Author/s:
Syahrudin, Syahrudin

Title:
Improving Spatial Metadata Usability for Spatial Data Users

Date:
2019

Persistent Link:
https://hdl.handle.net/11343/235618

Improving Spatial Metadata Usability for Spatial Data Users

By

Syahrudin

A thesis submitted to the University of Melbourne in total fulfilment of the requirements of the degree of Doctor of Philosophy

August 2019

Centre for Spatial Data Infrastructure, Department of Infrastructure Engineering,
School of Electrical, Mechanical & Infrastructure Engineering,
The University of Melbourne,
Victoria, Australia

This page is left intentionally blank

Abstract

Spatial metadata plays an important role in promoting spatial data sharing and re-use to support activities that require spatial data such as development planning. It contains information about geographic or spatial data descriptions that will help spatial data users to discover and select spatial from spatial data catalogue systems or geoportals.

However, the usability of spatial metadata for spatial data users is questioned. On one hand, metadata is created following standards originally designed for spatial data producers for data inventory purposes with the information mostly come from the producers' knowledge and understanding about the spatial data. On the other hand, spatial data users come from different backgrounds with different level of knowledge to spatial data that is not necessarily similar to the owners' knowledge.

The need for improving the usability of spatial metadata for spatial data users has necessitated an understanding of the interaction between data users and the metadata, as well as with the spatial data. Due to the necessities of gathering such behavioural data from actual spatial data users, in an environment they are familiar with, User-Centred Design (UCD) has been employed as an approach. Think-Aloud Protocol (TAP), interviews and surveys were implemented as empirical UCD methods for this research.

At the first stage, TAP, interviews and surveys were conducted to evaluate the usability and to identify the usability problems of the existing metadata and user interface, as well as to gather users' needs and expectations of the metadata based on their needs of spatial data. The aims are to provide the baseline or reference for usability improvement and to gather the required information for developing a user-oriented spatial metadata profile, based on the ISO 19115:2014 metadata standard. The profile was used as the basis for a synchronised metadata development approach, where it was simultaneously manifested into functional metadata and user interface design and capabilities, to be proposed as a solution for improving metadata usability. The TAP and interviews were conducted at the

final stage to test and evaluate the usability of the newly developed metadata and discovery system against the baseline or reference acquired at the first stage.

The findings provide a better understanding of how spatial data users employ metadata for spatial data discovery and selection, based on their activities with the spatial data. The first usability testing found that the usability of metadata is not solely determined by the information contained in the metadata records, but it is also depended on the usability and of the user interface. The results also found that information related to previous user experiences along with the live map viewer of spatial data provide a significant improvement to the effectiveness of the spatial data selection process. In the meantime, the user interface design following the user needs and requirements significantly increase the efficiency of the spatial data discovery and selection process.

Declaration

This is to certify that:

- The thesis comprises only my original work towards the PhD except where indicated.
- Due acknowledgement has been made in the text to all other material used.
- The thesis does not exceed 100,000 words.

Syahrudin

August 2019

Acknowledgement

I would like to express my special appreciation and thanks to my supervisors Dr. Mohsen Kalantari and Professor Abbas Rajabifard, you have been tremendous mentors for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice and support on both research and life have been invaluable. I would also like to thank my committee chair, Dr. Martin Tomko for not only chairing the committee but also for your supports and encouragements during my hardship in this PhD Journey.

I would especially like to thank Sam Amirebrahimi for you have been there to support me since day one until the end line of my PhD Journey. I would also like to thank all the members of the Centre for Spatial Data Infrastructures and Land Administration (CSDILA) and the Centre for Disaster Management for all the discussions and enjoyments. I would also like to thank my special friend and mentor, Arief Erman, for your endless supports. I would also thank to my friend Hilman Karyadi and Hardi Subagyo for your technical supports on the development of the website protype. I would also like to thank Ms. Eileen, you have been such tremendous person. I would also like to thank Balawyn Jones, for making this thesis possible to read.

Words cannot express how grateful I am to my mother, Nurjanah, and my father, Yadih, for everything. To my wife, Hastin, I am very grateful to you for all the sacrifices and prayers. My sunshines, Sofia and Liam, for their cuteness and pure hearts. Finally, I thank Alloh SWT, for letting me through all the difficulties. I have experienced Your guidance day by day. You are the one who let me finish my degree. I will keep on trusting You for my future. Alhamdulillah.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANZLIC – Australia New Zealand Land Information Council

CAD – Computer Aided Design

CEN – The European Committee for Standardisation

CSDGM – Content Standard for Digital Geospatial Metadata

CSW – Catalogue Service for Web

DIGEST – The Digital Geographic Information Exchange Standard

ESRI – Environmental System Research Institute

FGDC – Federal Geographic Data Committee

GIS – Geographic Information System

HCI – Human Computer Interaction

HTTP – Hypertext Transfer Protocol

ICA – International Cartography Association

IHO – International Hydrographic Organisation

INSPIRE – Infrastructure for Spatial Information in Europe

IoT – Internet of Things

ISO – The International Organisation for Standardisation

ISO/TC – ISO Technical Committee

NASA – National Aeronautics and Space Administration

OGC – Open Geospatial Consortium

PHP – Personal Home Page

SDI – Spatial Data Infrastructure

TAP – Think-Aloud Protocol

UNGGIM – UN Committee of Experts on Global Geospatial Information Management

UCD – User-Centred Design

VGI – Volunteered Geospatial Information

WMS – Web Map Service

XML – Extensible Markup Language

This page is left intentionally blank

1. Introduction

1.1. Background and Motivation

Geospatial or spatial data are location properties about geographic features or phenomena, which may include the location or area, relationships and other descriptive information (OGC, n.d.). They are important for evidenced based policy-making processes to manage, monitor and measure sustainability of a development and for supporting citizen services, land management and resource sustainability (Revi, 2014). The availability and the use of spatial data has been growing rapidly in the past decade and is expected to continue in the next decade (Craglia & Shanley, 2015b). This is due, but not limited to, the growing awareness of the importance of spatial data as well as technological advancement in spatial data production in areas such as; earth observation satellite; and information access related to computers and the Internet (Belward & Skøien, 2015; Craglia & Shanley, 2015a).

Most spatial data are made available over the networked environment such as Spatial Data Infrastructures (SDIs) to support these activities. With the vast amount of spatial data, the role of metadata for managing, locating and disseminating information about the data is vital (Göbel & Lutze, 1998; Göbel, 2001). Metadata is data about data. It contains information about the description of a specific resource, which might include bibliographic and other relevant information like title and abstract (Smits, 2009). Spatial metadata is the terminology used for metadata of spatial data or information. It contains specific information that can only be possessed by geographically related data or information, for instance, coordinates and spatial references.

In spatial data inventory, metadata has been acknowledged for its essential role in data management purposes such as data storage cataloguing, data maintenance and data retrieval (Tsou, 2002). Metadata is also acknowledged for its essential role in spatial data sharing, by providing information to data users to help them discover, select and acquire spatial data for their specific projects or applications (Chan et al., 2001; Smits, 2009), promoting spatial data use, and encouraging people to use spatial data as the basis for their daily activities (Olfat et al., 2012).

Current applications for spatial metadata are limitless, since the application of spatial data and information has been growing rapidly. The advancement of technologies such as Linked Data, Internet of Things (IoT), deep learning and machine learning, has brought the application of spatial data and information into the next level (Chen & Yang, 2019; Jiang et al., 2018; Yong et al., 2018). For example, farmers use the maps to overlay their fields with historical data and sensor data such as soil moisture, pH and other data to produce crops more efficiently (Specka et al., 2019). The sensors can provide information, including locating where the crops need irrigating, with the exact percentage of moisture, at the exact time. Farmers can also be connected to the rain forecast data to manage their irrigation schedules (Zenus, 2018).

Metadata plays a very crucial role in the new spatial data applications. With large amounts of data from objects or sensors, the communication and data exchange between the objects becomes a problem. Metadata can be used to identify the sensors and the data using standardised information, so the objects or sensors can find a suitable protocol to communicate with each other.

However, the discovery of spatial data from geoportals for average and general users is still challenging as there are differences between users' terms and the metadata (Lafia et al., 2019). With the limitless possible applications of the metadata, at the very least we need to know how usable the metadata for discovering and selecting appropriate spatial data for spatial data users. Moreover, with the growing number of spatial data users outside the spatial field with specific knowledge, the needs and requirements of spatial data users should also be considered. The metadata should make spatial data from specific domains become visible and discoverable (Chafiq et al., 2018; Kaláb et al., 2019).

Most of spatial metadata records contain descriptions about spatial data or datasets following a metadata standard such as ISO 19115. The standard provides a framework that dictates how metadata should be created, for example, by metadata elements and classes of information. This standard has been used by many organizations or authorities to create and develop metadata for managing their spatial data inventories. The same metadata is published to spatial data users via numerous web catalogue systems or geoportals that might help users to find, select and acquire spatial data for their projects or applications.

However, the usability of the metadata for data users is questioned, as the standards were originally designed for spatial data producers for data inventory purposes rather than for spatial data users (Goodchild, 2007; Comber et al., 2007; Goodchild, 2008).

The need for an investigation into the usability of metadata for data users is crucial given the importance of spatial metadata for enabling spatial data sharing and to support the spatial based decision-making process, thus advancing spatially enabled society. Since spatial data users interact with metadata through user interfaces, the effectiveness and efficiency of metadata to deliver appropriate information about spatial data to users depends on the user interfaces. Therefore, the design and development of spatial metadata should be synchronised with the interface to ensure that the information about spatial data is delivered and presented appropriately. The study is expected to give insights and propose solutions to improve the usability of spatial metadata for spatial data users.

1.2. Research Problems

Despite its important role in spatial data sharing and management, spatial metadata is less usable for spatial data users to discover and select spatial data for their purposes. Less or no user involvement in spatial metadata creation, implementation of data producer-oriented standard for metadata creation, and loosely coupled development of user interface for spatial metadata with both the users and metadata creator contribute to the usability of spatial metadata. As the usability of metadata is defined by both the metadata records and the user interface, the current approach in spatial metadata improvement do not address all the issues. Therefore, a synchronized improvement approach is required to increase the usability of spatial metadata.

1.3. Hypothesis

Implementing a user-oriented spatial metadata profile as the basis for a synchronised development approach for metadata and user interface will increase the usability of the spatial metadata for data users.

1.4. Research Aim

The aim of this research is to improve the usability of spatial metadata for spatial data discovery and selection by proposing user-oriented spatial metadata profile as the basis for synchronised development approach for the metadata and the user interface.

1.5. Research Scope

Spatial metadata usability is subject to specific groups of users, i.e. spatial data owners, spatial data users, and computers. Since each user group have their own specific purposes in using spatial metadata, the usability of spatial metadata is different from one group to another. This research is focused on the usability of spatial metadata to spatial data users who mainly use metadata for spatial data discovery and selection from geospatial data portals or web catalogues.

1.6. Research Questions and Objectives

This, therefore, leads to the following major research question:

> How can user-oriented spatial metadata profile be developed and implemented in synchronised spatial metadata development to improve the usability of spatial metadata?

The question is divided into the following sub-questions:

1. What are the current spatial metadata usability and problems?
2. What do users need and expect from spatial metadata?
3. How can the usability problems, user needs and user expectations be used to develop the user-oriented metadata profile?
4. How can the profile be used to develop user-oriented metadata and user interface?

5. How do users perceive and react to the newly proposed metadata and user interface?

The research objectives are, therefore, as follows:

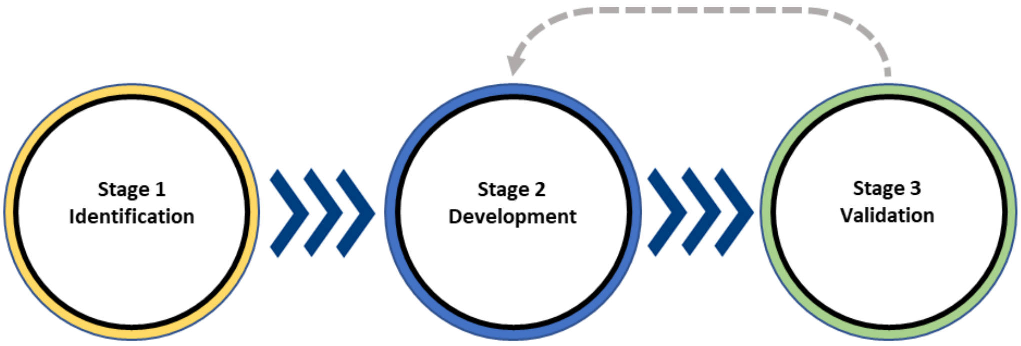
1. To investigate spatial metadata usability, identify problems and user experience of spatial metadata.
2. To identify the user needs and expectations of spatial metadata and compliance of the metadata standard, the ISO 19115-1:2014, to these needs.
3. To propose a synchronised metadata development based on the user-oriented metadata profile.
4. To test and evaluate the profile.

1.7. Research Approach

This research employs User-Centred Design (UCD) as an implementation of mixed method approach, with the emphasis on the qualitative. This unique mixed method is designed specifically for this research to achieve the objectives, thus answering the questions. The mixed method approach is also implemented to gain the breadth of the research and to to triangulate data collection for validation purposes. As a qualitative approach, the research focuses on the exploration of the phenomenon, in this case exploring spatial data users' insights into their interaction with spatial metadata as the basis for their decisions for discovering and selecting spatial data.

Three different UCD methods are used as the main data collection methods; Think-Aloud Protocol usability testing (TAP), interviews, and survey. The analysis methods for each collected data were selected accordingly to ensure the extraction of resourceful results and findings. The detailed explanation of the methodology selection is presented in Chapter 3, Research Methodology.

The research is designed consisting three major stages, starting with the identification stage, followed by the development stage and the implementation, and concluding with the validation stage, as illustrated in Figure 1.1.



Research objective(s) Objectives 1 Objectives 2	**Research objective(s)** Objectives 3	**Research objective(s)** Objectives 4
Research activities Determining usability of established spatial metadata Identifying user needs and expectations of spatial metadata	**Research activities** Developing the user oriented metadata profile Implementing the profile into metadata and user interface prototypes	**Research activities** Evaluating the usability of the prototypes
Expected outputs Usability status of the established spatial metadata A list of user needs and expectations of spatial metadata	**Expected outputs** User-oriented spatial metadata profile as an extension of ISO 19115:2014 spatial data discovery metadata Metadata database prototypes Spatial data portal prototypes	**Expected outputs** Usability status of the prototypes User acceptance of the prototypes

Figure 1.1: Research Approach

The identification stage starts with investigations on spatial metadata developments and presentations by reviewing the literature and established catalogue systems, followed by a usability evaluation of established spatial metadata presented in data portals or web catalogue systems. The former is meant to establish an understanding on how the metadata has been created and delivered to spatial data users and to define the scope of the research. The latter is designed to determine the usability of publicly accessible spatial metadata, to identify the problems appearing in spatial metadata use and to gain user experience and perceptions on spatial metadata to help them discover and select spatial data. This yielded the usability status of the established metadata and a list of usability problems. It also captured spatial data user experiences of spatial metadata, including the factors that affected their decision-making during spatial data discovery and selection. Further, the evaluation yielded users' initial needs and expectations of spatial metadata as well as their approach to discovery and decision-making relating to appropriate spatial data. The results were then used as the basis for designing a questionnaire to be circulated in a survey for gaining wider perspectives of the needs and the expectations of users.

The development stage is divided into two sub-stages. The first sub-stage is the identification of user requirements and expectations of spatial metadata for spatial data discovery and selection. The second sub-stage is the development of the user-oriented spatial metadata profile, as an extension of discovery metadata of ISO 19115:2014. It starts by interpreting and analysing spatial data users' needs and expectations of spatial metadata from the survey and the usability testing. This yields a list of metadata elements and their presentation styles for metadata delivery as expected by the users. The results are incorporated with the metadata discovery of ISO 19115:2014 and used for extending the discovery metadata and to create the user-oriented metadata profile.

The last stage of the research is the validation stage. Is starts with the development of the metadata and user interface prototype. Employing system development process, the profile is used for identifying the requirements and functionalities of the metadata records and user interface. The metadata records and the user interface are created and developed accordingly. It is tested and evaluated using a practical approach by spatial data users using the prototype. TAP and interviews were conducted to evaluate the usability as well as to gather user experiences and user acceptance of the prototype.

As the approach is qualitative, to demonstrate validity, this research declares the biases the researcher brings to the research. As for maintaining the reliability, data triangulation will be implemented by conducting different types of data collection: usability testing (TAP), interviews, and surveys. For demonstrating generalisability or transferability of the results, analysis and presentation of the results will use thick and rich description of the spatial data users' perceptions about the spatial data requirements when they search and select spatial data and their expectations of spatial metadata. Reliability was also maintained by including negative or contrary information to the analysis and report.

1.8. Thesis Structure

The next chapter will outline the principal theories and concepts related to spatial metadata usability and an explanation of the importance of spatial metadata usability to support spatial data sharing. It starts with the introduction of fundamental spatial metadata concepts. This includes spatial metadata definitions, roles, users, standards and systems. It then addresses the elements in metadata systems that support the full functionalities of the metadata in supporting spatial data sharing. It also briefly explains spatial data infrastructure as an initiative for facilitating spatial data sharing, including concepts, elements and current developments, as well as recent implementations of spatial data and metadata, for example, Smart City and IoT. Exploration of metadata standards as well as the metadata system will be emphasised in the interaction between metadata and spatial data users. Usability, as a concept, and its implementation for spatial metadata will then be explored and described. Chapter Two concludes with the identification of usability concepts, including the elements, that will appropriately portray the use of spatial metadata by spatial data users to be the basis for designing the research methodology.

Chapter 3 outlines the reasoning for selecting the research methodology for this research. It describes the process of developing the strategy for answering the research questions as well as achieving a set of objectives, provided in Chapter 1. The subjective nature of spatial data discovery and selection is embraced to demonstrate the objectivity of this study. The methodologies for studying subjective phenomena are then outlined as well as the appropriate methods, with User-Centred Development (UCD) is selected as the research approach. The initial usability problems were identified by implementing UCD data collection methods, Think-Aloud Protocol (TAP) and semi-structured interviews. Both methods are also expected to gain insights from spatial data users on their interactions with spatial metadata as well as their expectations. The findings were then used to propose a solution that addresses the problems accordingly. A survey is conducted to get wider user requirements on specified problems and the findings used to develop a user-oriented metadata profile. The profile is used as the basis for synchronised development of spatial metadata records and the user interface prototype. TAP and semi-structured interviews are used in the validation stage to test and evaluate the usability of the prototype.

Chapter 4 presents a detailed explanation of the usability testing utilising TAP and semi-structured interviews. The testing is conducted to identify existing usability problems in spatial metadata use as well as to gain initial insights about spatial metadata-spatial data user interactions. The chapter starts with explanations about the selection of TAP as the method by contrasting it to other UCD methods. It then explains the user selection process, followed by the development of the test design and implementation. Data analysis, findings and results are presented and explained. The outcomes of this chapter are crucial for the development of the user-oriented metadata profile.

Chapter 5 reports the results of the survey conducted to gather wider spatial data users' needs and expectations of spatial metadata. It starts with an explanation about the importance of conducting the survey and the process of designing the questionnaire. The results from the survey are described after an explanation of the survey implementation, including participants' selection. This chapter concludes with the user-oriented metadata profile as an extension to discovery metadata of ISO:19115-1:2014 spatial metadata standard. Key information related to spatial data characteristics used to determine the criteria for spatial data discovery and data fitness for purpose for spatial data selection are explained.

Chapter 6 discusses the validation process of the proposed user-oriented spatial metadata profile. The validation is conducted through development of spatial metadata and user interface prototype, implementing a synchronised approach created based on the profile, and usability evaluation. A series of usability tests employing TAP and interviews are conducted employing selected spatial data users as participants. This chapter then discusses the results and findings against the usability results from the identification stage.

Chapter 7 concludes the study by addressing the aims and objectives, answering the questions and highlighting the contributions to knowledge and discusses future developments or directions of the research.

1.9. Chapter Summary

This chapter lays the foundations of this research as it introduces the motivation for conducting the study, research questions, aim, objectives, approach, and the structure of the research.

What motivates this study is that increasing use of spatial data from various backgrounds outside the spatial field followed by increasing amounts of spatial data available via the Internet or networked environment, is not necessarily followed by metadata development that will increase the discoverability and select-ability of the available spatial data. As stated in the research problem, spatial metadata are less usable to data users as they were originally designed to help data producers or owners to manage their data inventories. Spatial data users are not necessarily familiar with information presented in metadata and neither the user interface. Consequently, spatial data users find it difficult to discover and select appropriate spatial data for their purposes from web-based spatial data catalogues or geoportals that base their systems on metadata.

As a response to the problem, the research is designed to put forward four objectives. Firstly, it investigates the current spatial metadata usability and identifies the problems faced by spatial data users in their attempts to use spatial metadata. Secondly, it explores and identifies the spatial data users' requirements for spatial data discovery and selection. Thirdly, it designs and develops a user-oriented spatial metadata profile as an extension to discovery metadata of the ISO 19115:2014. The user-oriented metadata profile is meant to be a tool for an integrated approach that will improve the metadata usability of both the metadata and the user interface.

Finally, a validation process is implemented by first implementing the profile into metadata and user interface prototypes using a synchronised development approach and testing them against usability criteria as used in the previous usability testing.

The next chapter provides fundamental concepts of spatial metadata and their importance in supporting spatial data sharing. It includes the definition, the roles and functions, as well the standards and systems of spatial metadata, and recent applications. It also includes the description of usability concepts and how the concepts can be implemented and are important for spatial metadata.

This page is left intentionally blank

2. Literature review on spatial metadata

2.1. Introduction

Spatial data is captured and processed following specific purposes for specific users. Even generic spatial data, such as topographic datasets, are not generic for every user. From that specific spatial data, characteristics of the data are captured and collected to create spatial metadata. Spatial metadata are delivered to spatial data users to help them discover the data and understand the characteristics of the spatial data. These users might belong to various communities which are different from the specified users whom the data was created for.

