by indicating the varied experiences and perspectives that explained the different approaches. Examples are found in the following statements from one of the interview participants, who explained that they worked with the authors' keywords and that they only transcribed what the author had sent to them. The participant continued to indicate that they encountered challenges, e.g. at times they felt that they could add more subject metadata, but they were limited by poor knowledge of the keywords and terminology of different disciplines. This is why they found it relevant to transcribe only what the author provided. This approach suggested that less subject analysis was performed by some metadata creators, because the submitted subject representation were accepted without further modifications. Based on these findings, it was necessary to investigate whether the prominent practice of creating subdivisions, shown in the findings, was performed as retrospective enhancement for existing ETD descriptive records.

Different explanations were found in literature. For example, Jensen and Carr (2020), Zavalina, Zavalin and Miksa (2016), and Tarver et al (2015:37) pointed to different reasons, including changes in practices; standards; and the goals of the institutions, which often necessitated retrospective enhancement of descriptive metadata.

The addition of subdivisions is important to enhance the specificity of the metadata. The Dublin Core schema and other standards, like the LCSH, recommend specific descriptions of geographic place names. However, it is not expected to be the only form of subject metadata that is created by the professional library staff. The professional library staff can improve on metadata produced from other sources by creating subdivisions to provide more specific subject descriptions. However, professional metadata creators are expected to be involved in performing subject analysis to provide primary subject representation with added subdivisions when necessary.

In addition, the experiences and perspectives shared in the interviews showed that different factors impacted on the perceptions of subject description and its processes, including the influence in instances where geographic subject metadata was not assigned. Despite the value ascribed to geographic subject metadata, they tend not to be assigned at all times. It was assumed that this situation may partly result from the explanations given by the metadata creators, who tended to avoid what was perceived as complex activities and to assign the types of subject metadata considered as easier. The perceived complex nature of geographic subject metadata creation affects the way the activity is performed. As a result, the current ETD subject metadata in ETD repositories may not be considered comprehensive in terms of coverage of ETD content, including the geographic subject descriptions.

The examination of existing records in the different ETD repositories, through the content analysis method, helped to show how the conception of subject analysis revealed in the primary quantitative and qualitative findings was made practical in the creation of the geographic subject metadata. The findings of the content analysis were integrated with the survey questionnaire and the interviews findings to reflect the existing state of subject metadata during the identified study period.

In addressing the first research question, the observed ETD subject metadata records from the university repositories revealed the actual practices, and the presence and extent of use, in the five years period, 2014 to 2018. The findings of the content analysis helped to confirm the use of this type of subject descriptions from the ETD descriptive records.

The evidence found in terms of the extent of use was that an improvement occurred and more subject metadata was created between 2014 and 2018, while the geographic subject metadata also increased. An increase of more than one quarter in the assigned subjects was experienced between 2014 and 2018. Further studies for the period after the year 2018 should help to investigate if the upward trajectory is maintained consistently.

Generally, the records analysis findings revealed that geographic subject metadata was created over the five-year period and that the geographic subject metadata existed on the ETD descriptive records contained in the different university libraries' institutional repositories.

The findings of the content analysis confirmed the results of the first two phases of this study, where the metadata creators indicated that they participated in subject metadata creation. In addition, the content analysis findings offered the best form of evidence of usage extent by using an unobtrusive method of study to help reveal concealed realities. The findings also supported the value that the participants attached to the geographic subject metadata creation function. The significant levels of use and the positive perception of geographic subject metadata highlighted the importance of these descriptive representations to enhance access to ETD content. Despite the evident increase in the use of geographic subject metadata over the years, the number of eligible records (records with place names in their titles) to geographic subject metadata was not assigned, remained concerning. The content analysis results provided a different approach to show usage and its extent.

Overall, the qualitative findings supported the quantitative phase outcomes, where all metadata creators confirmed that they used geographic subject metadata to describe ETDs.

Clarity was also provided once this was done, indicating that they created both primary descriptions and added subdivisions when necessary. The quantitative findings yielded quantitative measures of the responses, while the qualitative data from the interviews supported the in-depth analysis of the behaviour and attitudes that impacts on the current practices. Contextual differences in terms of policies, standards, staff expertise levels, available staff and time formed part of the key explanations for the different perceptions and practices.

The conclusion drawn from the findings is that the use of geographic subject metadata and the extent of use was affirmed. The findings aligned with the recommendations of the IFLA Working Group on Subject Access (International Federation of Library Associations and Institutions (IFLA). Working Group on Guidelines for Subject Access by National Bibliographic Agencies 2011), which show the importance of geographic subject metadata to facilitate access to information resources.

The question on the use and extent of geographic subject metadata was answered satisfactorily through the combination of the quantitative findings and the clarifications from the qualitative findings. The different instruments used in this study provided important information on different geographic subject metadata creation aspects.

The articulation of subject metadata creation among the participants was considered relevant, although different practices were observed. The responses were relevant to address the research question, but different practices were confirmed in the findings. More studies and engagements will assist in the endeavour to improve the effectiveness of the ETD subject metadata and to keep up with the trends.

The next section discusses the different practices regarding the kind of terms that are used and what is perceived as suitable ways to describe ETD content. These aspects are discussed in detail in the next section to indicate the nature of use of geographic subject metadata.

7.5.2 Nature of usage of subject metadata for ETDs

Different meanings in the findings on the use of subject metadata. The questionnaire survey, the interviews and the content analysis findings were used to draw interpretations on who contributed to or assigned subject metadata; the type of terms used to represent ETDs subjects; the suitable categories of metadata to enhance the discovery of subject content; and the sources in which suitable metadata can be found.

The perspectives on suitable categories of subject metadata, in relation to their sources of origin, were identified from the findings of the first quantitative phase. Subject metadata created by the library staff was found to be the most suitable approach by more than two thirds of the questionnaire respondents. The participants were more agreeable about the suitability of this approach, which was demonstrated by the highest ratings for this approach as always being suitable and general positive views on it. This result was consistent with the qualitative findings, which similarly showed that ETD metadata creators in the South African university libraries were mostly professional librarians.

This finding suggested that the creation of ETD subject metadata was mainly considered as a professional job and that subject analysis based on professional principles was necessary. A theoretical foundations was regarded as important for professional practices and this study examined how such a foundation could assist in the creation of appropriate subject metadata for ETDs.

The second highly preferred source of origin for subject metadata was the keywords in full-text or analysis of the entire ETD content. The keywords supplied by the authors, who were the producers on the works, were naturally expected to be the best form of descriptions, although it was not always the case. Less than a third of the respondents considered keywords as always being suitable and some considered them as rarely suitable. This finding suggested that author keywords were considered as important suitable forms of subject representation. However, although they were viewed as important, they were not always found to be describing the ETD content, with less than one third of the respondents viewing them as rarely suitable.

This finding was consistent to perspectives in literature (Tera et al 2021:2; Husic 2014:2) regarding the importance and common use of author submitted keywords, while recognising that they were often characterised by shortcomings of inconsistencies, inaccuracies and ambiguities.

The questionnaire survey findings indicated that tags contributed by users were considered the least suitable. A possible reason was found in the interview explanations, where uncertainty was expressed about the relevance of this approach. The reviewed literature (Mayernik 2020; Hjorland 2017a; Marshall 1998) consistently indicated that user-contributed tags may be appropriate to broaden the perspective of different uses of an information resource, although their relevance remained questionable. More studies may help metadata creators to reduce existing uncertainties regarding this approach.

The questionnaire findings could not establish why there were insignificant differences in preference of some the suitable categories of the subject metadata. There were common variations between what was considered always suitable and sometimes suitable for the different types of subject metadata categories.

However, the evident lack of clear priority for specific practices and the sources where the subjects are derived can be explained from the suggestions made during the interviews that the individual metadata creators' judgement prevailed regarding the adopted practices. Some participants explicitly stated that their libraries did not have a guiding policy and that the metadata creators used their own judgement to determine which subject metadata to assign. Consistently, the quantitative findings showed that almost two thirds of the participants indicated that there were no specific policy guidelines for geographic subject metadata in their institutions. This suggested that there was no intentional disregard for implementing procedures on suitable subject metadata, but that such guidance was lacking.

Furthermore, the qualitative findings clarified the different ways used to analyse the ETD subject content for geographic coverage. Some of the interview participants stated that they always verified the areas or scope covered by the research and assigned the geographic subject metadata. Some participants only explained the procedures they followed, without indicating how geographic coverage was accommodated. It was not conclusive that all subject metadata creators used the appropriate sources to determine the necessary geographic subject representations. Therefore, it was assumed that this could be one of the explanations for cases in which geographic subject metadata were not assigned, as shown by the content analysis findings.

Only more than half of the participants showed the kind of subject terms that they used to be in controlled form. Differently, all the interview participants indicated that they used existing subject headings standards. Inconsistencies of practice could be found in individual institutions in terms of applying standards. Despite these inconsistencies are not being desirable, it is not unusual that they may occur, because the existing standards for basic functional metadata creation provided a framework for the procedures of subject metadata creation, but the nature of application remained the responsibility of those applying them and, therefore, it was influenced by their contextual circumstances. The findings from the different phases of the study showed that standards were applied differently and possibly inconsistently within individual institutions. This situation carried the possibility of affecting the type of the geographic subject metadata that was created.

It is concluded that the discussions in Sections 7.5.1–7.5.2 demonstrated how the findings of this study provided answers to the question on the extent and nature of usage of the ETD subject metadata. Previous studies in the context of South African libraries do not address the use of geographic subject metadata or ETDs subject metadata in general. This study provides the practical and theoretical context in a single study by using the mixed methods for an in-depth investigation.

The three instruments used in this study collectively provided in-depth answers to support the use of subject metadata as a form of ETD content description. Despite the noted differences in approaches, it could be concluded that subject metadata was significantly used to describe the ETD subject metadata. Consistent with the findings of this study, Steele and Sump-Crethar (2016) established that ETD records in university institutional repositories used subjects as a field of description. The study provided empirical evidence from the content analysis findings in support of the continued use of subject metadata to describe ETD content. The content analysis findings revealed that averagely, almost 60% of subject metadata usage was established, with geographic subject metadata use at less than 50%. The next section discusses the subject analysis and metadata creation approaches used specifically for geographic subject metadata.

7.6   ETD geographic subject metadata creation approaches

The second question of this study aimed at examining the approaches followed to create geographic subject metadata for ETDs. The findings are examined to identify the experiences and perspectives related to subject analysis and geographic subject metadata creation. Additionally, the information will be used to explain the theoretical basis of subject analysis. The findings on how geographic subject metadata for ETDs were determined reflect the perspectives of the practitioners involved in geographic subject metadata creation. Key issues that needed clarity were identified from the questionnaire findings, clarified through the interview findings, while the content analysis findings provided further confirmation. The findings of the current practices with regards to the kind of terms used, the perceived adequacy and suitability of the assigned geographic subject metadata are discussed. In addition, findings regarding quality control are also discussed.

The questionnaire findings revealed that the approaches in metadata creation varied. The qualitative findings confirmed that the responsibility for subject metadata creation mostly lied with professionally trained library staff, although their levels of experience and perspectives varied. However, there dominant practices found among the different universities that participated in the study. These are discussed in the next sections.

It is generally observed that the extent and nature of use of geographic subject metadata need to be improved to enable important scientific information about specific geographic areas to be collated in retrieval from the ETD repositories.

7.6.1 Standards used for ETD subject metadata

Different vocabulary standards and schema elements are used to describe the subjects of ETDs. Reliable and interoperable subject metadata requires standardised approaches within and across different platforms. As discussed in the next subsections, issues of controlled vocabulary and the use of description schema elements emerged as important for geographic subject metadata creation.

7.6.1.1 Controlled vocabulary

The theoretical principles of subject metadata guide the use vocabularies when assigning subject representations. The kind of vocabulary used influences the decisions taken about the terms that are assigned as subject representations (Hjorland 2017a:56; Mai 2005:600; Lancaster 2003:9; Hjorland 1997:39; Albrachtsen 1993:219). Assigning subject representations is the concluding step of the subject analysis process. The representations can be assigned either in the form of controlled subject headings or uncontrolled terms.

Controlled vocabulary was one of the dominant themes that emerged from the questionnaire findings and the interview responses. This could be related to the background of the metadata creators, which was found to be influenced by their cataloguing experience. The questionnaire findings identified LCSH as the most common standard used in university libraries participating in the study. The quantitative findings showed that the LCSH was used by almost all the participating libraries. In addition, only a few participants confirmed the use of other geographic specific standards, e.g. the National Geospatial-Intelligence Agency database (Geonet names server).

The qualitative findings confirmed the use of the LCSH and its preference by all the participants. The purpose of using controlled vocabularies is to maintain consistency and accuracy in allocating descriptive subject terms of information resources.

The results confirmed the conclusions drawn from a survey conducted by Steele and Sump-Crethar (2016), in that controlled vocabulary is important for subject metadata.

However, it remained unclear why the questionnaire findings showed that the uncontrolled metadata schema element, for example "dc:subject", was used the most. It was, therefore, necessary to confirm if the standards were applied in practice. In response to questions on the use of standards, some of the interview participants indicated the challenges that they experience in finding standardised terms for South African place names from the Library of Congress Subject Headings and Library of Congress Name Authorities (LCNAF). They understood this sometime being caused by delays in including new place names in the controlled vocabulary and authority lists. Additionally, the problems raised in the interviews concerned time, staffing, novices who are inexperienced in metadata creation and the complexity of assigning geographic subject metadata. The issues implied that contextual differences may affect the use of controlled vocabulary. Furthermore, the content analysis findings revealed that, in almost all the years between 2014 and 2018, uncontrolled subject metadata elements were more than double the number of the controlled subject metadata.

Consistent with the quantitative findings, there was a high indication of the use of standards observed from the experiences and perspectives shared in the interviews. It was perceived that controlled vocabulary leads to effective ETD geographic subject metadata. The inconsistencies noted from the quantitative questionnaire survey findings, pointing to more use of free text subject descriptions schema elements, "dc:subject", needed further clarity. This did not relate to the perceptions in the qualitative interview findings about the use of standardised vocabularies.

Despite the fact that controlled vocabulary was not significantly used at the time of the study, almost all the participants viewed it as critically important. It may be necessary to investigate further why importance is attached to controlled vocabulary, while it is not applied in practice. This will provide further clarity to the conflicting views about the use of controlled vocabulary found in literature.

The advantages of both controlled and uncontrolled vocabulary are highlighted by, among others, Zavalina (2014:87) and McCutcheon (2011:67). Furthermore, Maurer and Shakeri (2016:214) point out that there is sufficient proof of support for the blended or mixed use of keywords and controlled vocabulary. The preference and recommendations for use of controlled vocabulary are considered rational by the researcher, in that they provide for standardised practices that facilitate the use of metadata across different disciplines and different systems in a controlled and meaningful manner. A theoretical understanding could assist decisions on the choices of a specific kind of vocabulary or how to apply the blending.

A different finding from the reviewed institutional repositories records showed flexibility in the use of both forms, with dominance of free text subject descriptions. According to the analysed records, the extent of usage of controlled vocabulary subject description metadata schema elements (e.g. dc:subject:LCSH) in practice, for the period 2014 to 2018, was found as being outweighed by the use of uncontrolled elements. In the last year (2018), the observation was that more than half of the subject metadata, including the geographic representations, were assigned to the uncontrolled subject fields.

In concluding discussions of the findings in Section 7.6, it was observed that, despite the indicated high preference for controlled vocabulary, there were possible hindering this approach in practice. Possible reasons were mentioned in the previous paragraphs. Moreover, it was concerning that the existing subject keywords created from the full text or keywords submitted by the authors were considered by less than half of the study participants as always suitable. It is recommended that solutions should be sought to address the limiting factors that impede on the implementation of what is perceived by the metadata creators as best practice.

Furthermore, in the overall findings, the use of both controlled and free text subject descriptions was found to be the best approach to represent ETD geographic subject metadata. Zavalina (2014) made similar observations in a study that compared the use of free-text and controlled vocabularies in three institutional digital repositories. Another important finding is that the library metadata creators were chiefly responsible for ETD subject metadata creation.

Further studies are recommended to investigate how to improve guiding mechanisms for ETD authors to help them create appropriate subject representations, as uncertainty was established about following this approach as the best way to provide subject descriptions. Additionally, the library staff should relate the revealed perceptions of inclination towards standardisation with the choice of descriptive schema elements. Although the use of some metadata element may not be mandatory, the use of controlled vocabulary is recommended to enhance the discovery of ETD content. This recommended approach should be emphasised for metadata to meet the required levels of appropriateness in terms of geographic access.

7.6.1.2 Descriptive metadata schema elements

The questionnaire findings showed inconsistencies in the form of descriptive schema elements that were being populated and the data value standards used. The decisions taken about the descriptive schema elements to be used for subject representations and the vocabulary standards like the LCSH were interrelated. The findings identified Dublin Core as the most used metadata description schema for ETD repositories for the libraries that participated in this study. The schema caters for the use of controlled and uncontrolled subject elements, e.g. "*Subject: (LCSH)" or "Subject"* respectively. It was, however, unexpected that the uncontrolled "dc:subject" would be the commonly used description element. This was in contradiction to the finding where the metadata creators indicated that they mostly used terms from controlled vocabularies. More than half of the respondents indicated that they followed the controlled vocabulary approach to represent ETD subjects, as opposed to the keywords obtained from ETD texts.

The qualitative findings indicated the likelihood that the descriptive schema elements commonly used in an institution influenced decisions on what type of subject metadata to assign for ETDs. It was indicated in the interviews that the metadata creators focused on the fields that they believed should be populated. Confirmation from the content analysis findings showed a stronger focus on the uncontrolled subject elements fields. The different schema commonly recommend the use of mandatory and recommended fields to improve accessibility of the described resources.

Maintaining flexibility in the use of both controlled and uncontrolled vocabulary is recommended to enhance geographic subject representations for ETDs. However, recommended use of specific elements, e.g. the elements that are used for controlled data values, should be encouraged. This recommendation for use of controlled vocabulary aligns with the findings of the studies conducted in other countries, e.g., the study conducted by Maurer and Shakeri (2016:214). In addition, the metadata schema should not be used in isolation, but in line with relevant theoretical frameworks for subject description practices, e.g. the quality principle of consistency. Relevant subject analysis theoretical frameworks should be adopted to enhance the usefulness of schema elements in ETD geographic subject metadata creation.

7.6.2   Perspectives on the adequacy of geographic subject metadata

The findings of the questionnaire indicated that the metadata creators' perceptions were that the level of adequacy of geographic subject metadata attached to the different universities library's ETD records was low. This was the finding, despite the different kinds of terms that were indicated as being used. In alignment with these findings, McCutcheon (2011), Thompson et al (2019) and Flynn and Ahrberg (2020) discuss the general inadequacies in ETD subject metadata. More than half of the participants' responses considered adequacy to be below the high adequacy level. The interviews supported these findings by indicating the challenges in geographic subject metadata creation and the need for improvement and control. A number of statements explained the challenges. One participant said that there were times they felt like adding more terms, because they regard some missing keywords as significant. This finding is similar to the that involved in the study conducted by Steele and Sump-Crethar (2016:24), where metadata creators felt they could do things differently concerning the ETD metadata levels of adequacy, if they had the opportunity to do so. The second statement showed perceptions that there was always room for improvement, although it was not always easy, because there was no supervision to direct the metadata creators.

These explanations suggested that contextual factors, like poor clarity on the correct procedures to be followed and poor guidance on how to maintain appropriate descriptions, were viewed as limitations that affected the level of adequacy of the

subject descriptions. The continued use of different standards and practices, which are contextual, determines quality of metadata creation (Jensen & Carr 2020:568). However, positive sentiments expressed by the metadata creators were that the situation could be improved. Interventions in the form of clear policies or guidelines and quality control are possible intervention that may improve the existing situation.

The findings of the content analysis revealed additional information, which showed that, for the period covered in this study, the number of records without subject were lower, whereas the number of records not allocated with geographic subject metadata were considerably higher. The findings suggested that the absence of geographic subject metadata on most ETD metadata records may not be totally ascribed to novice metadata creators in the different university libraries. This argument is based on the finding that more than two thirds of the total number of participants in this study had been involved in ETD metadata creation for more than five years.

Based on the findings, it can be that assigning geographic subject metadata was more challenging for most metadata creators than other forms of subject metadata. The recommendation is, therefore, that interventions for knowledge improvement should include specific guidelines on effective subject analysis approaches for creating complex types of subject metadata, including geographic subject metadata. The guidelines should be continually updated to reflect advances in theory and practice.

It was overall observed that respondents' perspectives indicated that confidence levels about the current adequacy levels of the geographic subject descriptions for ETDs in their collections were not completely positive. Less than half of the respondents said they were highly adequate. It was important to understand the reasons for the existing situation, since it was shown that professional library staff responsible for creating ETD subject metadata. It was expected that the adequacy levels would be high, considering the fact that the professional library staff members were indicated as playing the major role and are considered as the main source of origin of the existing geographic subject metadata. However, the conclusion is that adequacy levels for the geographic subject metadata needed to be improved.

It is recommended that more studies are conducted on adequacy levels, through records analysis to inform the need for improved training on ETD geographic subject metadata and retrospective enhancements.

7.6.3 Influence of understanding the theory of subject analysis

The primary quantitative research findings indicated positive views on the need to understand the theory of subject analysis as a basis for geographic subject metadata creation. No negative views were expressed. These views were consistent with those expressed by Rondeau (2014:15), in that theory is important to provide a basis to address the complexities of the contemporary knowledge organisation environment.

However, differences were found in terms of the perceived levels of influence of theory. Despite the general positive views, the quantitative findings showed that less than half of the questionnaire respondents believed that the theory of subject analysis had major influence on geographic subject metadata creation. This position did not confirm that all current subject metadata creation practices could be viewed as significantly being influenced by the application of subject analysis theories, because the geographic subject metadata creators did not perceive theory as chiefly influential. Issues of the theoretical basis for information resource description remains debatable. As observed by Mayernik (2020:705), different opinions about the issues remain a point of discussion within the library community.

Although the interview participants explained that the theoretical basis of subject analysis was important, some the participants indicated the need to improve this knowledge. More emphasis was placed on the need to educate new staff. A related study conducted by Chu and O'Brien (1993) showed related challenges experienced by novices in conducting subject analysis at an advanced level, which may be linked to the challenges of relating theory and practice. The differences in the knowledge levels may be understood as the reason for the varied understandings on the importance of the theoretical foundation of subject analysis.

It is concluded that both the questionnaire and interview phase findings assisted to reveal the need to improve the understanding on the influence of theory. It is recommended that further training and continuous development programmes are supported by the library management, with emphasis on creating awareness of the benefits of theoretical principles to support practice. Specific attention should be placed on new staff coming into the role of ETD subject metadata creation to close any theory knowledge gaps. Other factors learned from the findings revealed the existing approaches presented in the next section.

7.6.4 Other contributory factors to effective ETD geographic subject metadata creation

The quantitative findings revealed additional dominant factors linked to subject analysis and the creation of geographic subject metadata for ETDs. The findings of the additional factors are summarised and discussed in the next sections.

7.6.4.1 Perspectives on policies and guidelines

The quantitative findings from the questionnaire showed that more than half of the study participants indicated that they received guidance from policies or guidelines to create geographic subject metadata and the other half did not. Additionally, almost half of the participants indicated that policies did not exist in their libraries. The interview participants clarified and confirmed the absence of policies to guide best practices and consistent ETD subject metadata creation. The nonexistence of policies was contrary to practices in similar environments in other countries, as revealed by the empirical findings of Flynn and Ahrberg (2020), where policies or guidelines were found to exist in almost all libraries and are being updated to facilitate the upgrading of ETD metadata.