This chapter reviews the literature on spatial metadata utilisation as the background for this research. It starts with a review of fundamental concepts and principles of spatial metadata including the roles and functions of spatial data sharing. It is then followed by a review of spatial metadata systems, from their earliest development to their delivery to and use by users. To better understand the use of spatial metadata by users, literature on usability as a concept for describing product-user interaction is also reviewed in this chapter.

2.2. Spatial metadata

Metadata is data about specific data. A metadata record contains information about the description and characteristics of specific data. The information is structured in pre-defined elements representing specific attributes of the data, that is not specified in the data itself (Danko, 2011). The information provided in metadata can be used to find and locate specific resources or data (Smits, 2009) as well as to retrieve, use or manage the data (Timpf et al., 1997). The term was initially used in computer science literature in the late sixties (Aalders, 2005; Howe et al., 1996). A more technical definition of metadata is:

> structured data about an object that supports functions associated with the designated object (Greenberg, 2003).

In information systems, metadata is used extensively and comes in many applications or software packages used in everyday life (e.g. photo browsers, music players, and video browsers). Metadata plays a key role to the functionality of these applications to manage content, enable users to find items of interest, and even to record and share items with others (NISO, 2014).

In the spatial community, metadata was first known in the paper-map era as the marginalia containing information about properties and characteristics of a map (Aalders, 2005). Metadata, as known in the context of information systems, was first used when digital spatial data was introduced with the introduction of Computer Aided Design (CAD), and geographic information system (GIS) used for data analysis (Olfat, 2013). However, it was not until the early 1990's that the term metadata found its place in the geospatial community, when the USA introduced an initiative to develop the National Spatial Data Infrastructure and the Federal Geospatial Data Committee (FGDC), followed by the first FGDC metadata standard in 1994 (Aalders, 2005).

The FGDC – as adopted by ISO/TC211 – was the technical committee responsible for preparing standards in the areas of geographic information and geomatics, defining metadata as data about content, quality condition, and other characteristics of data. To distinguish metadata for geospatial data or resources from, and to prevent confusion with, other disciplines, the term geospatial or spatial metadata was introduced in the spatial community. The International Cartographic Association (ICA) in their technical meeting in Den Haag, The Netherlands (1996) adopted the spatial metadata definition as follows:

> Data that describe the content, data definition and structural representation, extent (both geographic and temporal), spatial reference, quality, availability, status and administration of a geographic dataset (Cooper et al., 2013).

As an information object, spatial data have three features that should be reflected through spatial metadata, as follows (Gilliland, 2008):

- *Content*. Information about what the data is about, intrinsic to the data.
- *Context*. Information that indicates any aspects associated and extrinsic to the data.

- *Structure*. Information about formal associations for an individual or group of data, both intrinsic and extrinsic.

Therefore, spatial metadata are created in a structure containing sets of element types to describe different aspects of spatial data or resources and their respective metadata (Aalders, 2005) – as illustrated in Figure 2.1.

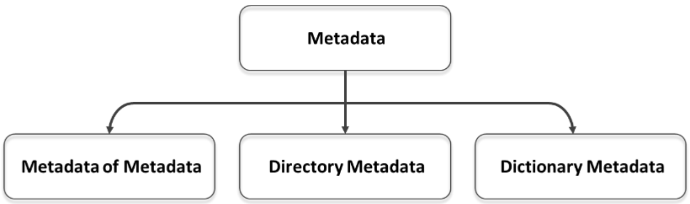


Figure 2.1: Spatial metadata element types (Adapted from Aalders, 2005)

'Metadata of metadata' is a type of metadata element containing information to describe the metadata as an entity including the organisation of the metadata, the language used, and the reference system for semantic, geometric and temporal extent. 'Directory metadata' are used to describe the spatial data characteristics including the identification, data quality, data extent, data origin, and data administration. Lastly, 'dictionary metadata' is used to describe the semantics of data and the schema for the data model to allow search engines to access and interpret the data from different multiple metadata catalogues following the semantics based on thesauri for the metadata management (Albertoni et al., 2018).

Spatial metadata elements set in the metadata types contain information that can be used to describe the following, but are not limited to, characteristics of spatial data (FGDC, 1998):

- *Data identification*. Basic information about the spatial data that can be used to identify the data including the data citation, data description, production and maintenance date, data status, spatial and/or temporal extent, theme, contact, etc.

- *Data quality*. Information about the results and process of a general assessment of the quality of the data including positional accuracy, attribute accuracy and lineage.
- *Spatial data organisation*. Information about the mechanism for representing spatial information in the spatial data, including vector and raster representation mechanism.
- *Spatial reference*. Information about spatial reference frame used in the spatial data including horizontal and/or vertical coordinate system and associated definitions.
- *Entity and attribute*. Detailed information about the content of the spatial data including the entity types, attributes, and domains of the attributes.
- *Distribution*. Information about the responsible party for the distribution of the spatial data including contact information and options for obtaining the data.
- *Metadata reference*. Information about the metadata including the currentness and responsible party for maintaining the metadata.

This section established what information is provided in the spatial metadata, the next section will build on this, discussing the roles and importance of spatial metadata for the spatial community.

2.2.1. Spatial metadata roles and benefits

Spatial data or information has been gaining importance in society, especially for evidence-based, decision-making processes in the context of planning, managing, monitoring and measuring the development. Spatial data are also important in smart city development for supporting citizen services, land management and maintaining the sustainability of resources (Revi, 2014). Moreover, proliferation of spatial information systems and spatial applications has inevitably led to dramatically increased spatial data production. Comprehensive research conducted by Geospatial World in 2018 reported that the geospatial industry is projected to reach US$439.2 billion by 2020, growing at 13.6% during 2017-2020, significantly higher than the growth rate of 11.5% during 2013-2017. With surveying and earth observation constituting more than 60% of the market share, the amount of spatial data continues to skyrocket (GeoBuiz 2018 Report - Global Geospatial Outlook 2018, 2018).

The plethora of spatial data available from this extensive spatial data production requires more effort towards data management, and the importance of spatial metadata is becoming apparent. For instance, Synthetic Aperture Radar (SAR) images are readily available, but the sensor specific nature of the data made them hardly discovered by potential users (Pierce, 2019). Big volume of available other satellite images also requires a tool for inventory management and integration between distributed databases (Fan et al., 2018). Therefore, the main role of spatial metadata is to help data owners to catalogue their spatial data in a structured way so that they are able to locate and retrieve the data effectively and efficiently (Limbach et al., 2003; Tsou, 2002). Beside the role of managing spatial data inventories internally, metadata is also essential as the enabler of spatial data interoperability (Danko, 2011) for introducing and making spatial data known to external users, who might need access to the data for their activities, should the owners want to share the data (Batcheller et al. , 2009). If we look at the capabilities of metadata in the information systems mentioned in the earlier section, e.g. in audio data management (Woodcock et al., 2018) or in education (Rautenbach et al., 2019) spatial metadata can play a similar role in these spatial communities by enabling spatial data users to discover and select spatial data of interest, or even to organise their own virtual database of available spatial data and share this selected data with other data users. Spatial metadata can also be utilised for specific applications, e.g. for integration and evaluation of heterogeneous spatial planning documents (Iwaniak et al., 2016) or for enhancing photogrammetric process in spatial data production (Dhanda et al., 2018). In disaster management, metadata can also play an important role such as in the development of evacuation plan of Merapi Volcano disaster in Indonesia (Yuliadi et al., 2011).

Metadata can also be used for sensor discovery and interoperability as implemented by NASA (Di et al., 2009; Yong et al., 2018) and for seamless integration and connection for multi-sourced sensors in natural disaster management (Alamdar et al., 2016) or agriculture (Specka et al., 2019). With vast development of Linked Data and Internet of Thing (IOT) in geospatial communities, applications for spatial metadata are limitless. Sensors and devices are connected and employed to support spatial applications. For example, farmers can now use maps to overlay their fields with historical data and sensor data such as soil moisture, pH and other data to produce crops more efficiently. These sensors can provide information, including locating where crops need irrigating, the exact percentage of moisture and the exact time. Farmers can also be connected to rain forecast data to manage their irrigation schedules. In the middle of these activities, spatial metadata play an important role to ensure sensors and devices are connected and functioned correctly (Zenus, 2018).

Unlike other data, spatial data have specific characteristics which data users might acquire in order to discover and select the data by accessing and working with the data directly (e.g. data resolution or scale, spatial reference, geographic extent or coverage). However, spatial data largely remain inaccessible as data are normally presented in large size files and not on a free basis (De Martino & Albertoni, 2011). Spatial metadata have the capability to help users to discover and assess spatial data by providing information about the data characteristics without directly accessing or downloading the data. They even can be used for cross-border spatial data sharing if there is common semantic descriptions of the used vocabularies (Herrera et al., 2019).

There are further benefits of spatial metadata that can assist either spatial data owners or data users, adapted from FGDC CSDGM (FGDC, 1998) and ISO 19115 (ISO, 2014), as follows:

- To enable spatial data owners to maintain and manage their spatial data inventory effectively.
- To facilitate spatial data sharing and re-use by providing information about the spatial data inventory to data catalogues or brokerages.
- To help and facilitate spatial data users to discover spatial data from data catalogues.

- To help spatial data users to determine spatial data fitness for use for their specific needs.
- To help spatial data users to access spatial data by providing information to acquire the data.
- To help spatial data users to process, use and transfer spatial data.

With such detailed information about spatial data characteristics, spatial metadata allow users to integrate and compare spatial data from different sources or organisations. Therefore, spatial data users are able to find, select and obtain the most appropriate and most accessible spatial data for their purposes (Chan et al., 2001), including geospatial intelligence for marketing or geo-marketing (Guarda et al., 2019; Guarda & Augusto, 2019).

Considering spatial metadata's significant benefits and potential uses, spatial metadata plays an important role in the development and implementation of spatial data sharing platforms, i.e. Spatial Data Infrastructure (SDI) (Batcheller, 2008). Without metadata it is difficult to manage large amount of spatial data and at the same time enable spatial data users to effectively and efficiently discover spatial data from these sources. Spatial metadata can act as a surrogate for the spatial data that can be managed and shared efficiently within metadata catalogues or clearinghouses.

SDIs are designed for spatial data and information sharing, yet they represent more aspects regarding spatial data and information, ranging from data production, management, dissemination and utilisation. Spatial data availability and access are crucial for spatial data sharing, but SDIs do not stop there. Most of the spatial data and information are underutilised not because they are unavailable or inaccessible, but it is because the users or potential users do not understand the functions or the benefits of the data for their activities, e.g. development planning at local government level in Indonesia (de Vries, 2006). Asseng et al. (2018) in their study confirmed it by assessing the determinant factors for local SDIs in Indonesia, where they found that the organisational and technological aspects are the dominant factors, with less human resources nor data aspects affecting the development of local SDIs in Indonesia. Moreover, Sutanta et al. (2014) obtained similar findings in their study for determining the SDI readiness index for the local SDI development in Indonesia.

Therefore, the role of spatial metadata can be expanded as a tool for increasing the utilization of spatial data by providing information about the functions or benefits of the data. Further, with a standardised structure such information can be easily exchanged or transferred between metadata catalogues. Spatial data users only need to access the data catalogues and interact with spatial metadata to discover available spatial data from different sources or organisations that reside on the network (Phillips et al., 1999). They might also access and integrate spatial data from open big data platforms (Brovelli et al., 2019) or from volunteer geographic data catalogues (Imran, 2019). With its role and importance for spatial data sharing, spatial metadata is defined as one of the fundamental indicators for evaluation of spatial data sharing platforms (Najar et al., 2007).

Having established the roles and benefits of spatial metadata for spatial data sharing platforms, the next section will explain the process through which spatial metadata is created, managed, delivered and presented to spatial data users. These explanations are important to define the interactions between spatial data users and spatial metadata, and to identify any issues or challenges that might arise from these processes – which will form part of the background of this research.

2.2.2. Spatial metadata system

Spatial metadata as data about data is designed, created, managed and distributed in a process called 'spatial metadata system'. The most common metadata system to support spatial data sharing platform, i.e. Spatial Data Infrastructure (SDI), is geospatial data cataloguing as proposed by Nebert (2004), illustrated in Figure 2.2.

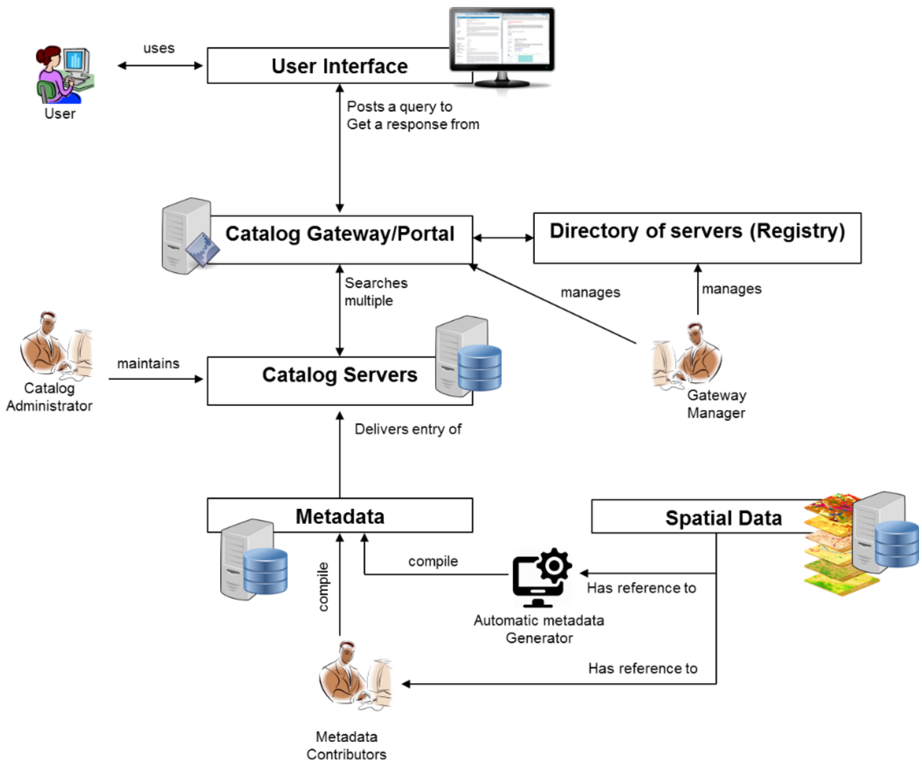


Figure 2.2: Example of spatial metadata system (Adapted from Nebert, 2004)

Within a geospatial data catalogue system, spatial metadata is created based on information about characteristics of spatial data by metadata contributors or by automatic metadata generators. Compilations of spatial metadata records are stored in metadata databases to be harvested by metadata catalogue servers maintained by catalogue administrators. The catalogue administrator(s), using a set of rules or criteria, ensures that metadata entries from the metadata database comply with the rules, before sending them to the catalogue servers. The catalogue administrator(s) are also responsible for updating and maintaining the catalogue servers by adding or deleting metadata entries. The gateway manager is responsible for developing, hosting and maintaining the distributed search capabilities for users. The capabilities will be implemented in the user interface for users to interact with the metadata catalogue servers by posting queries and receiving responses as defined in the catalogue gateway.

Following this system, there are three stages to be completed before spatial metadata reaches the spatial data users:

- Catalogue service development – the area where spatial metadata is created, stored, managed and maintained in the metadata catalogue server.
- Catalogue gateway and access interface development – the area where search capabilities or functionalities and the presentation of the user interface is designed and developed for users to access and interact with the metadata from registered catalogue servers.
- Catalogue service registration – the area where the catalogue server is registered so they can be known in a community and be a searchable by the catalogue gateway.

Leaving the catalogue registration as a permission makes the spatial metadata catalogue service accessible. Making spatial metadata accessible enables spatial data users to discover and select spatial metadata, depending on the condition of spatial metadata records and the functionalities of user interface to retrieve, manipulate and present the information provided in the metadata.

The next section explains spatial metadata creation to identify possible issues that could influence the condition of spatial metadata records from spatial data users' perspectives.

2.2.3. Spatial metadata creation

Most spatial metadata in spatial data catalogues in SDIs are created by metadata experts (Olfat et al., 2010), following widely used standards, i.e. ISO 19115. There are also crowdsourced metadata, but they are created mainly for open data initiative, e.g. OpenStreetMap (Neumaier et al., 2018). The latest version of the ISO 19115 was established in 2014, and the latest version of the FGDC Content Standard for Digital Geospatial Metadata (CSDGM) was published in 1998. Although spatial metadata creation can be conducted automatically or semi-automatically, manual creation is still the dominant practice in the geospatial community (Olfat et al., 2013). The reason lies in the nature of the characteristics of the information related to spatial data, which can be categorised into two groups: inherent and contextual (Hill, 1999).

Inherent information is information about data characteristics that is embedded in the data, e.g. the spatial bounding box, spatial reference, feature type, spatial data representation and other information that is part of the data. These are sometimes also called objective or implicit metadata elements (Jokela, 2001). This category of information can be extracted automatically from the data using machine algorithms or applications such as ESRI ArcGIS software, and therefore requires no or less human interference.

On the other hand, explicit metadata elements or contextual information is information about data characteristics which is given or assigned to spatial data by metadata authors, e.g. title, abstract, purpose, price, and contact point fall into this category. This type of information is mostly entered into metadata manually and therefore is highly dependent on human interference. Most metadata elements fall into this category, as reviewed in the next section regarding spatial metadata standards.

Since manual data creation is highly dependent on human intervention, the following issues may emerge:

- Metadata creation is considered time-consuming and costly, therefore most organisations are reluctant to create extensive metadata and tend to create only the very minimum metadata for their purposes, mainly for data storage and retrieval (Guptill, 1999; West & Hess, 2002; Mathys, 2005).
- The quality and relevance of such information is highly dependent on the authors' knowledge or information collectors' knowledge. As proposed by (Greenberg, 2003), metadata authors can be grouped into four categories: professional metadata creators, technical metadata creators, content creators and community subject enthusiasts. Each category of authors has a different level of metadata knowledge, with professional creators having the most sophisticated knowledge. Not every organisations can afford to assign professional metadata creators who have sophisticated knowledge and capacity related to metadata classification system or standards. Even with professional authors, information presented in the explicit metadata is still highly dependent on the knowledge possessed by the particular author, which is not necessarily the same between one author to another.

- As the result of the two above-mentioned problems, inconsistency in spatial metadata, as well as missing or incomplete information, in spatial metadata is common (Kalantari et al., 2009). The inconsistency is also found extensively that hamper and limit accessibility of the spatial data (Schauppenlehner & Muhar, 2018).

The fact that spatial metadata is mostly created following a standard, does not guarantee that spatial metadata is consistent or that its content will be relevant to spatial data users. The first reason is that the metadata creation process is usually disconnected from the users, either professional or casual users (Kalantari et al., 2014). The second is the content of the standard itself, which was originally designed for data owners for data inventory and management purposes (Goodchild, 2007; Comber et al., 2007; Fisher et al., 2010). Therefore, to check the quality of the metadata, two approaches can be implemented based on who the users are, i.e. following ISO 19157 – spatial data quality for spatial data owners (Ureña-Cámara et al., 2019) and following user perspective on geospatial data quality (Bielecka et al., 2014). To explore these issues further, the next section reviews the most widely known and used metadata standards in geospatial community.

2.2.4. Spatial metadata standards

Metadata standards are developed to ensure that all aspects of spatial data are covered, to maintain consistency, and to facilitate interoperability between metadata catalogues (Aalders, 2005; Litwin & Rossa, 2011). By implementing standards, it is expected that comparisons between spatial data can be made quickly and easily as the result of consistent metadata, both in content and presentation (Kalantari et al., 2009). Within the geospatial community, there are two groups of standards for metadata. The first group establishes the rules to create metadata and the second establishes the rules for metadata catalogues or services.

Within the first group, three known standards are used widely for creating spatial metadata: ISO 19115, FGDC CSDGM and Dublin Core. ISO 19115, the latest version having been released in 2014, is widely used in most countries worldwide, including Australia, European and Asian countries, while FGDC is mainly used in the USA. Dublin core is only used specifically for communities that require simple and general metadata.

The second group includes ISO/TS 19139:2007 for implementation of the XML schema for encoding the spatial metadata from ISO 19115, OGC Catalogue Service for Web (CSW) and OGC CSW Application Profile for ISO 19115. These standards provide a set of rules to create and maintain catalogue services and provide capabilities for discovery and publishing so the metadata within the catalogues could be registered and discovered within a community, i.e. spatial data infrastructures (SDIs).

The following sections review FGDC CSDDM and ISO 19115 as the most widely used standards for spatial metadata creation in order to identify their characteristics and determine metadata conditions as one of the key factors for spatial metadata usability.

FGDC Content Standard for Digital Geospatial Metadata (CSDGM)

Within the geospatial community, the Content Standard for Digital Geospatial Metadata (CSDGM) is one of the earliest standards used to create metadata for geospatial resources, along with other regional standards, e.g. ANZLIC Core Metadata Elements, Canadian CEN Env 12657, IHO S 57 (Aalders, 2005). Developed in 1994, with the latest version published in 1998 by the Federal Geospatial Data Committee of the USA, this standard was designed to provide common terminologies and definitions for documenting digital spatial data for the National Spatial Data Infrastructure (FGDC, 1998).

It was developed from the users' perspective to help them discover, determine fitness for use, access and transfer spatial data. Accordingly, a set of metadata elements and information about the content of these elements were developed for these purposes. The standard is accompanied by complimentary documents to help create the metadata, i.e. CSDGM Workbook 2000 and metadata quick guide 2006, created by FGDC. and CSDGM essential metadata elements. Table 2.1 presents the main sections/chapters of the standard and metadata elements covered in the CSDGM metadata standard (FGDC, 2000).

Table 2.1: CSDGM metadata sections

Metadata Section	Main Sections	Supporting Sections
Metadata	Identification Data Quality Spatial Data Organisation Spatial Reference Entity and Attribute Distribution Metadata Reference	Citation Time Period Contact

As per the above table, the CSDGM contains three primary sections/chapters: metadata, main and supporting. The 'metadata section' is the starting point and is composed of the 'main sections' of the metadata. The 'main sections' are the body of the standard and contain all spatial data aspects. The 'supporting sections' supplement the main sections by providing a set of methods to define citation, temporal aspects and contact information to be used in the respective main sections, e.g. contact for metadata reference, distribution and data quality.

Moving on to consider each of the 'main sections', the CSDGM assigns an option for every element: Mandatory (M), Mandatory-if-applicable (M-A) or Optional (O). Elements which are assigned M must be provided in the metadata. Contrastingly, O titled elements are provided based on the discretion of the metadata author. M-A is the title used for elements that must be provided if the data exhibits specifically defined characteristics. In the main sections, only identification and metadata reference information are assigned as mandatory. The information contained in these sub-sections is considered mandatory. Consequently, their supporting sections are also assigned as mandatory as they contain follow-on information relevant to the identification and metadata reference sub-sections. The rest of the elements in the main sections are mandatory-if-applicable. The optionality status of the CSDGM sections is illustrated in Figure 2.3.

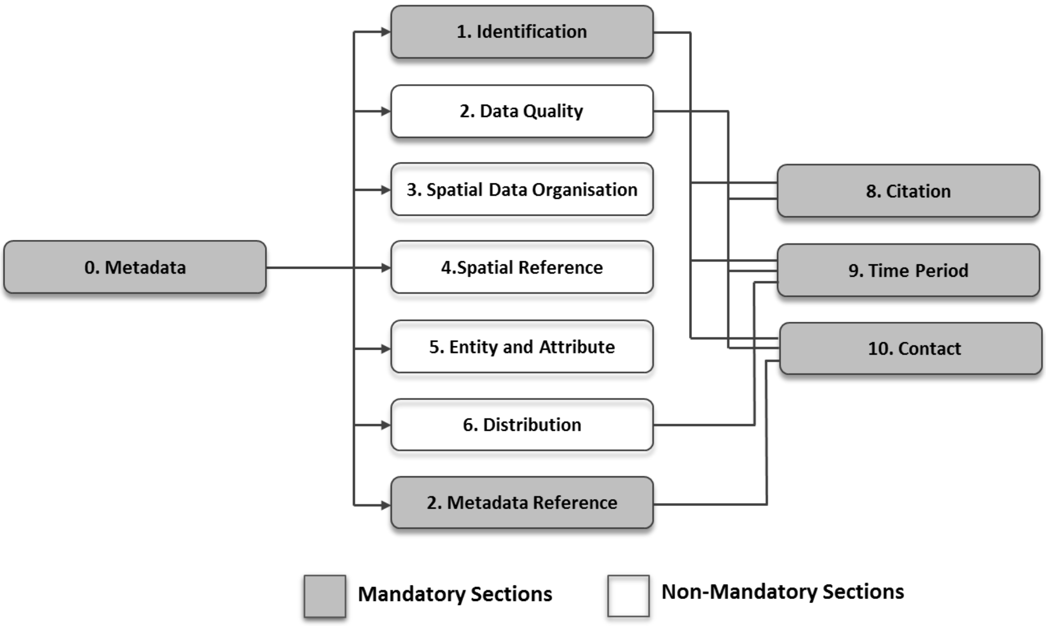


Figure 2.3: FGDC CSDGM Optionality of Metadata Sections (FGDC, 2000)

Implementation of the standard for creating metadata is another challenge. That is why FGDC has published complimentary documents to be used as guidelines for implementation. Following the essential metadata elements document, there are core and/or minimum metadata elements that can be used as the template for organisations to commence standard implementation. Examples of these elements are presented in Table 2.2.

Table 2.2: Examples of the essential metadata elements from CSDGM

Identification	Data Quality	Entity and Attribute Information
Originator	**Logical_Consistency_R**	**Entity_and_Attribute_Overview**
Publication_Date	**Completeness_Report**	**Entity_and_Attribute_Detailed_Description**
Title	**Process_Description**	
Abstract	**Source_Online_Linkag**	
Purpose	*Cloud_Cover*	
Currentness_Reference	Attribute_Accuracy_Rep	
Progress	Horisontal_Positional_Accuracy_Report	
Maintenance_and_Update_Frequency	Vertical_Positional_Accuracy_Report	
Use_Constraints		
Point_of_Contact		
Online_Linkage		

The bold text in the Table 2.2 above represents the mandatory sections and the italic text represents the optional ones. The normal text represents elements which are mandatory if applicable. In the identification section, since that section is mandatory, all of the elements included in the section are mandatory. This is distinct from the mandatory elements in the 'data quality', 'entity' and 'attribute information' sections, as these only become mandatory if the information is applicable. Optional elements remain optional regardless of their position in the sections or the applicability of the information, based on the discretion of the metadata author.

CSDGM provides detailed information about every element in the standard including how to collect relevant information and write information about the elements. Followings are examples of metadata elements descriptions provided in the standard:

1.2. Description -- a characterisation of the data set, including its intended use and limitations.
Type: compound
Short Name: descript

1.2.1. Abstract -- a brief narrative summary of the data set.
Type: text
Domain: free text
Short Name: abstract

1.2.2. Purpose -- a summary of the inventions with which the data set was developed.
Type: text
Domain: free text
Short Name: purpose

The above examples are descriptions of Abstract and Purpose elements. The two elements belong to the 'description' section, used for describing the characterisation of the spatial data set. These sub-sections are presented with a definition, short name, type and domain of the element, which is free text. This means that the metadata creator can type any text associated to the definition, e.g. brief narrative summary of the data set. With that broad definition, the metadata creator can put almost anything into the abstract. To help create relevant and consistent information for metadata elements, CSDGM published the CSDGM Workbook and metadata quick guide, as presented in examples below.

Section 1: Identification

Citation (1.1)

Originator (8.1)

It is recommended that you indicate the party responsible for the data set. While that is most commonly the organization that developed the data set, in some cases, it is not. For example, if a county planning department hires a contractor to build a street centerline road file, the planning department, not the contractor should be identified as the Originator. The contractor should be fully cited using the Data_Set_Credit (1.11) element, e.g. 'this data set was developed for the Wayne County Planning Department by Smith Engineering, Inc'.

Title (8.4)

- Minimum – where, what, when
- Best practice – who, why, resolution, filename, source e.g. Aquifer Systems and Recharge Potential in Louisiana from LDEQ source data, Geographic NAD83, LOSCO (1999) [aqrgeog3dpdeq]

Abstract (1.2.1)

Be sure to include

- General content and feature
- Data set form (GIS, CAD, Image, Dbase)
- Geographic coverage (county/city name)
- Time period of content (begin and end date or single date)
- Special data characteristics or limitations

It is expected that spatial metadata can be created consistently with relevant information as described in the quick guide and workbook. However, the domain remains free-text and there are no restrictions whatsoever related to the length of the information, e.g. in the abstract. This might be an issue for maintaining the consistency and relevance of the information in metadata, especially when manual creation is implemented.

FGDC CSDGM allows organisations or communities to make some modifications to the standard and provides a guideline to do so. Such modifications can be made to the development of profiles, extended metadata elements or both. A profile might be developed when the optionality or repeatability of non-mandatory elements is required to meet the requirements of an organisation or a community. Creating extended metadata elements is allowed when metadata elements required by an organisation or a community are not provided by the standard.

ISO 19115 Geographic information – Metadata

ISO (the International Organisation for Standardisation) published the first ISO 19115 metadata standard for geospatial resources in 2003. The latest version was published in 2014 as a revision of the 2003 version.

The standard was prepared by Technical Committee ISO/TC 211, Geographic information/geomatics, responsible for standardisation in the field of digital geographic information (ISO). In the process of developing the standard, ISO/TC211 carried out a series of studies on a number of standards available from different regions such as the Australia-New Zealand standard (1995), Canadian standard (1995), European CEN Env 12657 (1996), USA FGDC CSDGM (1994), DIGEST (1994), IHO S 57 (1995) (Aalders, 2005). That is the reason behind the adoption of ISO 19115 by many international geospatial communities, including ANZLIC in Australia and New Zealand, and INSPIRE in the European Union.

ISO 19115 is considered as the most comprehensive and the most complicated metadata standard for geospatial resources (Tsou, 2002). Unlike FGDC CDDGM, which focuses on metadata contents, ISO 19115 provides the schemas for cataloguing and transferring metadata. Although it is designed for geospatial resources, ISO 19115 is applicable to other types of resources, e.g. software, hardware, non-geographic datasets. This is due to the extensive set of the metadata elements it provides. In the 2014 version, the standard is defined and provided by 13 packages and three additional packages. Within the 13 packages, there are two minimum packages that should be used for a complete metadata set, i.e. metadata and identification, as illustrated in Figure 2.4.

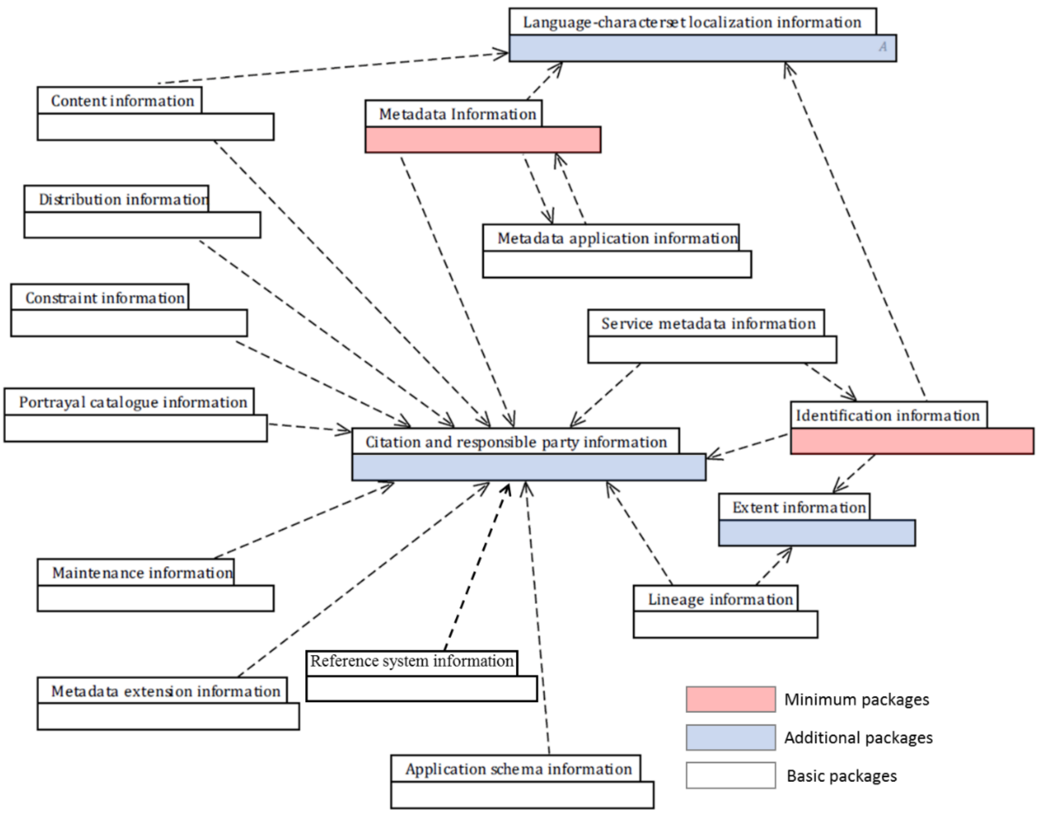


Figure 2.4: ISO 19115:2014 metadata packages (ISO 19115:2014)

Each metadata package in Figure 2.4 contains one or more metadata classes and associated attributes. Full metadata is provided by 12 classes from metadata and identification packages, as illustrated in Figure 2.5. Except for data quality, other classes are provided in the ISO 19115:2014, i.e. identification information, constraint information, maintenance information, application schema information, portrayal catalogue information, content information, distribution information, lineage information, spatial representation information, reference system information and metadata extension information. Data quality is provided in ISO 19157.

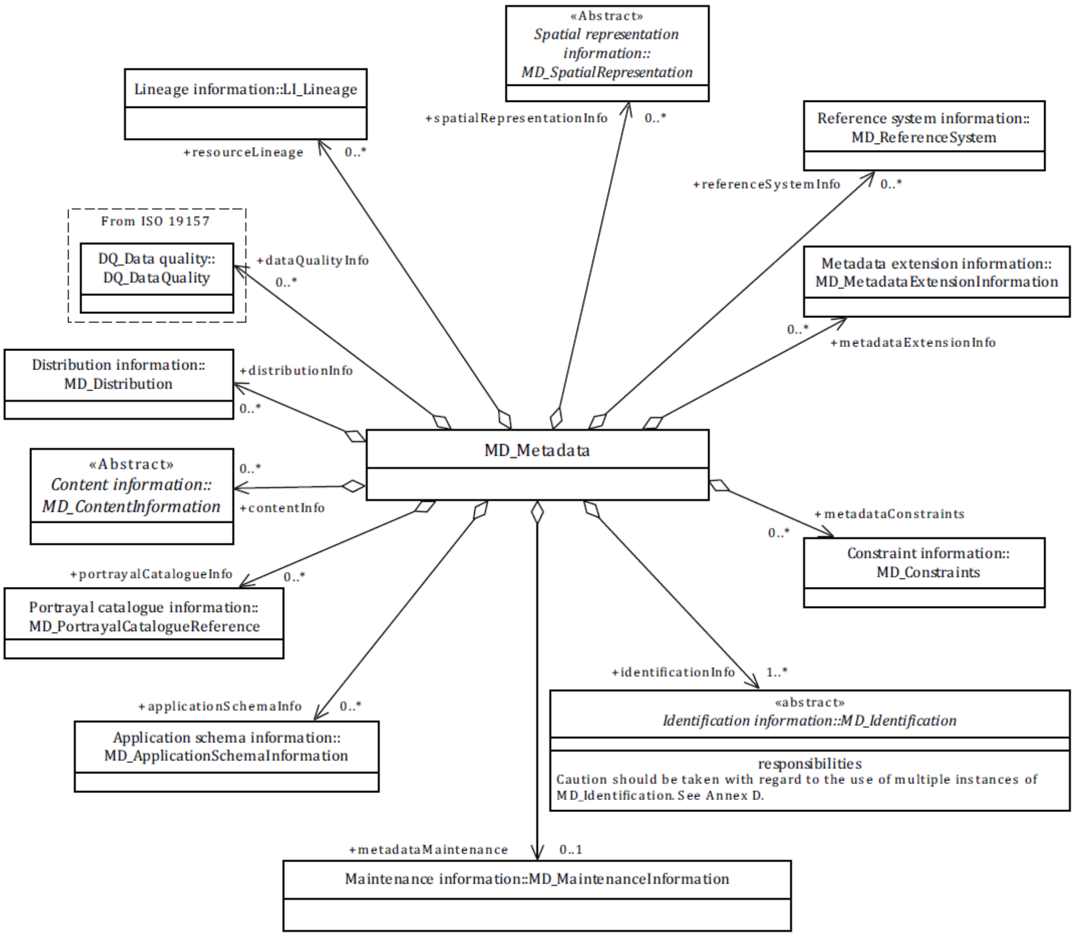


Figure 2.5: ISO 19115:2014 Metadata classes (ISO 19115:2014)

Each sub-class in metadata class contains one or more metadata elements designed to describe specific information related to spatial data characteristics. Table 2.3 illustrates examples of metadata elements provided in the ISO 19115:2014.

Table 2.3: Examples of metadata element in ISO 19115:2014

	Name / Role name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
42.	***MD_Identification***	**basic information required to uniquely identify a resource or resources**	**Use obligation from referencing object**	**Use maximum occurrence from referencing object**	**Aggregated Class (MD_Metadata) <<Abstract>>**	**Lines 43-62** **NOTE Caution regarding the use of multiple instances of this class. See Annex E.**
43.	citation	citation for the resource	M	1	Class	CI_Citation (Table B.16)
44.	abstract	brief narrative summary of the resource	M	1	CharacterString	Free text
45.	purpose	summary of the intentions with which the resource was developed	O	1	CharacterString	Free text
46.	credit	recognition of those who contributed to the resource	O	N	CharacterString	Free text
47.	status	status of the resource	O	N	Class	MD_ProgressCode <<CodeList>> (B.3.25)
48.	pointOfContact	identification of, and means of communication with, person(s) and organisation(s) associated with the resource	O	N	Class	CI_Responsibility (Table B.16.1)
49.	spatialRepresentationType	method used to spatially represent geographic information	O	N	Class	MD_SpatialRepresentation TypeCode <<CodeList>> (B.3.29)
50.	spatialResolution	factor which provides a general understanding of the density of spatial data in the resource or describes the range of resolutions in which a digital resource may be used NOTE This element should be repeated when describing upper and lower range.	O	N	Class	MD_Resolution <<Union>> (Table B.3.4)
51.	temporalResolution	smallest resolvable temporal period in a resource	O	N	Class	TM_Duration (B.2.4)
52.	topicCategory	main theme(s) of the resource	C / is metadataScope/ resourceScope equal "dataset" or "series"?	N	Class	MD_TopicCategoryCode <<Enumeration>> (B.3.30)

NOTE 1: The UML model for this table is shown in Figure 6.

NOTE 2: SV_ServiceIdentification is introduced in Table B.14.

Each metadata element consists of role name, definition, obligation, maximum occurrence, data type, and domain. Obligation refers to the optionality of the element – in this standard these are mandatory, optional or conditional. A category of optionality is assigned to each element or class. Except for the 'conditional' category, the obligations are similar to the options used in FGDC CSDGM. Mandatory (M) elements in optional classes only become mandatory if the class is used. Optional (O) elements are optional regardless of the class. Conditional (C) classes or elements can be mandatory if one of the following three conditions occurs:

- There are two options where at least one of the options must be documented.
- There is an associated element which dictates that this C element must be documented when the associated element is documented.
- There is an associated value of another element which dictates that this C element must be documented when the associated value is documented.

The data type and domain in the ISO metadata standard are similar to those in FGDC CSDGM. From the examples in Table 2.3, abstract and purpose have the same free text domain, and even the same definition. Unlike FGDC, ISO/TC211 does not publish complimentary documents as guidelines for metadata creation. Organisations or authorities that create metadata following the ISO standard create and distribute guidelines for their own use or for use by other organisations under their jurisdiction, e.g. ANZLIC (ANZLIC, 2007) and INSPIRE (European Commission Joint Research Centre, 2013). ANZLIC metadata guideline for the 'Purpose' element can be seen in Table 2.4.

Table 2.4: Purpose description and guidance in ANZLIC Metadata Profile Guideline Version 1.2

Name (Number)	Purpose
Definition	Summary of the intentions with which the resource(s) was developed
Data type	CharacterString
Domain	Free text
Meaning & Purpose	This purpose element is similar to the abstract element in that it amplifies the results of resource searches to permit users to obtain a better appreciation of the resource and assists in determining the fitness for purpose. This element will not be used as a search criteria
Guidance	This element allows for a free text description of why the resource was developed

In accordance with such guidance, ANZLIC would expect that spatial metadata under their jurisdiction can be created consistently with relevant information as described in the guidelines. However, with the free text domain and no restrictions whatsoever related to the length of the information maintaining the consistency and relevance of the information in metadata is difficult, especially for manual metadata creation.

With extensive metadata elements (more than 400 elements) provided in the ISO 19115:2014, it is expected to cover most information related to the characteristics of spatial resources, either for vector spatial datasets, images, grid data or for spatial services. The standard also provides a set of elements for discovery purposes, called discovery metadata for geographic resources. Discovery metadata is similar to the core metadata elements in the previous version, however discovery metadata is designed to facilitate spatial data discovery in catalogue services and clearinghouses.

The standard also provides guidelines for creating metadata extensions and community profiles. An extension can be made by adding a new element, creating a new code-list to replace the free text domain, expanding a code-list, imposing a more stringent obligation on an existing metadata element or by imposing a more restrictive domain on an existing element. A community profile is recommended when new information is extensive, or where many new elements need to be added.

Having reviewed the spatial metadata creation process and the standards commonly used by organisations and authorities to create metadata, the next section will review the delivery and use of spatial metadata by spatial data users to identify other possible issues that may influence the usability of spatial metadata for spatial data users.

2.3. Spatial metadata delivery and presentation

As explained in Section 2.2., spatial metadata is delivered to and accessed by users via a user interface. This user interface is usually a geoportal – a website that can be used as an entry point to metadata records from databases and catalogue services. These portals are mostly embedded with search engines to enhance the discovery of the spatial data within spatial data infrastructures (Corti et al., 2018).

There are many geoportals available and accessible on the internet. To name a few, the INSPIRE portal, Australian open data portal and US open data portal are available for the public to search and access metadata, illustrated in the following figures.