Furthermore, Potvin and Thompson (2016:102) highlight that the advent of ETDs necessitated the creation of new policies to guide the management and access of ETD content. In addition to policies, guidelines were recommended for efficient practices in metadata creation, particularly for new library staff who are involved in creating subject metadata. In addition, Potvin and Thompson (2016) point to the use of guidelines by the University of Houston libraries for managing the ETD metadata in their repository.

Notably, the metadata creators generally showed a high regard for the importance of policies. The quantitative findings showed a significantly high percentage in acknowledgement of the need for policies to guide ETD subject analysis. In affirmation, the answer to the question whether policies on ETD subject metadata were necessary was dominantly a definite "yes" during the interviews. However, as seen in one of the responses, the use of policies should not be regarded as the ultimate solution in assisting effective geographic subject metadata creation.

The comments about the limitation of policies in the interviews are aligned to the observation made by Rondeau (2014:27). Other influential factors that may improve the existing situation included offering instructions to help enhance the understanding of intellectual aspects of metadata creation. Similarly, Fourie (2002) discourages the "recipe approach" of relying on rules and guidelines only in creating information resource content representations. The conclusions based on these findings were that policies were not specific on subject analysis approaches and the importance of geographic subject metadata. It is recommended that university library polices should be designed to give specific guidance, in order to address the challenges regarding these aspects.

7.6.4.2 Perspectives on quality control

The reviewed literature revealed the importance of maintaining quality for descriptive metadata. Quality control for ETD subject metadata was found as being less prevalent in most libraries that participated in this study.

The questionnaire responses only showed an insignificant number of respondents indicating their involvement in quality control. The explanations of the metadata procedures given in the interviews did not reflect common quality control practices. The interviews responses provided further clarity, e.g. one participant indicated that the cataloguers' judgement prevailed in decisions on how to assign geographic subject metadata. It was further indicated that, due to workload, there was no peer review for theses and dissertations metadata. Quality issues were discussed in detail in Chapter 3, indicating the literature support for the significance of maintaining metadata quality.

Additional quality-related issues concerned consistency and specificity in creating the geographic subject metadata. Problems associated with consistency of practices were observed in the findings. The questionnaire responses revealed many inconsistent practices, while no absolute commonality existed among the participants. The importance of consistent subject metadata creation practices and the enhancements of quality through consistent practices is highlighted by Zavalina, Zavalin and Miksa (2016), Steele and Sump-Crethar (2016:47) and Tmava and Alemneh (2013:858).

In addition, the researcher endorses the views linking consistency to quality for subject determination, as discussed by Chu and O'Brien (1993:440). The authors indicate that maintaining different levels of specificity causes inconsistencies and poor quality levels. This is a particularly important principle for assigning geographic subject metadata for digital items (Dublin Core Schema). Furthermore, the intents that South African university libraries are currently discussing regarding the shared metadata platforms for ETDs will require consistency in the specificity of description of ETD subject content. This will assist to appropriately delineate geographic places.

Specificity is an additional way of ensuring quality subject metadata that enhances access to geographic related information from ETDs. While adding geographic terms as subdivisions to the existing subject headings was indicated as the most performed function, as discussed in Section 7.5.1, issues relating to specificity of the subdivisions were not possible to determine from the questionnaire findings. However, the qualitative interview findings provided clarity. Some explanations indicated that the subject metadata creators did not have the resources, time and staff to create more specific geographic subject metadata.

Contrarily, some sentiments were positive. For example, some respondents indicated that, because of being used to working with the LCSH, they followed standards and rules to create geographic subject metadata and started from the country and proceeded to smaller areas. This statement implicitly indicate efforts to address the principle of specificity, which is important for effective geographic subject metadata analysis.

The specificity levels can be further investigated through different study methods like the records content analysis to learn more about the quality of ETD geographic subject metadata. Describing specific geographic areas provides a good basis for creating links between subject metadata and other geographic databases.

The three instruments used in this study produced results that addressed the quality-related questions satisfactorily. The questionnaire findings showed less quality control practices, which was confirmed by the detailed explanations of processes given during the interviews.

The content analysis findings reflected that quality of existing records in different ETD repositories needed to be improved to ensure that geographic subject metadata is assigned when it is required. Additionally, consistent decisions need to be taken regarding the use of standards for controlled vocabulary.

More records studies – particularly investigation into quality issues – are recommended to obtain deeper insight into quality control for digital resources subject metadata. Furthermore, it is recommended to include quality control in the ETD metadata workflows to improve current practices. Tarver et al (2015:37) indicate the importance of workflows to facilitate appropriate metadata creation. The interview participants explained and confirmed the different subject metadata creation workflows that were in use in the different university libraries. Moreover, the background information on the study participants indicated that more than 40% of the metadata creators' experience in ETD metadata creation was in the one to five years category. Further investigation on how quality is currently maintained will help to provide the basis for development and to achieve effective metadata creation and improve ETD discoverability. The next section discusses the findings regarding the perception on how adequate descriptions of geographic places facilitate linking to other geographic databases.

7.6.4.3   Use of linked data and geographic subject metadata

Other important findings were established, which helped to understand the current geographic subject metadata creation practices and how they revealed techniques that shaped current subject metadata creation practices for ETDs and will continue to

impact future practices. The questionnaire findings showed general sentiments that technology had not made subject analysis for geographic subject metadata redundant, with almost two thirds of the y participants shown to agree with this view. However, a concerning number, almost a quarter, indicated that they were not sure about the impact of technology. This finding was unexpected, due to the major role played by technology in subject metadata creation.

However, the impact of technology and the linked data impact on the discovery of ETD geographic subject content was expressed during the interviews. Despite a few questions allowing general comments on factors that contribute to effective subject analysis, there were fewer pointers to how the changing landscape in subject metadata creation, mainly driven by technology, was understood to impact on geographic subject metadata creation. This was unexpected, due to the visible impact and the ongoing important discussions on the issue of linked data for the repositories and other information provision platforms.

A few of the interview participants, who shared detailed experiences and perspectives, demonstrated more knowledge of how technology has introduced transformations that cannot be overlooked in the metadata creation environment, including linked data. Albrachtsen (1993) and Tennis (2005) indicate how technologies contribute to new meanings for subject analysis. Linked data, a technique that involves creating links to various information resources and platforms, has found prominence in metadata creation practices. Links to geographic platforms and related resources, with coordinate search capabilities, enhance geographic subject descriptions. Common search capabilities, linked to common platforms like Google Maps are increasingly being introduced in library systems.

Appropriate geographic subject metadata can serve as best linking elements in ETD systems with such capabilities. Most descriptive schema allow the creation of links to other sources outside the ETD repositories. In addition, continuing work done through forums like the Linked Data Interest Group of the American Library Association shows the importance of linking library metadata for the semantic web. Linking place names to other data sources, e.g. the Geonet names server, provides a useful way of enhancing access through geographical subject metadata.

The interpretations and discussions in Section 7.6 generally demonstrate that underlying individual judgements about subject analysis and geographic subject metadata creation influence the appropriateness of s geographic subject metadata. Less was explicitly revealed about how the theoretical principles applied.

It is recommended that the metadata creators deliberately prioritise theoretical approaches. Continuous development programmes can assist to follow the debates and developments in theoretical frameworks and guide continued review of the subject analysis practices in line with the changes in the information provision landscape.

7.7 Effect of analysis approaches on ETD geographic subject metadata

The third research question of this study asked how and why the current subject analysis approaches followed during geographic metadata creation for ETDs affect the process. The aim was to establish the perceptions on the influence of the different approaches, mainly the analysis approaches, on the appropriateness of the geographic subject metadata. The study findings provided information on the awareness of the effects of existing procedures, familiarity or level of understanding of the importance of subject analysis and its impact on effective subject metadata creation. Perspectives on important aspects when determining what a resource is about and how subject analysis knowledge was acquired to facilitate the understanding of subject analysis basis. The study findings further showed how the experiences and perspectives were influenced by theoretical knowledge.

7.7.1 Awareness of the impact of current procedures

Awareness of the impact of procedures followed for creating geographic subject metadata for ETDs may assist metadata creators in making changes to enhance the effectiveness of their approaches. However, metadata creators may not always be aware of the impact of their subject analysis approaches. The question was asked if metadata creators were aware of the effect of their current subject analysis approaches and how they facilitated the effective description the ETD subject content.

It is important for the subject metadata creators to be aware of and understand how the procedures followed in their different libraries influence the appropriateness of the geographic subject metadata.

The findings revealed that varied approaches were followed to create subject metadata currently existing in ETD repositories. The quantitative results showed less than half of the respondents regarding the appropriateness of subject metadata in ETD repositories as highly adequate. This suggested that there was some awareness of shortcomings in the type of geographic subject metadata being produced. Related to this finding, a study conducted by Wolverton, Hoover and Fowler (2011) indicates the provision of different subject metadata for ETDs. It is noteworthy that the creation of geographic subject metadata is suggested by the use of resources for geographic place names. However, awareness for shortcomings was noted, which may imply challenges including the creation of geographic subject metadata. Despite the possible limitations, like the lack of training and other contextual factors that were identified, clarity on how the adequacy levels were associated with the existing practices was sought through the interviews.

The interviews findings showed that the general perspective of most participants was that the current procedures did have an impact, although it was not always positive for adequate subject metadata creation. Some explanations showed satisfaction with the existing local practices and that adequacy was judged on those grounds. One participant paraphrased "… what we have been doing so far has been a success, it might not be at higher a level like other institutions are but it is seen as sufficient that when information is searchable it can be retrieved, so far so good for our database". The statement seems to suggest that there was satisfaction, as long as some level of access was enabled, even if it was not the best forms of description when compared with other libraries. This view suggested concern for what was considered as relevant in that individual library's context only.

Additionally, some participants expressed that they could do more, but that challenges like poor subject specific knowledge were a hindrance. This perspective again showed an awareness of the need to improve current practices.

Recommendation for remediation or revision projects for ETD metadata are made by Thompson et al (2020), who support the need for improvement. It was also indicated during the interviews that novice subject metadata creators possibly lacked the appropriate knowledge. These assertions indicated an implicit awareness that current practices by some of the new staff could be impacting negatively on the kind of geographic subject metadata being created. Moreover, the findings revealed the general belief that the existing situation could be improved. Intentions to match the advanced levels of performance in other libraries were also implicit in the findings.

It was observed that there was often no conscious decisions made on the appropriate attitudes to be maintained for geographic subject metadata creation. One interview participant explained that they were not consciously or specifically looking for the place name, which may be something lacking in the attitude of the metadata creators in those institutions, which can be addresses to determine whether a place name in a thesis was of importance. It is normally expected that, if the place names in the titles, abstracts or contents are of significant value, a conscious decision must be made to identify geographic access points. However, the experiences and perceptions explain the different practices shown in the quantitative findings exists. The absence of clear methods and attitudes often lead to inconsistencies of practice. The findings corresponded with those about the adequacy levels of existing geographic subject metadata, which were perceived as low and that sometime adequacy was judged, based on individuals' perceptions, rather than on general acceptable practices.

It is realised that, although questions that explicitly enquired about impact of practices, most effects were implied in the responses. The interviews provided the best information to address this question from the experiences and perspectives shared by the metadata creators. Overall, the findings suggested that, in some cases, more concentration was placed on the product and on processes, with little concern for the significance of the metadata creators' actions. A similar finding is found in the study by Rondeau (2012:61), where the participants indicated that they only though about the headings and that they did not separate the subject analysis process from the way in which they determined the subject headings. Alertness of metadata creators' theoretical basis was also not conclusive.

Additional information in the interviews revealed awareness of other contributory factors to the existing position, e.g. one participant responded that, due to the complex nature of geographic subject metadata creation, the subject metadata creators often overlooked it.

Furthermore, the content analysis indicated that less geographic subject metadata was allocated between 2014 and 2018. In view of the findings, subject analysis is not seen as fully assisting the creation of geographic subject metadata in all instances, where it is generally expected to do so.

Overall, the findings revealed the need to adopt globally acceptable practices, based on a reliable theoretical foundation. Theory influences should be acknowledged to encourage adoption of theoretical principles as a basis for effective subject metadata creation. In support, Hjorland and Albrachtsen (1995:148) highlight the need for a theoretical outlook to adopt a broader perspective in subject analysis. It is concluded that awareness of the effect of current procedures seems to exist among the metadata creators. The metadata creators' perspectives are that improvement of current practices could increase the levels of efficiency and that it is important for all libraries to adopt best practices. Further investigations on how the metadata creators maintain awareness of the effect of their practices may help to establish a knowledge base for improvement of existing approaches and to accommodate new trends in subject metadata creation. Theory knowledge could help the metadata creators to be aware of the implicit principles of the methods that they choose to determine the subject representations for ETDs.

The next section reports on the findings on the familiarity with theoretical approaches. The effect of the levels of familiarity with theoretical principles is examined.

7.7.2 Familiarity with the theoretical principles of subject analysis and their impact

The questionnaire survey results indicated that most of the metadata creators followed common practices and used commonly accepted standards for subject descriptions. The findings showed that there was general familiarity, with just over half of the

participants being very familiar with the theoretical principles of subject analysis. However, the revealed different levels of familiarity or understanding were concerning. The findings showed that more than a third of the participants were less familiar with the theoretical principles. As a result, inconsistencies in subject metadata can be expected – particularly because there is less quality control being performed.

In addition, there was an observed association between the levels of familiarity with theory and the Library and Information Services training received. Almost all the participants in the questionnaire survey indicated that they had gained familiarity with the subject analysis theoretical principles through this mode of training. It appeared that, despite the formal training on the theoretical principles of subject analysis, a gap still exists in the knowledge retained. In a study that they conducted, Chu and O'Brien (1993) observed that the levels of familiarity with the process of conducting subject analysis at an advanced level varied, despite the training received by the research participants.

On-the-job training also played a role to create awareness, but to a lesser extent. The questionnaire findings showed that almost half of the respondents did not receive on-the-job training on theory integration into practice. Slightly more than a third of respondents showed uncertainty regarding its importance. However, it was observed that some of the common approaches followed by the participants showed an implicit association with the approaches advanced by the theories of subject analysis. The findings revealed examples of how they found the relevant geographic terms and the aspect that they considered important during analysis, to be discussed in the next section. It was assumed that that those who did not receive the training, would not be familiar with the theories.

A better understanding was acquired from the qualitative phase findings regarding the levels of familiarity of the theoretical principles and how they influenced how subject analysis is conceived. The findings showed that, in some cases, the lack of knowledge on foundational principles or a gap in the application for subject metadata creation was perceived to be prevalent among new ETD subject metadata creators.

In such cases, the inexperience was perceived as likely leading to the creation of subject metadata of inferior quality, while adequate experience was perceived as being influential on the creation of appropriate geographic subject metadata. Familiarity with the appropriate theoretical principles at the early stages of practice could provide a firm foundation for creating appropriate geographic subject metadata for ETDs. Familiarity with the influence of theory is important to guide practice by metadata creators of all levels of experience. Notably, Rondeau (2014:15) argues that philosophical issues concerning the understanding of information organisation are consequential for effective subject analysis. This view is supported by Hjorland (2021), who supports the consideration of a theoretical foundation to enhance the discoverability of information resources.

In addition, the different experiences and perspectives from the qualitative interviews findings showed how subject metadata creators understood the intensities of impact of theory. An unexpected finding showed that most interview participants considered geographic subject metadata creation as being an easy process. However, those who thought so, could not give clear explanations on how the subject analysis process facilitated the process and less theoretical implications were mentioned by them. There were no clear indications about how the theoretical principles helped to make the subject analysis process easy. Most explanations emphasised the use of standards to maintain consistency. The perceptions of the process being easy may suggest less understanding and less importance attached to possible guidance from theory.

Contrarily, some of the interview participants explicitly raised issues that pointed to the complexity of the process. In such cases, more explanations were give on how subject analysis was conducted. Similar conversation about complexities of geographic subject representations are continuing to take place on various subject description forums. The ongoing debates indicate the need to familiarise with theory. The IFLA Subject Analysis and Access Section leads some of the discussions that affect the international practices. The findings showed that the participants, who appeared more knowledgeable about the processes, were more conversant with its complexities and different standards, compared to those who only concentrated on explaining procedures and not what informed them.

Additionally, the quantitative findings indicated that the majority of the study participants indicated that they followed common practices and standards for assigning ETD subject descriptions. However, evidence from the record analysis shows that, in most cases, geographic subject metadata was not allocated and the use of controlled vocabulary was not being observed. A common finding by White (2012:124) shows challenges in allocating geographic subject terms as subject description of information resources. The general observation in this study was that there were inconsistencies concerning what metadata creators conceived as best practices in subject analysis, resulting in efficient subject metadata. The content analysis findings reflect inconsistent application of specific theoretical principles and practice.

Moreover, an observation from the content analysis phase was that, while most respondents indicated that they had a good understanding of the theories and their impact, the existing geographic subject metadata on the studied records showed less evidence of consistent theory guided implementation. Apart from the other challenges that were mentioned, the expressed high levels of knowledge regarding subject analysis practices seemed debatable, or knowledge application was problematic, as most analysed records found to be deserving geographic subject metadata descriptions had not been assigned with such.

It is expected that the expressed knowledge of theories should assist in creating more geographic subject metadata than is currently the case, unless there are other reasons for the existing low levels of assigned geographic subject metadata can be known. One of the hindrances found from the interviews was that geographic subject metadata was considered as a daunting task that required more knowledge of different types of subject descriptions.

The quantitative and qualitative findings suggested that, despite the observed levels of knowledge possessed by the subject metadata creators, there was less eagerness to participate in geographic subject metadata creation, presumably due to the complex nature of this type of subject descriptions.

Insufficient knowledge of best subject analysis approaches can be ascribed as an additional reason for this evasion. Strategies to create more awareness of the best theoretical approaches should include refresher on-the-job training on the relevance of theoretical principles, in order to enhance confidence to tackle complex forms of subject metadata, like the geographic type.

The knowledge of theoretical subject analysis principles may help to inform the understanding of and decisions about what focus to maintain during subject analysis to achieve maximum impact in different contexts. Geographic subject metadata can also be assigned as secondary topics, which, according to Chu and O'Brien (1993) and White (2012), were found challenging to assign, depending on the experience of the person conducting subject analysis or creating subject headings.

The next sections present the findings on how the current perspectives about the focus maintained during subject analysis influenced the processes of subject analysis. Additionally, the findings on the subject analysis orientations will be analysed against the theoretical principles from literature, particularly from the model of conceptions of subject analysis that is used as a framework for the interpretation of the findings.

7.7.3   Important aspects when determining what a resource is about

The question on the effect of current subject metadata practices was further addressed by investigating the basis used for subject analysis. The different focal points for analysis and their perceived impact on the appropriateness of the geographic subject metadata were examined. The participants were asked questions about aspects that they regarded as important when determining the subjects of ETDs.

The findings pointed to the existence of different analysis orientations and clarity is provided on metadata creators' perspectives of the impact of the different approaches to subject analysis. The varied analysis orientations are aligned to those outlined by, amongst others Hjorland (2021), Albrachtsen (2015), Fourie (2008) and Tennis (2005).

The model of conceptions, as discussed in Section 7.3, provided a framework for the interpretation of the findings on the analysis focus for the ETD geographic subject metadata. The model of conceptions of subject analysis and its variants are discussed in detail in Chapter 4. The model also provided a basis to examine the approaches associated with the attributes of the model of conceptions of subject analysis. Additionally, the identified conceptions were used to determine the association between the theoretical and practical perspectives.

Both the quantitative and qualitative findings pointed to different kinds of prioritisation or focus during analysis. The analysis of existing records also suggested different approaches, revealed by the different products. The main analysis approaches identified in the findings were categorised as content, domain, user and author focused. Both the quantitative and qualitative findings pointed to different kinds of prioritisation or focus during analysis conducted for geographic subject metadata creation for ETDs. The next paragraphs discuss the findings and their association with the different analysis conceptions.

7.7.3.1 Content-focused approach

According to the questionnaire survey findings, the content-focused approach was the dominant approach to subject analysis applied by the metadata creators. The questionnaire findings showed that more than two thirds of the participants considered the approach of focusing on the entire content of an information resource as very important and the others attached a fair importance to this approach. However, it was found that other, different approaches were also rated as notably important for effective subject analysis, making it difficult to confirm the definite focal point. The interviews confirmed the dominance of the content-focused approach through the participants' descriptions of how they determined the subjects. Significant indications were made about focus on examining the entire content of ETDs. As illustrated in Figure 6.11, most respondents mentioned this approach.

The reviewed literature confirms that the content-focused subject analysis approach is common in libraries (Zavalina 2012:142; Albrachtsen 1993:221). However, the approach poses several challenges, some of which are identified by the mentioned scholars.

The first challenge is the potential subjectivity in terms of judgement on content relevance, depending on the perceived user needs and the metadata creators' background. Secondly, change in the relevance of information over time is a possibility because of new, emerging user needs, which often result from their new circumstances. This may result in different purposes fulfilled by the content of an information resource. Thirdly, a limitation of the content analysis approach is that implicit meanings are often not captured when focus on content is a dominant aspect during analysis. Due to these limitations, retrospective enhancement of the subject metadata is often conducted to accommodate change and different purposes.

The findings provided no indication as to whether retrospective enhancement was done or considered. An improved understanding of the subject analysis theoretical principles may help to identify the strengths and shortcomings of this approach and inform how to improve the effectiveness of ETD geographic subject metadata. However, a library that adopts a content-focused approach can benefit from taking cognisance of these stated issues and provide for mitigating measures. This approach is recommended as useful when used in combination with others that are relevant to accommodate all current and potential uses of the content of an ETD.

7.7.3.2 Domain-focused approach

In this study, the domain-focused approach refers to specific subject or discipline-focused method of analysis. Prioritisation of focus on users in specific subject areas was revealed as the second highest approach applied, after the content approach. The quantitative findings showed that more than two thirds regarded this approach as very important, with the remaining participants also rating this approach positively. The qualitative findings supported this finding, with the second highest reference made to this approach, when the participants explained their experiences and perceptions about subject analysis.

The main idea behind specific subject or discipline-focused or domain-focused subject analysis is that individuals belong to knowledge domains and have knowledge and information needs that are common to the collective (Hjorland 2017a:59; Hjorland & Albrachtsen 1995).

Seen in the context of ETD subject metadata, this view suggests that the focus of analysis is specific subject disciplines in which the ETD content was produced and where the potential information needs are likely to emanate from. In addition, the approach suggests that people in different subject specialisations have unique knowledge and information needs to which the ETD descriptions should align.

It is for university libraries to place the focus of descriptive metadata on the researchers and their information needs, which are related to their research subject areas. As a result, subject analysis that specifically focuses on the specific subject areas is often considered favourable. However, in these research findings, this approach was considered second in importance. The qualitative findings helped to identify the difficulty experienced by the metadata creators in understanding specialised topics in the different disciplines, which often limited the depth of analysis being performed in a domain-focused context. The experiences and perspectives of the metadata creators were supported by the reviewed literature. The work of Maurer and Shakeri (2016:222) confirms the perspectives that metadata creators' lack of disciplinary specialisations or specific subject knowledge affects their subject analysis capabilities. These problems result in challenges in specific subject-focused subject analysis.

An important finding from the quantitative findings was that the overall perceptions showed support for balancing the focus on specific subject areas, with a focus on all the potential users. When asked to share additional views, particularly on geographic subject metadata, more than two thirds of the participants disagreed with the idea that specific subject approach could be used as the only focus to analysis.

A conclusion that can be drawn from the findings is that, although the specific subject or domain approach is considered appropriate for ETD subject metadata creation, it should not be used as a single approach. Overall, the approach was considered as appropriate by more than three quarters of the participants. Additionally, it was shown that its application was mainly hindered by the lack of subject expertise among metadata creators. The approach is recommended, but not to be used in isolation.