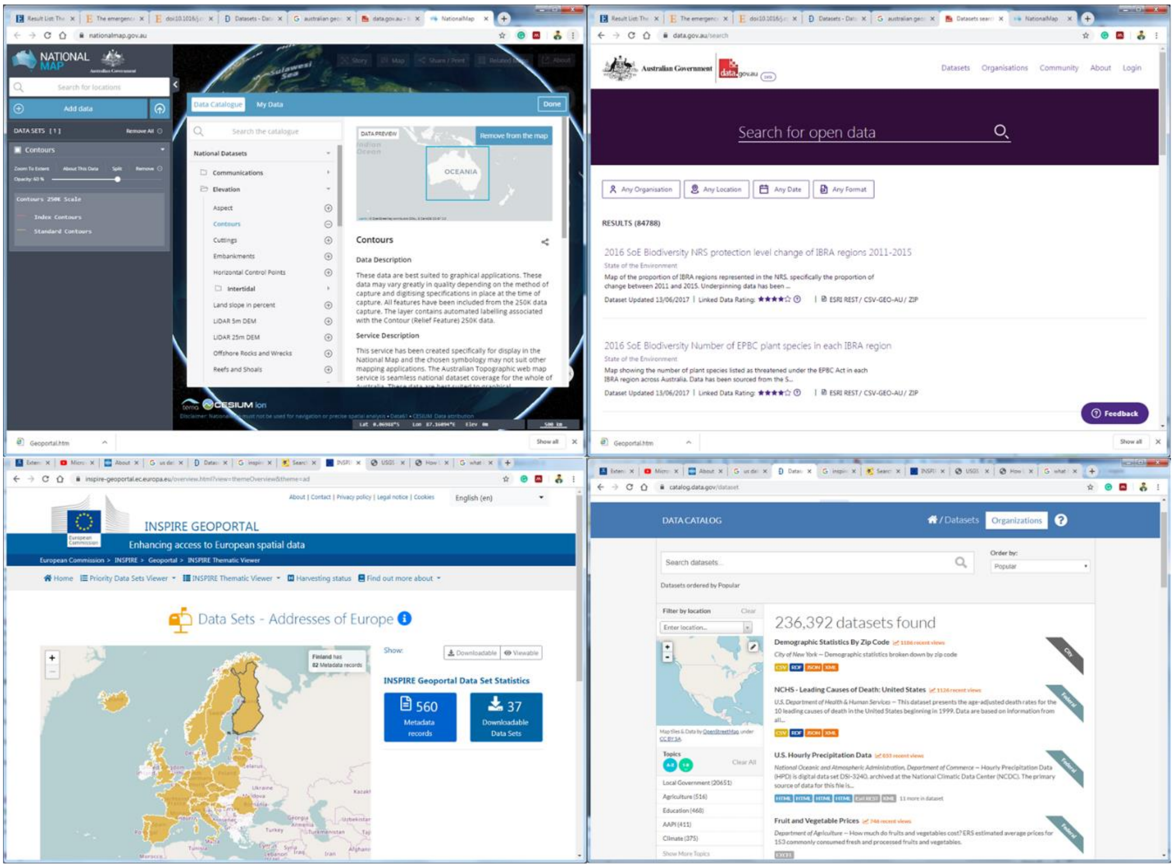


Figure 2.6: Examples of geoportals available on the internet

Most geoportals have been transformed from geospatial-based portals into open data portals. Increasing awareness of the importance for integrating spatial data with non-spatial data has influenced authorities and organisations to integrate their spatial and non-spatial portals into a single data portal. Three years ago, Australia had FIND as a geospatial portal for discovering national spatial data, but now the portal has been replaced by the Australian open data portal. Australia has also developed and established a national map as the portal for visualisation of the spatial data. INSPIRE (European Union Countries) still maintains their spatial portal, however they have brought other non-spatial data onto the portal.

These user interfaces play a major part in determining the usability of spatial metadata. Functionalities of the user interface, including the presentation of the information from metadata records need to meet user requirements and interests. Moreover, these user interfaces must ensure that important and relevant information from the metadata records are delivered and presented to spatial data users. Any changes in the user interface should anticipate impacts on their users. The following section reviews studies and developments on spatial metadata related work to identify the challenges for usability and potential improvements.

2.4. Spatial metadata usability related studies

Spatial metadata usability as explained in the previous sections is a combination of the usability of the metadata records (i.e. completeness and relevance of the information) and usability of the user interface (i.e. geoportals/websites). This section investigates the major research and development activities regarding spatial metadata that contributes to usability improvements.

2.4.1. Spatial metadata creation

In the spatial metadata creation domain, spatial metadata is mainly created manually by those who are assigned as the metadata authors. This creates inconsistency including incompleteness of information presented in metadata (Kalantari et al., 2009). Automated metadata creation was proposed as the solution to address this problem. Several methods have been proposed to achieve automatic creation of metadata, e.g. regular expression, machine learning (Han et al., 2003). Norlund (2010) proposed metadata creation based on a proprietary software by developing an algorithm for extracting information from spatial data and storing the information into a metadata template from INSPIRE. In addition to this, several automatic information extractions for updating metadata have been proposed, including using GML based spatial data (Olfat, 2013) or using WFS data services (Kalantari et al., 2017).

Most challenges for automatic metadata creation and updating relate to the type of information included in spatial data characteristics. Only a few pieces of information about spatial data can be automatically extracted and stored into metadata, e.g. the reference system, projection, date of last update, and lineage statement. All other information (e.g. abstracts and data quality descriptions) still rely on manual input by humans. Another challenge is that extraction can only be effective for new spatial data where metadata can be collected at each step in the spatial data production process. Automatic extraction for existing data can only be made if modifications or updates are made to the data. Moreover, as in manual metadata creation, the proposed automated metadata creations do not closely involve or consider users in the process.

There have been attempts to involve users in the creation of metadata, both direct or indirect. For example, direct involvement has been conducted by directly engaging users to provide information for metadata creation. A study by (Kowal & Prbreveidal, 2012) illustrated significant success in direct user involvement for metadata creation. By engaging the public in a crowd-sourced participatory mapping project, they managed to obtain information about bounding box coordinates for their paper maps inventory. Kalantari et al. (2014) also implemented crowd-sourcing and tagging with both direct and indirect user/public involvement to create metadata for Volunteered Geospatial Information (VGI). Direct or explicit involvement was implemented in these cases by allowing users to provide their comments on the spatial data through a tagging tool provided on the VGI web application. On the other hand, indirect or implicit user involvement is conducted by collecting and analysing users' search words and assigning the results as tags for the spatial data. For example, Olfat et al. (2013) has conducted indirect user engagement for the purposes of automatic metadata enrichment.

These studies have demonstrated the potential of user involvement in the creation of spatial metadata. However, they have faced similar challenges to creating complete and consistent metadata with all of the required information. Another issues with respect to user involvement is data validity, as users might exhibit biases and, as such, expert validation is imperative (Salk et al., 2016).

2.4.2. Spatial metadata user interfaces

Studies on spatial metadata interfaces can be grouped into two categories: visualisation and manipulation. The first category focuses on techniques for visualising information from metadata elements such as utilisation of color intensity (Ware, 1988), 3D-graphics (Jung et al., 1999) and scatterplots (Limbach et al., 2003). The second category focuses on the development of algorithms or mathematical models to be applied to metadata to help users better interpret the information from the metadata, such as for assessing the suitability or fitness for use of the data (Ivánová et al., 2013; Devillers et al., 2007).

Despite their potential for improving the usability of spatial metadata by providing eye-catching presentation and increased effectiveness of metadata delivery to users, current visualisation studies have only focused on specific elements of spatial metadata, e.g. data accuracy (Ahonen-Rainio & Kraak, 2005), geographic space and temporal attributes, and quantitative attributes (Jung et al., 1999). Moreover, (Ahonen-rainio, 2005) and (Boin & Hunter, 2008) assert that there is little evidence that visualisation techniques can accurately communicate metadata information, especially with respect to presenting data quality and margins of uncertainty.

On the other hand, the manipulation approach has demonstrated potential improvements in spatial metadata usability, e.g. focusing on a single characteristics such as temporal to be the main determinant (Corti & Lewis, 2017) or based on a specific case reasoning (Li et al., 2017). However, existing studies have only focused on specific group of users, e.g. spatial experts (Devillers et al., 2002), and are highly dependent on the condition of the metadata.

Beside the visualisation and manipulation approaches, there have been other studies that have explored the usability of spatial metadata, but they are not limited to and not always relating to user interface, i.e. focusing instead on semantic or linked metadata. These studies have predominately been about creating interlinks between metadata by transforming metadata elements from XML format into linked data format, e.g. RDF and associated spatial data using the triplets assigned in the RDF form of metadata (Patroumpas et al., 2014; Athanasiou et al., 2015; Patroumpas et al., 2015; Kyzirakos et al., 2018) or by developing ontology-based dictionaries and implementing web ontology services to facilitate the interoperability and link-ability of metadata (Arenas et al., 2014; Lacasta et al., 2007). These studies potentially improve the usability of metadata by providing context to the spatial data by linking the spatial data based on ontology-based terminology. Other efforts for making the spatial metadata is more usable are by trying to implement machine learning for intelligent geospatial to help users find the most relevant spatial data (Jiang et al., 2018).

Having reviewed the relevant studies on spatial metadata with respect to usability of metadata, the following section reviews usability concepts and studies about the usability of spatial metadata.

2.5. Usability

Usability is a concept about how well and how easily users can use all the functionalities of a product or a concept of the ease of use of a product for users (Nielsen, 1993). The term was coined in the early 1980s to replace the previous term 'user-friendly' that was considered vague and subjective (Bevan, 2009a). The definition of usability is derived from a number of views of usability. Following are four views that should be considered when defining the usability of a product (Bevan, 2009b):

- Ergonomic attributes – product oriented;
- User attitudes – user oriented;
- Interaction with the users – user performance oriented;
- Functions to a particular task in a particular environment – context oriented.

Nielsen (1993) suggests that usability is an attribute of an information system or product that is equally as important as utility for a product to be useful to users, as illustrated in Figure 2.7.

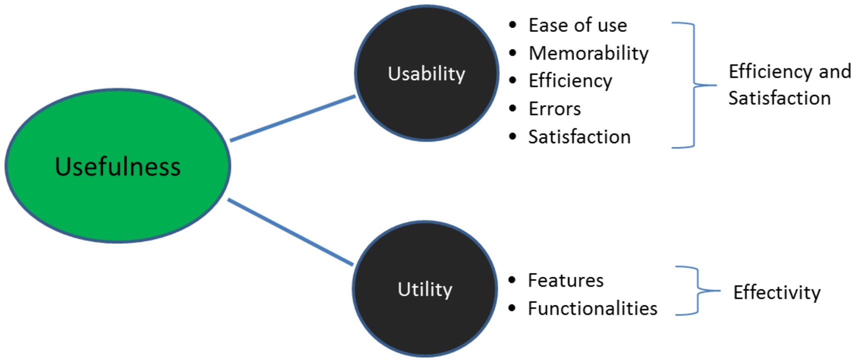


Figure 2.7: Relationship between usefulness, usability and utility

Utility is defined as the functionality or the extent to which a product provides the features users need for particular tasks. It is related to the effectivity of the product to help users to perform specific tasks. On the other hand, usability is related to ease of use and user satisfaction in using the product. Reflecting these views, ISO 9241 (ISO, 1998) adopted a definition of usability as:

> the extent to which a product or system can be effectively and efficiently used by specified users to achieve specified goals in a specified context of use with satisfaction.

In this definition, usability emphasises six keywords: specified users, specified goals, specified context, effectivity, efficiency, and user satisfaction. Therefore, usability of spatial metadata is defined as the extent to which spatial data users can use spatial metadata to successfully and accurately find, locate or assess data which is fit for use, and obtain, appropriately use or transfer spatial data (effective) for their specified purposes (specified context) in a reasonable time (efficient)and in a pleasant manner (ease of use and user satisfaction). As pointed out by Kim & Kim (2008), usability attributes are inter-dependent, and can be evaluated as a single attribute or as a group, depending on the purpose of the intended improvement and nature of user requirements.

Usability is arguably as important as the product itself and is a necessary condition for the survival of the product (Nielsen, 1993). For instance, when people need specific information and search for it using the internet, they will go to several websites that can potentially be useful for them. If they find it difficult to use the website, they will exit and no longer use it. If they get lost or the information presented on the website is hard to read or does not answer their key questions, they will leave.

In this sense, usability of the user interface is as important as the metadata itself. For example, when environmental researchers need specific spatial data for their project and decide to search for the data using a geoportal on the internet, if they find that the spatial metadata is difficult to understand because the information is presented in a way that makes it hard to read etcetera they are likely to give up on the geoportal and simply find another way to get the information.

2.5.1. Usability Evaluation and Improvement Methods

Since usability is determined by the quality of the interaction between people (or users) and a system (or a product), it requires a user-oriented approach to measure and improve it. There is an approach in the Human-Computer Interaction (HCI) domain that is specifically used to design a system (or a product) based on user requirements, namely Human/User-Centred Design (UCD).

UCD is an approach to designing and developing a product (or a system) where user requirements – needs, characteristics, abilities and wishes – are taken as the starting point and fulfilling these user requirements is the ultimate goal of the product design (Vredenburg et al., 2002). Various methods are available for implementing UCD at every stage in the design and development process depending on the purpose of the UCD (Abras et al., 2004). There is a list of UCD methods that can be implemented in three design stages based on the purpose of each method: at the beginning of the design project, early and mid-point in design cycle and final stage of design cycle (Lee, 2005), as illustrated in Table 2.5. According to the list, the usability evaluation of an established (or existing) product is placed at the final stage of the design process.

Table 2.5: UCD methods following product development process (Lee, 2005)

Implementation Stage	Purpose of UCD	Available Methods
At the beginning of Project Design	Collecting user needs and expectations	Interviews and Questionnaires
Early in the design cycle	Collecting data for product functionalities	Interviews and Questionnaires
	Gathering wider requirements and issues of the product	Focus Groups Discussion
	Collecting data about the environment of the product use	On-site observation
Mid-point in the design cycle	Evaluation of alternative designs and gathering additional requirements	Role playing, walkthroughs and simulation
Final stage of design cycle	Evaluation of product's usability	Usability testing
	Evaluation of user satisfaction with the product	Interviews and questionnaires

2.5.2. User in Usability

As mentioned above, usability is a concept related to the ease of use of a product that centres on users (Nielsen, 2010). Therefore, the product should be designed and created based on user requirements in order to provide the best possible user experience with the product. Following this logic, spatial metadata should provide information and present the information following users' requirements. As this research is focused on the usability of spatial metadata, the potential users of spatial metadata need to be clearly identified.

Regarding the implementation of UCD methods, identifying the potential users who may be involved is important, as different type of users will have different needs and, therefore, require different design strategies and outcomes. There are three types of users as proposed by Abras et al. (2004): primary, secondary and tertiary. Primary users are those who actively use the product; secondary users are those who occasionally use the product; and tertiary users are those who actually do not directly use the product but are affected by the use and the purchase of the product and, therefore, may indirectly use the product. For instance, in spatial data for the mining industry, primary users are GIS staffs, secondary users are GIS managers, and tertiary users are executive managers or directors.

As mentioned earlier, spatial metadata users include both spatial data providers and spatial data users. Both users have different purposes for spatial metadata based on their own interests. Spatial data providers use spatial metadata for managing their spatial data inventory as well as for sharing the data. On the other hand, spatial data users use the metadata for searching, selecting and obtaining the spatial data for their own purposes. In this research, the usability of spatial metadata is focused on the secondary category – spatial data users.

Spatial data users are likely to possess different levels of knowledge about spatial data as well as varying competency in using networked based systems (or internet-based portals) (Boin & Hunter, 2008). Following the user definition proposed by Abras et al. (2002), it is necessary to identify who the primary users of spatial metadata in this research are.

Boin & Hunter (2008) proposed three categories of spatial data users: experts, moderate users and inexperienced beginners. Moreover, Boin (2008) discovered that the majority of primary users are likely to be moderate spatial data users – who are likely to use spatial metadata for searching their required spatial data via SDIs or geoportals. Moderate spatial data users are those who work and are familiar with spatial data using one or more types of software. Moderate users have sufficient knowledge to work with and analyse spatial data for their intended use. These users are experts in their fields (e.g. disaster management, spatial planning, environment) and know exactly what to do with spatial data for their own purposes – they are not necessarily spatial data experts but nor are they inexperienced beginners.

2.5.3. Spatial metadata usability studies

Limited research has been undertaken on the usability of spatial metadata (Bulens et al., 2013). There has been some research focusing on how to increase the benefits of spatial metadata, as mentioned in section 2.4, for example by developing specific tools to analyse spatial metadata based on specific intended applications or by developing tools to address multilingual issues. That research implies that spatial metadata is not easy to use for users and that additional efforts are required to achieve optimal benefits from spatial metadata. Yet, only a handful of research has focused on evaluating the usability of spatial metadata or improving the spatial metadata based on usability evaluations.

Research and studies regarding the usability of information products such as spatial data and metadata are relatively few compared to research relating to the usability of interfaces or applications (Brown et al., 2013). Among these limited studies, Boin (2008), in her study about fitness for use from spatial data users' point of view, found that spatial metadata users did not actually use spatial metadata as the main source for evaluating the fitness for use of spatial data. In other study, Ahonen-Rainio (2005) found that spatial data users are mostly interested in only a few elements of spatial metadata, e.g. classification or semantic information (e.g. title, abstract, and contents), up-to-dateness (e.g. production date and last update) and geographic coverage of spatial data to find potential data for their specific applications. Regarding aspects of quality, users found that spatial metadata does not give them the confidence they need to determine the fitness for use of data for their applications (Boin, 2008). They prefer to look for other external sources, e.g. previous users of the data, data creators, or by looking at the reputation of the producers or spatial data experts' opinions (Ahonen-Rainio, 2005; Boin & Hunter, 2008; Ivanova et al., 2013). This phenomenon is considered to be a utility problem of spatial metadata caused by perception differences of fitness for use between spatial data users and producers (Comber et al., 2007; Fisher et al., 2010).

Further, Fraser & Gluck, (1999) assert that the presentation of spatial metadata (e.g. sequence or order of information, text size, layout, length of information) highly affects how users interact and use spatial metadata. However, the user-interface is not necessarily designed to present metadata as expected by users, as found in several data portals in European countries (Bulens et al., 2013). Moreover, information presented in spatial metadata following a given standard, e.g. ISO 19115, may also be confusing for spatial data users when they only want to find specific information which is relevant to their needs (Ahonen-Rainio, 2005; Fisher et al., 2010). Lastly, Bielecka (2015) found that spatial metadata is not readily usable for users to determine the fitness for use as most of the information related to data quality presented in the metadata is not relevant to users.

2.6. Chapter summary

This chapter reviewed the underlying concepts and principles of spatial metadata including its roles and benefits in the geospatial community, and spatial metadata systems including the FGDC CSDGM and ISO 19115:2014 as the most commonly and widely used metadata standards for geospatial resources.

This chapter also reviewed research and development related to spatial metadata assessing the benefits and usability of spatial metadata for users. In the last section, concepts and principles of usability were explained, as well as its implementation in the context of spatial metadata, followed by a review of research on the usability of spatial metadata.

A number of issues and problems were highlighted in this chapter that will be used as the basis of this research:

- Spatial metadata is an information product and users need a tool or platform to interact with and use it. Consequently, usability of spatial metadata is a combination of usability of metadata as information and the user interface.
- In spatial metadata systems, most of the spatial metadata is created manually by humans who does not necessarily possess similar knowledge with the users or even between other metadata authors in different organisations. This may lead to inconsistencies, incompleteness and/or irrelevant information in metadata.
- The commonly used standards do not necessarily meet user requirements, as most of the external information (e.g. abstract, data quality description) required by users have been assigned with free-text domain, and differing

obligations for inclusion. This can create inconsistency, incompleteness and irrelevance in the metadata with respect to user needs.

- Existing research and developments in the area of spatial metadata currently do not address metadata in relation to user interfaces. Moreover, research and studies about spatial metadata usability indicate a low level of spatial metadata usability.
- Lastly, UCD is a potential available approach to investigate and improve usability of spatial metadata.

The next chapter explains the selection process of the research methodology used to carry out this research.

3. Research methodology

3.1. Introduction

Having usability of spatial metadata for spatial data users as the focus of this research brings some challenges to the selection of research approach and methodological foundation. Working with user perceptions and intentions on spatial data is complex, since there are various types of spatial data with their own characteristics and numerous groups of users with their specific applications and varying levels of spatial knowledge and expertise.

This chapter explains the exploration and selection process of the methodology and methods for this research. It begins with an explanation of the conceptual design frameworks, followed by the reasoning for research approach selection. The chapter then discusses the available approaches and methods, and how those methods could be implemented in the research design development to achieve all the objectives of this research. Finally, the chapter justifies the quality judgments for this research.

3.2. Conceptual research framework

Metadata is often irrelevant for users and highly dependent on the context of use and the underlying purposes of use (Bishop & Grubesic, 2016). They are not necessarily useful for spatial data users to help them for finding suitable spatial data for their purposes (Comber et al., 2007; M. F. Goodchild, 2007). Therefore, understanding the process of spatial metadata creation and delivery to its end-users is as important as understanding how the users interact and use metadata to define a conceptual research framework. This framework is essential for setting up research, as it drives the research by identifying the variables and the relationships between variables within the investigated phenomena (Wigton & McGaghie, 2001). As explained in Chapter 2, spatial metadata cannot be treated as a stand-alone product. Instead, information provided in metadata is delivered and presented to users as part of an information dissemination system, for example, a digital catalogue system, as illustrated in Figure 3.1. The most common digital catalogue systems are internet websites or applications, such as, spatial data portals, open data portals and web catalogues.

Figure 3.1 illustrates the variables of spatial metadata usability and the relationships between the variables.

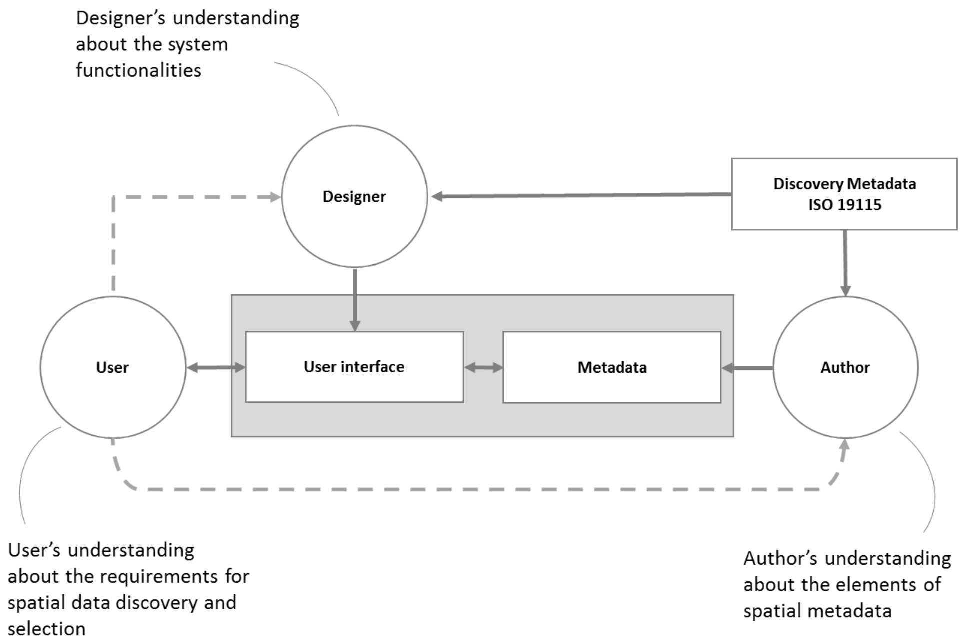


Figure 3.1: Spatial metadata information dissemination system

Metadata is expected to give spatial data users the required information they need to discover and select appropriate spatial data for their purposes. However, the author uses their knowledge and understanding to record and create metadata records following a well-known standard, ISO 19115, which does not necessarily meet user requirements. Moreover, the designer develops the user interface based on their knowledge of how the metadata should be presented to users following a list of functionalities given to them, including which information is to be presented, which again, does not necessarily meet user requirements.