Furthermore, it is recommended for metadata creators to have theoretically based insights to inform their views about this approach. This should help the metadata creators with a basis for the appropriate application of domain-focused analysis in a multidisciplinary environment to facilitate maximum usage of the geographic information contained in ETDs. This recommendation is based on the understanding that geographic information is applicable and useful across different disciplines and contexts.

A related user-focused approach is discussed in the next section.

7.7.3.3   User-focused approach

The findings of the questionnaire phase consistently showed a strong focus on the needs of the general user or all the potential users in subject analysis. More than half of the participants strongly agreed that the traits associated with the general user needs were very important. This finding aligned with the views of Soergel (2016:6), who opines that the representation of the document content should consider the potential users and use. However, it is noted that this position does not seems to disagree with the importance attached to focus on the user information needs within specific disciplines, as discussed in the previous section, rather that the two can be complementary. Despite the possible challenge in trying to balance the two approaches, focus on the user within and outside the subject domains, remained the second most preferred approach, after the content-focused approach.

A challenge was raised in the interviews concerning the difficulty of establishing potential user needs with certainty. The existing literature recommends studying real users or their information request patterns may assist in providing knowledge of the user needs (Lancaster 2003:99). Furthermore, Rondeau (2014:21) indicates that specificity and relevance are guided by knowledge of the users and their requirements. However, very little indications were made regarding the importance of studying the information use behaviour to provide a basis for subject analysis. One interview participant commented that determining how the user will go about retrieving information would be difficult, although surveys could be helpful.

A related view is raised by Soergel (2016:6), who indicates the problem of possible changes in relevance of descriptions over time and within different contexts. This suggests that the usefulness of results of the user studies is limited to the period and context in which they have been obtained. It is concluded that the user-focused approach can be a useful when complemented by continuous observation and the study of user behaviour. As a result, it is recommended for ETD management systems functionalities to extract information on user requests and to use the retrieval results of empirical studies to supply the information that can be used to support this user-focused approach to subject analysis.

The next section presents the findings of the author-focused approach to analysis.

7.7.3.4 Author-focused approach

Another approach that can be applied as a focus for subject analysis is that of author's intent. The questionnaire survey findings revealed that the author's intentions for producing a work were considered as very important by more than two thirds of the study participants. It is common for authors to attach keywords to the abstracts or with their submissions, with the purpose of describing the content or indicating the intention of their works. However, almost half of the participants considered the author supplied keywords only suitable at times. An insignificant number of the questionnaire respondents attached minor importance to this approach. It was noted that, despite that the author's intentions being perceived as very important, it was believed that the ETDs authors did not always supply the most suitable geographic subject representations for their subject content. Contrarily, the literature shows that maintaining focus on author-created descriptions is commonly used, based on the understanding that they are best suited to describe their own work (Tera et al 2021:2; Husic 2014:2). The different outlooks could be ascribed to how the ETD authors perceive geographic subject metadata.

Consistent with the quantitative phase findings, the qualitative findings from the interviews also pointed to some importance attached to the author-focused approach, but with less mention of the author's intent as a focus for analysis. A number of challenges were raised in the interviews concerning this approach.

Although the situation seemed to differ among them, the interview participants indicated that some of the submitted ETDs did not come with author supplied keywords. One interview participant stated that sometimes the titles, abstracts and the author-supplied keywords did not correspond. Another indication given in the interviews was that the author-supplied keywords were misleading the novices in subject metadata creation. This perception indicated a possibility of inappropriate subject representations sometimes being assigned for ETDs, by relying on author supplied keywords that may not be suitable. Moreover, the reviewed literature affirms that the author-supplied keywords present common problems like being semantically ambiguous and lacking standardisation (Terra et al 2021:1).

The author-supplied keywords are naturally associated with specific subject disciplines and, therefore, if the authors' submitted keywords were preferred, the subject descriptions could possibly be more subject specific. However, the reviewed literature indicates a challenge of implicit information contained in an information resource not being explicitly conveyed by the authors' keywords (Albrachtsen 1993:221). This finding suggested that more of the existing subject metadata based on this approach are possibly subject-specific, because the authors mostly assign keywords relevant to their subject specialisations. There is a possible association between the author supplied keywords and the domain-focused practices.

In addition, the content analysis results revealed that uncontrolled metadata elements were the most populated with subject metadata. It is common that keywords supplied by authors are captured in uncontrolled subject metadata elements. The content analysis findings aligned to previous findings of a study that Zavalina (2011) conducted among three digital collections repositories, where keywords were found to be predominant. However, the different metadata schema allow different forms of subject descriptions that accommodate both the controlled and uncontrolled subject terms.

Based on the findings, it is concluded that the author-focused approach may not provide the best subject analysis basis, if used by employed on its own, because ETD content may have wider usage than intended by the author.

It is recommended that, if the author-supplied keywords are used, they could be enhanced with the intervention of professional subject metadata creators, who mediate to perform quality control and add more relevant terms from the controlled vocabularies. Further studies may help to confirm whether the keywords that are currently used for subject descriptions have mainly been supplied by the authors or from other sources to inform decisions on retrospective enhancement.

An overall conclusion on the findings in this section is that all the three instruments used in this study revealed the prominent focal points in subject analysis. The questionnaire and interviews were more suitable for this purpose. The content-focused and domain-focused approaches were generally viewed positively, compared to the other approaches modelled in the theoretical framework of this study. However, the discrepancies on preferences noted in the findings were best clarified through the interviews. Additionally, it was found that the results helped to determine that the effect of the current approaches on the quality and adequacy of the geographic subject metadata is not always positive as seen in the existing products. The findings in this section resulted in the conclusion that mostly, implicit understandings were gathered, indicating that the effects of the focus to analysis impact both positively and negatively on the creation of subject metadata. Weaknesses and strengths were noted with the different approaches. Moreover, the conclusion is based on the findings of the content analysis, where empirical evidence showed limited representation of geographic subject metadata in the existing records.

Therefore, it is recommended that the hybrid approach is maintained in terms of role players (sources of metadata, e.g. authors and library staff), vocabulary, and focus during subject analysis. The recommendation is based on the understanding derived from the boundary object theory on integrating different elements and domains for collaborative approaches. Based on this understanding, subject metadata is further conceptualised as a boundary object between different subject disciplines. The findings indicated that subject metadata could serve as boundary objects facilitating interdisciplinary and wider access to ETD content by varied potential users across different contexts.

Furthermore, it is concluded that the way in which metadata creators did not make an explicit association of the effect of their approaches and the influence of theory knowledge on the appropriateness of ETD subject metadata that is produced. The reviewed literature indicates that the theoretical conception of how to determine what an information resource is about, informs how the principles of subject analysis are comprehended (Hjorland 2017a:60). It is important for the metadata creators to be familiar with how theory helps with appropriate conceptions of subject analysis. Appropriate subject analysis conceptualisation is important because, as McCutcheon's (2011:64) asserts, subject analysis is required to provide fullest ETD access. The next section examines the perspectives on theory knowledge. Appropriate perspectives on the required knowledge for a specific function are critical for its success and advancement.

7.7.4   Perspectives on the required theoretical knowledge for subject analysis

This section discusses the findings of the perceptions on the subject analysis theory knowledge required by the metadata creators and how this knowledge affects the appropriateness of geographic subject metadata. Theory basis is important to guide the effective implementation of subject analysis for ETDs. This section involves investigating perspectives on how the theory basis influences practice and the appropriateness of the resultant geographic subject metadata. The reviewed literature suggests that effective subject analysis requires an appropriate theory base and identifies the possible theoretical approaches (Hjorland 2017a & 2017b; Lancaster 2003; Hjorland & Albrachtsen 1995; Albrachtsen 1993).

The questionnaire findings showed that almost all participants received formal Library and Information Science training, where they mastered the theory of subject analysis. Furthermore, almost half of the participants were positive about the importance of on-the-job training to refresh or acquire theory knowledge. However, a few of the participants were negative about the relevance of on-the-job training to gain theory knowledge for subject analysis. The findings revealed evident gaps in knowledge, in that almost half of the participants did not express full confidence about their familiarity with theoretical knowledge.

The participants of this study indicated that they realised the need to understand the theory of subject analysis to provide a foundation for geographic subject metadata creation. Almost half of the participants found theory knowledge as highly important, whereas the remaining numbers were also positive, although to a lesser degree. It was unexpected that the quantitative findings demonstrated participants attaching less value to theoretical knowledge. This finding contradicted views in the literature on the subject, e.g. Mayernik (2020:705) indicates the need for professional metadata creators to understand the theoretical principles clearly that support their operations. Consistent with the literature, the experiences or perspectives from the qualitative findings affirmed the assertions on the importance of theory knowledge. All responses were positive towards the role of theory and its importance in subject analysis and metadata creation.

The general importance attached to theory was found to correlate to the value ascribed to training. Despite the indication that theory training was received and its value being fairly affirmed, the current levels of familiarity with theoretical principles among metadata creators was considered unsatisfactory. Only more than half of the participants indicated that they were very familiar with theory, which pointed to a gap between training and the levels at which the acquired theory knowledge was retained for effective application on the job. There was an indication that training was received, but, in spite of the training, those who had been trained, demonstrate lower familiarity with theory. The qualitative findings showed that most of the participants believed that the theory knowledge gap can be filled through on-the-job and other forms of continued training, so as to improve the integration of theory and practice.

Additionally, an implicit idea gathered from the interview findings on subject analysis theories was that the lack of adequate knowledge of theoretical principles often lead to less effective subject analysis practices. This was revealed when the need to address the challenges of inadequate knowledge was raised during the interviews. One interview participant explained that they used their discretion to decide what is correct or incorrect, as long as the resources were made accessible. This explanation suggested that there was no clear theoretical basis and that best practices were not followed consistently.

These challenges experienced in practice reveal a gap that could be filled by understanding the theoretical based principles and how they assist to promote best practices. A change of perceptions on the importance of theory is essential to improve subject analysis practices and achieve more than basic subject descriptions by the metadata creators. It is considered essential to acquire or promote theoretical knowledge to support appropriate geographic subject metadata creation in general.

Despite that the quantitative findings showing that the majority of the metadata creators' theoretical knowledge was acquired through studies for a LIS qualification, it was notable that a few of the professional library staff indicated that they did not receive this training. This was not surprising, because of the varied curricula of different library schools, with some having discontinued bibliographic description training from their subject offerings. Furthermore, both the quantitative and qualitative findings showed that the experience levels of the metadata creators differed considerably. This suggested varied training received in different eras and a possibility of different experiences resulting from library contexts that could lead to different levels of theoretical knowledge.

Additionally, the experienced metadata creators were perceived as an important link to transfer the theoretical knowledge to new librarians. For example, one interview participant explained that new staff were guided on what to look for when analysing ETDs content to determine their subject representations. Another participant stated the need for experienced metadata creators to transfer knowledge and skills, as soon as possible. These experiences and perspectives indicated that on-the-job training remained an important initiative to transfer knowledge on theoretical principles. A recent work by Mayernik (2020:705) supports the findings of this study about different ways of acquiring knowledge on the principles of metadata creation, including being taught as part of the professional curriculum or through continuous self-development activities.

Generally, the quantitative findings from the questionnaire revealed varied levels of importance being attached to training to help acquire subject analysis theoretical knowledge.

Explanations from the findings of the interview phase of this study revealed that the respondents attached high value to training to acquire subject analysis theoretical knowledge. However, in a few cases, the interview participants could not come up with clear explanations on how theory was infused into practice and gave very short answers like "yes" or "no", even after further probing.

Furthermore, less clarity was obtained about the knowledge gained through on-the-job training, which confirmed the uncertainty that was shown in the quantitative findings. Instead, it was observed that a common approach to improve knowledge on the job was through informal local discussions, workshops and conferences. The general finding was that the differences in ability to articulate theoretical influences might be understood to indicate the varied levels of theoretical knowledge or its transfer to practice.

Comparing the quantitative and qualitative findings shows an existing theoretical knowledge gap and its influence is not perceived very highly. Therefore, it is recommended to give more attention to formal and continuous education interventions on theory education and its application. It is also suggested that structured on-the-job training and continuous development programmes are formally supported in different libraries to bridge the existing gap on the theoretical subject analysis knowledge and its integration into practice. This is more necessary in cases where it was not confirmed that theoretical training was received, either through formal or informal interventions, as shown among a few respondents of the questionnaire survey. This suggestion aligns with the literature. For example, the findings of a study conducted by Park (2009:225) confirm the need for structured continuing education programmes to bridge the knowledge gap between theory and practice. In addition, Rondeau (2014:31) suggests that those involved in subject analysis need to engage more to address challenges related to subject descriptions.

Concluding Section 7.7, it is, therefore, recommended to follow the hybrid approach in terms of role players (sources of metadata, e.g. authors and library staff), vocabulary, (controlled or uncontrolled) and focus during subject analysis.

The recommendation is based on the understanding of the boundary object theory on integrating different ways for different libraries to enhance geographic subject metadata creation collaboratively, in a way that is suitable for different information needs across different contexts. Based on this understanding, subject metadata is further conceptualised as a boundary object between different subject disciplines within the universities. The findings indicated that subject metadata could serve as boundary objects facilitating interdisciplinary and maximum access to ETDs content by varied potential users.

7.8   Chapter conclusion

The collective findings of this study adequately addressed the problem of subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations. The findings contribute new knowledge about the significance of theory on the practice of ETD subject metadata creation, which was not covered in literature reviewed in this study. Consistent with the pragmatist view of research aiming to contribute practical solutions to inform future practice (Saunders, Lewis & Thornhill 2016:143), the findings of this have the potential to inform future directions in subject metadata practice by using merged and hybrid approaches that are theoretically based. The current effort by the South African universities to establish a national ETDs database may benefit from the empirical findings of this study about descriptive metadata best practices.

The study used tree data collection instruments – questionnaire, interviews and content analysis – in a sequential explanatory mixed methods design that was supplemented with an additional third phase. The instruments yielded quantitative and qualitative data that was integrated for analysis and interpretation. This practice is consistent with the requirements of the mixed methods design, which involves the integration of the analysis of research results, the interpretations and reporting of the findings. Conclusions were drawn from the findings of the three phases involved in this study.

Investigating subject analysis and geographic subject metadata creation was approached from a practical and theoretical point of view. The first research question was adequately addressed by the three instruments used in this study. The questionnaire results indicated fair usage of geographic subject metadata, varied procedures and different subject analysis approaches. However, the perceptions of the metadata creators were that the existing subject metadata was inadequate representations of the ETD geographic subject metadata.

Furthermore, the analysed ETD records confirmed fair use of geographic subject metadata and les use of controlled vocabulary. Additionally, the content analysis findings revealed the hybrid use of controlled or uncontrolled geographic subject metadata. The interview findings and the content analysis generally confirmed the questionnaire findings. However, uncertainties were revealed with the focus or prioritisation during subject analysis. It was not clear on what the metadata creators primarily concentrated during subject analysis.

The second research question on the current approaches followed to create geographic subject metadata for ETDs was best answered by the explanations in the interviews. The explanations clarified the questionnaire findings, indicating a general lack of policies, of standards being used, although not adequately; and unsatisfactory descriptions. Additionally, there were fair levels of positivity about the influence of theory on practice, although there were contextual challenges hindering implementation of theoretical principles.

The answers were confirmed by the results of the content analysis on the dominant use of other forms of subject metadata, compared to the geographic type and uncontrolled vocabulary compared to uncontrolled vocabulary. High perceptions existed regarding the value of controlled vocabulary, but application was less satisfactory. Similar findings were revealed in the study conducted by Zavalina (2014:70), where free text subject descriptions were found to be dominant in the explored digital collections. However, the author suggests that the use of both controlled and free text subject metadata adds value to descriptions. The positive views on the use of standards for headings control expressed by the subject metadata creators in the interviews were not confirmed by the content analysis findings.

Consistent with the findings of this study, Mayernik (2020:707) and Zavalina (2014:88) indicate that the lack of resources is perceived as the cause for compromise on quality of assigned metadata.

The interviews successfully indicated how the practices were experienced and the perceived impact of the way in which the processes were carried out. The experiences and perspectives shared during the interviews revealed varied levels of understanding and awareness regarding the influence of the existing practices and the impact of theory on practice. Almost half of the questionnaire respondents confirmed the high impact of theory and the remainder were positive, but not highly so. It could not be confirmed beyond doubt that the role of theory in practice was considered highly. The varied understandings and practices suggested that varied subject analysis approaches produced different forms of descriptions.

The third research question was answered by the results of the questionnaire survey findings, which gave a broad understanding of the effect of analysis approaches on the produced geographic subject metadata. The responses indicated more inclination to content-focused analysis approach, followed by user-focus subject analysis procedures. However, the question was best clarified through the experiences and perceptions shared in the interviews. The findings were interpreted, based on the approaches reflected in an existing model of conceptions of subject analysis. Subject analysis experiences, perceptions and practices were found to have an association with the theoretical approaches mentioned in the conceptual subject analysis model. The main focus on ETD content analysis was confirmed. The second confirmation on the questionnaire findings was the dominant focus on users within specific subject domains. The theoretical approaches to subject analysis was implicit. The metadata creators did not explicitly link their experiences and perceptions with their understanding of theoretical principles.

The implications for practice were presented by highlighting the impact of varied practices and perceptions of the metadata creators concerning subject analysis and ETDs geographic subject metadata creation. Generally, challenges were experienced across the different libraries when creating geographic subject metadata for ETDs.

Geographic subject metadata was viewed as complex to work with, compared to other forms of subject descriptions. Inconsistent subject analysis practices were common. Information searches based on subject metadata, conducted within individual repositories and across different libraries, could be adversely affected by the inconsistent practices. Continuous education and proper guidelines could assist the different libraries to adopt consistent practices that align to theoretical principles. Additionally, it is important to improve the ETD metadata creators' attitudes about the impact of theory, in order to have an appropriate foundation for practice.

The reviewed literature does not show a similar study based in the context South African libraries. Furthermore, it could not be established that the link of practice and theory knowledge has previously been investigated by using the same methods used in this study. The subject analysis conceptual model and the boundary object theory provided a basis on which the study was conducted and the findings analysed to address the research questions. Geographic subject was found to serve as a boundary object facilitating knowledge sharing across different contexts.

The findings indicated that subject metadata can be considered boundary objects that could enable cross-disciplinary subject descriptions and wider access to ETD content by varied potential users. Collocation of research collections on different geographic location housed in ETD repositories can be facilitated through subject analysis aimed at facilitating wide usage. Thinking about bridging boundaries for current and future use of geographic subject metadata to accommodate the changing needs is important. The boundary object theory can provide a good basis for research aimed at improving diversified usage of research information contained in ETDs.

Additionally, the findings suggested that geographic subject metadata could serve as a boundary object allowing the use of different role players in subject metadata creation and different types of vocabularies that can be used together to serve the same purpose of improving discoverability and access within the different libraries' ETD repositories.

Additionally, the use of automation is increasingly researched as an alternative source (role player) in subject analysis. Tennis (2005:115) indicates that the principles on which subject analysis is built should be explicit to provide theoretical foundations. Further research into the theoretical principles will avail empirical evidence that may improve awareness of metadata creators to justify their approaches. In addition, the information will serve as a framework for automated system developers to build appropriate systems to assist subject analysis.

The metadata creators' approaches and perceptions about their effect were adequately revealed in this study. However, the perceptions about the theory link to the subject analysis practices was mostly implicit. The conceptual subject analysis model helped to induce meanings about the application of theory. Further studies are recommended to provide more information on how theory and practice of subject analysis should be integrated to improve subject description for the digital formats of information resources.

In addition, subject metadata is viewed as a boundary object between different role players in metadata creation. The findings indicated the different responsibilities of contributors, the professional library staff, the authors, the users and automated functions in ETD subject metadata creation. This is consistent with an indication by Chaudhry (2016:93) that networking and digitisation introduced multiple role players to knowledge organisation functions. However, in line with the findings of this study, it is acknowledged that the professional library staff continue to play an important role in intellectual analysis, standardisation and quality assurance. As a result, the theoretical knowledge should assist them to build a balance around subject metadata creation.

The next chapter presents a proposed model to facilitate theory-based subject analysis for ETD geographic subject metadata, based on data interpretation and analyses.

CHAPTER 8: PROPOSED MODEL

8.1   Introduction

This chapter presents a proposed model to facilitate theory-based subject analysis for ETD geographic subject metadata. A summary of the findings and conclusions drawn on the effects of subject analysis theories on ETD subject metadata creation is presented as a background to the proposed model. Based on the reviewed literature and the findings of this study, a model is proposed with the aim of promoting effective subject analysis approaches to ETD geographic subject metadata creation.

The boundary object theory (Star & Griesemer 1989) is used in combination with the conceptual model for subject analysis (Albrachtsen 1993) to suggest a conceptual model. The model addresses the problem of theory integration during subject analysis to facilitate the creation of effective geographic subject metadata that promotes wide usage of ETDs to address varied information needs. The created geographic subject metadata can act as boundary objects, sitting at the boundaries of different geographic information-related needs of people from different social worlds. They make the information available not only to researchers at universities, but also to other user groups, who may use ETDs information in different ways.

8.2   Summary of the overall findings

The reviewed literature and the results of the investigation indicate that various factors influence the approaches used for subject analysis and geographic subject metadata creation. In addition, a gap was found between theory knowledge and practice. However, the findings indicated a consensus on the need for theory knowledge and implementation to facilitate the creation of effective ETD geographic subject metadata. The information provide a basis for the proposed model.

At this point, the research findings and conclusions on the research can be summarised as follows:

- Subject metadata is generally used for the description of subject content of ETDs. Geographic subject metadata is created by different libraries, but they are not provided in all instances where they are applicable. The records analysis revealed that some of the research on ETDs focuses on specific geographic areas, but the bibliographic descriptions do not include the geographic subject metadata to describe their contents.

- The findings revealed that professional librarians were currently the primary subject metadata creators and that metadata creators based their approaches on implicit knowledge, primarily gathered from past cataloguing experience. Practices in geographic subject metadata creation varied, because implicit knowledge is not shared by all metadata creators and was only applied according to an individual's understanding. It appeared not to address the problems of inappropriate subject analysis and inconsistent practices.

- There were different understandings among subject metadata creators on how theoretical principles should apply to subject analysis and influence geographic subject metadata creation. Analysis focus was often based on different conceptions held by the actors in metadata creation.

- Although the empirical findings of this study were limited to the libraries that participated in the study, they do have the potential to be transferred to the context of the different university libraries in South Africa.

Hybrid approaches emerged as the preferred option to ETD geographic subject metadata creation. The proposed model promotes this view and suggests a framework that can serve as a basis for the subject analysis approaches and the creation of the ETDs geographic subject metadata.

8.3 Rationale for the model

Different reasons can be advanced to justify the significance of the proposed model. The following issues are addressed by the model:

- The boundary object theory is used to illustrate the nature of the outcome of subject analysis in the form of geographic subject metadata as boundary objects, with the potential for use to facilitate information discovery to address varied information needs.

- The model will help to outline the conceptions of subject analysis and the effect of subject analysis on the creation ETD subject metadata. The improved knowledge will raise awareness of the effect of theory by explicitly imparting how the theoretical principles can reconcile the different subject analysis approaches.

- Collaboration on subject metadata creation and how it can be achieved among different actors can be learned from the model. In addition, the model highlights the mediation role, while demonstrating the value added by professional librarians to maintain integrity. Albrachtsen (1993:223) alludes to the importance of maintaining consciousness of the impact of our profession to mediating knowledge. Advances in digital publications introduced the possibility of different collaborations in subject metadata creation. The model will show how the different collaborations can be managed; particularly during subject analysis.

- The essential aspects that characterise subject analysis for ETD geographic subject metadata creation are illustrated by the proposed model. The impact of theory is indicated to indicate the effect on the effectiveness and efficiency of the current systems.

- The model will provide a basis to benchmark approaches to geographic subject metadata creation and other descriptive metadata and how automated systems can be developed and implemented to facilitate effective subject metadata creation.

- The presented model can shed information to help decision-makers in policy design by demonstrating the need to integrate theory in ETD subject metadata creation.

Key findings from the literature review, the shared practitioners' perspectives and the content analysis findings revealed the need to build on the theoretical knowledge to provide a basis for enhanced subject analysis and subject metadata creation practices.

The model illustrates alternative approaches to geographic subject metadata for theses and dissertations in digital format. The model provides a representation of how theory can be integrated throughout the different steps of subject analysis and geographic subject metadata creation and how the theoretical principles can be applied to enhance current practices. Additionally, the model is aimed at providing a basis for a shared understanding of subject analysis approaches and support effective and consistent practices among metadata creators across the different university libraries in South Africa.

8.4   Background for the model

This section presents an overview of the model, which addresses Research Objective 4 (RO4):

To establish what type of model can promote effective ETD geographic subject metadata creation and maintenance in South African university libraries.

A model is a theoretical description of a process or system to help understand or explain how it works. It is a theoretical representation that can be applied to different contexts against which ETD subject metadata creation processes can be evaluated. A good model should be flexible and implementable in different contexts (Zeng, Zumer & Salaba 2010:44).

The proposed model aims at providing a basis for practice and, although its application in different situations remains optional, the model is adaptable to the changes in the information provision landscape. The model is not a workflow process plan, but a simplified framework that can be developed further and integrated into business processes to improve the processes for professional subject metadata creators and to guide them in the integration of contributions by other types of actors and potential actors.

Furthermore, the model shows how the different stages of the geographic subject metadata development are characterised by integration of theory and how different boundary objects can be used to break the barriers to facilitate collaboration in the process of creating the geographic subject metadata. The different social contexts, different viewpoints and the outlying boundary objects that characterise the processes are also outlined.

Consistent with existing bibliographic models, the conceptualisation of "subject" as all the potential uses of a resource informs the viewpoints expressed in the proposed model. The concept itself is considered a boundary object that is understood differently by people from different backgrounds. Theory knowledge supports the understanding of the different perspectives and practices that are coordinated in an effort to create effective ETDs subject metadata, with the potential for broad usage.

8.5   Application of the subject analysis conceptual model

8.5.1   Introduction

The subject analysis conceptual model (Albrachtsen 1993:220) identifies the implicit and explicit theoretical viewpoints that may need to be negotiated during the analysis phase of the process of creating the subject metadata. The viewpoints expressed in the form of analysis focus are theorised in this study as boundaries to be bridged during subject analysis. Furthermore, the different analysis focuses or orientations are understood to lead to the development of subject metadata that transcend different information use boundaries. The subject analysis can also serve as a boundary object, because the process is often interpreted differently by different people and overlapping forms of practices are evident in the literature and in the findings of this study. The process sits at the boundaries of different theoretical viewpoints and interpretations.

The outlined subject analysis conceptions presented in the proposed model are not entirely new. They are associated with the conceptions outlined in the conceptual model of subject analysis discussed in literature by, among others, Hjorland (2021), Albrachtsen (1993; 2015), Fourie (2008) and Tennis (2005). The different conceptions are presented to clarify how theory and practice merge to support varied analysis practices.

The conceptual analysis model was used as a basis for the interpretation of where concentration is based during subject analysis and the different perceptions about the appropriate focal points.

The four conceptions of content-focus, author-focus, user-focus and domain-focus were found to be currently used by metadata creators, with the first approach being dominant. The experiences, perspectives and existing ETD descriptive records generally revealed that it was mainly the content-focus and uncontrolled terms that were dominant. However, a hybrid approach was considered suitable for the creation of appropriate geographic subject metadata for ETDs.

The gap between explicit and implicit knowledge in subject analysis theory basis and the effect in ETD geographic subject metadata creation are addressed through the model. The findings of this study revealed that metadata creators based their analysis approaches primarily on implicit knowledge, commonly gathered from past cataloguing experience. According to Bieber et al (2002:13), implicit knowledge exists in the minds of community members and can be shared with others through various means, including being expressed in the form of explicit concepts and being made explicit by being applied in real practice. The model aims to improve awareness on how theoretical knowledge is made explicit in subject analysis and subject metadata creation and its effect on the processes.

The research findings showed that metadata creators considered that theory knowledge can enhance subject analysis and subject metadata creation. Furthermore, it is considered that theory knowledge can help to mediate the analysis practices of different library staff and other metadata contributors or actors and facilitate collaboration in ETD subject metadata creation. In the existing literature, librarians are theorised as being boundary staff (Montoya 2017) who manages the integration of different practices that are adopted within the libraries.

In addition, theory-based standards and guidelines that are developed and managed by the professional metadata creators have a fundamental effect on the geographic subject metadata creation process and the outcomes of the process.

The findings of this study showed that more awareness needs to be created among the current library metadata creators about the need to bridge the theory and practice gap, in order to enhance the analysis conducted for the creation of ETD geographic subject metadata.

The proposed model embraces different actors in subject metadata creation. The current and potential contributors to the creation of the theorised boundary objects in the form of the ETD geographic subject metadata are indicated. Furthermore, it is suggested that metadata is a product of actors whose perspectives and practices differ, to the different social worlds where the actors are grounded. ETD metadata is primarily created and managed by the libraries of the different universities where the works are produced.

The study revealed that professional librarians were mainly involved in the creation and management and maintenance of subject metadata. Other contributors included the authors who submitted keywords with their ETDs. Additionally, users were viewed as potential contributors (Earle 2014) of metadata to describe library items, which is a view illustrated in the model as a potential additional way for creating ETD subject metadata. The use of automated systems to support keyword searches is also outlined as an alternative way of making subject metadata accessible.

The divergent perspectives about subject analysis and ETD geographic subject metadata are managed and synthesised by applying the principles of boundary object theory as a basis for the recommended model in this study.

8.5.2   Application of the boundary objects theory

The boundary object theory is used as a theoretical framework to explain the boundaries that geographic subject metadata bridges to enhance discoverability of ETD content. It is employed to indicate that the geographic subject metadata can serve as boundary objects to enhance the discovery of ETD content to satisfy information needs within different contexts.

Star and Griesemer (1989:393) define *boundary objects* as "… objects that are both plastic enough to adapt to local needs and the constraints of several parties employing them, yet robust enough to maintain a common identity across sites". Boundary objects are understood as being concrete or abstract and the boundaries can be tangible or intangible, real or imagined, and have different meanings within different social worlds (Huvila et al 2016:7; Star & Griesemer 1989:393). Star and Griesemer (1989:388) use the concept "social worlds" to identify the different groups or communities of practice and their varied backgrounds.

Objects are defined as something people act toward and with (Star 2010:603). Knowledge and appropriate implementation of subject analysis theories and associated theories help to create geographic subject metadata that crosses the different boundaries to make ETD content discoverable across a broad user community.

Boundaries can be regarded as objects that facilitate cooperation by fostering mutual understanding (Huvila et al 2016:2) and flexible enough to accommodate different practices. Montoya (2017:217) and Star (2010:602) define a *boundary* as the negotiated or shared space connecting different social worlds. The bridging of different boundaries is necessary to create functional metadata. Geographic subject metadata, which is the outcome of different practices and assistive technologies, is addressed in the model as boundary objects with a potential for varied uses. The understanding of subject metadata as boundary objects that are open to different interpretations or "interpretative flexibility" (Star 2010:602) reveals the potential for varied applications across different contexts. In addition, other boundaries, e.g. between human created and automatic facilitated subject metadata, are addressed in the proposed model.

The proposed model is not prescriptive, but primarily aims at showing how subject analysis theory and practice gaps can be addressed to make the current subject metadata creation approaches relevant in the context of ETDs. Its applicability can be tested across different libraries involved in ETD metadata creation and management. New, emerging boundaries are anticipated, due to developments that allow different social worlds to cooperate around ETD subject metadata creation, which may necessitate expansion or changes on parts of the proposed model.

Huvila et al (2016:4) state that varied artefacts, including activities and products, are theorised in literature as boundary objects. The author's viewpoint supports the stance in this study about the relevance of the items, the geographic subject metadata, which are suggested as the main boundary objects. These items are the products of the subject analysis process.

The analysis and subject metadata development processes are also considered boundary objects. The proposed model illustrates how boundaries are bridged during subject analysis and throughout the geographic subject metadata creation cycle. The arguments in the model are aimed at supporting the discovery of ETDs to satisfy the information needs of different people from varied social worlds.

8.6   Presentation and discussion of the proposed model

The proposed model is composed of different connecting parts that illustrate theory integration within the subject analysis and ETD geographic subject metadata creation cycle. Additionally, the model aid the understanding of the concepts, components and processes of subject analysis and ETD subject metadata creation. The inclusion the different components is also aimed to demonstrate the potential application of the model to the complete cycle. This is a basic model that can apply across the university libraries to provide a framework for best practices in ETD subject metadata creation.

The model was adapted from Star and Griesemer (1989:390). The key components of the boundary objects theory that are used in this model are: Interpretative flexibility, translation (key activities of standardisation of methods) and development of boundary objects (Star 2010:601; Star & Griesemer 1989). The use of these concepts in the study was informed by the meanings used by the above-mentioned authors. The concepts are discussed and explained in Sections, 8.6.1–8.6.4. The concepts in the boundary object theory are used in the model, which illustrates coherence in coexistence of varied forms of geographic subject metadata and cooperation between different allies working together for a common purpose of providing subject access to ETD content.

The components involved in the proposed model illustrated in Figure 8.1 (p. 311) are discussed under separate topics in this section. The model relates to other existing, globally accepted conceptual models for bibliographic and authority data, e.g. the IFLA FR family of conceptual models and the consolidated Library Reference Model (LRM).
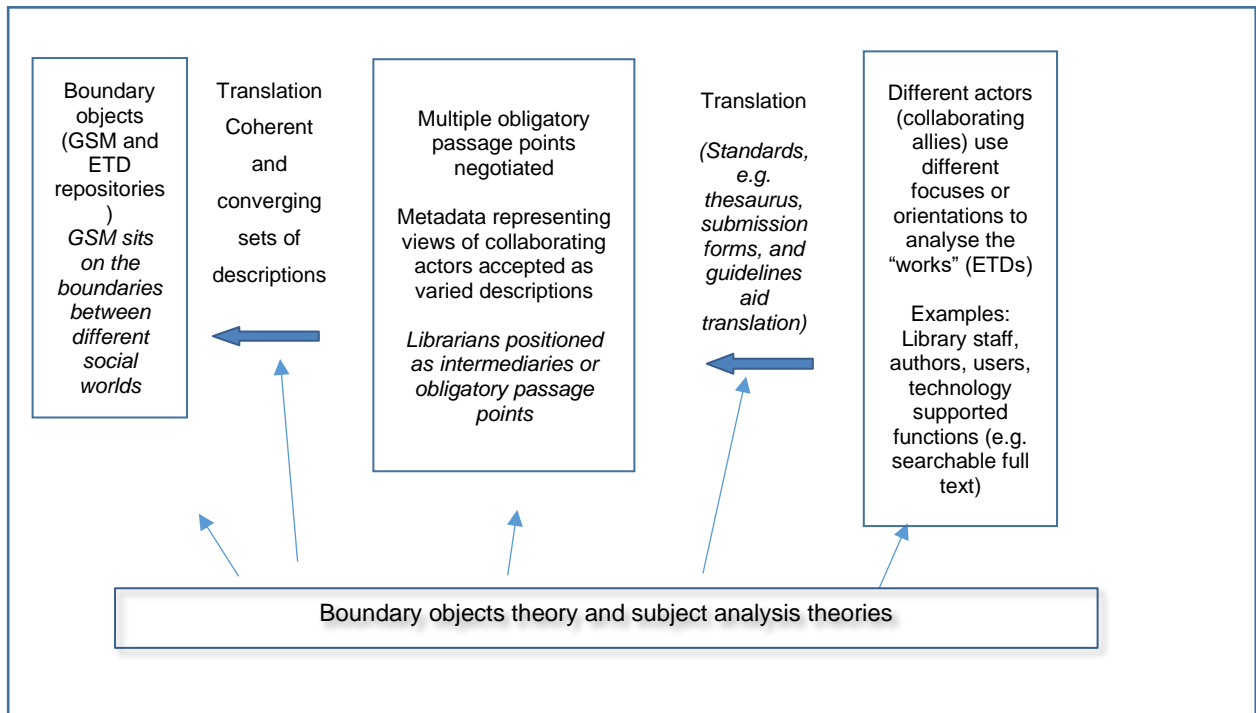
Figure 8.1: Flipped diagram of development of boundary objects and integration of theory

The boundary object theory, combined with the subject analysis conceptual model, provides a basis for the proposed model. The blending demonstrates integration of theory and practice during the different phases of subject analysis and the development of boundary objects. The interpretation of geographic subject metadata occurs from different points of view in the different communities that use them to search for information.

Furthermore, the model suggests solutions for the development and management of the ETD subject metadata. The geographic subject metadata is used as an example with the view that the proposed solutions can apply to different forms of subject metadata. The model illustrates that geographic subject metadata serves as boundary objects between different user groups with potential for varied "use" purposes. Boundary objects can be shared or are sharable across boundaries (Kertcher & Coslor 2020:83). For example, the metadata can be used to bridge the boundaries within different discipline domains in an academic environment and facilitate interdisciplinary use of the research outputs. Metadata can be characterised as boundary objects with different meanings for different people in different worlds (Huvila et al 2016:2), which serves as a basis for facilitation of different purposes.

Finally, the model reflects the different stages in the ETD subject metadata creation phases, starting with subject analysis and the influencing theoretical perspectives and ending in the creation of subject metadata that serves as boundary objects, with a potential for wide usage. The concept of subject analysis itself is considered as a boundary object that is interpreted differently by different participants and performed in different ways to facilitate subject metadata creation. The analysis process sits at the boundaries of the different theoretical perspectives that need to be coordinated to accommodate the varied focuses or orientations of subject analysis.

8.6.1 Interpretative flexibility and subject metadata creation

The concept of interpretative flexibility is applied to the works, actors and subject analysis process and the geographic subject metadata, which are considered to serve as boundary objects of importance in the illustrated model. Huvila et al (2016: 4, 6) present an analysis of literature that reveals different types of boundary objects, including activities, methods, concepts and documents.

8.6.1.1 The "work" as a boundary object

The model presents the "work", which is interpreted as a boundary object with potential for varied uses. In this study, the "works" are the ETDs. Star and Griesemer's (1989:396) assertion that scientific publications are boundary objects affirms the notion that works can be considered as boundary objects.

The content contained in the works may hold different meanings to different people and address different information needs. The subjects of ETDs sit at the intersections or boundaries of different subject disciplines and can be of relevance to different communities of practice, including the general public. Any work may have more than one theme, as their subjects and content can be used to address different purposes (IFLA 2009, Riva, Le Boeuf & Zumer 2017:88*).* Place is an example of entities that can be the subject of a work that can be used to create geographic subject metadata.

8.6.1.2 Actors or allies in subject analysis

The concept "actors" was used by Preminger, Rype, Adland, Massey and Talleras (2020:131) and the IFLA Working Group on Metadata for Digital Resources (International Federation of Library Associations and Institutions. Working Group on Metadata for Digital Resources 2009:10) to identify creators of metadata. The actors or allies in subject metadata creation can be human or nonhuman (technology facilitated). Some of the existing and potential metadata creators can be part of different social worlds. For example, an author can be aligned to different subject disciplines and perceptions during subject analysis and subject metadata creation will be influenced by the multidisciplinary affiliations. The librarian can also serve as a boundary object responsible to facilitate subject descriptions for multiple user communities.

In addition, relationships are formed between the metadata creators performing the "analysis" to create subject metadata. For example, library staff can collaborate within their organisation or with other libraries or external allies to produce subject metadata (boundary objects) (Albrachtsen & Jacob 1998:310). The proposed model indicates the possibility of similar collaborations for the creation of geographic subject metadata. It is important for librarians to understand the role of other, potential metadata creators (actors) and to adapt to the changes in metadata creation approaches, mainly driven by technology. The analysis approaches can always change, based on contextual changes and/or trends, e.g. the emergence of electronic resources has enabled manipulation of documents and full-text keyword searches.

Collaboration of the different contributors or actors in subject metadata creation begins in the analysis stage. The subject analyses model of conceptions helps to explain the focuses that are maintained during the analysis stage and how they facilitate the creation of subject metadata that serves as boundary objects. The theoretical perspectives and backgrounds of the different actors influence their analysis practices. An understanding of the theoretical basis for different actions is important for the mediating librarians.

8.6.1.3   Subject analysis process as a boundary object

This model suggests that the subject analysis process is a boundary object, due to the different interpretations that it involves. Subject analysis is open to interpretative flexibility. The findings of this study indicated that a hybrid approach to analysis orientations was preferred as a best practice for ETDs for subject description. Different approaches to the process of subject analysis may need to be negotiated within a single group of subject metadata creators from a common environment or among collaborating communities.

Subject analysis sits at the boundaries of the different actors' conceptions of "subject". The different conceptions are informed by the theoretical positions, which are critical to influence the nature of focus maintained during subject analysis. The conceptions of what a "subject" is and what the work is about have an effect on the analysis focuses or orientations. The findings of this study confirmed that the different focuses that are maintained during subject analysis have implicit and explicit links to theory. Understanding the effect of theories may help to manage the collaboration and to facilitate mediation between the different actors and the different focuses they maintain during subject analysis.

The results of this study indicated that currently, one of the dominant local practices was to focus the analysis for subject descriptions of ETDs to support the information needs within the different subject domains at a university. However, the perspectives that other forms, including the general user-focused analysis, are important have also emerged out of the findings. The boundary objects theory can help to manage the coexistence of the different practices to enhance the discovery of ETD content.

Furthermore, the model can benefit the professional metadata creators from its description of processes and their outcomes from the perspective of the boundary objects theory, to enhance their role as mediators for convergence and coherence of the different perspectives, practices and varied subject metadata. It can also have an influence on the directions to be followed in the development of subject analysis policies, tools and systems.

Effective management of integration of theory and practice can help to bridge the different boundaries and provide the basis for translation through the use different theory-informed standardisation mechanisms among different libraries and other environments.

8.6.2   Role of translation in the ETD subject metadata creation cycle

Translation can play an important role in standardising the different methods used by metadata creators from different social worlds and to control the quality and integrity of metadata of different origins. Translation may involve management of the interpretation of the meaning of a "subject" and how it is to be represented in the form of subject metadata.

The different metadata creators' viewpoints can be managed through the use of standards. For example, Albrachtsen and Jacob (1998:294) suggest the use of an end-user thesaurus. Thesauri or other forms of controlled vocabularies can facilitate translation through the use of common vocabulary. Applying controlled vocabulary is a form of standardisation that can enhance consistency of subject descriptions. Standards mediate between different groups (Preminger et al 2020:138). However, a challenge may arise concerning how the different actors respond to the translation or standardisation efforts. Standards can also serve as boundary objects that can be applied differently by different communities of practice (Preminger et al 2020:130). It is of importance to this study to understand that standards incorporate theory. Their adoption suggests the use of theory during subject analysis.

In the course of translation, it becomes important for the librarians to have knowledge of the theoretical principles that support the different analysis viewpoints and translation standards, in order to play an effective mediation role and apply the standards appropriately. The librarians interpret the theory knowledge on behalf of other actors by simplifying standards that can be applied by non-experts. As indicated by Star and Griesemer (1989:387), some of the efforts in effecting mediation involve translation and simplification. Standard forms for submitting subject metadata are an additional way of simplifying the metadata creation process and facilitating the merging of different viewpoints.

Standardised methods can facilitate interactions and cooperation amongst the diverse actors. Standards promote consistency, which is one of the important quality standards for functional subject metadata identified the literature review and the findings of this study. Guidelines on standard forms of subject description can be used as a means of translation. However, they sit at the different boundaries of interpretations by individuals and groups from different social worlds.

8.6.3 Professional metadata creators as obligatory passage points for ETD subject metadata

Despite the important role played by the translation processes, the need to maintain integrity of the descriptions necessitates the establishment of obligatory passage points (Star & Griesemer 1989:388) to standardise them. Library staff members serve as obligatory passage points, using their expert knowledge for mediation to facilitate convergence or reconciling the different subject analysis approaches. Their implicit or explicit theoretical perspective on "aboutness" provides a basis for the mediation role. Additionally, they support the judgements made about appropriate analysis and outcomes. Their mediation role allows the coexistence of different forms of analysis focuses and descriptions (share boundaries), e.g. both controlled and free text forms of heading can be used for subject description. The professional metadata creators serve as obligatory passage points, where viewpoints are negotiated (Star & Griesemer 1989:390) to facilitate integration (mixing) of descriptions from different cooperating social worlds.

The free-text descriptions, which are commonly created by actors outside the library environment and do not always comply with recognised standards, are authorised by the library staff as descriptions, coexisting with standardised forms. Control is exercised as to what is considered functional geographic subject metadata.

Librarians play an important role in conducting quality control. Their role is not to be restrictive, but rather an intermediary one to support quality and functionality. Star and Griesemer (1989:406) express the view that work done by non-experts can be open to analysis by professionals.

Although simple metadata is encouraged over authoritative forms in the digital world, the importance of quality and authority control in improving discoverability and access cannot be overemphasised. The view maintained in this study was that the blending of ideas regarding simplicity and control of subject descriptions is necessary.

The boundary objects theory alludes to the possibility of multiple translation points (Star & Griesemer 1989:392). The proposed model reflects these multiple translations at the different points in the metadata creation process. After successfully passing the obligatory passage points, translation occurs to ensure that the different negotiated analysis viewpoints are reconciled into functional subject access points that serve the common goal of geographic subject descriptions. The resultant, merged approaches are, as described by Kertcher and Coslor (2020:79), the products of resolved conflicts. The ETDs repositories analysed in this study reflected the use of the different forms of subject metadata that coexists within the records.

At the end of the cycle illustrated in the proposed model, co-created geographic subject metadata, from actors with different viewpoints and interests, is produced and can be accepted as hybrid access points. A related proposal is made by Alemu (2016), who recommends a metadata enriching metadata theory, which integrates librarian-created and user generated metadata.

8.6.4 Geographic subject metadata as boundary objects

The created geographic subject metadata serves as boundary objects between different use purposes. As boundary objects, geographic subject metadata is open to interpretative flexibility. The metadata can be interpreted differently and be used for different purposes.

ETD geographic subject metadata can serve as boundary objects that help the discovery of information of geographic nature to satisfy the information needs of people from different social worlds. These can be considered as boundary-objects-in-use (Huvila et al 2016:6).

It is important that the reliability of subject analysis and subject metadata be maintained in the midst of diversity. Mediation is important, since the actors from the different social worlds may not have the same expertise and may not be required to understand the different professional standards. Many other outlying categories of boundary objects are possible to emerge with developments in digital publishing and emerging metadata creation practices.

8.7 Conclusions and recommendations on the proposed model

The proposed model is not aimed at being prescriptive, but to contribute ideas that can assist further developments regarding geographic subject creation and the effective subject analysis practices. The model can be adapted to provide a framework for effective ETD subject description within different contexts. A theoretical basis is provided to aid the synthesis of varied approaches that aims at a common purpose of improving the discovery of the ETD subject contents. The proposed model is related to other conceptual models, like the IFLA LRM model.

The model clarifies the mapping of theory and practice and application in different contexts. The key orientations identified in this study and the theories informing them build on the knowledge about subject analysis. The different analysis conceptions serve as boundary objects aligning different analysis positions, including the beliefs about automated processes. They facilitate the creation of geographic subject metadata of varied nature that is synthesised into functional description of ETD content. The findings of this study showed a strong perspective for a hybrid approach of using different types of subject metadata that originates from varied sources and bridges the boundaries of different use purposes. The proposed model aligns with the hybrid activities and various geographic subject metadata.