Therefore, the conceptual research framework should be developed to meet the requirements for evaluating and improving spatial metadata usability for spatial data users by addressing all the problems that emerge from the metadata system. Figure 3.2 illustrates the conceptual research framework by showing the connections and relationships between the phenomenon being observed (usability of spatial metadata), research aims and questions, objectives, methodology and methods, and expected outcomes of this research.

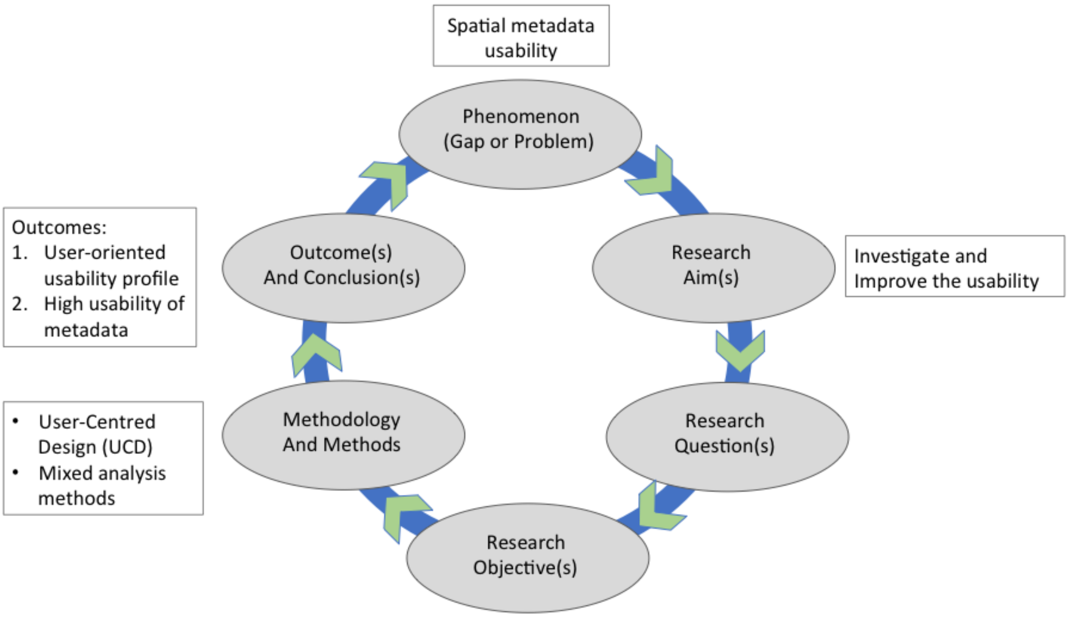


Figure 3.2: Conceptual research framework

Since the research aims to improve the usability of spatial metadata for spatial data users, there are different facets regarding spatial data users that need to be considered. Firstly, differences in their professional backgrounds. Secondly, various levels of spatial knowledge. Thirdly, different intentions and goals for spatial data use. These should be considered in the development of the research questions, thus requiring a research approach which allows these different facets to be treated accordingly.

Understanding the nature and development of the research questions is essential for selecting a research approach able to answer the questions. The first question aims to determine the usability status of current established spatial metadata and to explore the user experiences with the metadata. It is primarily qualitative in nature as the approach emphasises observations made of the interactions between the users and the metadata. The second question aims at identifying the user requirements of spatial metadata for spatial data discovery and selection. It can be answered by using both qualitative and quantitative approaches consecutively, with the emphasis more on the qualitative. The third and last question is qualitative in nature, while the fourth question would suit a prototyping approach.

The above questions can only be addressed by investigating and exploring the interaction between the metadata and the users. Moreover, improving the usability of metadata requires an approach that also involves the users in the re-design and redevelopment process. To meet the requirements, this research implements user-centred design (UCD). UCD mainly employs qualitative data collection methods or techniques, such as think-aloud protocol usability testing (TAP), interviews, focus group discussions and field observations. UCD can also be implemented using quantitative data collection methods, such as surveys for this study. The selection and implementation of these techniques is unique and based on the reasoning and purpose of this study (Hackos & Redish, 1998). A combination of the quantitative and qualitative UCD techniques are implemented in this research, as explained in the research design. The next section describes the selection of the research methodology in detail.

3.3. Research approach

Research approach is a plan or procedure for conducting research spanning from the conceptual stage to detailed data collection, analysis and interpretation methods (Creswell, 2014). This plan should be developed uniquely to study a specific topic and to address a specific research problem or answer specific research questions.

This research is developed based on the problem that spatial metadata usability is more into spatial data producers, for data inventory management, rather than spatial data users, and the aim is to investigate and improve the usability. The hypothesis is that the usability can be improved by implementing a user-oriented metadata profile in the spatial metadata design and development process, and the research questions and objectives are developed accordingly. The approach should be able to provide step-by-step methods for data collection, data analysis and interpretation to achieve the objectives as well as to answer the questions. It should also be able to provide the means or tools for evaluating the usability of the new metadata as well as gathering users' reactions or acceptance to the new metadata.

Following are the available research approaches and an explanation of the selection of the methods for conducting this research.

3.3.1. Quantitative approach

Quantitative research is an approach for conducting research based on systematic empirical investigation of specific phenomena using statistical, mathematical or computational methods. The objective is to test theories or hypotheses by defining variables and examining the relationships among the variables (Creswell, 2014). A quantitative approach is implemented by measuring and calculating quantifiable variables to determine the relationship between the variables using mathematical or statistical models.

Since it focuses on measurable and quantifiable data, a quantitative approach is very effective at answering direct questions of "what" or "how" of the investigated phenomena, such as what percentage? How many? (Goertzen, 2017).

Quantitative approach works well when research deals with a set of isolated variables, but it does not work well when applied to complex scenarios. In the context of the spatial field, experiments for assessing effectiveness of visualizations of a single metadata attribute have been conducted successfully using quantitative approach (Evans, 1997; Kardos et al., 2005). But when similar experiments were conducted testing more complex scenarios, such as data uncertainty, a quantitative approach was not able to get satisfactory results (Boin, 2008).

A quantitative approach has strengths and weaknesses as summarized in Table 3.1. (Choy, 2014).

Table 3.1: Strengths and weaknesses of quantitative research approach (Choy, 2014)

Strengths	**Weaknesses**
Reliability by critical analysis	No human perceptions and beliefs
Short time frame to administer survey	Lack of resources for large scale research
Facilitates numerical data for group and extent of agree or disagree from respondents	No in-depth experience description

In the context of this research, quantitative methods are implemented to measure usability attributes; effectiveness, efficiency, and user satisfaction. Effectiveness can be calculated by measuring the percentage of users who successfully complete a given task. Efficiency can be obtained by measuring the time spent by users to complete a task related to spatial data discovery and selection. User satisfaction can be measured by giving questionnaires to users, related to their experiences with spatial metadata. The quantitative method is also used to collect user requirements for metadata, by conducting a survey of spatial data users. However, the quantitative measurement methods leave a large gap related to the cause of the results achieved, and the relationship between those causes. In addition, the quantitative approach is unable to gather users' perceptions and expectations for metadata. To fill the gap, another approach is required to identify the problems and to provide deeper explanations related to the numbers generated from the quantitative method. Qualitative approaches are used in this type of situation where research phenomena need to be explored and understood and the important variables are either unknown or impractical to isolate (Creswell, 2014) .

3.3.2. Qualitative Approach

As opposed to a quantitative approach, qualitative research focuses on observations of specific phenomena to obtain non-numerical data (Babbie, 2014). Instead of depending on statistical data to answer the question, “How often?”, a qualitative approach emphasises descriptions, characteristics, meanings or definitions of the variables to answer the “why” and “how” a phenomenon may occur (Berg & Lune, 2012). A qualitative approach is mainly used in social settings for exploring and understanding individual or group insights or beliefs on a specific problem. Despite criticism related to its validity and reliability, a qualitative approach is still powerful and can produce information-rich data for analysis (Creswell, 2014). It can even be implemented to test hypotheses if the theoretical proposition can be supported by evidence (Gray, 2014).

The strength of a qualitative approach lies within its methods, where the data is mainly collected in natural settings involving direct interactions between the researcher and the phenomenon; individuals or groups. Therefore, the data is highly contextual and contains rich information that can be extracted to gain a deep understanding about meanings, beliefs or relationships forming the phenomenon (Maxwell, 2004). However, it is important to understand that the researcher should acknowledge their own biases and the effect of their presence into the research environment and take this into account in data interpretation (Boin, 2008). The methods include case studies, interviews, field observations, where data collection and analysis may occur simultaneously, consecutively or iteratively. The critical difference of qualitative approach in contrast to quantitative is that the research is not intended for drawing statistical conclusions, rather to explore the phenomenon being studied and to identify concepts, themes, or in the context of this research, usability problems; effectiveness, efficiency, user satisfaction and user expectations.

Choy (2014) summarises the strengths and weaknesses of a qualitative approach, illustrated in Table 3.2, that can be used to consider the applicability of the approach for this study.

Table 3.2: Strengths and weaknesses of qualitative research approach (Choy, 2014)

Strengths	**Weaknesses**
View of homogeneous exploration	No objectively verifiable result
Raises more issues through broad and open-ended inquiry	Skill required of interviewers
Understanding behaviours of values, beliefs and assumptions	Time consuming during the interview process and intensive categorisation process

In the context of this research, the spatial metadata usability problems, user needs, and user expectations can be inductively identified from the data collected through an observation of the interaction between spatial data users and spatial metadata, as well as from interviews with the users. Surveys with open-ended questions can also be conducted to explore users' opinions as well as their expectations of spatial metadata for spatial data discovery and selection.

As the intention is not to draw statistical conclusions, the research implementing a qualitative approach should be judged differently. If a quantitative approach may be judged by rigours such as validity, reliability and generalisability, alternative rigours related to the intention of qualitative research are used to increase the understanding of a phenomenon (Creswell, 2014). Parameters such as thick descriptions, credible and plausible insights can be used to judge the quality of qualitative research.

3.3.3. Mixed Methods Approach

Mixed methods research is an approach to inquiry involving collecting both quantitative and qualitative data, integrating the two forms of data, and using distinct designs that may involve philosophical assumptions and theoretical frameworks (Creswell, 2014). Mixed methods have strengths that cannot be achieved by implementing a single quantitative or qualitative approach (Creswell, 2014; Teddlie & Tashakkori, 2009):

1. Allowing research questions that cannot be answered by a single approach.
2. Providing better (stronger) inferences.
3. Providing the opportunity for presenting a greater diversity of divergent views.

As explained in the earlier sections, this research employs both quantitative and qualitative methods for achieving research objectives and answering research questions. Consequently, this research employs mixed methods approach and the research design is developed following the mixed methods approach.

As can be seen in Table 3.3, the types of mixed methods are selected based on the requirements and rationale of the research design. The selection of mixed methods implementation for this research follows the requirements for increasing the depth and breadth of the results and to maintain the research quality (validity).

Table 3.3: Purposes of mixed methods evaluation design for this research adapted from Gray (2014)

Purpose	Rationale	This research
Triangulation: Seeks convergence, corroboration, correspondence of results from different methods	To increase the validity of constructs by counteracting or maximising the heterogeneity of irrelevant resources of variance attributable to inherent bias. Inquirer bias or biases in inquiry context	✓
Complementarity: Seeks elaboration, enhancement, illustration, clarification of results from the other method	To increase the meaningfulness and validity of constructs by capitalising on inherent method strengths and counteracting inherent method biases	✓
Development: Seeks to use the results of one method to inform the other method	To increase the validity of constructs and inquiry results by capitalising on inherent method strengths	✓
Initiation: Seeks the discovery of paradox and contradiction, new perspectives, the recasting of questions or the results from one method with questions or results from the other method	To increase the breadth and depth of inquiry results and interpretations by analysing them from the different perspectives of different methods and paradigms	×
Expansion: Seeks to extend the breadth and range of inquiry by using different methods from different inquiry components	To increase the scope of inquiry by selecting the methods most appropriate for multiple inquiry components	✓

Implementation of the mixed methods approach in research stages should also be considered for designing this research. Depending on the requirements, quantitative and qualitative methods could be mixed based on any particular stage of the research. In the context of this research, the combination is required in both data collection and data analysis, as well as in the design of research questions.

Table 3.4: Types of mixed methods implementation in this research design, adapted from Gray (2014)

Types of mixing	Comments	This research
Two types of research questions	One fitting a quantitative approach and the other qualitative	✓
The manner in which the research questions are developed	Pre-planned (quantitative) versus participatory/emergent (qualitative)	×
Two types of sampling procedures	Probability versus purposive	×
Two types of data collection procedures	Surveys, etc. (quantitative) versus focus group, interviews, etc. (qualitative)	✓
Two types of data analysis	Numerical versus textual (or visual)	✓
Two types of data analysis	Statistical versus thematic	✓
Two types of conclusions	Objective versus subjective interpretations	×

3.4.4. User-centred Design (UCD) approach

Having the reasoning, questions, objectives, and the selection of methods to be used, this research implements the mixed methods employing UCD approach for a number of reasons: (1) this research has usability as the main focus; (2) this research aims at exploring the relationship between user spatial data and metadata to evaluate usability metadata status, identify existing problems, and identify user needs and expectations; and (3) UCD offers a number of methods that allow us to obtain all inquiries in this research in a limited time using a small number of participants, as explained in the next section.

UCD is a method in the Human-Computer Interaction (HCI) domain that is specifically used to design a system (or a product) based on user requirements. It is an approach in designing and developing a product (or a system) where user requirements; needs, characteristics, abilities and wishes, are taken as a starting point, and fulfilling these user requirements is the ultimate goal of the product design (Mao et al., 2005). Various methods are available for implementing UCD at every stage in the design and development process, depending on the purpose of the UCD: (1) at the beginning of the design project; (2) at early and mid-point in design cycle; and (3) at the final stage of design cycle, as illustrated in Table 3.5. According to the list, usability evaluation (testing) of spatial metadata, as an established (or existing) product is placed at the final stage of the design process.

Table 3.5: UCD methods implementation throughout development process (Lee, 2005)

Implementation Stage	Purpose of UCD	Available Methods
At the beginning of project design	Collecting user needs and expectations	Interviews and questionnaires
Early in the design cycle	Collecting data for product functionalities	Interviews and questionnaires
	Gathering wider requirements and issues of the product	Focus groups discussions
	Collecting data about the environment of the product use	On-site observations
Mid-point in the design cycle	Evaluation of alternative designs and gathering additional requirements	Role playing, walkthroughs and simulations
Final stage of design cycle	Evaluation of product's usability	Usability testing
	Evaluation of user satisfaction with the product	Interviews and questionnaires

Table 3.6: UCD methods implementation (Mao et al., 2002)

UCD Methods	Number of usages	Average Importance Ranking *
Iterative Design	65	2.15
Usability Evaluation	43	2.39
Task Analysis	34	2.61
Informal Expert Review	31	3.28
Field Studies	28	2.00
Focus Groups	16	2.79
Formal Heuristic Evaluation	15	2.86
Prototype without user testing	15	3.07
User Interviews	11	3.00
Surveys	9	3.17
User Requirement Analysis	7	2.00
Participatory Design	7	3.40
Card Sorting	5	3.33

Among the UCD methods, the most commonly used methods are iterative design, usability evaluation, task analysis, informal expert review, and field studies (Mao et al., 2005). Excepting informal expert review, which has low impact, these methods have the highest impact on the improvement of organization products/systems usability. Informal expert review is widely used only for its inexpensive cost. On the other hand, although it has the highest impact on the improvement of product/system usability, field studies are not the most widely used method because they are cost intensive and time consuming. Iterative design and usability evaluation are the two most commonly used methods as they offer the best cost-benefit balance. Table 3.6 summarises the results of the study. Accordingly, Nielsen (1993) suggests that usability testing is the most useful method for usability evaluation of a fully developed product, such as published metadata, that can reveal acute usability problems with a relatively small number of participants and relatively low-costs. Nielsen also refers a specific usability testing method, namely think-aloud protocol (TAP) for the usability evaluation.

There has been criticism of TAP producing less accurate data due to the interference of participants' cognitive process when they have to talk and express their thoughts in verbal form (Nisbett & Wilson, 1977). However, Ericsson & Simon (1984) as in Austin & Delaney (1998) found the opposite when they conducted their experiments, and the results are supported by other results from a series of experiments conducted by Henry et al., (1989), suggesting that there are no effects on participants' performance regardless of their level of expertise.

3.5. Research design

The research design is developed based on the nature and sequence of the questions and the reasons (purposes) of the mixed-methods, and the methods are incorporated accordingly. This research employs a mixed methods approach to create triangulation to maintain the validity of the research, produce complimentary data to maintain the breadth and depth of the results, and to understand the spatial data users' perspective on spatial metadata.

Table 3.7: Mixed methods design recommendations (Creswell, 2014)

Reasons for choosing	Expected Outcomes	Recommended Design	This research
Triangulation to draw different perspectives from quantitative and qualitative data	Merging the two databases to show how the data converges or diverges	Convergent parallel	✓
Complementarity allows quantitative results to be considered with qualitative data	A more in-depth understanding of the quantitative results	Explanatory sequential	✓
Developing better measurement instruments	A test of better measures for a sample of a population	Explanatory sequential	×
Understanding experimental results by incorporating perspectives of individual	An understanding of participant views within the context of an experimental intervention	Embedded	✓
Developing an understanding of needed changes for a marginalized group	A call for an action	Transformative	×
Understanding the need for an impact of an intervention program	A formative and summative evaluation	Multiphase	×

Therefore, following Creswell's recommendations for the selection of a mixed methods design in Table 3.7, this research implements the embedded style design with convergent parallel and explanatory sequential in the process, illustrated in Figure 3.3. This selection is suitable to fulfil the requirements of this research, to answer the questions and maintain the quality of the research, as discussed later.

This research consists of three stages: identification, development, and evaluation. Each stage implements qualitative and quantitative methods uniquely suitable for the purposes and the expected outputs.

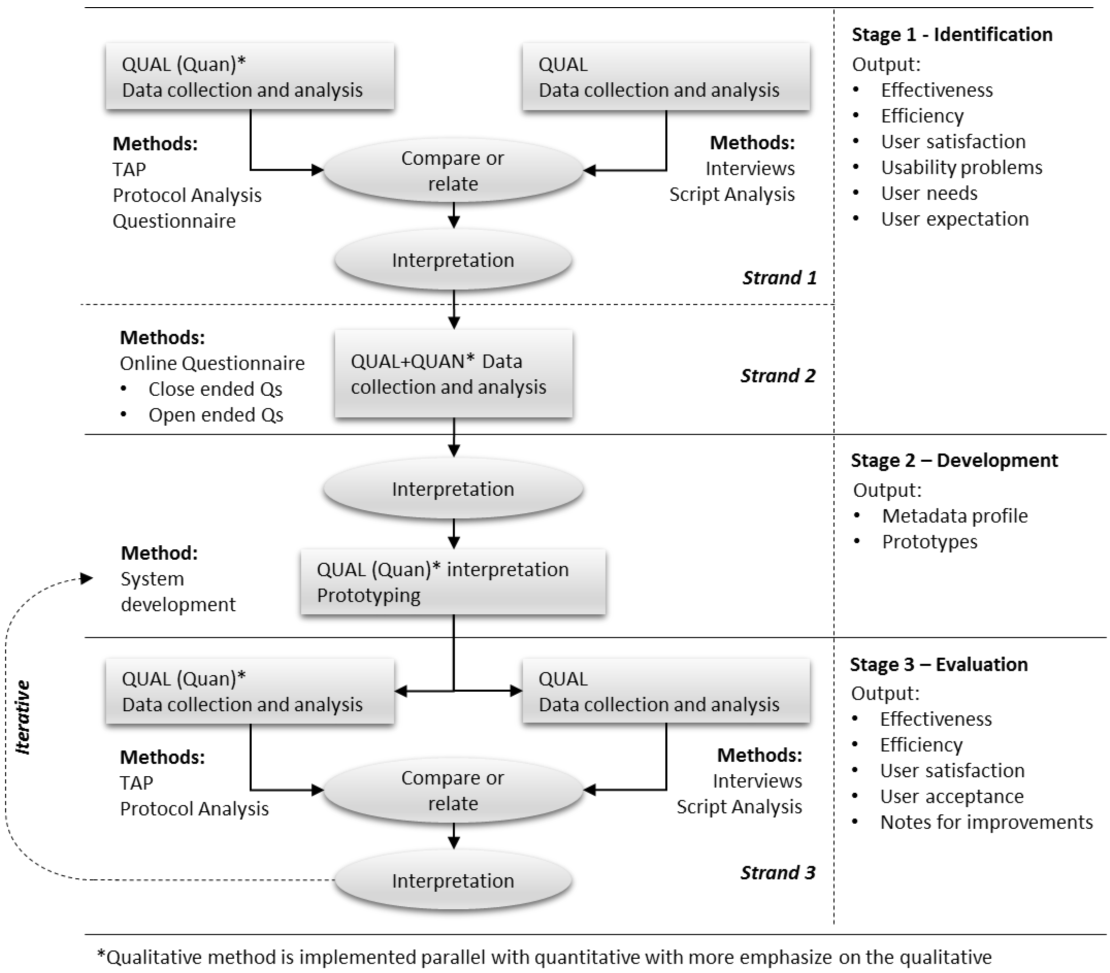


Figure 3.3: Research Design employing UCD approach

3.5.1. Stage 1 – Identification

The identification stage is designed to evaluate the usability status of the existing spatial metadata, identify usability problems and gather user expectations of spatial metadata. The results will be used as a baseline for designing the user-oriented spatial metadata and the user interface. In this stage, embedded mixed methods, consist of two strands; convergent parallel and explanatory sequential models. These strands are implemented by conducting four data collection methods: qualitative think-aloud protocol, quantitative measurements, quantitative questionnaire and qualitative semi-structured interviews. The first strand, the convergent parallel, is conducted by implementing TAP, measuring for effectiveness and efficiency, undertaking user satisfaction questionnaires, and semi-structured interviews. The purpose is to increase the validity of constructs by seeking the convergence or correspondence from different methods (triangulation) (Creswell, 2014) and to increase the meaningfulness and validity of the construct by seeking enhancement and clarification from the other method (complimentary). The second strand, explanatory sequential, is conducted by using the results from the convergent parallel to develop questionnaires for online surveys. The purpose of the second strand is to increase the scope of the results (Gray, 2014).

Selection of spatial metadata

This stage begins with the selection of spatial metadata to be evaluated and spatial data users as participants of the experiment. The metadata selection is conducted following a case study approach, where the selected metadata should represent the global setting based on a set of criteria.

Spatial metadata from two established spatial data portal or websites, Australian Spatial Data Directory (FIND) and Victoria Open Data Portal (VDD), are chosen for the following reasons:

- Australian spatial metadata has been designed and created following the internationally recognized and used standard, ISO 19115 – metadata standards.
- Australian spatial metadata has been distributed and delivered to spatial data users using well established technology that is implemented world-wide, e.g., spatial data directory, open data platform, thematic web map services, and closed community-based data directory.
- FIND and VDD fit the requirements of the research, that is, spatial metadata for general spatial data users, not thematic or community-based users.

Selection of participants

Selecting the appropriate samples and the sample size is crucial in research design. Sampling selection should consider characteristics of the methods, as can be seen in Table 3.8.

Table 3.8: Mixed methods sampling characteristics used for this research (adapted from Gray, 2014)

Reasons for choosing	This research
Overall purpose of sampling	To generate a sample that will address all research questions
Intended outcomes	For some strands/research questions intended outcome is external validity; for other strands intended outcome is transferability
Rationale for selecting cases/units	For some strands the focus is on representativeness; for other strands, the focus is on information rich cases
Sample size	For some strands there will be a large number of cases/units; for other strands, there may be one case or a few cases
Depth /Breadth of information per case/unit	Focus on both depth and breadth of information across all research strands
Form of data generated	Both numeric and narrative data are generated

In Strand 1, the objective is to determine the usability of established spatial metadata and to identify usability problems and user expectations of the metadata. Five spatial data users are targeted as participants for the mainly qualitative experiment. Hence, QUAL (Quan) data collection is implemented where TAP and interviews are the main qualitative methods and questionnaires, the main quantitative method. Since the rationale is aiming at the depth, intensity and holistic contextual view of the phenomena involving interaction of the user or groups of users with the metadata, this size suffices the requirements. Moreover, five participants as the participants for TAP and interviews has two advantages. Firstly, it can detect 80% of acute usability problems (Nielsen, 1992). Secondly, the data should be easier to extract and analyse. This number has been tested by Virzi as cited in Caulton (2001) conducting a usability test of a voice mail system in which he did a Monte Carlo procedure to build an average accumulation curve and concluded that 80% of acute usability problems can be detected with four or five users, and new problems or information are less likely to be found by adding more users. However, additional participants are anticipated in this stage if the saturation of the data is not obtained.

Any spatial data users above 18 years old from any geographic regions or any field of expertise can be selected as the participants so long as the interviews and other data collection methods adhered to the requirements of the School of Engineering Human Ethics Advisory Group (EHEAG) at The University of Melbourne. The requirements dictate that the participants must be shown a plain language statement before conducting the TAP and being interviewed, and that they must provide their written consent. The selected participants are identical for both the TAP and interviews.

Strand 2 in this stage is designed for increasing the scope and breadth of the results by expanding the geographic scope and number of spatial data users world-wide. For this purpose, an online survey as data collection method is implemented. Purposive sampling is implemented for the sample selection by conducting snowball sampling world-wide. Spatial data users known to the researchers are contacted and asked to be participants via email communication based on their qualifications, that is, active users of spatial data or have been working in spatial related projects or organisations. They are also asked to pass the questionnaire to their associates who share similar qualifications. This sampling method maintains the validity of participants. The questionnaire is also circulated by utilizing social media, i.e., LinkedIn and Twitter targeting the researcher's networks that are mostly those who work in the spatial domain. In order to maintain the validity of participants from this open distribution method, a number of questions are designed to obtain their background and experiences which can be used to detect whether the participants are suitable for this survey or not.

Designing a scenario and a testing laboratory for TAP

TAP method for spatial data usability testing requires a natural setting for the participants to interact with the spatial metadata, or in a simulation designed to be closely related to the participants' context. Since the targeted participants had been working in disaster related projects within Australia, a scenario for a disaster response project in Australia was developed.

It is a simple scenario. Each participant is assigned as a project member who has a responsibility for preparing spatial data for a flood evacuation plan for the greater Melbourne Area and East Coast of Australia (Queensland and New South Wales). They are instructed is to find and select required spatial data to be used as the major data source for developing the evacuation plan, and they should use the selected websites. The scenario is accompanied by a set of instructions issued before the commencement of the TAP and interviews in a written format. The instructions include the required spatial data for participants to find and select.

For simulation purposes, a set of computers and peripherals and a working room in the Centre for Spatial Data Infrastructure (CSDILA) are prepared and solely dedicated as a usability-testing laboratory during the TAP data collection. The participants choose whether they want to perform the TAP and interviews in their own setting or in the prepared space.

TAP and interviews schedule

The TAP procedure is started with the participants' knowledge that they are about to perform some tasks to find and select required spatial data as given in the scenario. The researcher will also remind them that this procedure is not a test for them but seeks their insights and thoughts about spatial metadata during the spatial data discovery and selection process. Therefore, they are asked to actively give their thoughts by continuing to talk and give their comments or expressions, verbal or gestures, during the process. Their verbal comments and the interviews are recorded. This is helpful, as the researcher can focus on specific moments or incidents for note taking, for example, measuring the time when a participant completed a task.

Each TAP session typically lasts 40 minutes, and towards the end of the session each participant is given a questionnaire asking for their opinions regarding their experience during the session. This questionnaire is used as the guiding questions for the semi-structured interview, which typically lasts 50 minutes. During the TAP data collection process, the researcher also conducts a quantitative data collection by measuring the rate of success of the tasks given to participants and the time spent by participants to complete the tasks.

In the interview sessions, besides guiding questions, the participants are also asked for any incidents or moments that they considered important or critical during the experiments and to give explanations about those incidents.

The participants are informed that they will be contacted and given a summary of their comments during the TAP and the interview for member-checking purpose. Since the participants are physically reachable, the researcher will bring the summary to the participants and discuss it with them.

Data analysis

Protocol analysis (PA) is a set of methods designed for extracting rich and reliable information from verbal data, in this case think-aloud data. Think aloud data is different from any other verbal data, since it comes specifically from the participant's mind based on their live and direct interaction with a phenomenon. Ericsson & Simon (1984) called this "concurrent verbal reports", where the participants think-aloud and verbalize their thoughts as if they were talking to themselves. There is no 'one size fits all' set of rules for implementing PA, so a researcher can have slight variations on the basic procedures (Austin & Delaney, 1998) depending on their requirements. However, typical analyses share the same processes, where the textual documents are segmented into blocks or phrases and a code is given for each block. The codes are developed based on the theory or the phenomenon being studied to identify the functional relationships (Potter, 1997). This procedure is used to obtain individual statements from long and continuous transcripts. After the segmentation, the researcher observes the segments with only the necessary context to make categorical decisions about the data. The coded blocks or phrases are then reassembled in their original order for analysis.

Using similar procedures, participants' verbalized comments are transcribed into textual documents from the tape recording, named transcription of data (ToD). The researcher uses ToD to highlight all the nouns and noun phrases and provides codes for the phrases in accordance with the spatial data discovery and selection process; a process called the referring phrase analysis (RPA). The coded phrases are then re-ordered following the respective tasks in spatial data discovery and selection process, and given assertion codes to identify the relationships between concepts, for example, intention, actions and decision during the process. This process is called assertions analysis (AA). The last procedure is to reassemble the asserted phrases or blocks and the researcher starts the analysis or interpretation procedure. The protocol analysis procedures used in this analysis are illustrated in Figure 3.4.

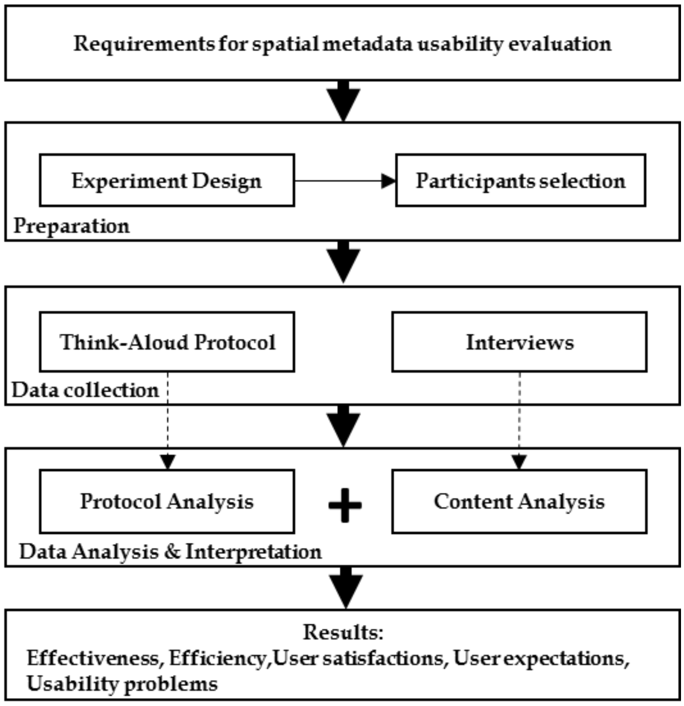


Figure 3.4: Protocol Analysis Framework (Adapted from Fonteyn ,1993; Austin and Delaney, 1998)

Interview data is analysed differently using simplified content analysis. The procedure starts with transcribing the data from the tape recording, and the researcher groups the data into blocks based on a specific task or sub tasks. The results from the protocol analysis are used to identify any phrases or blocks that refer to similar incidents or moments. The researcher re-orders the phrases or blocks and interprets them to extract any information regarding spatial data discovery and selection based on participants' experiences during the process.

Survey and questionnaire design

A survey, the second strand in this stage, is conducted as a follow up for gathering more information from wider spatial data users about their requirements and expectations of spatial metadata. The survey is also designed to get users' requirements for specific metadata elements such as the title and abstract, which appear in the TAP and interview to be the most important elements and require specific treatment.

The survey is intended to be online or web-based for following reasons:

- The participants are distributed word-wide.
- The participants can easily access the questionnaire via any device that has internet browser.
- The data is captured digitally allowing for easier and faster analysis.

The questions are developed by combining the previous results with the ISO 19115-1:2014. Specific metadata elements from discovery metadata for geographic resources and user-related metadata elements are chosen to form part of the questions. The objective is to obtain spatial data user's interests and opinions of these elements in spatial data discovery and selection.

The questions are divided into the five following sections:

1. Introduction section, where the plain language statement is presented.
2. Consent section, where an online consent form is presented, and users must give their consent digitally by selecting the 'yes' option before they can proceed.
3. Participant information, where participants provide their backgrounds to be used for spatial data user profiling.
4. Spatial data search and selection section, where the main questions related to spatial data (resources) discovery and selection are presented in two formats; closed ended and open ended.

5. Appreciation section, where the participants are thanked by the researcher and given the researcher's contact should they need information related to the study.

The questions are tested and refined internally and externally. Internal testing is implemented by conducting a series of evaluations testing the questions against the objectives, as well as ensuring worldwide accessibility using any Internet browser and any device. The questionnaire is also tested to ensure that responses are appropriately recorded. External testing is implemented by sending the questionnaire to the researcher's associates to get feedback and responses. The responses and feedback are incorporated to refine the questions and ensure that the researcher gets the expected responses or answers.

Survey data collection and analysis

The questionnaire is distributed world-wide in two ways; emails to spatial data users known to the researcher with an additional request to circulate the questionnaire to their networks and using social media platforms such as Twitter and LinkedIn. Close-ended questions will be used for gathering users' statements with regards their background information, such as their length of experience, whilst open-ended questions will be used to gain users' insights and ideas about certain aspects of spatial metadata related to their requirements of data discovery and selection. As part of an exploratory and qualitative study, the sample size of 10 to 30 for this survey will suffice the requirement as cited by Isaac and Michael in (Gray, 2014). The survey is conducted within a two months period.

The data is analysed by conducting qualitative and quantitative analysis methods. The aim is to identify user requirements for spatial data discovery and selection and their expectations of spatial metadata, as well as the discovery system (user interface). The findings are used for constructing the user-oriented spatial metadata profile and the functionalities of the user interface. Investigation of current spatial metadata standards is conducted as the basis for the profile. The user requirements are then contrasted to the standard to answer the research question, "Do the established spatial metadata standards meet the user needs?"

3.5.2. Stage 2 – Design and development

The results from Stage 1 are used to develop a user-oriented spatial metadata profile following the extension guidance in the ISO 19115:2014 geospatial metadata standards. The result should be an extension of discovery metadata for geographic resources of the ISO 19115:2014.

Following the UCD approach, the next step for usability improvement for spatial metadata is to re-design and re-develop the spatial metadata and the supporting system (user interface) towards the specified user needs and expectations, thus addressing the identified usability problems. A prototyping approach is implemented adapting the development system process introduced by (Nunamaker Jr et al., 1990).

The original process consists of five different stages: construct a conceptual framework; develop system architecture; analyse and design the system; build the system; and observe and evaluate the system. Figure 3.5 illustrates the system development process used for this research and the original process as suggested by Nunamaker Jr et al. (1990).

The results from Stage 1 are analysed and investigated to construct the requirements and functionalities of both the spatial metadata and the user interface as the discovery system for spatial data discovery and selection. The identified requirements and functionalities are then put into a structure where the relationships between them can be analysed and validated against the logical consistency of the data discovery and selection process. The spatial metadata database and user interface physical design can then be developed. Once the physical design is finished and validated, the researcher is able to start building the prototypes. The last stage in the process is to observe and evaluate the prototypes to ensure that the requirements and functionalities were integrated accordingly.

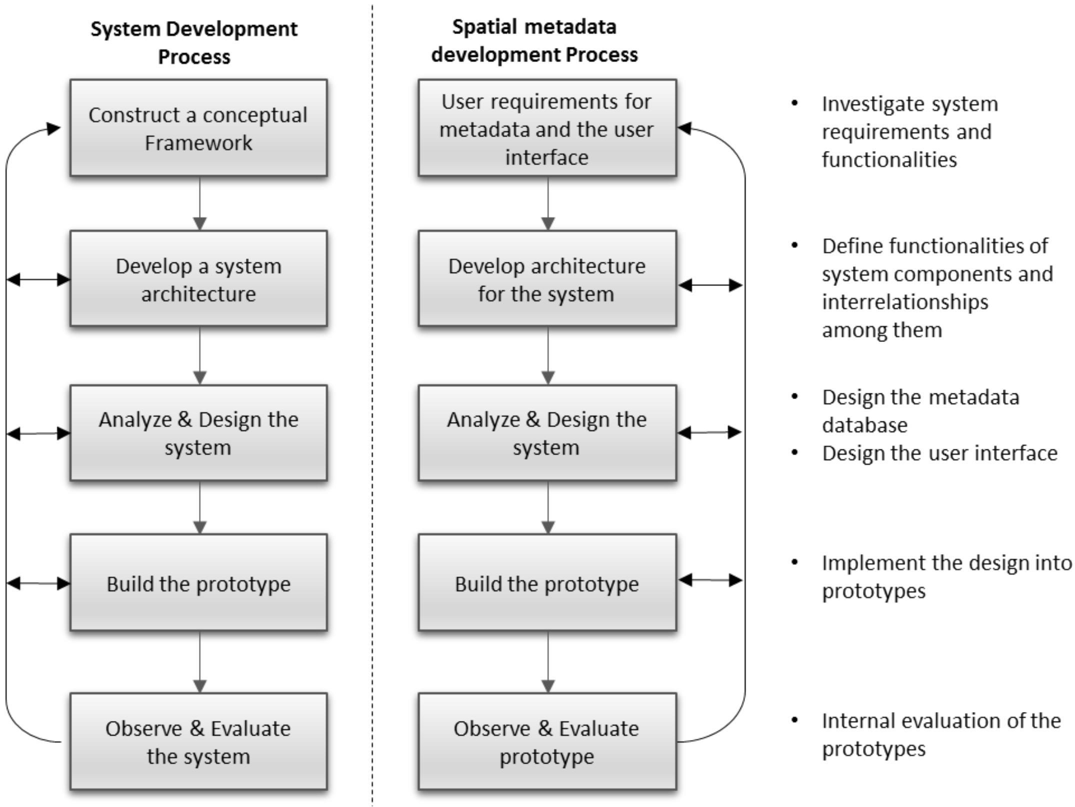


Figure 3.5: User oriented spatial metadata prototype development process

(Adapted from Nunamaker Jr et al., 1990)

3.5.3. Stage 3 – Evaluation

This stage aims to get spatial data users' acceptance and opinions of the user-oriented spatial metadata as well as to verify the improvement of spatial metadata usability. This is part of UCD requirements where the product is evaluated before it can be released as a final product. With respect to this research, this stage is specifically designed as the ultimate test of the hypothesis that usability of spatial metadata can be improved by incorporating spatial data users in the design and development process of the metadata. Thus, answering the research question, "How can spatial metadata usability for spatial data users be improved?"

The evaluation is conducted implementing similar methods to those used in the first stage, i.e., TAP and interviews and the same criteria or indicators are evaluated; effectiveness, efficiency and user satisfaction. The TAP is also designed to identify usability problems and user expectations of spatial metadata and the user interface, to be used as the basis for future improvements. It is important to keep in mind that this research follows UCD approach which is iterative in nature. The only difference in this stage compared to the Strand 1 of Stage 1 is the participants.

The participants for the evaluation stage are a combination of participants from Stage 1 and new participants. Again, using TAP and interviews as the data collection methods, this stage aims at five participants. Participants from different educational and professional backgrounds are targeted.

Having similar methods for data collection, the data is accordingly analysed by conducting similar analysis methods; protocol analysis and simplified content analysis or transcript analysis. The results will be compared to the results from Stage 1 and discussed.

3.6. Research quality

Due to the nature of typical qualitative research, which is based on subjective impressions, anecdotal evidence and lacks reproducibility and generalisability, the quality of this research cannot be judged quantitatively, by implementing statistics with a certain degree of confidence. Therefore, this research should be judged following the strategies and against the measures for qualitative data. Instead of having quantitative rigours such as reliability, validity and generalisability, this research uses alternate rigours namely trustworthiness and authenticity (Bryman & Cramer, 2004; Seale, 2004). Trustworthiness means that research is believable, confirmable and dependable, while authentic means that research is fair, presents both positive and negative viewpoints and allows participants to understand their stance as well as others and to make changes to the research (Boin, 2008).

Practically, these rigours can be obtained by implementing strategies as proposed by Creswell (2014) and Gray (2014), as summarized in Table 3.9.

Table 3.9: Techniques for maintaining validity, reliability and generalisability of qualitative research, adapted from Gray (2014) and Creswell (2014)

Technique	Description	Research quality aspect
Member checking	Checking data accuracy with participants (selected spatial data users) during requirement gathering	Validity
Data collection triangulation	Gathering data (requirements) using two or more different methods	Reliability
Rich and Thick description	Presenting deep and intense description about spatial data users' insights and their interaction with spatial metadata	Transferability/ Generalisability
Include negative information	Including negative or contrary information which emerges during data collection in analysis process	Transferability/ Generalisability

To demonstrate its validity, this research will implement a member checking method by confirming the findings with participants during the data collection process and will mechanically record the data, that is, the participant's comments and answers using audio recorder and notes (Gray, 2014). As for maintaining the reliability, data triangulation will be implemented by conducting three different types of data collection: TAP, interviews and questionnaire (Gray, 2014). In order to demonstrate generalisability or transferability of the results, analysis and presentation of the results will use thick and rich description of the spatial data user behaviours, their interactions with and expectations of, spatial metadata and by including negative or contrary information to the analysis and report (Creswell, 2014).

3.7. Chapter summary

This chapter outlines the methodological foundation for this thesis. The first part is dedicated to explaining the reasoning of the selected research approach available in the research community. It also provides the description and reasoning for choosing UCD, the implementation of the selected mixed methods approach, as well as selection of UCD methods; TAP, interviews and questionnaires, for data collection in this research.

The second part contains a step-by-step explanation of the research design, starting with the reasoning for mixed-methods types selections against the objectives and the research questions. The research design implements the embedded style mixed with convergent parallel and explanatory sequential style. The design is developed following the sequence of the research questions and to maintain the quality of the research by having triangulation in data collection.

The research contains three stages; identification, development and evaluation. The identification stage consists of two strands and is designed to achieve Objectives 1 and 2 of this study. Strand 1 employs five spatial data users as participants for TAP, interviews and questionnaire as the data collection methods. A survey is circulated for Strand 2, targeting spatial data users globally.

Data from both strands are used as the basis for the development stage. Following the extension guidance in ISO 19115:2014 metadata standard, the data is used to develop a user-oriented spatial metadata profile as an extension of the discovery metadata for geospatial resources of the ISO 19119:2014 metadata standard. The user-oriented metadata prototype and user interface prototype are developed based on the profile and evaluated in Stage 3 by conducting TAP and interviews with spatial data users.

The last part of this chapter is dedicated to explaining the quality judgments for this research, as a qualitative research. Alternative rigours for displaying the quality were explained; trustworthiness and authenticity, as well as the practical methods for maintaining the quality; member-checking, data triangulation, rich and thick description and including negative results or point of views into the interpretation.

The next chapter will investigate the process, results and findings of Strand 1 of the first stage. The strand consists of convergent parallel type of mixed methods by implementing TAP and measurements (QUAL(Quan)); interviews (QUAL) for data collection methods; and respective analysis methods for the data, protocol analysis and simplified content (transcript) analysis.