8.8 Chapter summary

The chapter presented the findings of the study as a basis for the discussion of the proposed model. The model was based on the conceptions described by the subject analysis conceptual model and the principles of the boundary objects theory.

Different phases of the illustrated model showed how theory was integrated in the ETD geographic subject metadata creation cycle. The final outcomes of subject analysis in the form of the geographic subject metadata, which can serve as boundary objects, were discussed.

As the last chapter, Chapter 9 presents the final conclusions and recommendations of the study.

CHAPTER 9: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

9.1 Introduction

This chapter presents a summary of the findings and conclusions drawn on the effects of subject analysis theories on ETD subject metadata creation as a background for the concluding discussions of this study. The chapter also outlines the limitations and implications of the findings; presents the contributions to practice and theory development for effective subject analysis and ETD geographic subject metadata creation; and makes suggestions for further research.

9.2 Summary of the findings and conclusions

The aim of this study was to investigate the approaches that are being followed during subject analysis and how subject analysis theories are applied to facilitate geographic subject metadata creation for ETDs. The results of this study indicated that varied approaches are followed for subject analysis and geographic subject metadata creation. In addition, a gap was found between theory knowledge and practice during subject analysis. However, the findings indicated a recognition of the need for theory knowledge and implementation to facilitate the creation of effective ETD geographic subject metadata. A summary of the findings and conclusions based on the study problem is presented.

9.2.1 Usage of ETD geographic subject metadata

The first research question was: *To what extent and in what ways is geographic subject metadata used for description of ETDs in South African university libraries?*

The findings of this study confirmed that geographic subject metadata is used for description of ETDs. However, the extent and ways of usage varies among different university libraries in South Africa. Both the quantitative and qualitative findings confirm varied usage approaches. Geographic subject metadata is not created in all instances where it is applicable.

The conclusion is that the current subject analysis approaches lead to inadequate practices in the use of geographic subject metadata to describe the subject content of ETDs. The current description practices need to be improved to increase the extent to which geographic subject metadata is assigned when it is necessary to do so. The theories of subject analysis support the assignment of geographic subject descriptions to enhance the discoverability and access to information resources. There is a gap between theory and practice, in that some ETDs that are focused on specific geographic localities have not been assigned geographic subject metadata. This was confirmed by the analysed records, which revealed ETDs are not always assigned geographic subject metadata even when it was necessary.

The reviewed literature also indicates that, after the translations stage of subject analysis, subject terms are assigned from a controlled vocabulary or from natural language. This is consistent with the empirical findings of this study, which revealed that this hybrid approach is practiced in the different universities, with more dominance of uncontrolled terms used. However, geographic subject metadata continues to be inadequately assigned to ETDs.

Access to the geographic information contained in the ETDs is currently not optimally supported. It is important to implement ways of improving the current practices to enhance the discoverability of the ETD content. The extent of usage and the nature of assigned subject terms appropriate levels of descriptions for ETDs need to be supported by policies and guidelines that are informed by relevant theories to improve current subject analysis and ETD subject metadata practices.

9.2.2 Current approaches for ETD geographic subject metadata

The second research question was: *What are the current approaches being followed to create the geographic subject metadata for ETDs in South African university libraries?*

The findings revealed that professional librarians are currently the primary subject metadata creators. This is consistent with the previous findings reported in the reviewed literature.

Furthermore, it was found that metadata creators base their approaches on implicit knowledge, primarily gathered from past cataloguing experience. Other common subject metadata (keywords) contributors include the authors of the theses and dissertations. The findings show support for professionally created metadata for value added descriptions. There is variation in terms of the use of controlled vocabulary and natural language or free text descriptions by the different libraries in South Africa.

The conclusion drawn is that the findings are indicative of the varied practices within and across the institutions that participated in this study. Knowledge of subject analysis theory is implicit and awareness of its effect cannot be confirmed among all metadata creators. Due to inconsistent practices, quality and integrity of the current description may not be guaranteed in all cases in which the subject metadata is created among different institutions. Collaboration across libraries and interoperability across systems may be hampered by the varying practices.

9.2.3 The effect of current analysis approaches

The third research question that was answered by the findings of this study was: *How and why do the current analysis approaches being followed during geographic metadata creation for ETDs in South African university libraries affect the process?*

The findings demonstrated comprehensively that the current analysis approaches have an effect on the success of subject analysis and the quality of geographic subject metadata created for ETDs. The current metadata creators are aware of the effect of their approaches. However, it was found that the metadata creators do not critically reflect on the effect of subject analysis theories during their metadata creation. There are different understandings among subject metadata creators on how theoretical principles impact on subject analysis and how they influence geographic subject metadata creation. As a result, analysis focus is often based on the different conceptions of subject analysis held by the metadata creators. Furthermore, implicit knowledge is dominantly used as a basis for the varied practices, although the knowledge is not made explicit to guide novices and to assist consistency of practices.

It was generally found that the experiences, perspectives and existing ETD descriptive records are linked to theoretical models in literature (Hjorland 2021; Albrachtsen 2015; Fourie 2008; Tennis 2005). The association of current approaches to theory is confirmed by this finding, although it was generally implicit. The findings showed that the content-focus and uncontrolled or free text terms are dominantly used. However, a hybrid approach for the creation of appropriate geographic subject metadata for the ETDs is regarded as the most appropriate approach by the metadata creators.

It is concluded that the current approaches have an effect on subject analysis for ETD geographic subject metadata creation. The current approaches are based on implicit theory knowledge, which is not sufficient to enhance current practices and improve ETD subject metadata creation in general. The research question was answered, indicating that there is a noticeable need for improvement of current approaches to have a meaningful impact on the appropriateness of the created geographic subject metadata. This was confirmed by the existing ETD records that were analysed. Although the findings are limited to the libraries that participated in this study, they have the potential of being transferable to the context of the different university libraries in South Africa and other different contexts.

Based on the major findings of the three research methods employed in the study, the general conclusion can be made that geographic subject metadata is a highly significant descriptive element to improve the discoverability of ETDs. Additionally, policies need to be developed, reviewed and enforced to promote consistency. Subject analysis approaches for ETDs were found to be varied among different libraries and sometimes even within the same institution. Furthermore, it is concluded that, despite the preference for founding subject analysis on appropriate theories, it was not explicitly shown how theories were integrated in practice. The findings on the subject analysis approaches gave an indication of the theoretical positions adopted by the ETD geographic subject metadata creators.

Awareness and knowledge of the theoretical foundation and the different conceptions of subject analysis need to be improved among ETD subject metadata a creators, so as to promote consistency and effectiveness of the metadata.

The research problem of the study was addressed sufficiently by using the different methods, which allowed different perspectives to be gathered on the same issues. The problem involved this study was to investigate the way in which subject metadata creators analyse information resources content to determine geographic subject metadata for electronic theses and dissertations – in particular the way in which subject analysis theories were applied and the role that they played in subject description. The recommended model illustrates how theories can be integrated for subject analysis and effective subject metadata creation.

9.2.4 Recommended model

The study also had to answer the fourth research question*:*

How can the findings of this study be applied to develop a model for subject analysis and creation of geographic subject metadata for ETDs in South African university libraries?

Based on the research findings of this study, a model was recommended in Chapter 8, based on the subject analysis conceptual model and the boundary objects theory. The model is considered as a useful basis for improving subject analysis and integrating theory to improve the quality of ETD geographic subject metadata.

9.3 Implications of findings

The use of mixed methods to conduct the study resulted in findings that gave an in-depth understanding of the problem involved in the research. Optimal discoverability of digital collections, including ETDs, is one of the key performance standards of university libraries. Theory of subject analysis needs to be reviewed continuously in light of the changes affecting common description practices. This work has implications for practice and theory development.

The research findings of the study contributed to the theoretical foundations of subject analysis. It was found that metadata creators were aware of the theoretical implications, although they did not always consciously link theory and practice during subject analysis. As a result, the direct impact of theory was not explicit.

This practice may lead to poor discoverability of ETD content; particularly in the collocation function in terms of content of geographic relevance.

The empirical findings facilitated the discussions on varied approaches to subject analysis and geographic subject metadata and informed the focus of the interpretations and recommendations on the management of diversity and cooperation. Subject metadata can be used across different contexts and by a broad user community. Therefore, metadata is considered as boundary objects, created by different actors and crossing the boundaries of different use interests.

The different Interpretations (interpretative flexibility) are considered to account for the differences in practice found in this study. Consensus may not be expected in all cases, but integrity remains a crucial measurement criterion for appropriate outputs in the form of the geographic subject metadata.

9.4   Contributions of this study

The implications or contributions to practice and theory development for effective subject analysis and ETD geographic subject metadata creation are discussed in this section.

9.4.1   Implications for theory development

This study contributed to subject analysis and ETD subject metadata creation theory by highlighting the importance of theory as an underlying basis for practice. The synthesised approach of using the subject analysis conceptual model and the boundary objects theory successfully demonstrated how integration of theory into subject analysis can be managed. The approach also helped to interpret the different metadata creators' experiences and perspective regarding subject analysis and ETD subject metadata. The information can help the review of theories and conceptions of subject analysis to align with the digital environment.

In addition, the findings indicated the need to focus LIS research on how different types of aspects are considered as boundary objects. Staff, practices, standards and technological systems can be viewed from the perspective of helping to improve subject analysis and subject metadata creation, e.g. staff may continuously need new

skills sets to operate in hybrid environment of print and digital library collections management.

Theoretical studies on subject analysis and boundary objects will benefit from the findings of this study by using the empirical evidence to interpret new findings. This study also has implications for decision-making and policy implementation in hybrid environments by highlighting alternative approaches to ETD subject metadata creation. Although the proposed model has not been tested in practice, it does hold the potential for successful theory development for subject metadata creation and other digitisation projects.

The recommended model is anticipated as being of help to bridge the gap between theory and practice. The theories used helped to illustrate the processes and explain how theory integrates in the different phases of subject analysis and subject metadata creation.

9.4.2   Contributions to policy and practice

Possible policy and practical implications of the results of this study are presented in this section. The implications are viewed in the context of ETD subject metadata.

The findings of this study provided important information to inform policies and practices in subject analysis and subject metadata creation in general. These findings can be used to inform policy development and reviews for South African university libraries. It was found that policies were non-existent in some of the participating libraries. Additionally, existing policies were inadequate in terms of guiding geographic subject metadata creation for ETDs. Policies should outline the levels of descriptions, as well as the necessary information to include in subject description, based on the adopted schemes involved in effective subject content description, e.g. the Dublin Core and the associated standards like the RDA and LCSH.

The findings supported the view that different metadata creators' contributions can be coordinated to improve ETD geographic subject and subject metadata creation in general. Coordinated effort in university libraries and across different libraries may help to achieve a coherent way of subject metadata creation.

The coordination of efforts across boundaries is viewed a viable problem-solving initiative (Star & Griesemer 1989:392). The librarians are considered obligatory passage points or intermediaries translating between different viewpoints and other boundaries to facilitate effective geographic subject metadata creation on behalf of the information users. People and objects can serve as boundary objects (Huvila et al 2016:10).

Collaboration between professional metadata creators and other subject metadata contributors can be improved to remain relevant to changes and developments in digital resources description and information user behaviour. However, professionals are expected to apply theory, while other partners will benefit from their intervention, e.g. from making updated standards like controlled vocabulary easily accessible and translated to simple forms for novices, such as new authors and for the general public.

Coherence of divergent approaches can be achieved and maintained through standardisation and development of boundary objects (Star & Griesemer 1989:392), which, in the case of this study, are ETD geographic subject metadata. Libraries can apply the proposed model to make theory knowledge explicit to facilitate the development of standards and to aid consistency of practices among the metadata creators in the different libraries, across different libraries and other categories of metadata contributors.

Subject metadata is often built by using terms from a controlled vocabulary. Different schema used for ETDs recommend the use of this approach. The research findings will guide the way in which metadata description schema fields should be set up in local libraries to accommodate the varied forms of metadata. In addition, ETD metadata should clearly indicate the nature of the subject terms used to facilitate decisions for fields to be harvested for shared metadata. Interoperability, which involves sharing across different systems, requires consistency to support metadata sharing. Making theories explicit can facilitate decision-making to aid consistency of practices amongst the metadata creators.

Linked data is a current form of value added data that can benefit from well controlled subject metadata; particularly geographic subject metadata. Assigning geographic subject metadata from controlled vocabularies and authority subjects can improve the use of linked data functionalities in subject descriptions. The findings of this study indicated that the application of controlled subject metadata is considered as helping to improve the discoverability of ETD geographic information. Studies from the users' perspective should contribute to learning more about the user's perspective.

Furthermore, practical implications for ETD repositories management involve bringing subject created metadata created from different sources together under one platform, which, in turn, will support collaborative metadata creation. This will also help the current practices to adapt to current digital publishing and the new possibilities it creates and where collaborative work of different forms is made possible.

The improved awareness of theoretical knowledge will assist making informed decisions on how to improve the current forms of author-supplied keywords. The authors may not be expected to have expertise in subject metadata creation, but the professional metadata creators can use their theory knowledge to offer different forms of translation assistance. The importance of basic author-contributed metadata is confirmed in the literature (Maurer, McCutcheon & Schwing 2011). The identified gap of subject analysis theoretical knowledge needs to be addressed to improve practice and to maintain and enhance effective discovery of the important universities' research outputs in the form of ETDs.

The prospects of enhancing the broad usage of geographic subject metadata has been found as being highly valued among current ETD subject metadata creators. This has implications for the review of current practices. All potential ETD information users from different contexts should be assisted creating appropriate subject metadata to find geographic and other types of information from ETDs easily.

In addition, the contributions of this study are considered to have implications for ETD metadata theory literature in general.

9.5   Limitations

The research design resulted in findings that addressed the research problem of this study in a comprehensive manner. Apart from theoretical knowledge, other facts were found to have an effect on ETD subject metadata quality. Common contextual factors, including staffing and volume of work, emerged as some of the impeding factors for appropriate practices. There is no consistency among the libraries with regard to resources – particularly staffing – invested to provide value added metadata. Although these factors are important, they do not have negative implications on the findings of this study, as they were not the focus of this study.

The findings are limited to the libraries that participated in this study, but they have the potential to be transferable to the different university libraries in South Africa. The libraries that did not grant permission were not included in the study. As a result, not all South African university libraries were included in the study.

Investigating the needs of the different communities of metadata creators may yield useful information to clarify the requirements-oriented approach of subject analysis and the implications of theory to metadata creators outside the library environment. The scope of this study did not include the user needs analysis.

9.6   Recommendations

Recommendations for practice and further research are given, based on the research findings of this study and information obtained from the reviewed literature.

9.6.1   Recommendations for practice

A summary of recommendations for practice is supported by the reviewed literature and the empirical findings of this study. It is recommended that the extent and nature of use of ETD geographic subject metadata is improved to enable the collation of important scientific information on specific geographic areas during the retrieval from ETD repositories.

Theoretical knowledge should be improved through continuous education and current practitioners have to be consistent in the extent and nature of the use of the geographic subject metadata.

Other recommendations include the following:

- Policies should be reviewed from time-to-time to reflect support for new approaches, including the hybrid methods in ETD metadata creation.
- The principles of the boundary object theory should be infused in the policies and standards to promote collaboration.

Appropriate subject metadata practices will improve the success of collective efforts in subject metadata creation and the nature of metadata, which has the potential for varied uses. There may not be formal collaboration agreement among the different groups of metadata creators, but translation mechanism by the professional metadata creators will help to provide mechanisms for collaboration, e.g. through the use of a thesaurus and structured submission forms. The concept of "socially constructed boundary objects" – used by Worrall (2014:57) – is considered appropriate to describe the proposed hybrid subject metadata from different communities with the potential to contribute towards ETD subject metadata creation.

Interest is shown in the possibility to collaborate and benefit from the geographic subject metadata contributed by different sources. Automated or automatic aided systems are a viable future alternative for information related services. However, future development of the systems will continue to profit from subject analysis theory, informed human interpretation and input from expert knowledge on information organisation and retrieval.

9.6.2   Suggestions for further research

Recommendations for further research are presented as possible follow-up to the present study. Further research can extend the research findings of perceptions and practices of the various stakeholders, other than library staff, on their understanding of subject analysis and ETD subject metadata.

The research results and findings can also be extended to comparative research with other countries that are at the different stages of ETD implementation.

Additionally, further research can be conducted to explore the use of the findings to improve subject metadata creation workflows among the different university libraries in South Africa. Uncertainties regarding automation of subject analysis processes can be further researched, so as to investigate how they can contribute to the process.

Records studies can be conducted by using a different study period to establish new trends in ETD subject metadata creation practices. The study of titles and content in languages other than English is recommended to improve how the subject content and place names in different languages can be described for improved access of the ETD content.

Furthermore, further research into the principles of boundary object theory and how to infuse them in metadata creation workflows is recommended. Theory plays a critical role as the foundation for subject analysis approaches that support the creation of appropriate ETD geographic subject metadata. The implicit theoretical knowledge is not sufficient to enhance current practices and improve ETD subject metadata creation in general. Further research may help to expose how to make this knowledge more explicit.

BIBLIOGRAPHY

Al Salmi, JMY. 2014. *Electronic theses and dissertations programmes in the Arab Gulf States: exploring factors affecting their adoption and development.* Doctoral thesis, Victoria University of Wellington, New Zealand.

Albrachtsen, H. 1993. Subject analysis and indexing: from automated indexing to domain analysis. *The Indexer,* 18(4):219–224.

Albrachtsen, H. 2015. This is not domain analysis. *Knowledge Organisation,* 42(8):557–561.

Albrachtsen, H & Jacobs, EK. 1998. The dynamics of classification systems as boundary objects for cooperation in the electronic library. *Library Trends,* 47(2):293.

Alemneh, D & Philips, M. 2016. Indexing quality and effectiveness: an exploratory analysis of electronic theses and dissertations representation. *Proceedings of the Association for Information Science and Technology,* 53(1):1–4.

Alemu, G. 2016. A theory of metadata enriching and filtering: challenges and opportunities to implementation. *Qualitative and Quantitative Methods in Libraries,* 5(2):311–334.

Alhojailan, MI. 2012. Thematic analysis: a critical review of its process and evaluation. *West East Journal of Social Sciences,* 1(1):39–47.

Anderson, JD & Perez-Carballo, J. 2001. The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part 1: Research and the nature of human indexing. *Information Processing and Management,* 37:231–254.

Angadi, M, Bandi, S & Koganuramath, M. 2012. *Unlocking scholarly access: migrating traditional dissertations into electronic theses and dissertations (ETDs): a case study of Tata Institute of Social Sciences, Mumbai.* Paper presented at the National Conference on Empowering Libraries for Life Long Learning Skills, Excellence, Creativity and Challenges, Gulbarga, Kamataka, August 2012.

Association of African Universities. 2017. *Database of the African Theses and Dissertations-Research (DATAD).* https://aau.org/current-projects/database-of-african-theses-and-dissertations-research-datad-r/ (Accessed 30 January 2017).

Bake, BR. 1974. *Determining geographical subject headings for local history collections.* Research paper, Brigham Young University, Utah. http://files.eric.ed/fulltext/ED102971.pdf (Accessed 6 July 2016).

Bangani, S, 2018. The history, deployment and future of institutional repositories in public universities in South Africa. *The Journal of Academic Librarianship,* 44(1):39–51.

Baro, EE & Etiode, PG. 2014. Electronic theses and dissertations and institutional repositories: roadmap to research visibility in Africa. *Preservation Digital Technology & Culture,* 43(3):114–126.

Bates, MJ. 1986. Subject access in online catalogs: a design model. *Journal of the American Society for Information Science,* 37(6):357–376.

Bates, MJ. 2005. Information and knowledge: an evolutionary framework for Information Science. *Information Research,* 10(4). http://informationr.net/ir/10-4/paper239.html (Accessed 19 February 2022).

Beghtol, C. 1986. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation,* 42(2):84–113.

Beuhler, MA. 2013. *Demystifying the institutional repository for success.* Oxford: Chandos.

Bezuidenhout, R & Cronje, F. 2014. Qualitative data analysis. In *Research matters,* edited by F Du Plooy-Cilliers, C Davis and R Bezuidenhout, pp. 228–293. Claremont: Juta.

Bibliographic Services Task Force. 2005. *Final report, December 2005: rethinking how we provide bibliographic services for the University of California.* http://libraries.universityofcalifornia.edu/groups/files/bstf/docs/Final.pdf (Accessed 1 February 2017).

Bickman, L & Rog, DJ (eds). 1998. *Handbook of applied social research methods.* Thousand Oaks, CA: SAGE.

Bidney, M & Clair, K. 2014. Harnessing the geospatial semantic web: towards place-based information organization and access. *Cataloging & Classification Quarterly,* 52(1):69–76*.*

Bieber, M, Engelbart D, Furuta R, Hiltz SR, Noll J, Preece J, Stohr, EA, Turrof, M & Van De Walle B. 2002. Toward virtual community knowledge evolution. *Journal of Management Information Systems,* 18(4):11–35*.*

Biesta, G. 2010. Pragmatism and the philosophical foundations of mixed methods research. In *SAGE handbook of mixed methods in social & behavioural research.* 2nd ed., edited by A Tashakkori and C Teddlie. Thousand Oaks, CA: SAGE. http://dx.doi.org/10.4135/9781506335193

Bothmann, RL. 2011. Evolving standards: making the jump to RDA in historical context. In *Practical strategies for cataloging departments*, edited by RL Lubas, pp. 1–17. Santa Barbara, CA: Libraries Unlimited.

Brannen, J & Halcomb, EJ. 2009. Data collection in mixed methods research. In *Mixed methods research for nursing,* edited by S Andrews and EJ Halcomb, pp. 67–80. http://0-lib.myilibrary.com.oasis.unisa.ac.za/Open.aspx?id=230604 (Accessed 12 October 2017).

Braun, V & Clarke, V. 2006. Using thematic analysis in research. *Qualitative Research in Psychology.* 3(2):77–101.

Braun, V & Clarke, V. 2020. One size fits all? What counts as quality practice in (reflexive) thematic analysis. *Qualitative Research in Psychology,* 18(3):328–352.

Bryman, A. 2006. Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research,* 1(1):2–82.

Buckland, MK.2012. Obsolescence in subject description. *Journal of Documentation,* 68(2):154–16.

Burns, E. 2010. Developing e-mail interview practices in qualitative research. *Sociological Research Online,* 15(4). www.socresonline.org.uk/15/4/8.html#gunter2002 (Accessed 5 October 2017).

Campos, MLA & Gomes, HE. 2017. Ontology: several theories on the representation of knowledge domains. *Knowledge Organisation,* 44(3):178–186.

Chan, LM. 2000. *Exploiting LCSH, LCC and DDC to retrieval networked resources: issues and challenges.* http://www.loc.gov/catdir/bibcontrol/chan_paper.html (Accessed 2 April 2015).

Chaudhry, AS. 2016. Re-conceptualization of knowledge organization: imperatives of networked resources and digitization. *International Journal of Knowledge Content Development Technology,* 6(2):93–108.

Chowdhury, GG & Foo, S. 2012. *Digital libraries and information access: research perspectives.* London: Facet.

Chu, CM & O'Brien, A.1993. Subject analysis: the critical first stage of indexing. *Journal of Information Science,* 19(6):439–454.

Chung, E. 2006. *A framework of automatic subject term assignment: an indexing conception-based approach.* Doctoral thesis, University of Northern Texas, Texas.

Chung, E & Hastings, SK. 2006. Conceptions-based approach to automatic subject term assignment for scientific journal articles. *Proceedings of the American Society for Information Science and Technology,* 43(1):1–21.

Chung, E, Miksa, S & Hastings, SK. 2010. A framework for automatic subject term assignment for text categorisation: an indexing conception based approach. *Journal for the American Society for Information Science and Technology,* 61(4):688–699.