This page is left intentionally blank

4. Spatial metadata usability evaluation

4.1. Introduction

This chapter reports the results of an experiment for investigating the usability of spatial metadata from two metadata delivery systems: (1) Victorian spatial open data portal, and (2) Australian geospatial data portal, by employing two UCD methods, namely Think-Aloud Protocol usability testing (TAP) and semi-structured interviews. TAP records the verbal expressions of participants' thoughts during the process of task completions and researcher's notes taken while observing the process. The semi-structured interviews are used to reveal detailed information of participants' experiences by encouraging them to recall specific incidents during the task completions. Both Data captured by TAP and semi-structure interviews are then analysed and extracted to determine the usability of the spatial metadata, identify the usability problems, describe the user experience, and reveal the user expectations.

The chapter starts with an explanation of the data analysis method, followed by descriptions of experiment design, reasoning for material and participants selection, results, discussion, and chapter summary.

4.2. Data analysis methods

The method used in the analysis process is adapted from Protocol Analysis (PA) and content analysis. PA is capable for producing deep and accurate analysis from TAP data, as these methodologies are qualitative in nature and designed to obtain rich and deep data from small quantities of samples (Gray 2014; Fonteyn et al., 1993; Nielsen 1993). PA consists of several progressive steps as shown in Figure 4.1.

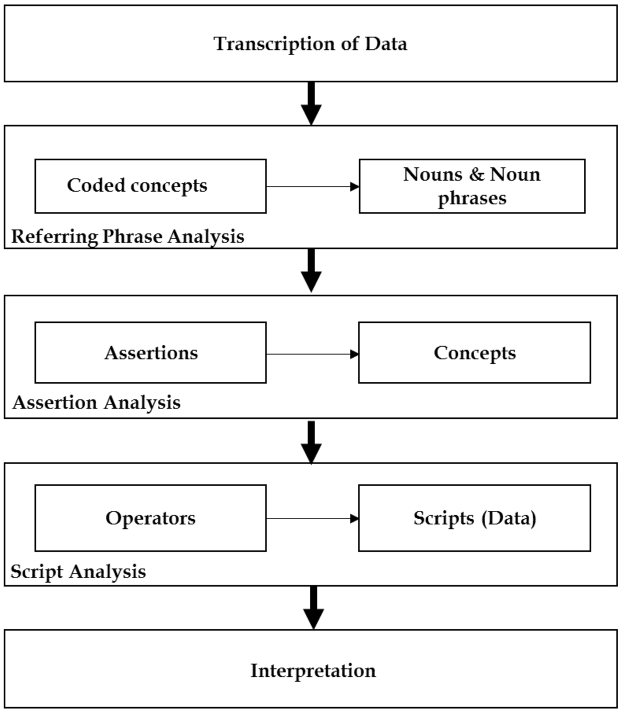


Figure 4.1: Protocol Analysis Framework (Fonteyn et al., 1993)

The analysis process begins with Transcription of Data (ToD), where recordings and researcher's notes during the testing are transformed into textual transcripts. The next step is Referring Phrase Analysis (RPA). It is done by highlighting all the nouns and noun phrases identified in each participants' verbal data and attaching codes to those phrases in accordance with code references within the spatial data search activities. Based on the provisional coding results of the RPA, the investigator then groups each code according to each relevant action during the spatial data discovery and selection. The next step is Assertions Analysis where the investigator identifies the set of assertions made by subjects to determine how relationships were being formed between concepts during problem solving. As part of the analysis, the purposes of subjects' assertions are identified. The last step is Script Analysis to identify and provide an overall description of the reasoning process during problem solving and allowing the investigator to illustrate the problems identified by the subjects during the problem-solving process.

4.3. Experiment Design

The objective of this experiment is to collect and analyse data to answer research questions based on a selected usability evaluation method. The experiment was designed based on the usability evaluation framework, TAP. The TAP data collection method utilises semi-structured interviews and post-questionnaires, and spatial metadata usability evaluated using three criteria: *usefulness, effectiveness* and *user satisfaction*. The design was based on a task-oriented activity where participants were given a set of tasks concerned with spatial metadata, following a scenario to achieve specified goals. During the completion of the tasks, they were asked to actively and continuously express their thoughts and opinions verbally, as well as answering questions given by the researcher. At the end of the experiment, the participants answered some questions in semi-structured interviews to provide the researcher with detailed and deeper information related to the given tasks and based on their experiences during the TAP.

4.3.1. Participants

For this study, five spatial data users from different educational background were selected; civil engineering, spatial science and computer science. They were working on disaster management and had been regularly working with spatial data for their projects. The number is optimal for usability testing method using TA protocol, since it is able to identify the most usability problems with optimum benefit-cost ratio (Nielsen, 1993). In addition to that, the study is a qualitative and exploratory in nature, focusing on the extraction of user's insight about spatial metadata. This type of study usually needs a small number of participants with a long time of participant-researcher interaction, as opposed to a quantitative study which focuses on statistics, using a large number of participants and little if any participant-researcher interaction (Gray, 2014). The number of participants was expected to grow if the saturation of the data was not achieved.

4.3.2. Materials

Spatial metadata records were selected from two data directory or portals: Victoria Data Directory (VDD) and Australian Spatial Data Directory (FIND). The portals were selected based on the following criteria. The metadata is: (1) created in accordance with international metadata standards (ISO 19115); (2) available publicly and accessible via the internet; and (3) the VDD represents open data platform and VDD represents spatial data catalogue. VDD is a state level portal that connects and shares spatial metadata repositories with official institutions in Victoria State, while FIND is a national level portal consisting of metadata nodes from national institutions such as Geoscience Australia and Bureau of Meteorology.

Other materials were a set of questions designed as guide questions for semi-structured interviews and written instructions for the participants containing scenarios, tasks and objectives. Another important piece of equipment for this study was an audio recorder for recording participants' verbal expressions and opinions during the completion of the test.

4.3.3. Scenario

The scenario should be designed to use as many types of spatial data from as many sources as possible, in order to evaluate spatial metadata from a wide range of spatial data applications and sources. This scenario should also be applicable and reasonable, meaning that it is based on real activity and can be conducted in a limited time.

For this experiment, the scenario was to develop an evacuation map for Greater Melbourne and the East Coast of Australia (Queensland and New South Wales). The set of spatial data required for this activity follows:

- Elevation dataset with minimum scale of 1:25.000
- Hydrographic datasets (e.g. rivers, dams, ponds, etc.) with minimum scale of 1:25.000
- Meteorology Datasets of information (e.g. rainfalls)
- Road network or transportation network datasets with minimum scale 1:10.000 for urban area and 1:100.000 for rural area
- Soil type maps with minimum scale of 1:100.000

To acquire the appropriate data, participants searched the Victorian Data Directory (VDD) and Australian Spatial Data Directory (FIND) websites.

4.3.4. Tasks

Following the given scenario, the participants were allocated three main tasks, to:

1. find the required spatial data from discovery systems (VDD and FIND);
2. determine the suitability of spatial data for the given scenario; and
3. determine how to obtain and appropriately use the spatial data.

The first task is designed to investigate the usability of metadata, as well as discovery systems, in helping users to find and locate the required spatial data. The second is for participants to use the information from metadata records to assess the suitability of the spatial data for the given purpose. They were asked to verbally express their thoughts about the tasks and spatial metadata in detail.

4.3.5. Procedure

The usability test was conducted in a prepared laboratory environment using the following procedures:

1. Orientation Session – Each participant attended orientation prior to task performance; data discovery and selection. During the orientation, the participants were given an explanation about the purpose of the study, the

scope and expectations of the experiment, and the tasks to be performed. Written instructions were also given. In this session, participants were allowed to ask questions related to the study and the tasks, and at the end of this session participants were asked to sign a consent form. This session lasted five minutes.

2. Task performance – Each participant performed and completed three tasks as described in the task section.
    - The first task was searching and locating the spatial metadata required for the given scenario from the spatial data directory. Participants were given 30 minutes to search and find the required spatial data from the discovery systems. They were asked to stop the task when the time was finished. At that stage, success rate was observed.
    - The second task was to decide whether the data was suitable for the given project or not, by examining the spatial metadata records obtained from the first task or, if the participant had been unable to find relevant records, the pre-selected metadata records prepared by the researcher.

3. Interviews – After completing the tasks, participants were asked to fill out a questionnaire and to explain their answers to the researcher. The questions were designed to get participant's opinions about their experiences working with the spatial metadata during the data discovery and selection process.

During the experiment, the participants were accompanied by the author. The author's roles were to give instructions and explanations about the test during the orientation session, answer questions raised by the participants during the performance of the tasks, and to make sure that the participants stayed focused on the objectives. The researcher also asked participants the prepared questions when an interview stalled or to elicit more detailed insights from participants about the metadata.

4.4. Results and Interpretation

Two data sources were collected from the experiment: audio recordings and semi-structured interviews based on a set of questions. As an exploratory study, analysis methods used were mainly qualitative. TAP analysis was used to analyse the TAP data. The interviews data was analysed using content analysis.

4.4.1. Think-Aloud Protocol (TAP)

During the spatial data discovery and selection process, participants' verbal expressions were recorded. Each record was then transcribed and analysed separately using the protocol analysis. The transcriptions were analysed by identifying referring phrases (RPA) and each phrase was given codes based on the reference codes listed in Table 4.1. Examples of the phrases, and the codes which resulted from this process, are presented in Table 4.2.

Table 4.1: Reference codes used in Protocol Analysis (Fonteyn, 1993)

Concept	Definition
Action	The manner or method or performing; a thing done
Element	A piece of information describing certain characteristics of spatial data
Sign	Objective information indicative of status of the information or tools
Time	A chronological reference
Needs	Information, tools or actions required to discover and select data
Value	A rating of usefulness, importance, or worth
Goal	A specific goal for an action

From Table 4.2, we can see phrases and respective concepts (codes) resulted from the participants' verbal expressions during the spatial data discovery and selection. These phrases and codes identify all the vocabulary and concepts the participants' focused on during the process.

Table 4.2: Example of RPA results from TAP for spatial metadata usability testing

Phrases	Coded concept(s)
✓ We need *topographic or elevation map of Melbourne*	Data
✓ (I did the *search* for *topography*, so) I see some *results*	Action, Element, Signs
✓ If I *don't find the results,* I can *search* other *terminology*	Sign, Action, Element
✓ (and I see*) eight pages of the results*	Signs
✓ *Why can't short it* based on the *most relevant ones*	Signs, value
✓ *The title doesn't mention* the *scale*	Element, Signs, Needs
✓ We need a *scale* here	Needs
✓ The *scale* is not *mentioned* here	Needs, Signs
✓ *It could be good* to put *the scale (in the title)*	Value, Needs
✓ (*If we use*) *elevation map*, *37 datasets* were found	Action, Element, Signs
✓ (*By applying) the most relevant*, it gives me *1300*	Action, Signs
✓ I *tried to put the newest back*, and it gives me *2000* results	Action, Signs
✓ This might be related to *technical issues* of the *system*	Signs, Needs
✓ We *go with the newest first*, but we cannot go around *3000* results	Action, Signs
✓ I think users will *look at the title* on the results	Action, Element
✓ *the title* is *not adjusted very well*	Element, Signs
✓ and when I tried *to search* using *region* (as a keyword) I *cannot find exact data* for Melbourne	Action, element, Signs
✓ If I *want to find* a specific region in Melbourne, like *Southbank*, I *cannot find it easily*	Needs, Element, Signs
✓ *The system doesn't provide* any *advance search*	Needs, Signs, Needs
✓ It *doesn't have* the *filters*	Signs, Needs
✓ But *it has* like *license, wms, shapefile output, vegetation*, etc.	Signs, Element
✓ So maybe it's better to *try these tags* (to see what's going to happen)	Needs
✓ Let's see this *HY waterpoint*, *I don't know what it stands for*, maybe *hydro*	Element, Sign, Element
✓ I think the *filters (for the search) are not enough* and *the data are not well tagged*	Needs, Sign, Data, Sign
✓ because it *doesn't have enough tags* and the *categories* are *lacking*	Sign, Needs, Needs, Sign
✓ *I don't know how* to *filter* the *scale*	Sign, Action, Element
✓ *No tools* to *choose* the *scale*	Sign, Action, Element
✓ Some of the *titles have scale* in them, and *some don't*	Element, Sign, Needs, Sign
✓ We can *spend a lot of time* ~~for~~ *searching* each data	Sign, Action
✓ We can *use* either different *keywords*, and maybe after *half an hour or more*, we *can find* the one you look for	Action, Element, Time, Sign
✓ But the *efficiency* of *the search tools* and the *database* should be *higher than that*	Needs, Needs, Value
✓ I need *to find* the (potential) data in *one minute*, Otherwise, I could *spend a lot of time*	Action, data, time, Sign

The phrases and the codes were then further analysed to determine the assertions of each phrase. These assertions were then used to understand how the relationships were being formed during the spatial data discovery and selection. Table 4.3 shows the phrases and their respective assertions.

Table 4.3: Example of assertions analysis of the TAP data

Data	Assertion
✓ *I don't know how* to *filter* the *scale*	Causal***
✓ *No tools* to *choose* the *scale*	Causal
✓ Some of the *titles have scale* in them, and *some don't*	Connotative**
✓ We can *spend a lot of time* for *searching* each data	Causal
✓ We can *use* either different *keywords*,	Indicative*
✓ and maybe after half an hour or more, we can find the one you look for	Causal
✓ But the *efficiency* of *the search tools* and the *database* should be *higher than that*	Connotative
✓ I have to be able *to find* the (potential) data in *one minute*, Otherwise, I could *spend a lot of time*	Causal

*Indicative assertions form relationships of significance of an action
**Connotative assertions form relationships of meaning or state
***Causal assertions form relationships of cause and effect

The final stage of the analysis process was conducted by implementing script analysis, which assigns a set of operators to the phrases to identify the process of reasoning during spatial data discovery and selection. Three operators were used; suggest, study, and conclude. These operators represent the interactions in the reasoning process of spatial data discovery and selection. Table 4.3 provides an explanation of the operators.

Table 4.4: Example of script analysis of the TAP data

Data	Operator
✓ we need a *scale* here	Suggest*
✓ The *scale* is not *mentioned* here	Study**
✓ *It could be good* to put *the scale (in the title)*	Suggest
✓ (*If we use*) *elevation map*, *37 datasets* were found	Study
✓ (*By applying) the most relevant*, it gives me *1300*	Study
✓ I *tried to put the newest back*, and it gives me *2000* results	Study
✓ this might be related to *technical issues* of the *system*	Conclude***
✓ I think users will look at *the title* on the results	Suggest
✓ *the title* is *not adjusted very well*	Study
✓ and when I tried *to search* using *region* (as a keyword) I *cannot find exact data* for Melbourne	Conclude
✓ If I *want to find* a specific region in Melbourne, like *Southbank*, I *cannot find it easily*	Conclude
✓ The system *doesn't provide* any *advance search*	Study
✓ It *doesn't have* the *filters*	Study
✓ I could use (filters) to filter the data	Suggest

*Suggest means to consider an action or information to achieve a specific goal
**Study means to observe or consider information or facts carefully
***conclude means to decide on the significance, value or results of an action

From the tables presented above, we can identify the participants' activities, goals, and expectations, as well as relationships between them and their reasonings when they try to discover and select spatial data using the metadata via the user-interfaces.

The results of script analysis were then used as the basis for identifying the effectiveness, efficiency and user satisfaction, as well as identifying usability problems and user expectations for spatial data discovery and spatial data selection.

Spatial data discovery

In spatial data discovery users use the user interface to interact with metadata to discover the data for the given purpose according to the scenario. Effectiveness, efficiency, user satisfaction, problems and user expectations were then identified by interpreting the results.

Effectiveness

Effectiveness of the spatial metadata is a combination between the discovery systems or searching tools and the metadata records. To this end, effectiveness was measured by the successfulness of the participants in finding and discovering potential spatial data for the given scenario from the two discovery systems: VDD and FIND. From the script analysis results, we can identify the successful rate of the discovery by looking at the conclude operator, which is related to significance, value or results of an action.

To acquire a comprehensive result, conclude phrases were examined and grouped based on the tones of the phrases; positive, neutral and negative tones. Positive is when a phrase indicates a successful result and negative if it indicates otherwise. Neutral is a tone where a phrase is neither positive or negative.

> "I see some *results"* – Positive conclude
>
> "I *cannot find exact data* for Melbourne" – Negative conclude
>
> "can we assume that we are *failing* to *discover* the *map*?" – Neutral conclude

By only analysing the conclude phrases throughout the analysis document, it is hard to make a clear statement about the effectiveness of the spatial metadata for data discovery. Nevertheless, the concludes provide indications of ineffectiveness. The example concludes phrases given above, are a participant's verbalisations during the discovery process of the Melbourne Topographic Map. Meteorology data provides other examples of conclude phrases:

> "so, the *meteorological didn't respond that much*"
>
> "if you look for *the rain*, you only *get 1 result*"

Most of the conclude phrases do not provide a clear indication that the data discovery process can be successfully completed or not. In the examples above, the participant tried to finish the searching with a doubtful decision. It is interesting to investigate why this participant came to that decision. To find out, we look at other phrases; study and suggest, which, together with conclude, form the whole process of reasoning, as illustrated in Table 4.4.

Table 4.5: Script analysis data for a topographic data

Data	Operator
✓ we need a *scale* here	Suggest
✓ The *scale* is not *mentioned* here	Study
✓ *It could be good* to put *the scale (in the title)*	Suggest
✓ (*If we use*) *elevation map*, *37 datasets* were found	Study
✓ (*By applying) the most relevant*, it gives me *1300*	Study
✓ I *tried to put the newest back*, and it gives me *2000* results	Study
✓ I *cannot find exact data* for Melbourne	Conclude
✓ Some of *the titles* have *scale* in them, and *some don't*	Study
✓ We can *spend a lot of time* ~~for~~ *searching* each *data*	Conclude

Looking at the whole reasoning process, the participants have negative concludes because they could not find the scale in the title. One participant faced 37 initial results, 1300 results after clicking the most relevant and 2000 after another click on the newest results. This participant found that the results, represented in a list of titles, should have included a scale. A scale would have greatly assisted them to locate the required spatial data, as the participant did not want to spend a lot of time searching titles by clicking on them one by one.

Another participant showed a different attitude when facing the same situation, as they just clicked on a title that they thought was the most relevant to the required data.

> "There're *some problems with this*. I type *elevation*, and this one is the *most related*, but it *doesn't come the first*"
>
> "I think *accuracy is enough*, I think if we *have this*, the *accuracy will be enough*"
>
> "so, can *I look* for *another one*?"

The participant ended up spending a lot of time, more than the time given to do the search, by clicking and opening each and every title in the result list to find the relevant data for the given scenario.

There were recurring groups of phrases; positive, negative and neutral from the data during the spatial data discovery process, which indicate that spatial metadata effectiveness is subject to the quality of the metadata. The quality of the metadata is determined by the following: completeness, consistency and relevance of the information presented to users; the intuitiveness of the user interface against the user needs; detailed criteria for data search filters; and willingness of users to do further detailed search within the results. The results indicate that spatial metadata, along with the searching tool, can be used to find and discover spatial data, but users found some obstacles that hampered them from performing the task efficiently, as explained in the next section. These obstacles might prevent users from reaching the task goal, as they tended to stop the searching right away when they found that the search was not going as they expected. They would try other ways to find the data, such as finding information about the data from an internet search engine and contacting the producer directly.

Efficiency

Spatial data users can search the required potential data from the websites, but some effort and time was necessary to find and discover the data. The following concludes indicate that the data discovery was not a straightforward process, where participants could not get the expected results in the given time:

> "We can *spend a lot of time* for *searching* each *data"*
>
> "and maybe *after half an hour or more*, we can *find* the *one* you look for"
>
> "But the *efficiency of the search tools and the database* should be *higher than that"*
>
> "I have to be able to *find* the (potential) *data* in *one minute*, Otherwise, I could *spend a lot of time"*

During the process, similar concludes were recurring to the point that most of the users did not want to continue their current data searches and asked to move to another search.

> "If I cannot *find it* in like *one minute*, it means I have to *move* on to the *next search*"

The researcher's notes during the process of this task also indicate that none of the participants could discover all the required data in the given time.

These results indicate that the spatial metadata is able to help users to discover spatial data from the websites, but there are efforts required to increase the efficiency of the process. What needs to be done to enhance efficiency may be determined by first revealing the problems and user expectations, as explained in the next sections, and addressing the problems by enhancing the metadata and the user interface simultaneously.

Usability problems

Participants faced several problems in their attempts to discover the required spatial data using the websites. The problems are related to both the website (user interface) and the metadata (content). These problems can be identified by examining the concludes in the results in the following examples. Figure 4.2 illustrates some of the problems identified during spatial data discovery as mentioned by the participants, e.g., unmatched keywords and results, tags/filters that are not easily recognised nor sufficient to participants' needs.