Collis, J & Hussey, R. 2014. *Business research: a practical guide for undergraduate and postgraduate students.* 4th ed. Houndmills: Palgrave Macmillan.

Cox, NS. 2015. How did we get here: binding of print theses and dissertations to processing of electronic theses and dissertations? *North Carolina Libraries,* 73(1):20–24.

Creswell, JW. 2014. *Research design: qualitative, quantitative and mixed methods approaches.* Los Angeles, CA: SAGE.

Creswell, JW & Plano Clark, VL. 2011. *Designing and conducting mixed methods research.* 2nd ed. Los Angeles, CA: SAGE.

Creswell, JW. & Plano Clark, VL. 2018. *Designing and conducting mixed methods research.* 3rd ed. Los Angeles, CA: SAGE.

Cruikshank, K, Daniels, C, Meissner, D, Nelson, NL & Shelstad M. 2005. *How do we show you what we've got? Access to archival collections in the digital age.* Ann Arbor, MI: MPublishing.

Cutter, CA. 1904. *Rules for a dictionary catalog.* 4th ed. London: The Library Association.

David, RH & Thomas, D. 2015. Assessing metadata and controlling quality in scholarly Ebooks. *Cataloging & Classification Quarterly,* 53(7):801–824

De Groat, G. 2009. *Future directions in metadata remediation for metadata aggregators.* Digital Library Federation. https://old.digilib.org/aquifer/dlf110.pdf (Accessed 30 May 2016).

Dean, RJ. 2004. Development of simplified headings for metadata. *Cataloging & Classification Quarterly,* 39(1–2):331–352.

Diao, J & Hernandez, MA. 2014. Transferring cataloguing legacies into descriptive metadata creation in digital projects: cataloger's perspective. *Journal of Library Metadata,* 14(2):130–145.

Dimec, Z, Zumer, M & Riesthuis, GJA. 2005. Slovenian cataloging practice and functional requirements for bibliographic records: a comparative analysis. In *Functional requirements for bibliographic records (FRBR): hype or cure-all?* Edited by P Le Boeuf, pp. 207–227. Philadelphia, PA: Haworth Information Press.

Dousa, TM. 2010. Facts and frameworks in Paul Otlet's and Julius Otto Kaiser's theories of knowledge organisation. *Bulletin of the American Society of Information Science and Technology,* 36(2):19–25.

Dousa, T & Ibekwe-Sanjuan, F. 2014. Epistemological and methodological eclecticism in the construction of knowledge organization systems (KOSs): the case of analytico-synthetic. In *Knowledge organisation in the 21st century: between historical patterns and future prospects,* pp. 152–159. Krakow, Poland:

Ergon Verlag. https://hal.arvchives.ouvertes.fr/hal-01066220 (Accessed19 February 2022).

Du Plooy-Cilliers, F & Cronje, J. 2014. Qualitative data collection. In *Research matters*, edited by F Du Plooy-Cilliers, C Davis and R Bezuidenhout, pp. 147–172. Claremont: Juta.

Dublin Core Metadata Initiative (DCMI). 2018. http://Dublincore.org (Accessed 1 June 2018).

Dublin Core Metadata Initiative (DCMI). *DCMI glossary*. 2018a. Sv "Information resource". https://www.dublincore.org/specifications/dublin-core/usageguide/glossary/ (Accessed 1 June 2018).

Dublin Core Metadata Initiative (DCMI). DCMI glossary. 2018b. Sv "Metadata". https://www.dublincore.org/specifications/dublin-core/usageguide/glossary/ (Accessed 1 June 2018).

Dublin Core Metadata Initiative (DCMI). *DCMI glossary*. 2018c. Sv "Controlled vocabulary". https://www.dublincore.org/specifications/dublin-core/usageguide/glossary/ (Accessed 1 June 2018).

Dulle, FW. 2010. *An analysis of open access scholarly communication in Tanzanian public libraries*. Doctoral thesis, University of South Africa, Pretoria.

Dutta, B & Dutta, C. 2013. Concept of "subject" in the context of Library and information Science from a new angle. *Annals of Library and Information Science,* 60:78–87.

Earle, EF. 2014*. Crowdsourcing metadata for library and museum collections using a taxonomy of Flickr user behavior*. Master's dissertation, Cornell University, Ithaca. http://hdl.handle.net/1813/36010 (Accessed 20 February 2022).

Evans, G & Schwing, T. 2016. OhioLINK recent developments at a United States academic library consortium. *Interlending and Document Supply,* 44(4):172–177.

Ezema, IJ & Igbo, HU. 2016. Electronic theses and dissertation and the promotion of scholarship in African universities: prospects and challenges. *International Information & Library Review,* 48(1):21–30.

FAST Project. 2018. http://www.oclc.org.research/themes/data-science/fast.html (Accessed 20 April 2018).

Fattahi, R. 1998. Library cataloguing and abstracting and indexing services: reconciliation of principles in the online environment. *Library Review,* 47(4):211–216.

Fetters, MD, Yoshioka, T, Greenberg, GM, Gorenflo, GW & Yeo, S. 2007. Advanced consent in Japanese during prenatal care for epidural anesthesia during child birth. *Journal of Mixed Methods Research*, 4:333–365.

Fidel, R. 1994. User-centered indexing. *Journal of the American Society for Information Science,* 45(8):572–576.

Fink, A. 1995. *How to design surveys.* Thousand Oaks, CA: SAGE.

Flynn, EA & Ahrberg, JH. 2020. Electronic theses and dissertations (ETDs): metadata policies, practices and workflows: a survey of the ETD metadata lifecycle at United States academic institutions. *Journal of Library Metadata,* 20(2–3):91–110.

Fourie, I. 2002. How can we take a socio-cognitive approach in teaching indexing and abstracting? *The Indexer,* 23(2):83–85.

Fourie, I. 2005. Conceptualisation: what is it all about? In *Indexing for Southern Africa: a manual compiled in celebration of ASAIB's first decade, 1994–2004,* edited by JA Kalley, E Schoeman and M Burger, pp. 20–28. Pretoria: Unisa Press.

Fourie, I. 2008. Book indexing: a reflection on the contrasting complexities and ease of conceptualisation and how we can deepen our understanding. *Mousiaon,* 25(1):111–125.

Fox, EA. 2021. Building and using digital libraries for electronic theses and dissertations. *The Journal of Electronic Theses and Dissertations,* 1:1–30.

Fox, EA, Eaton, JL, McMillan, G, Kipp, NA, Weiss, L, Arce, E & Gyer, S. 2004. National digital library of theses and dissertations: a scalable and sustainable approach to unlock university resources. In *Electronic theses and dissertations: a sourcebook for educators, students and librarians,* edited by EA Fox, S Feizabadi, JM Moxley and CR Weisser, pp. 34–48. New York: Marcel Dekker.

Fox, EA, McMillan, G & Eaton, JL. 1999. *The evolving genre of theses and dissertations*. Paper presented at the Hawaii International Conference on System Sciences, Hawaii, January 1999.

Frank, I & Rowe, WC. 2004. Indexing and accessing electronic theses and dissertations: some concerns for users. In *Electronic theses and dissertations: a sourcebook for educators, students and librarians,* edited by EA Fox, S Feizabadi, SJM Moxley and CR Weisser, pp. 343–353. New York: Marcel Dekker.

Fuller, S, Hjorland, B, Ibekwe-Sanjuan, F, Ma, L Mai, JE, Tennis, J & Werner, J. 2013. Uncovering epistemological assumptions underlying research. *Association of Information Science and Technology,* 50:1–4.

Gerolimos, M. 2013. Tagging for libraries: a review of the effectiveness of tagging systems in library catalogs. *Journal of Library Metadata* 13(1):36–58.

Golub, K, Soergel, G, Buchanan, D, Tudhope, G, Lykke, M & Hiom, D. 2016. A framework for evaluating automatic indexing or classification in the context of retrieval. *Journal of the Association for Information Technology*, 67(1):3–16.

Greenberg, J. 2005. Understanding metadata and metadata schemes. *Cataloging & Classification Quarterly*, 40(3–4):17–36.

Greene, JC, Caracelli, VJ & Graham, WF. 1989. Towards a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis,* 11(3):255–274.

Gross, T, Taylor, FG & Joudrey, GN. 2015. Still a lot to lose: the role of controlled vocabulary in keyword searching. *Cataloging & Classification Quarterly,* 53(1):1–39.

Guest, G, & Fleming, P. 2014. Mixed methods research. In *Public health research methods,* edited by G Guest and E Namey, pp. 581–610. Thousand Oaks, CA: SAGE.

Guest, G, Namey, EE & Mitchell, ML. 2013. In-depth interviews: a field manual for applied research. In *Collecting qualitative data: a field manual for applied research,* pp. 1–58. http://methods.sagepub.com/book/collecting-qualitative-data/i510.xml (Accessed 10 October 2017).

Guimaraes, JAC. 2017. Slanted knowledge organisation as a new ethical perspective. In *The organisation of knowledge: caught between global structures and local meaning,* edited by J Anderson and L Skovig, pp. 87–102. Emerald Publishing.

Hadar, I & Soffer, P. 2006. Variations in conceptual modeling: classification and ontological analysis. *Journal of the Association of Information Systems,* 7(8):568–592.

Hall, S, Hoover, L & Wolverton, RB. 2005. Administration of electronic theses/dissertation programs: a survey of U.S. institutions*. Cataloging & Classification Quarterly,* 22(3):1–17.

Hall B, & Howard, K. 2008. Conducting mixed methods research with typological and systemic design considerations. *Journal of Mixed Methods Research*. 2(3):248–269.

Henning, E, Van Rensburg, W & Smith, B. 2004. *Finding your way in qualitative research*. Pretoria: Van Schaik.

Hider, P. 2012. *Information resource description: creating and managing metadata*. London: Facet.

Hirwade, MA. 2011. A study of metadata standards. *Library Hi Tech News,* 28(7):18–25.

Hjorland, B. 1992. The concept of subject in Information Science. *Journal of Documentation,* 48(2):172–200.

Hjorland, B. 1997. *Information seeking and subject representation: an activity-theoretical approach to Information Science*. Westport, CT: Greenwood Press.

Hjorland, B. 1998. Theory and metatheory of Information Science: a new interpretation. *Journal of Documentation,* 54(5):606–621.

Hjorland, B. 2001. Towards a theory of aboutness, subject, topicality, theme, domain, field content … and relevance. *Journal of the American Society for Information Science and Technology,* 52(9):774–778.

Hjorland, B. 2002. Epistemology and the socio-cognitive perspective in Information Science. *Journal of the Association for Information Science and Technology,* 53(4):257–270.

Hjorland, B. 2004. Empiricism, rationalism and positivism in Library and Information Science. *Journal of Documentation* 61(1):130–155.

Hjorland, B. 2009. Concept theory. *Journal of the American Society for Information Science and Technology,* 60(8):1519–1536.

Hjorland, B. 2013. Theories of knowledge organisation, theories of knowledge. *Knowledge Organisation,* 40(3):169–181.

Hjorland, B. 2017a. Subjects (of documents). *Knowledge Organisation,* 44(1):55–64.

Hjorland, B. 2017b. Reviews of concepts with knowledge organisation. *Knowledge Organisation,* 44(6):436–463.

Hjorland, B. 2021. Information retrieval and knowledge organisation: a perspective from the philosophy of science. *Information,* 12(3):135.

Hjorland, B & Albrachtsen, H. 1995. Toward a new horizon in Information Science: domain-analysis. *Journal of the American Society for Information Science,* 46(6):400–425.

Ho, J & Stokes, C. 2019. Core metadata elements recommendations for institutional repositories in Texas A&M university libraries. *Journal of Library Metadata,* 19(3-4):187–214.

Hoffman, GL. 2009. Applying the user-centered paradigm to cataloguing standards in theory and practice: problems and prospects. *NASKO,* 2(1):27–34. http://journals.lib.washington.edu/index.php/nasko/article/viewFile/12812/11294 (Accessed 1 September 2017).

Howarth, LC & Olson, HA. 2016*. Subject access principles in the new world: procrustean or procreative?* Paper presented at the IFLA Conference: Subject Access: Unlimited Opportunities, Columbus, 2016.

Husic, G. 2014. Enhancing an open-access linguistics journal archive with Library of Congress-like metadata: a case study of the effectiveness for improving discovery. *Kansas Working Papers in Linguistics,* 35:1–8.

Huvila, I, Anderson TD, Jansen E, Mc Kenzie, P & Worrall A. 2016. Boundary objects in Information Science. *FIMS Publication,* 156. https://ir.lib.uwo.ca/fimspub/156?utm_source=ir.lib.uwo.ca%2Ffimspub%2F156

&utm_medium=PDF&utm_campaign=PDFCoverPages (Accessed 18 February 2022).

International Federation of Library Associations and Institutions (IFLA). 2013. *IFLA/UNESCO Manifesto for Digital Libraries: bridging the digital divide: making the world's cultural and scientific heritage available to all*. https://www.ifla.org/publications/ifla-unesco-manifesto-for-digital-libraries/ (Accessed 9 September 2016).

International Federation of Library Associations and Institutions (IFLA). 2016. *Access and opportunity for all: how libraries can contribute to the United Nations 2030 Agenda*. http:/www.ifla.org/publications/node/10546 (Accessed 2 February 2017).

International Federation of Library Associations and Institutions (IFLA). Working Group on Guidelines for Subject Access by National Bibliographic Agencies. 2011. *Guidelines for subject access in national bibliographies.* https://www.ifla.org/wp-content/uploads/2019/05/assets/classification-and-indexing/subject-access-by-national-bibliographic-agencies/nba-guidelines-for-approval.pdf (Accessed 10 June 2017).

International Federation of Library Associations and Institutions (IFLA). Working Group on Metadata for Digital Resources. 2009. *Metadata for digital resources.* https://www.ifla.org/wpcontent/uploads/2019/05/assets/cataloguing/reports/metadata-for-digital-resources_2009.pdf (Accessed 16 February 2022).

International Organization for Standardization (ISO). 2015. *Quality management principles.* https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100080.pdf (Accessed 6 June 2017).

Jackson, AS. 2011. Moving ahead with metadata: adding value through grant-funded projects. In *Practical strategies for cataloging departments,* edited by RL Lubas, pp. 95–105. Santa Barbara, CA: Libraries Unlimited.

Jansen, E, Orland, B, Hartel, J & Huvila, I. 2016. *Domain analysis: assumptions and new techniques.* https://works.bepress.com/jenna_hartel/44 (Accessed 23 August 2017).

Jensen K & Carr, C. 2020. Retrospective enhancement of thesis cataloguing and the match rate of authors in national authority files. *Cataloging & Classification Quarterly,* 58(6):566–590.

Johnson, RB & Onwuegbuzie, AJ. 2004. Mixed methods research: a research paradigm whose time has come. *Educational Researcher*, 33(7):14–26.

Kagoro, KC, Khayundi, FE & Ayelana, AA. 2016. *Constraints faced by academic libraries in digitisation of theses and dissertations in a few selected African Countries.* http://krepublishers.com/02-Journals/JSS/JSS-49-0-000-16-Web/JSS-49-1,2-000-16-Abst-PDF/JSS-49-1,2-011-16-2120-Oyelana-A-A/JSS-49-1,2-011-17-2120-Oyelana-A-A-Tx[2].pdf (Accessed 7 June 2017).

Kertcher, Z & Coslor, E. 2020. Boundary objects and the technical culture divide: successful practices for voluntary innovation teams crossing scientific and professional fields. *Journal of Management Inquiry,* 29(1):76–91.

Khoo, M & Hall, C. 2013. Managing metadata: networks of practice, technological frames and metadata work in a digital library. *Information and Organization,* 23:81–106

Koonin, M. 2014. Validity and reliability. In *Research matters,* edited by F Du Plooy-Cilliers, C Davis and R Bezuidenhout, pp. 252–261. Claremont: Juta.

Krejcie, RV & Morgan, DW. 1970. Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3):607–610.

Krippendorff, K. 2013. *Content analysis: an introduction to its methodology*. Los Angeles, CA: SAGE.

Lancaster, FW. 2003. *Indexing and abstracting in theory and practice.* 3rd ed. London: Facet.

Larson, RR. 2011. The decline of subject searching: long-term trends and patterns of index use in an online catalogue. *Journal for the Society of Information Science,* 42(3):197–215.

Laskowski, MS. 2016. When good enough is not good enough: resolving cataloging issues for high density storage. *Cataloging and Classification Quarterly,* 54(3):147–158.

Leedy, PD & Ormrod, JE. 2013. *Practical research: planning and design.* 10th ed. Boston: Pearson.

Library of Congress. 2004. *Thesaurus for graphic materials: subject terms (TGM II)* http://www.loc.gov./rr/print/tgm2/toc.html (Accessed 10 June 2017).

Library of Congress. 2016. *Library of congress subject headings.* 38th ed. Https://www.loc.gov/aba/publications/FreeLCSH/lcshintro.pdf (Accessed 28 June 2016)

Liebetrau, P. 2010. *Metadata.* In *Managing digital collections: a collaborative initiative of the South African framework,* edited by P Liebetrau, pp. 30–34. Pretoria: NRF.

Lopez, KD, Willis, DG. 2014. Descriptive versus interpretive phenomenology: their contributions to nursing knowledge. *Qualitative Health Research*, 14(5):726–735.

Lubas, RL. 2009. Defining best practices in theses and dissertations metadata. *Journal of Library Metadata,* 9(3–4):252–263.

Lubas, RL. (ed). 2011. *Practical strategies for cataloguing departments*. Saint Barbara, CA: Libraries Unlimited.

Ma, 2006. Managing metadata for digital projects. *Library Collections, Acquisitions and Technical Services,* 30:3–17.

Mai, J. 2005. Analysis in indexing: document and domain centered approaches. *Information Processing and Management,* 41:599–611.

Mai, J. 2010 Classification in a social world: bias and trust. *Journal of Documentation,* 66(5):627–642.

Mao, N. 2004. *Automatic extraction of author self contributed metadata for electronic theses and dissertations*. University of North Carolina, University Libraries, Carolina Digital Repository. https://doi.org/10.17615/swwp-wx58 (Accessed 16 February 2022).

Marc, DT. 2016. *Assessing metadata quality and terminology coverage of a federally sponsored health data repository*. Doctoral thesis, University of Minnesota, Minnesota.

Margatipoulos, T, Margatipoulos, M, Mavridis, I & Manitsaris, A. 2008. *A conceptual framework for metadata quality assessment.* DCMI 08: Proceedings of the International Conference Dublin Core and Metadata Applications, Singapore, September 2008, pp. 104–113.

Marshall, CC. 1998. *Making metadata: a study of metadata creation for a mixed physical-digital collection.* Proceedings of the third ACM conference on Digital Libraries, Pittsburgh, June 1998, pp. 162–171.

Martin, KE & Mundle, K. 2014. Positioning libraries for a new bibliographic universe: a review of cataloguing and classification literature, 2011–2012. *Library Resources & Technical Services,* 58(4):233–249.

Matanji, PHM. 2012. *In-house indexing of periodical literature: a study of university libraries in Kenya.* Masters dissertation. University of South Africa, Pretoria.

Matizirofa L & Ramalibana K. 2015. *NRF OA statement survey report back workshop.* http://hdl.handle.net/10907/128. (Accessed 2 September 2016).

Maurer, M, McCutcheon, S & Schwing, T. 2011. Who is doing what? Findability and author supplied metadata in the library catalog. *Cataloging & Classification Quarterly,* 49(4):277–310.

Maurer, MB & Shakeri, S. 2016. Disciplinary differences: LCSH and keywords assignment for ETDs from different disciplines. *Cataloguing & Classification Quarterly,* 54(4):213–243.

Mayernik, MS. 2020. Metadata. *Knowledge Organization*, 47(8):696–713.

Mayoh J & Onwuegbuzie AJ. 2015. Towards a conceptualization of mixed method phenomenological research. *Journal of Mixed Methods Research,* 19(1):91–107.

McCutcheon, A. 2007. *Ohio ETD survey.* http://www.tdl.org/wp-content/uploads/2009/06/mccutcheon_survey.pdf. (Accessed 7 June 2017).

McCutcheon, AM. 2010. *Impact of publishers' policy on electronic theses and dissertations (ETD) distribution options within the United States.* Doctoral thesis. Ohio University, Ohio.

McCutcheon, S. 2011. Basic, fuller, fullest: treatment options for electronic theses and dissertations. *Library Collections, Acquisitions & Technical Services,* 35(2–3):64–68.

Mcnamara, J. 2018. Content analysis. In *Mediated communication,* edited by P. Napoli, pp.191–212. Boston: De Gruyter Mouton. Doi:10.1515/9783110481129-012 (Accessed 17 September 2022).

Mendez, E & Van Hooland, S. 2014. Metadata typologies and metadata uses. In *Handbook of metadata, typologies and ontologies*, edited by MA Sicilla, pp. 9–40. Singapore World Scientific Publishing.

Merriam-Webster dictionary. 2021. www.webster.com/dictionary. Sv. "model".

Middleton, CC, Jean, JW & Gilbertson, MA. 2015. A process for the original cataloguing of theses and dissertations. *Cataloging & Classification Quarterly,* 53(2):234–246.

Miller, S & Perkins, J. 2015. Introduction. *Journal of Library Metadata,* 15:3–4:135–141.

Montoya, RD. 2017. Boundary objects/boundary staff: supporting digital scholarship in academic libraries. *Journal of Academic Librarianship,* 43:216–223.

Moran-Ellis, J, Alexander, VD, Cronin, A, Dickinson, M, Fielding, J, Sleney, J & Thomas, H. 2006. Triangulation and integration: processes, claims and implications. *Qualitative Research*, 6(1):45–59.

Morgan, DL. 2014. *Integrating qualitative and qualitative methods: a pragmatic approach.* Los Angeles, CA: SAGE.

Morse, J. 2010. Procedures and practice of mixed method design: maintain control, rigor and complexity. In *SAGE handbook of mixed methods in social & behavioural research.* 2nd ed., edited by A Tashakkori and C Teddlie, pp. 1–18. http://dx.doi.org/10.4135/9781506335193

Moulaison, H & Dykas, F. 2016. High quality data and repository staffing: Perceptions of United-States based OpenDOAR participants. *Cataloguing and Classification Quarterly,* 54(29):101–116.

Mouton, J. 2001. *How to succeed in your master's and doctoral studies: a South African guide and resource book.* Pretoria: Van Schaik.

National ETD Portal (NETD): South African theses and dissertations. 2018. www.netd.ac.za (Accessed 19 May 2018).

Networked Digital Library of Theses and Dissertations (NDLTD). 2018. *ETD_MS: an interoperability metadata standard for electronic theses and dissertations.* https://ndltd.vt.domains/wp-content/uploads/2021/04/etd-ms-v1.1.html (Accessed 4 June 2018).

Neuman, WL. 2014. *Social research methods: qualitative and quantitative approaches.* 7th ed. Harlow: Pearson.

Ngulube, P, Mokwatlo, K & Ndwedwe, S. 2009. Utilisation and prevalence of mixed methods research in Library and Information Science research in South Africa, 2002–2008. *South African Journal of Library and Information Science*, 75(2):105–116.

O'Cathain, A, Murphy, E & Nicholl, J. 2010. Three techniques for integrating data in mixed methods studies. *British Medical Journal*, 341(7787):1147–1150.

Ochoa, X. 2014. Metadata quality. In *Handbook of metadata, semantics and ontologies*, edited by MA Sicilia, pp. 63–88. Singapore: World Scientific Publishing.

ODLIS: online dictionary of Library and Information Science (ODLIS). 1996. Sv "subject analysis". https://www.library.ucsb.edu/research/db/1182 (Accessed 18 February 2022).

Oh, JS & Park, ON. 2018. Topics and trends in metadata research. *Journal of Information Science Theory and Practice,* 6(4):39–53.

Ohio LINK. Electronic Theses & Dissertations Center (ETD). 2018. https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/1?clear=0,1,5,10,20,21,1001 (Accessed 20 April 2018).

Onwuegbuzie, AJ & Combs, C. 2010. Emergent data analysis techniques in mixed methods research: a synthesis. In *Handbook of mixed methods in social and behavioural research*, 2nd ed. edited by A Tashakkori and C Teddlie, pp. 1–35. Thousand Oaks, CA: SAGE. http://dx.doi.org/10.4135/9781506335193 (Accessed 20 September 2022).

Pal, JK. 2016. Resolving the confusion over metadata-creation in digital archives. *Annals of Library and Information Science,* 63:110–116.

Palavitsinis, N. 2013. *Metadata quality issues in learning repositories.* Doctoral thesis, Universidad de Alacala, Madrid, Spain.

Panage, BM & Bonde, H. 2016. Digital library: definitions and interpretations. *International Journal of Innovative Knowledge Concepts,* 2(3):34–38.

Park, JR. 2009. Metadata quality in digital repositories: a survey of current state of the art. *Cataloging & Classification Quarterly,* 47(3–4):2013–228.

Park, J & Tosaka, Y. 2010. Metadata creation practices in digital repositories and collections. *Information Technologies and Libraries,* 29(3):104–116.

Pickard, AJ. 2013. *Research methods in information.* 2nd ed. London: Facet.

Potvin, S & Thompson, S. 2016. An analysis of evolving metadata influences, standards, and practices in electronic theses and dissertations. *Library Resources and Technical Services,* 60(2):99–114.

Preminger, M, Rype, I, Adland, MK, Massey D & Talleras, K. 2020. The public library metadata landscape: the case of Norway, 2017–2018. *Cataloging & Classification Quarterly,* 58(2):127–148.

Pretorius, M. 2005. Environmental issues. In *Indexing for Southern Africa: a manual compiled in celebration of ASAIB's first decade, 1994–2004*, edited by JA Kalley, E Schoeman and M Burger, pp. 226–232. Pretoria: Unisa Press.

Proquest. 2017. https://www.proquest.com/ (Accessed 28 May 2017).

Punch, KF. 2005. *Introduction to social science research: quantitative and qualitative approach.* 2nd ed. London: SAGE.

Rasuli, R, Alipour-Hafezi, M & Solaimani, S. 2016. *The identification of critical success factors in the development of the national ETDS programs.* 19th International Symposium on the Electronic Theses and Dissertations (ETD2016): "Data and dissertations", France, July 2016.

Riley, J. 2004. *Understanding metadata: what is metadata and what is it for? A primer.* Baltimore, MD: National Information Standards Organization https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf (Accessed 30 June 2017).

Riley, J. 2017. *Understanding metadata: what is metadata and what is it for?* Baltimore, MD: National Information Standards Organization (NISO). https://groups.niso.org/higherlogic/ws/public/download/17446/Understanding%20Metadata.pdf (Accessed 20 September 2022).

Riva, P, Le Boeuf, P & Zumer, M. 2017. *IFLA Library Reference Model.* https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf (Accessed 9 June 2021).

Robinson, K, Edmunds, J & Mattes, SC. 2016. Leveraging author-supplied metadata, OAI-PMH and XSLT to catalog ETDs: a case study at a large research library. *Library Resources & Technical Services*, 60(3):191–203.

Rondeau, S. 2014. The life and times of aboutness: a review of the Library and Information Science literature. *Evidence Based Library and Information Practice,* 9(1):13–35.

Rubin, RE. 2010. *Foundations of Library and Information Science*. 3rd ed. New York: Neal Schuman.

Sandy, HM & Dykas F. 2016. High-quality metadata and repository and staffing: perceptions of United States-based OpenDOAR participants. *Cataloging & Classification Quarterly,* 54(2):101–116.

Saunders, M, Lewis, P & Thornhill, A. 2016. *Research methods for business students.* 7th ed. Harlow: Pearson.

Sauperl, P. 1999. *Subject determination during the cataloguing progress*. Doctoral thesis, University of North Carolina, Chapel Hill, Northern Carolina.

Sauperl, P. 2004. Catalogers' common ground and shared knowledge. *Journal of the American Society for Information Science and Technology,* 55(1):55–63.

Sauperl, P. 2005. Subject cataloging process of Slovenian and American catalogers. *Journal of Documentation,* 61(6):713–734.

Shannon-Baker, P. 2016. Making paradigms meaningful in mixed methods research. *Journal of Mixed Methods Research,* 10(4):319–334.

Shubert, SB. 1996. *Subject access to museum objects: applying the principle of the subject approach to information from Library and Information Science to the*

documentation of Humanities Museum objects. Doctoral thesis, University of Toronto.

Soergel, D. 2009. Knowledge organization for digital libraries. In *Semantic digital libraries*, edited by SR Kurk and B MacDaniel, pp. 9–39. Berlin: Springer.

South African Geographical Names Council (SAGNC). 2017. http://dat.gov.za/sites/default/files/Handbook%20on%20Geographical%20names.pdf (Accessed 21 September 2022).

Southwick, SB, Lambert, CK & Southwick, R. 2015. Preparing controlled vocabularies for linked data: benefits and challenges. *Journal of Library Metadata.* 15(3–4):177–190.

Star, SL. 2010. This is not a boundary object: reflection on the origin of the concept. *Science Technology and Human Values,* 35(5):601–617.

Star, SL & Griesemer, JR. 1989. Institutional ecology: translations and boundary objects: amateurs and professional in Berkely's Museum of Vertebrate Ecology, 1907–39. *Social Studies of Science,* 19(3):387–420.

Steele, T & Sump-Crethar, N. 2016. Metadata for electronic theses and dissertations: a survey of institutional repositories. *Journal of Library Metadata,* 16:53–68.

Szostak, R. 2014. Classifying for social diversity. *Knowledge Organisation,* 41(2):160–170.

Tarver H, Phillips, M, Zavalina, O & Kizakhethil, P. 2015. *An exploratory analysis of subject metadata in the digital public library of America.* Proceedings of the International Conference on Dublin Core Metadata application, Sao Paulo: Brazil, 1–3 September, pp. 30–40.

Tashakkori, A & Teddlie C. 2003. *SAGE handbook of mixed methods in social & behavioural research.* Los Angeles, CA: SAGE.

Teddlie, A & Tashakkori, C. 2009. *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences.* Los Angeles, CA: SAGE.

Teddlie, C & Yu, F. 2007. Mixed methods sampling: a typology with examples. *Journal of Mixed Methods Research,* 1(1):77–100.

Tennis, JT. 2005. *Conceptions of subject analysis: a metatheoretical investigation.* Doctoral thesis, University of Washington, Washington DC.

Terra, AL, Lacruz CA, Bernades O, Fujita, MSL & De la Fuente, GB. 2021. Subject-access metadata on ETD supplied by authors: a case study about keywords, titles and abstracts in a Brazilian academic repository. *Journal of Academic Librarianship,* 47(1):102268.

Thompson, S, Liu, X, Duran, A & Washington AM. 2019. A case study of ETD metadata remediation at the University of Houston libraries. *Library Resources and Technical Services,* 63(1):62–74.

Tmava, AM & Alemneh DG. 2013. *Enhancing content visibility in institutional repositories: overview of factors that affect digital resources discoverability.* iConference 2013 Proceedings, pp. 855–859. https://www.ideals.illinois.edu/bitstream/handle/2142/42117/437.pdf (Accessed 20 February 2022).

Tolosana-Calasanz, R, Alvarez-Robles, JA, Lacasta, J, Nogueras-Iso, J, Muro-Medrano, PR & Zarazaga-Soria, FJ. 2006. *On the problem of identifying the quality of geographic metadata.* International Conference on Theory and Practice of Digital Libraries, edited by J Gonzalo and MF Verdejo, pp. 232–243. Berlin: Springer.

Ubogu, FN. 2001. Spreading the ETD gospel: a Southern African perspective: *International Information & Library Review,* 33:249–259.

Ubogu, F. 2002. *ETDs in South Africa: current status.* Paper presented at the Fifth International Symposium of Theses and Dissertations, Provo, Utah, May 2002.

United States Electronic Theses and Dissertations Association (USETDA). 2018. http://usetda.org (Accessed 12 May 2018).

University of South Africa (UNISA). *Policy on research ethics.* https://www.unisa.ac.za/static/corporate_web/Content/Colleges/CLAW/Research/Docs/Policy%20on%20Research%20Ethics%20-%20rev%20appr%20-%20Council%20-%2015.09.2016.pdf (Accessed 20 March 2018).

Vardakosta, I & Kapidakis, S. 2013. New trends for librarians in management of geographic information. *Procedia Social and Behavioural Sciences,* 73:794–801.

Venkatesh, V, Brown SA & Sullivan, YW. 2014. Guidelines for conducting mixed methods research: an extension and illustration. *Journal of the Association of Information Systems,* 17(7):435–495.

VIRGINIA TECH. 2017. *ETDs: Virginia Tech electronic thesis and dissertations.* https://vtechworks.lib.vt.edu/handle/10919/5534 (Accessed 28 May 2017).

Webley, L, Chipeperekwa, T & Suleman, H. 2011. *Creating a national electronic and theses portal.* Proceedings of the 14th International Symposium on International Theses and Dissertations, Cape Town, 13–17 September 2011.

Wells, CS & Wollack, JA. 2003. *An instructor's guide to understanding test reliability.* Madison, WI: Testing & Evaluation Services, University of Wisconsin. https://testing.wisc.edu/Reliability.pdf (Accessed 13 July 2022).

White, HC. 2012. *Organising scientific data sets: studying similarities and differences in metadata and subject term creation.* Doctoral thesis. University of Carolina, North Carolina.

Wolverton, RE & Hoover L. 2004. Historical perspective on the treatment and cataloguing of theses and dissertations. *Technical Services Quarterly*, 21(3):1–15.

Wolverton, RE, Hoover, L, Hall, SL & Fowler R. 2008a. ETDs from the beginning. *Technical Services Quarterly*, 25(S1):1–10. https://dx.doi.org/10.1080/07317130802127819

Wolverton, RE, Hoover, L, Hall, SL & Fowler R. 2008b. Cataloging and treatment of ETDs. *Technical Services Quarterly,* 25(S1):43–73. https://dx.doi.org/10.1080/07317130802127884

Wolverton, RE, Hoover, L, & Fowler R. 2011. Subject analysis of theses and dissertations: a survey. *Technical Services Quarterly,* 28(2):201–222.

Worrall, A. 2014. *The roles of digital libraries as boundary objects within and across social and information worlds.* Doctoral thesis, Florida State University, Florida.

Xie, I & Matusiak, K. 2016. *Discover digital libraries: theory and practice*. Amsterdam: Elsevier.

Yang, L. 2016. Metadata effectiveness in Internet discovery: an analysis of digital collection metadata elements and internet search engine keywords. *College and Research Libraries,* 71(1):7–19.

Yasser, CM. 2012. An experimental study of metadata training effectiveness on errors in metadata records. *Journal of Library Metadata,* 12(4):372–395.

Yiotis, K. 2008. Electronic theses and dissertations (ETDs) repositories: what are they? Where do they come from? How do they work? *OCLC Systems & Services: International Digital Libraries Perspective,* 24(2):101–115.

Zavalina, OL. 2010. *Collection-level subject access in aggregations of digital collections: metadata application and use*. Doctoral thesis, University of Illinois, Urbana.

Zavalina, OL. 2012. Subject access: conceptual models, functional requirements and empirical data. *Journal of Library Metadata,* 12:140–163.

Zavalina, OL. 2014. Complementarity in subject metadata in large-scale digital libraries: a comparative analysis. *Cataloging & Classification Quarterly,* 52(1):77–89.

Zavalina, OL, Kizhakkhethil, P, Alemneh, DG, Phillips, ME & Tarver, T. 2015. Building a framework of metadata change. *Journal of Information and Knowledge Management,* 14(1) 1–16.

Zavalina, OL, Zavalin, V, Shakeri, S & Kizhakkethil, P. 2016. Developing an empirically-based framework of metadata change and exploring relation between metadata change and metadata quality in MARC library metadata. *Procedia Computer Science,* 99:50–63.

Zavalina, OL, Zavalin, V & Miksa, SD. 2016. *Quality over time: a longitudinal quantitative analysis of metadata change in RDA-based MARC bibliographic records representing video resources*. 2016 annual meeting of the Association for Information Science and Technology (ASIST), Copenhagen, Denmark, 14–18 October 2016.

Zeng, ML, Zumer M & Salaba S. (eds). 2010. *Functional requirements for authority data (FRSAD): a conceptual model*. http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf (Accessed 1 February 2017).

APPENDICES

Appendix A1: Questionnaire

<div align="right">

University of South Africa
Department of Information Studies
Date: August 2018

</div>

Dear participant

I am a registered student at the University of South Africa and I am conducting a survey for a D.Litt. et Phil in Information Science study titled "Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations". I have obtained permission from UNISA and < …your institution…> to conduct the study. The purpose of the study is to investigate what approaches are followed during subject analysis and how subject analysis theories are applied to facilitate geographic subject determination for the ETDs. The questionnaire forms part of the research. Your kind response to the questionnaire will benefit the general community of subject metadata creators to enhance subject metadata creation.

In adherence to the ethics requirements, the confidentiality of participants' identities and that of their institution will be maintained. There are no known or anticipated risks to participation in this study. The results will be preserved and made available for reference and future use when needed by the respondents and the wider Library and Information Service community.

Participation in the study is voluntary and you may withdraw from the study at any time should you wish to do so. Your participation in this study is highly valued. It will take approximately 15 minutes to respond to the questionnaire and your time is highly valued. If you have any questions concerning the study, you may contact me directly or my supervisor. My contact details: e-mail - khomotso.maphopha@ul.ac.za. My supervisor's contact details: e-mail - Lindacloete@mweb.co.za

Thank you in advance for your cooperation in this study.

Consent

Your consent is requested to indicate your willingness to participate in the study, please read below and indicate your consent:

I voluntarily give my consent to participate in the study by providing information for this study. The objectives of the study have been made clear to me in the introductory letter. I have been informed that:

- The confidentiality of participants' identities and that of their institution will be maintained.
- There are no known or anticipated risks to participation in this study.
- Participation in the study is voluntary and I may withdraw from the study at any time should I wish to do so.

Please indicate your consent by clicking in the provided answer choice "Yes" below.

Yes	

INSTRUCTIONS

Questionnaire: The following questions are meant to assist with gathering data related to the application of subject analysis theories to facilitate the creation of geographic subject metadata for the Electronic Theses and Dissertations (ETDs). Subject analysis is defined in this study as "determining what a work is about and subsequently expressing the analysis as subject representation data".

All the questions in this survey, except for demographics, are related to the creation of subject metadata for the ETDs. Most of the questions are focused specifically on subject analysis and geographic subject metadata. Some questions may seem related, but please answer all of them to help achieve the objectives of the study.

Please select the appropriate options or fill in the spaces provided.

SURVEY QUESTIONS

A. DEMOGRAPHIC INFORMATION

1. Please state the name of your institution.

2. How many metadata creators in your institution are involved with creation of subject metadata for the ETDs? (Please indicate the number)

3. How would you describe your position? (Please click the box next to your chosen response):

Professional Librarian (1)	
Library Assistant (2)	
Library Technician (3)	
Other, please specify (4)	

4. How long have you been involved with metadata creation, specifically for ETDs? (Please click the button next to your chosen response):

0 – 5 years	(1)	
6 – 10 years	(2)	
11 – 15 years	(3)	
15 or more years	(4)	

5. Altogether, how many years of experience do you have in information resources subject description (Assigning subject headings or subject metadata creation) (Please click the button next to your chosen response).

0 – 5 years	(1)	
6 – 10 year	(2)	
11 – 15 years	(3)	
15 or more years	(4)	

6. Please indicate which one of the following best describes your role in metadata creation. (Please click the button next to your chosen response, click multiple items where applicable. Where relevant, specify your answer in the space provided)

Metadata creator	(1)	
Metadata editor	(2)	
Metadata quality controller	(3)	
Other, please specify:	(4)	

B. SUBJECT ANALYSIS BASIS

7. Does your institution have any policy or guidelines that specifically instruct on how to conduct subject analysis during metadata creation, (Please click the button next to your chosen response)

Yes	(1)	
No	(2)	
Not sure	(3)	

8. Please indicate the extent to which you feel subject analysis policy should be enforced. (Please click the button next to your chosen response)

Always enforced (1)	Sometimes enforced (2)	Enforced (3)	Seldom enforced (4)	Not enforced (5)

9. When you conduct subject analysis (determining what a work is about), what level of importance do you attach to the listed aspects (Please rate on a scale of 1 – 5 by clicking the button that closely represent your views)

Aspect	Very important (1)	Fairly Important (2)	Important (3)	Slightly important (4)	Not important (5)
Only identifying the keywords within an information resource					
Analysis of whole contents (All key topics covered within an information resource)					
Determining the author's intention for producing the work					
Focusing on all the potential needs to be satisfied by the subject content of the ETDs					
Only focusing on satisfying the information needs of a specific subject area					
Satisfying user needs across different disciplines					

10. How familiar are you with the theoretical principles that provide the basis for subject analysis? (Please rate on a scale of 1 – 5 by clicking in the box that closely represent your views).

Very familiar (1)	Fairly familiar (2)	Familiar (3)	Slightly familiar (4)	Not familiar (5)

11. Did you receive any training on the theoretical principles that support the different approaches on how to determine what a work is about during your Library and Information Science education training? (Please click the box next to your chosen response):

(1) **Yes**		(2) **No**	

12. If you answered "Yes" to the above question, how do you rate the importance of the theoretical knowledge that you received on the different subject analysis approaches during your training for Library and Information Science qualification(s)? (Please rate on a scale of 1 – 5 by clicking in the box that closely represent your views).

Very important (1)	Fairly important (2)	Important (3)	Slightly important (4)	Not important (5)

13. Did you receive any training on the theoretical principles that support the different approaches on how to determine what a work is about as part of on-the-job training? (Please click the button next to your chosen response).

(1) **Yes**		(2) **No**	

14. If you answered "Yes" to the question above, how do you rate the importance of the theoretical knowledge that you received on the different approaches to subject analysis, as part of on-the-job training. (Please rate on a scale of 1 – 5 by clicking in the box that closely represent your views).

Very important (1)	Fairly important (2)	Important (3)	Slightly important (4)	Not important (5)

15. If there is any other way that you learnt about the different approaches for subject analysis, other than in the ways stated above, (Please specify your answer in the space provided)

C. METADATA PROCEDURES

16. Please indicate who contribute or assign subject metadata for the ETDs at your institution: (Please click the button next to your chosen response, click multiple items where applicable. Where relevant, specify your answer in the space provided)

Theses and dissertation authors (1)	
Library metadata creators (2)	
Contributed by theses and dissertation authors and verified or edited by library staff (3)	
ETDs Users (tagging) (4)	
Other, please specify (5)	

17. Please indicate the forms of terms that are used at your institution to represent the ETD subjects: (Please click the button(s) next to your chosen response(s) or, where relevant, specify your answer, mark multiple items where applicable).

Keywords from text (1)	
Terms from natural language (2)	
Controlled subject headings (3)	
Other, please specify: ………….. (4)	

18. How often do you find the listed categories of metadata to be suitable for the discovery of the ETDs subject content? (Please rate on a scale of 1 – 5 by clicking the button that closely represent your views):

Metadata categories	Always suitable (1)	Sometimes suitable (2)	Not sure (3)	Rarely suitable (4)	Never suitable (5)
Keywords contributed by theses and dissertation authors					
Subject metadata contributed by library staff					
Subject metadata created through automated library processes					
Subject metadata created in the form of user contributed tags					
keywords in the full-text of information resources					

19. How often do you engage in the following activities during the creation of geographic subject metadata for the ETDs? (Please click the button next to your chosen response, mark multiple items where applicable).

Activity	Always (1)	Often (2)	Sometimes (3)	Rarely (4)	Never (5)
Creating completely new geographic subject metadata					
Adding geographic terms as subdivisions to the existing subject headings					
Using the keywords derived from the information					

Activity	Always (1)	Often (2)	Sometimes (3)	Rarely (4)	Never (5)
resource to create additional geographic subject headings					
Creating new geographic subject metadata or subdivisions using geographic terms not derived from the information resource itself					

20. What is the maximum number of subject headings that your institution allows to be assigned to each ETD metadata record? (Please mark with an "X" in the appropriate box)

1 – 5	(1)	
5 – or more	(2)	
Not sure	(3)	
No limit set	(4)	

21. Where do you usually find relevant geographic terms to be used as subject metadata? (Please click the box next to your chosen response):

Title of a work	(1)	
Abstract	(2)	
Contents of a work	(3)	
Other, Please specify:	(4)	

D. ETDs AND GEOGRAPHIC SUBJECT METADATA

22. Do the policies in your institution specifically guide on geographic subject metadata creation? (Please click the box next to your chosen response):

(1) **Yes**		(2) **No**	

23. Which standards do you use to control assigned geographic subject metadata for ETDs? (Please click the button next to your chosen response, click multiple items where applicable. Where relevant, specify your answer in the space provided).

LC subject headings	(1)	
NLM subject headings	(2)	
Sears list subject headings	(3)	
FAST headings	(4)	
Geographic names authority lists	(5)	
Others, please state:	(6)	

24. Which metadata description elements do you use to assign geographic subject metadata for ETDs? (Please specify the name of the standard and the elements that are used for your ETDs in the space provided below, e.g. Dublin Core subject, dc:subject).

25. How would you rate the current level of adequacy of geographic subject metadata attached to your institution's ETDs records? (Please rate on a scale of 1 – 5 by clicking the button that closely represent your views).

Very adequate (1)	Fairly adequate (2)	Adequate (3)	Slightly adequate (4)	Totally inadequate (5)

26. In your own opinion, how do you think understanding the theory of subject analysis influence geographic subject metadata creation? (Please rate on a scale of 1 – 5 by clicking the button that closely represent your views).

Major influence (1)	Moderate influence (2)	Not sure (3)	Minor influence (4)	No influence (5)

27. When assigning geographic subject metadata for the ETDs, how would you rate the importance of the following aspects? (Please rate on a scale of 1 – 5 by clicking the button that closely represent your views).

Aspects	Very important (1)	Fairly important (2)	Important (3)	Slightly important (4)	Not important (5)
The author's intentions for producing the work					
The overall contents of the ETD (topics coverage)					
The information needs of all potential users of the information resource					**Past**
The information needs of users within a specific subject field					
Researching the information users' requests history					

28. Which other factors are significantly contributing to useful subject analysis for the creation of geographic subject metadata for the ETDs?

F. Perspectives of subject metadata creators on geographic subject metadata analysis and creation.

The purpose of this section is to get your perspectives based on your experience with geographic subject metadata analysis and creation.

29. What are your views on ETDs geographic subject metadata? (Please rate on a scale of 1 – 5 by clicking the button that closely represent your views):

	Statement	Strongly Agree (1)	Agree (2)	Don't know (3)	Disagree (4)	Strongly disagree (5)
30	Geographic subject metadata is important for the description of electronic theses and dissertations subject content.					
32	Library staff involved with metadata creation assign the most suitable geographic subject metadata for the discovery of the ETDs subject content.					
33	ETDs' authors supply the most suitable geographic subject representations for the discovery of their subject content.					
34	It is important to consider the public information users during the creation of geographic subject metadata for the ETDs					
35	Standardised approaches are essential during the creation of geographic subject metadata for the ETDs to facilitate the sharing of metadata.					
36	Guidelines on how to conduct subject analysis to determine the geographic subject metadata are essential					
37	Training on how to conduct subject analysis is essential.					

	Statement	Strongly Agree (1)	Agree (2)	Don't know (3)	Disagree (4)	Strongly disagree (5)
38	Knowledge on subject analysis foundational theories is important to facilitate professional subject metadata creation.					
39	Technology has made subject analysis for the creation of geographic subject metadata to be unnecessary.					
40	The ETDs subject metadata creation approaches that are currently followed in my library are accommodative of the broad internet user community searching for geographical information.					
41	Subject analysis should only focus on satisfying the information needs within a specific subject field.					
42	Place names changes affect ETDs discoverability.					

39. Please indicate any other challenges that are commonly experienced with subject analysis and the creation of geographic subject metadata for ETDs in general.

Thank you for participating in the study

Appendix A2: Qualitative interview questions

The interview is meant to clarify and provide insight into the approaches followed during subject analysis for the creation of geographic subject metadata for the Electronic Theses and Dissertations (ETDs).

CATEGORY	INTERVIEW QUESTIONS
Metadata procedures	1. Please explain how you experience the process of metadata creation for the ETDs. 2. Please explain how you experience the process with specific relevance to geographic subject metadata creation.
Subject analysis theoretical basis	1. What is your view about the application of the subject analysis theories during ETDs subject content analysis? 2. Why do you think it is necessary to have knowledge on the subject analysis theories? 3. How did you acquire knowledge on the subject analysis theories? 4. Do you think that your training helped you to get a full understanding of subject analysis theory? 5. What approaches can be followed to facilitate the understanding of subject analysis theoretical basis.
ETDs and geographic subject metadata	1. How does the current metadata creation approaches that you are following impact on the creation of geographic subject metadata? 2. What measures can be taken to facilitate the creation of geographic subject metadata? 3. Would you prioritise the document keywords, the potential users' needs or the specific subject areas information needs during subject analysis for creation of geographic subject metadata for ETDs?

CATEGORY	INTERVIEW QUESTIONS
General	1. Please give any general comments about your experiences with subject analysis and geographic subject metadata creation for ETDs. 2. Would you say that your understanding of subject analysis theories had implications of how you created geographic subject metadata for ETDs? 3. What recommendations can you propose about suitable ways that can be used to gain knowledge on the theories that provides a basis for subject analysis approaches that can be used to improve geographic subject metadata creation for ETDs? 4. What recommendations can you propose for the development of a subject analysis approach that can be used to improve geographic subject metadata creation for ETDs?

Appendix B: Content analysis coding scheme

Coding form and descriptions

CODING FORM

Institution 1							
Record Creation date	Title	Place name in title	NETD subjects	NETD Geographic subject	IR Subjects	IR Controlled Subj field e.g.dc:subject (LCSH)	IR Uncontrolled subj field e.g.dc:subject
		Place name	Comply? ➢ Y/N (Total subjects	Comply? ➢ Y/N (Total Geog. Subjects)	Comply? ➢ Y/N (Total subjects)	Comply? ➢ Y/N (Total controlled fields)	Comply? ➢ Y/N (Total uncontrolled fields)
01/01/2014	Mining in China	China	Y / N	Y / N	Y / N	Y / N	Y / N
01/01/2018	Namibia today	Namibia	Y / N	Y / N	Y / N	Y / N	Y / N
15/12/2018	South African fish, birds and reptiles	South Africa	Y / N	Y / N	Y / N	Y / N	Y / N

Institution 2							
Creation date	Title	Place name in title	NETD subjects	NETD Geographic subject	IR Total Subjects	IR Controlled Subj field e.g. dc:subject (LCSH)	IR Uncontrolled subj field e.g. dc:subject
		Place name	Comply? ➢ Y/N (Total subjects	Comply? ➢ Y/N (Total Geog. Subjects)	Comply? ➢ Y/N (Total subjects)	Comply? ➢ Y/N (Total controlled fields)	Comply? ➢ Y/N (Total uncontrolled fields)
01/01/2014	Hotels in India	India	Y / N	Y / N	Y / N	Y / N	Y / N
01/01/2018	Limpopo rivers	Limpopo	Y / N	Y / N	Y / N	Y / N	Y / N
15/12/2018	Crime in Lina Park	Lina Park	Y / N	Y / N	Y / N	Y / N	Y / N

Measurement description summary

The title comply to the criteria that it must have a place name e.g. University students in **China** and **South Africa**		
Unit of analysis	Code: Comply/Not comply	Frequency (Total counts)
1. Geographic subject metadata (e.g. Does the record comply with the criteria of being assigned geographic subject metadata?)	• Yes (yes assigned = Positive/comply) • No (not assigned) = Negative/Does not comply	If Yes, how many assigned.
2. Controlled subject metadata e.g. Are the assigned subject in controlled form	• Yes (yes assigned = Positive/comply) • No (not assigned) = Negative/Does not comply	How may controlled subject elements used in the record?

Appendix C: Ethical clearance

UNISA | university of south africa

DEPARTMENT OF INFORMATION SCIENCE RESEARCH ETHICS REVIEW COMMITTEE

Date: 31 January 2018

Ref #:
2018_KAMaphopha_7103670_001
Name of applicant: KA Maphopha
Student #:X
Staff #:

Dear Mrs KA Maphopha,

Decision: Ethics Approval

Name: Title and name of principle applicant, address, e-mail address, and phone number

KA Maphopha, Unisa Information Science, 7103670@mylife.unisa.ac.za; and 015 268 2467

Proposal: Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations.

Qualification: PHD in Information Science

Thank you for the application for research ethics clearance by the Department of Information Science Research Ethics Review Committee for the above mentioned research. Final approval is granted for 5 years.

For full approval: The application was reviewed in compliance with the Unisa Policy on Research Ethics by the Department of Information Science Research Ethics Review Committee on 31 January 2018.

The proposed research may now commence with the proviso that:

1) *The researcher/s will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.*
2) *Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study, as well as changes in the methodology, should be communicated in writing to the Department of information Science Ethics Review Committee. An amended application could be requested if there are substantial changes from the existing proposal, especially if those changes affect any of the study-related risks for the research participants.*

UNISA | university of south africa

COLLEGE OF HUMAN SCIENCES RESEARCH ETHICS REVIEW COMMITTEE

01 March 2019

Dear Khomotso Amanda Maphopha

NHREC Registration #:Rec-240516-052
CREC Reference: 2019-CHS-0239
Name:Khomotso   Amanda   Maphopha
Student #: 7103670

Decision: Ethics Approval from 01 March 2019 to 28 February 2024.

Researcher(s): Khomotso Amanda Maphopha

Khomotso.maphopha@ul.ac.za

015 2682467

Supervisor (s): LM Cloete

Department of Information Science

0828227027

Research Title
Subject analysis theories and their application to geographic subject metadata for electronic thesis and dissertations

Qualification: Phd (Library and Information Science)

College of Human Science ethics chairperson hereby acknowledge your application for Research Ethics Certificate; approval is granted for five years.

The medium risk application was reviewed by a sub committee of URERC on 27 February 2019 in compliance with the Unisa Policy on Research Ethics and the Standard Operating Procedure on Research Ethics Risk Assessment. The decision was approved on 27 February 2018.

The proposed research may now commence with the provisions that:

1. The researcher(s) will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.
2. Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study should be communicated in writing to the CHS Research Ethics Committee.
3. The researcher(s) will conduct the study according to the methods and procedures set out in the approved application.
4. Any changes that can affect the study-related risks for the research participants, particularly in terms of assurances made with regards to the protection of participants' privacy and the confidentiality of the data, should be reported to the Committee in writing, accompanied by a progress report.
5. The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the specific field of study. Adherence to the following South African legislation is important, if applicable: Protection of Personal Information Act, no 4 of 2013; Children's act no 38 of 2005 and the National Health Act, no 61 of 2003.
6. Only de-identified research data may be used for secondary research purposes in future on condition that the research objectives are similar to those of the original research. Secondary use of identifiable human research data require additional ethics clearance.
7. No field work activities may continue after the expiry date (28 February 2024). Submission of a completed research ethics progress report will constitute an application for renewal of Ethics Research Committee approval.

Note: The reference number 2018-CHS-0239 should be clearly indicated on all forms of communication with the intended research participants, as well as with the Committee.

Yours sincerely,
Signature

CREC Chair
Dr Suryakanthie Chetty
Chetts@unisa.ac.za
0124296267

Executive Dean: CHS
Prof. AP Phillips
Phillap@unisa.ac.za
0124296825

Appendix D1: Permission to conduct research: University of the Witwatersrand

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG

OFFICE OF THE DEPUTY REGISTRAR

26 October 2018

Khomotso Maphopha
PhD Candidate
UNISA

TO WHOM IT MAY CONCERN

"Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations"

This letter serves to confirm that the above project has received permission to be conducted on University premises, and/or involving staff and/or students of the University as research participants. In undertaking this research, you agree to abide by all University regulations for conducting research on campus and to respect participants' rights to withdraw from participation at any time.

If you are conducting research on certain student cohorts, year groups or courses within specific Schools and within the teaching term, permission must be sought from Heads of School or individual academics.

Ethical clearance has been obtained (2018_KAMaphopha_7103670_001)

Nicoleen Potgieter
University Deputy Registrar

Appendix D2: Permission to conduct research: University of KwaZulu Natal

UNIVERSITY OF ™
KWAZULU-NATAL

INYUVESI
YAKWAZULU-NATALI

31 August 2018

Mrs Khomotsho Maphopha (SN 7103670)
Faculty of Information Science
University of South Africa
Email: khomotso.maphopha@ul.ac.za   lindacloete@mweb.co.za

Dear Mrs Maphopha

RE: PERMISSION TO CONDUCT RESEARCH

Gatekeeper's permission is hereby granted for you to conduct research at the University of KwaZulu-Natal (UKZN), towards your postgraduate studies, provided Ethical clearance has been obtained. We note the title of your research project is:

"Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations".

It is noted that you will be constituting your sample as follows:
- with a request for responses on the website. The questionnaire must be placed on the notice system http://notices.ukzn.ac.za. A copy of this letter (Gatekeeper's approval) must be simultaneously sent to (govenderlog@ukzn.ac.za) or (ramkissoonb@ukzn.ac.za).

Please ensure that the following appears on your questionnaire/attached to your notice:
- Ethical clearance number;
- Research title and details of the research, the researcher and the supervisor;
- Consent form is attached to the notice/questionnaire and to be signed by user before he/she fills in questionnaire;
- gatekeepers approval by the Registrar.

You are not authorized to contact staff and students using 'Microsoft Outlook' address book. Identity numbers and email addresses of individuals are not a matter of public record and are protected according to Section 14 of the South African Constitution, as well as the Protection of Public Information Act. For the release of such information over to yourself for research purposes, the University of KwaZulu-Natal will need express consent from the relevant data subjects. Data collected must be treated with due confidentiality and anonymity.

Yours sincerely

MR SB MOKOENA
REGISTRAR

Office of the Registrar
Postal Address: Private Bag X54001, Durban, South Africa
Telephone: +27 (0) 31 260 8005/2206 Facsimile: +27 (0) 31 260 7824/2204 Email: registrar@ukzn.ac.za
Website: www.ukzn.ac.za

1916 - 2016
100 YEARS OF ACADEMIC EXCELLENCE

Founding Campuses   ■ Edgewood   ■ Howard College   ■ Medical School   ■ Pietermaritzburg   ■ Westville

Appendix D3: Permission to conduct research: University of Fort Hare

University of Fort Hare

OFFICE OF UNIVERSITY REGISTRAR

Alice (main) Campus:
Private Bag X1314, King William's Town Road, Alice, 5700, RSA
Tel: +27 (0) 40 602 - 2501 • Fax: +27 (0) 40 602 - 2577
Email: nmabindisa@ufh.ac.za

University of Fort Hare
Together in Excellence

30 July, 2018

Ms. Khomotso Maphopha
Khomotso.maphopha@ul.ac.za
University of Limpopo

Dear Ms. Maphopha

Approval from the Registrar's Office to Conduct Research

Having consulted the Chairperson of the Research Ethics Committee, I hereby grant permission for Ms. Maphopha to conduct research relating to his thesis "Subject Analysis Theories and their Application to Geographic Subject Metadata for Electronic Thesis and Dissertations.

We look forward to reading the research report.

Kind regards

Prof M Somniso
University Registrar

Bhisho Campus: P. O Box 1153, KWT 5600, Independence Avenue , Bhisho, 5600, RSA
Tel: +27 (0) 40 608 - 3407 • Fax: +27 (0) 40 608 - 3408

East London Campus: Private Bag X9083, EL 5200, 50 Church Street, East London, 5201, RSA
Tel: +27 (0) 43 704 - 7000 • Fax: +27 (0) 43 704 - 7095
V/C Dial Up: +27 (0) 43 704 - 7143/ 7144

www.ufh.ac.za

Appendix D4: Permission to conduct research: University of Venda

Research and Innovation
Office of the Director

02 October 2018

Ms KA Maphopha

P O Box 20

Bendor Park

Polokwane

0700

Dear **Ms Maphopha**

Permission to conduct Research at the University of Venda

You are hereby granted permission to conduct research at the University of Venda.

The research will be based on your Doctoral research topic: *Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations.*

The conditions are that all the data pertaining to University of Venda will be treated in accordance with the Ethical Principles and that will be shared with the University. In addition consent should be sought by you as a researcher from participants.

Attached is our policy on ethics.

Thank you

Senior Prof. G.E. Ekosse

Director Research and Innovation

Cc: Senior Prof LB Khoza (Acting DVC Academic)

UNIVERSITY OF VENDA
DIRECTOR
RESEARCH AND INNOVATION
2018 -10- 0 3
Private Bag X5050
Thohoyandou 0950

Appendix D5: Permission to conduct research: Walter Sisulu University

---

:WSU
Walter Sisulu University

DIVISION OF ACADEMIC AFFAIRS AND RESEARCH
DIRECTORATE OF RESEARCH DEVELOPMENT

Nelson Mandela Drive *Buffalo City*
Mthatha Campus *Potsdam Campus*
Private Bag X1 *EAST LONDON*
MTHATHA 5117 *Tel: + 43 708 5444*
Tel: + 27 47 502 2947/2647 *Fax: + 43 708 5458*
Fax: +27 47 502 2185

---

28 February 2019

Mrs KA Maphopha
UNISA
South Africa

Dear madam

Re: Permission Letter to Conduct Research at WSU

Qualification: PhD in Information Science

Institution: University of South Africa

Permission is hereby granted for the study entitled **Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations.** Provided that copies of your completed study will be submitted to the Campus Rector of the campus in which the study will be conducted and the Directorate of Research Development

All data pertaining to Walter Sisulu University will be treated confidentially and you are required to abide by ethical principles at all times. It is your responsibility to seek consent from the participants.

Directorate: Research Development

WALTER SISULU UNIVERSITY
PRIVATE BAG X1 MTHATHA 5117
REPUBLIC OF SOUTH AFRICA
TEL: + 47502 2947 FAX: +47502 2185
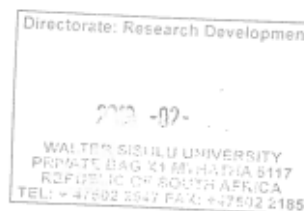
Regards

Prof W. Akpan
Senior Director
Directorate : Research and Innovation

Walter Sisulu University

Appendix D6: Permission to conduct research: Tshwane University of Technology

Appendix D7: Permission to conduct research: North West University

Private Bag X6001, Potchefstroom
South Africa 2520

Tel: +2718 299-1111/2222
Web: http://www.nwu.ac.za

Research Data Gatekeeper Committee

NWU RDGC PERMISSION GRANTED / DENIED LETTER

Based on the documentation provided by the researcher specified below, on 28/11/2018 the NWU Research Data Gatekeeper Committee (NWU-RDGC) hereby grants permission for the specific project (as indicated below) to be conducted at the North-West University (NWU):

> **Project title**: Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations.
>
> **Project leader**: Dr. L Cloete
> **Researcher**:   K. Maphopha
>
> **Ethics reference no**: 2018_KAMaphopha_7103670_001
> **NWU RDGC reference no**: NWU-GK-2018-056
>
> **Specific Conditions:**
>
> **Approval date**: 28/11/2018   **Expiry date**: 27/11/2019

General Conditions of Approval:

- The NWU-RDGC will not take the responsibility to recruit research participants or to gather data on behalf of the researcher. This committee can therefore not guarantee the participation of our relevant stakeholders.
- Any changes to the research protocol within the permission period (for a maximum of 1 year) must be communicated to the NWU-RDGC. Failure to do so will lead to withdrawal of the permission.
- The NWU-RDGC should be provided with a report or document in which the results of said project are disseminated.

Please note that under no circumstances will any personal information of possible research subjects be provided to the researcher by the NWU RDGC. The NWU complies with the Promotion of Access to Information Act 2 of 2000 (PAIA) as well as the Protection of Personal Information Act 4 of 2013 (POPI). For an application to access such information please contact Ms Amanda van der Merwe (018 299 4942) for the relevant enquiry form or more information on how the NWU complies with PAIA and POPI.

The NWU RDGC would like to remain at your service as scientist and researcher, and wishes you well with your project. Please do not hesitate to contact the NWU RDGC for any further enquiries or requests for assistance

Yours sincerely

Prof Marlene Verhoef
Chairperson NWU Research Data Gatekeeper Committee

Original details: (22351930) C:\Users\22351930\Desktop\test 2.docm
13 November 2018

Current details: (22351930) M:\DSS1\8533\Monitoring and Reporting Cluster\Ethics\Applications RDGC\2018\28 Nov 2018\1. NWU-GK-2018-56 Maphopha\FV\RDGC Permission Letter Maphopha.docm
3 December 2018

File reference: 1.1.4.3

1

Appendix D8: Permission to conduct research: Rhodes University

Human Resources Division
Office of the HR Director
Administration Building, Grahamstown, 6139, South Africa
PO Box 94, Grahamstown, 6140
South Africa
t: +27 (0)46 603 8114
f: +27 (0)46 603 8046
e:l.govender@ru.ac.za

www.ru.ac.za

RHODES UNIVERSITY
Where leaders learn

26 April 2019

Mrs Khomotso Amanda Maphopha
Department of Information Science
University of South Africa (UNISA)
Khomotso.maphopha@ul.ac.za

Dear Mrs Maphopha

REQUEST TO CONDUCT RESEARCH WITH RHODES UNIVERSITY STAFF AND/OR STUDENTS

This letter is to confirm that your request to conduct research on *"Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations."* topic has been approved by the Ethics Committee. In my capacity as HR Director, I do not have any objection should you wish to follow a coordinated approach by surveying and/or interviewing staff.

Yours sincerely

Dr L. Govender
HR Director

Appendix D9: Permission to conduct research: Stellenbosch University

UNIVERSITEIT · STELLENBOSCH · UNIVERSITY
jou kennisvennoot · your knowledge partner

INSTITUTIONAL PERMISSION:

AGREEMENT ON USE OF PERSONAL INFORMATION IN RESEARCH

Name of Researcher:	Khomotso Maphapha
Name of Research Project:	Subject analysis theories and their application to geographic subject metadata for electronic theses and dissertations.
Service Desk ID:	IRPSD 1088
Date of Issue:	2 November 2018

You have received institutional permission to proceed with this project as stipulated in the institutional permission application and within the conditions set out in this agreement.

1	WHAT THIS AGREEMENT IS ABOUT	
What is POPI?	1.1	POPI is the Protection of Personal Information Act 4 of 2013.
	1.2	POPI regulates the entire information life cycle from collection, through use and storage and even the destruction of personal information.
Why is this important to us?	1.3	Even though POPI is important, it is not the primary motivation for this agreement. The privacy of our students and employees are important to us. We want to ensure that no research project poses any risks to their privacy.
	1.4	However, you are required to familiarise yourself with, and comply with POPI in its entirety.
What is considered to be personal information?	1.5	'Personal information' means information relating to an identifiable, living, individual or company, including, but not limited to:
	1.5.1	information relating to the race, gender, sex, pregnancy, marital status, national, ethnic or social origin, colour, sexual orientation, age, physical or mental health, well-being, disability, religion, conscience, belief, culture, language and birth of the person;
	1.5.2	information relating to the education or the medical, financial, criminal or employment history of the person;

1
Institutional Permission Standard Agreement: 13 March 2017 V1

	1.5.3	any identifying number, symbol, e-mail address, physical address, telephone number, location information, online identifier or other particular assignment to the person;
	1.5.4	the biometric information of the person;
	1.5.5	the personal opinions, views or preferences of the person;
	1.5.6	correspondence sent by the person that is implicitly or explicitly of a private or confidential nature or further correspondence that would reveal the contents of the original correspondence;
	1.5.7	the views or opinions of another individual about the person; and
	1.5.8	the name of the person if it appears with other personal information relating to the person or if the disclosure of the name itself would reveal information about the person.
Some personal information is more sensitive.	1.6	Some personal information is considered to be sensitive either because:
	1.6.1	POPI has classified it as sensitive;
	1.6.2	if the information is disclosed it can be used to defraud someone; or
	1.6.3	the disclosure of the information will be embarrassing for the research subject.
	1.7	The following personal information is considered particularly sensitive:
	1.7.1	Religious or philosophical beliefs;
	1.7.2	race or ethnic origin;
	1.7.3	trade union membership;
	1.7.4	political persuasion;
	1.7.5	health and health related documentation such as medical scheme documentation;
	1.7.6	sex life;
	1.7.7	biometric information;
	1.7.8	criminal behaviour;
	1.7.9	personal information of children under the age of 18;
	1.7.10	financial information such as banking details, details relating to financial

2
Institutional Permission Standard Agreement: 13 March 2017 V1

Appendix D10: Permission to conduct research: Durban University of Technology

Directorate for Research and Postgraduate Support
Durban University of Technology
Tromso Annexe, Steve Biko Campus
P.O. Box 1334, Durban 4000
Tel.: 031-3732576/7
Fax: 031-3732946

18th March 2019

Mrs K.A. Maphopha
c/o Information Science
University of South Africa

Dear Mrs Maphopha

PERMISSION TO CONDUCT RESEARCH AT THE DUT

Your email correspondence in respect of the above refers. I am pleased to inform you that the Institutional Research and Innovation Committee (IRIC) has granted full permission for you to conduct your research "Subject analysis theories and their application to geographic subject metadata for theses and dissertations" at the Durban University of Technology.

The DUT may impose any other condition it deems appropriate in the circumstances having regard to nature and extent of access to and use of information requested.

We would be grateful if a summary of your key research findings can be submitted to the IRIC on completion of your studies.

Kindest regards.
Yours sincerely

PROF CARIN NAPIER
DIRECTOR (ACTING): RESEARCH AND POSTGRADUATE SUPPORT DIRECORATE

Appendix E: Turnitin digital receipt

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: K A MAPHOPHA
Assignment title: Complete dissertation/thesis submission for examination
Submission title: Thesis for examination (1)
File name: THESIS_FINAL.docx
File size: 670.99K
Page count: 371
Word count: 105,429
Character count: 616,522
Submission date: 27-Feb-2022 06:22PM (UTC+0200)
Submission ID: 1771939751

Thesis for examination (1)

Appendix F: Confirmation of editing

10 Jack Nicklaus Drive

Pecanwood Golf Estate

Hartbeespoort 0216

26 February 2022

TO WHOM IT MAY CONCERN

Please be advised that I, EM (Lucia) Geyer (ID Number 580425 0023 082), edited the doctoral thesis of Ms KM Maphopha (Student Number 7103670) entitled

> ***Subject Analysis Theories and their Application to Geographic Subject Metadata for Electronic Theses and Dissertations***

The editing exercise included the following:

- Language editing;
- Structuring;
- Formatting; and
- Bibliographic control: checking of text references and bibliographic entries.

I edited this dissertation to the best of my ability, based on my extensive experience as an academic and an academic editor.

I take no responsibility for the suggestions and changes that I made to the manuscript that the student has not accepted.

Sincerely

EM (Lucia) Geyer

lgeyer@gmail.com

Mobile: 081 368 9014