> *"No tools* to *choose* the *scale"*
>
> "Some of the *titles have scale in them*, and *some don't"*
>
> "Why *can't short it* based on the *most relevant ones"*
>
> "The *title doesn't mention* the *scale"*
>
> "(By applying) the *most relevant*, it gives me *1300*. I tried to put the *newest* back, and it gives me *2000 results"*
>
> "and when I tried to *search* using *region* (as a keyword) I *cannot find* exact *data* for Melbourne"
>
> "The *system doesn't provide* any *advance search"*
>
> "I think the *filters* (for the search) *are not enough* and the *data are not well tagged"*
>
> "Interesting part is the *titles are not very accurate*, I think they come from the title of database or datasets"
>
> "This *tag is helpful*, *otherwise* you see the *title might not be relevant"*
>
> "you have to *have* some sort of *scale*, so *you can filter*."
>
> "One (thing that) I could do is *open this one* to *find* out the *scale* because *it is not mentioned* here. Look *how many data to open to see the scale*."
>
> "if you're using *this system* after a while you find that *tags are not helpful*"
>
> "what is this *VIF 2015*? I *don't know*, I *have to open* it. I *have to see* the *data* to understand VIF, and *it takes time*."
>
> "That's why it's *important to put the name*, *not the abbreviations*."
>
> "the *title is about elevation*, but here *the search is land*."
>
> "when I *typed* the *keywords* and the same come out is *totally different from what I expected*"

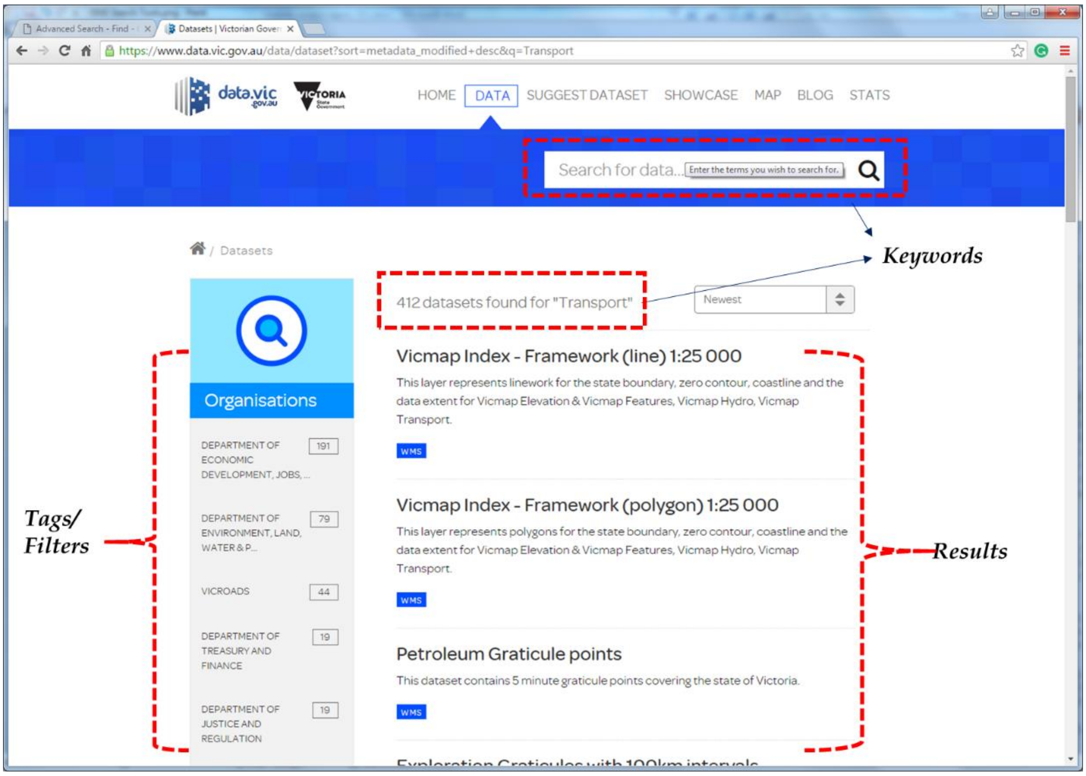


Figure 4.2: Problems identified with the VDD systems in spatial data discovery

Similar types of concludes are found in the results. From the lists the identified problems can be grouped into two types; user interface problems and metadata records problems. Table 4.5 summarise the problems that can be addressed to increase the usability of spatial metadata for spatial data discovery.

Table 4.6: Identified usability problems in spatial data discovery

User interface problems	**Metadata records problems**
✓ Missing important filter for data search e.g. scale is not provided (in VDD and FIND)	✓ Inconsistent information presented in titles e.g. scale, unknown abbreviations
✓ Tags are not necessarily relevant to the data (VDD)	✓ Keywords are different from the data description
✓ Inconsistency of the results when users change the newest to most relevant (FIND)	✓ Inconsistent abstracts e.g. contain superfluous information or too short and uninformative
✓ Irrelevant results with the submitted keywords (VDD and FIND)	
✓ Advanced search is not available (VDD)	

User satisfactions and expectations

User satisfaction is not a usability attribute that can be easily evaluated, since it is related to emotions, feelings and experiences of participants during the task completions. However, it could be detected by carefully observing the data (phrases) in conjunction with researcher's notes while observing the experiments. In the process of identifying the user satisfaction, user expectations on both the spatial metadata records and user interface are observed simultaneously, since the satisfaction is determined by the gap between the expectations and the reality. Following are some of the phrases that can be used to detect user satisfaction and expectations during the spatial data discovery.

> "I think, I *cannot find* the *scale*, probably *I need to open* one by one *to see* what *the scale is*"

> "People should *define some basic needs*, and they *have to face* with *this* kind of *search*"

Similar phrases were recurring during the spatial data discovery process that can be used to detect user satisfaction as well as identify the expectations. From the above phrases, an unsatisfactory result can be detected as the participant could not find the filter scale to narrow down the results. They expected that the searching tool would provide them with a set of basic criteria to narrow down their search.

The researcher's notes also indicate that participants were not enthusiastic when they opened the VDD website, since it only provides a simple search by submitting keywords. They did not notice the tags provided on the navigation window at the left-hand side of the web page. Even when they found out about the tags, they were not happy because when they tried to use them, the results were more irrelevant than they expected. Different reactions were given by the participants when they opened the FIND website. Participants were enthusiastic when they found the advanced search in the website. However, they started to complain about the website when they received results, especially because they could not find the way to narrow down their search using the scale. Table 4.6 summarises user expectations for spatial data discovery.

Table 4.7: Identified user expectations for spatial data discovery

User interface	**Metadata records**
✓ Simple (keywords) and advanced search based on basic criteria for spatial data, including the scale	✓ Consistent and informative titles e.g. containing the scale and other basic information
✓ Consistent and relevant tags for the filters	✓ Consistent and informative abstract e.g. detailed information about data contents presented in brief textual descriptions
✓ Relevant and consistent results presentation e.g. titles and abstracts	✓ Consistent and relevant keywords with titles

Spatial data selection

Spatial data selection is a process where participants try to understand the characteristics of spatial data by reading and interpreting information presented in spatial metadata to decide whether the spatial data is suitable or fit for the given scenario or not. In their attempts to read and interpret the information, participants use their knowledge and experiences which vary between them.

The effectiveness, efficiency, usability problems, user satisfaction and expectation for spatial data selection can be identified from the protocol analysis results.

Effectiveness

Conclude phrases related to this process were observed to identify the result of participants' attempts on data selection.

> "So, it is *not very easy* to *make a quick decision*"
> "After they give you *a sample* then maybe you have *to contact them* and *ask them* what *this data is about*"
> "Currently from this *metadata, you're not sure about data*"
> "I'm not sure of this. I think, I can say that *I am in the middle*"
> "*I'm confident*, *but I need more time to read*, probably *I need more experience*"
> "You need to *spend more time*, and you will be *more confident*"
> "I think *it is enough* for me *to decide* whether this is what I need or not"
> "*I think*, if can have *a look at the data itself,* I might be confident"

Again, similar phrases recurred during the experiments as thoughts by participants. As in spatial discovery, it is not easy to clearly say that spatial metadata is effective or ineffective for spatial data selection, as participants gave positive, negative and neutral concludes. Nonetheless, the phrases allow us to sense a degree of ineffectiveness. For example, a participant's verbalisations vacillated during the process. At some points, the participant would be confident in their decisions, but later, lack conviction when they had a different experience with the metadata.

Based on researcher's notes during the process, at first, participants looked at the metadata page as a whole to find information that might be critical to their decisions, including abstract, production date, update, producer and accuracy, and then felt confident that they could make any decision. However, after they had a detailed look at each element, they started to have doubts. For example, when they looked at the data accuracy, they were very critical and demanded detailed information. They complained about the unclear values of the accuracy, e.g. 1m to 10m, and asked for accuracy evaluation process. Participants had similar experiences with metadata from both websites, VDD and FIND. Moreover, all participants required additional information that was not provided in the metadata, including spatial data sample (access to the data), information regarding producer's reputation and reports on spatial data quality tests.

Efficiency

Similar to spatial discovery, participants were only able to read and to evaluate half or less of the metadata from the given tasks. They spent much of their time trying to find critical information within the metadata to assess the suitability of the data, for example, accuracy, to figure out whether the information was useful or not. They spent more time trying to interpret information that was not always relevant to their knowledge. The following phrases are examples of what participants had to deal with and what cost them time locating and interpreting information in the metadata.

"the *description* is *not so clear* and *too much for data source*"

"just have *a list* of maybe *who is the providers*"

"why put this in data source? *class shouldn't in data sources*"

"I think *they mess (mixed up*) the *information* (in this part of metadata)"

Accordingly, efficiency of metadata is due to be increased. Moreover, participants need more time to acquire additional information from different sources, including producers and data experts, to assist them to make clear decisions on the data.

Usability problems

Data analysis results provide us with various problems faced by participants during the spatial data selection process. These problems prevent them from making a clear decision about the suitability of the data for the given scenario, as well as hampering participants in making selections in the given amount of time. Missing important content, as in maintenance date or last update, discouraging presentation of data quality and abstracts, were also issues. However, the main problem, which all participants identified, was the irrelevance and inconsistency of information returned in abstracts and in accuracy. Figure 4.3 illustrates the problem found in an abstract.

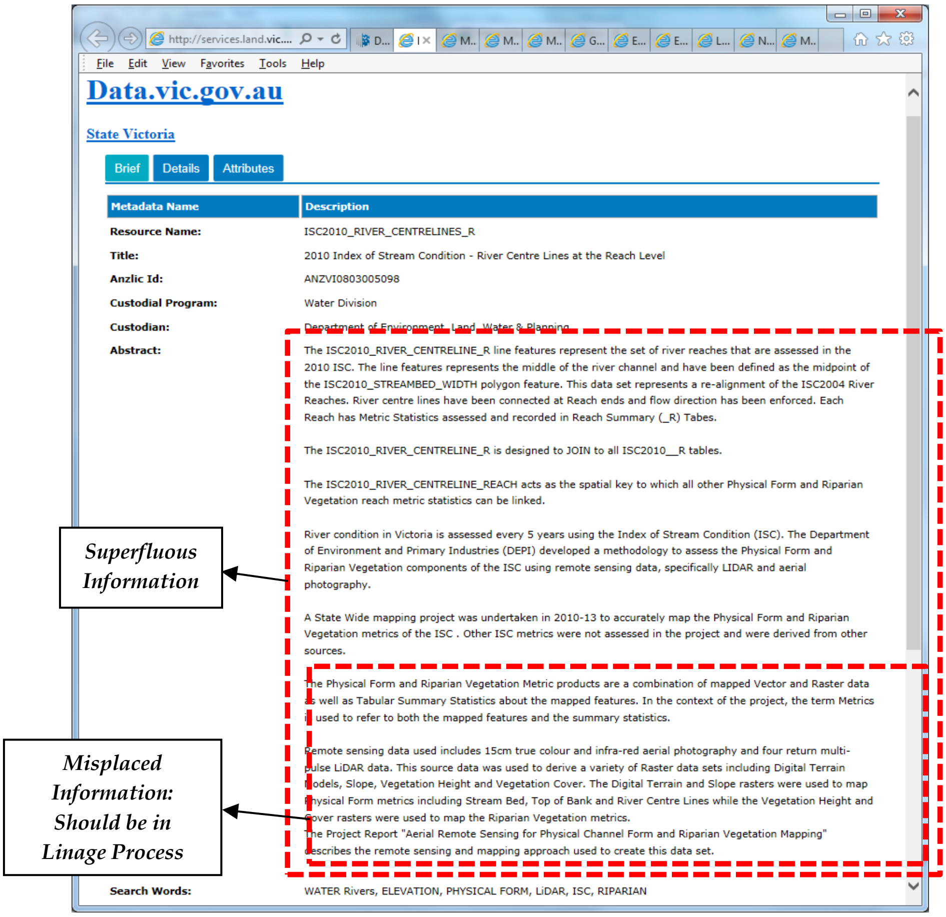


Figure 4.3: Superfluous information in Abstract

Participants agreed that abstract plays a crucial role in spatial data selection, as they would start the process and reasoning with it. A brief, clear, easy to understand and relevant abstract, with the terminologies they know, is essential to help them understand what the data is about and how the data would fit their purposes. A good quality abstract, according to users' perceptions, would encourage and guide them to find other information about the data within the metadata. Conversely, a messy abstract would confuse and discourage them from getting more information from the metadata. Table 4.7 summarises the usability problems identified from the experiment.

Table 4.8: Identified usability problems in spatial data selection

User interface	**Metadata records**
✓ Discouraging presentation e.g. plain textual presentation	✓ Missing metadata contents ✓ Inconsistent content between metadata records e.g. in abstract ✓ Unfamiliar terminologies e.g. in abstract and accuracy ✓ Inconsistent length of content e.g. in abstract and accuracy ✓ Lack of information in contents e.g. no validation test for accuracy

Besides the above listed usability problems, the data provides us with a meta-problem, that is, a problem beyond a usability problem found by participants in metadata records and user interfaces. This meta-problem was detected from the phrases below.

> "I think it's the *provider*, because we *need to know to judge*, we *can trust the data or not*"
>
> "to clearly *describe the data*, but I think *the best way is to have a sample* or something"
>
> "or they *give you a sample* then after that maybe you *have to contact them*"
>
> "if you can go and *search one sample*, not too much but *to see what kind of data* they have *to be sure*"
>
> "you can read and say -*definitely I need this*- *but when you, like, down load* or get down, you figure out that *this is not my data*"

The problem was related to confidence, where participants felt that information presented in metadata would only be useful to examine the information in metadata. It is not that they did not trust the metadata, but they thought that they could not base their decisions on the suitability of the spatial data soley on the information presented in the metadata. In order to make decisions, they required additional information, referred to as meta-information, including the reputation of the producers, their experiences with the data or the metadata system, that could only be obtained by live or direct interaction with the data. Similar reasoning might be found in the car sales business, where customers would look and read the brochure or advertisements to identify potential cars that would fit their criteria. However, deciding to buy or not to buy the car is a different story and requires different information that cannot be provided by the brochures or advertisements, that is, a live experience with the car itself. Hence, test driving. The needs and expectations of this meta-information are explained in the next section.

User satisfaction and expectations

During the spatial data selection process, participants identify specific information in metadata records that could be useful to assess the suitability of the data for the given project. In doing so, participants start with reading the abstracts to get a brief understanding about the data before they continue the process. However, participants found that most of the abstracts were problematic and discouraged them from reading and spending more time to understand the information. This led them to go to other metadata records based on their existing knowledge to identify the important information that might be useful for them to assess the suitability. Most of them went for accuracy information and found that information presented in the accuracy was not necessarily clear and comprehensible. Participants went for, amongst other things, information such as production date (the age of spatial data) and the boundary (geographic coverage). During the process, they kept verbalising their thoughts, which were mainly negative, about the metadata, as they found that metadata did not meet their expectations, as summarised in Table 4.8.

Table 4.9: Identified users' expectations for spatial data selection

User interface	Metadata records
✓ Alternative presentation style e.g. statistical or graphical presentation for data accuracy	✓ Brief and consistent abstracts with critical information for spatial data selection
✓ Dynamic/intuitive presentation/pages	✓ Use of familiar terminology to users' knowledge and experiences ✓ Access to data sample or preview for data self-assessment ✓ Access to data producers for additional information about the data

4.4.2. Semi-structured interviews

After completing the given tasks, participants were given a set of questions regarding the spatial data discovery and selection. The researcher used these questions as a guideline for the semi-structured interview to explore participants' opinions and experiences about working with spatial metadata. They were encouraged to recall any critical incidents during the process and give their explanations about the incidents.

Question 1: To what extent do you agree that you can find the required spatial data?

Participants gave various answers to this question as they had different experiences and problems. As can be seen in Table 4.9, for both spatial data discovery systems, participants' opinions vary, with no strong indications that they successfully located the required spatial data.

Table 4.10: Participants' opinions about the results of the spatial data discovery

Participants' responses	VDD		FIND	
	n	%	n	%
Strongly agree	0	0%	0	0%
Agree	2	40%	1	20%
Either agree or disagree	2	40%	1	20%
Disagree	0	0%	1	20%
Strongly disagree	1	20%	2	40%

Participants mentioned the lack of criteria as filters to narrow down their search from the earliest stage. They recalled how irritating it was when they had to look at so many listed titles and to read and check the titles one by one only to discover that the titles were inconsistently presented. When they were reminded about tags in the VDD website that could be used to filter the data, most of them were still not happy, as the tags were not easily recognised and did not provide the expected criteria, such as map scale and region (geographic coverage). They did not want to spend a lot of time searching for data that was not there. Hence, their need for complete filters for searching criteria to get immediate results or no results.

Question 2: To what extent do you agree that the spatial metadata (and the user interface) meet your expectations?

As can be seen in Table 4.10, participants' responses to the questions were mostly negative, as were their answers to the previous question. They found serious problem during the discovery process, e.g. the inconsistency between the keywords they submitted with the titles presented in the result. A participant who had experience with the discovery websites mentioned about how to improve the system, e.g., provide tags from previous users to help him identify appropriate data as they might have different terminology for the same data and share this with other users.

Table 4.11: Participants' opinions about their expectations in spatial data discovery

Participants' responses	VDD		FIND	
	n	%	n	%
Strongly agree	0	0%	0	0%
Agree	0	0%	1	20%
Either agree or disagree	1	20%	0	0%
Disagree	3	60%	2	40%
Strongly disagree	1	20%	2	40%

Another participant mentioned putting more effort into making the metadata records complete, clear and consistent, both within a record and between metadata records. They recalled a specific experience where they had encountered a title with abbreviations in it that has no explanation whatsoever in the abstract nor in the rest of the records, and they had to open the metadata because the title appeared on top of the most relevant result.

Question 3: To what extent do you agree that you can determine spatial data fitness for use to decide whether you will use the data or not?

Participants with a spatial information background tended to leave the metadata page once they had information about the contents/features, the age of the data (last update) and the geographic coverage (region covered by the data). They would look for opportunities to download and get the data and perform their own assessment on the actual data or data sample, to find out whether the data was fit for their purpose or not. Or, they preferred to contact the data producer directly to get the information they required to make decisions about the data, instead of reading the metadata and basing their decisions on the metadata. Those from non-spatial background, like the civil engineer, would prefer to look at the data producer and get information regarding the reputation of the producers to be sure about the data. The participants' explanations, as can be seen in Table 4.11, were therefore mostly negative.

Table 4.12: Participants' opinions about the spatial metadata for suitability assessment

Participants' responses	VDD		FIND	
	n	%	n	%
Strongly agree	0	0%	0	0%
Agree	1	25%	0	0%
Either agree or disagree	3	50%	1	25%
Disagree	1	25%	4	75%
Strongly disagree	0	0%	0	0%

Another participant, who was very confident with his spatial knowledge, preferred to read the report on spatial data processing, to find out information about the methodology for data collection and standards for quality validations.

Question 4: To what extent do you agree that the spatial metadata met your expectations for spatial data suitability assessment?

As can be seen from the participants' responses in Table 4.12, the metadata sufficiency for assessing suitability of spatial data for a certain purpose is subject to their experience with the data and their professional backgrounds. Most of the participants gave negative responses, as the metadata barely meet their expectations. Instead, they required additional information that was not part of the metadata. The next question is, to what extent the metadata could be improved to meet their expectations? Participants' responses were unclear, as they could not articulate what they expected from the metadata. Some of them, again, asked for spatial data access or sample. Some asked for previous usages by previous users as in reports or reviews. Some participants said that they would not decide the suitability of the data based on the metadata. They would only use it to discover the data availability and they would try to access or obtain and check the data suitability by working with the data directly.

Table 4.13: Participants' opinions about the spatial metadata related to their requirements for assessing suitability of spatial data

Participants' responses	VDD		FIND	
	n	%	n	%
Strongly agree	0	0%	0	0%
Agree	1	20%	1	20%
Either agree or disagree	2	40%	0	0%
Disagree	2	40%	3	60%
Strongly disagree	0	0%	1	20%

Question 5: To what extent do you agree that you are confident to use spatial metadata again in the future for spatial data discovery and selection?

This is the last question given to participants. As Table 4.13 shows, the responses are, again, mostly negative. But their explanations were not so negative. Most of the participants would still use and rely on the metadata, and data discovery systems, for searching potential spatial metadata for their future purposes. They would like to see the metadata and user interface improved by addressing identified problems and their requirements from this experiment. They still thought that metadata was important, but improvement should be made to make metadata more usable and useful to them. One participant would like to see metadata presented in different styles and presentations based on users' expertise levels; general, intermediate, and experts, using different terminologies in accordance with the level of expertise. Users could choose and use the discovery system and presentation suitable for them.

Table 4.14: Participants' confidence level using spatial metadata in the future

To what extent are you confident with spatial metadata (without any additional information)?	VDD		FIND	
	n	%	n	%
Strongly agree	0	0%	0	0%
Agree	1	20%	1	20%
Either agree or disagree	2	40%	0	0%
Disagree	2	40%	2	40%
Strongly disagree	0	0%	2	40%

4.5. Analysis

Since there was no different treatment nor weight was given to any participant with specific spatial knowledge whatsoever. Therefore, their responses during the TAP and interviews were treated equally, and the results and findings of the usability evaluation of established spatial metadata reveals that the metadata and the user interface remain problematic for spatial data users to discover and select spatial data, regardless of the participants' educational background or experience with similar user interface and metadata.

Problems such as inconsistent information found in most elements in the metadata records hampered spatial data users from effectively and efficiently discovering and selecting spatial data. Another problem that can be highlighted is the irrelevance of the information presented in the metadata to the users' knowledge. The problems found in the user interface add another difficulty to the discovery and selection process. Lack of searching criteria, irrelevant results from the submitted search, and discouraging textual presentation that is either too short or too long, make the discovery and selection process harder.

The fact that these metadata records were created following the same metadata standard ISO 19115 did not prevent the problems from happening. Most of the problems come from the free-text type of elements in spatial metadata standards such as title, abstract, keywords, data quality and data provenance. The information mainly comes from the metadata authors' understanding about the data. This indicates that the standard of capability remains lacking for maintaining the consistency of information presented in the metadata and the relevance of the information to the spatial data users' knowledge.

Another finding in spatial data selection reveals that spatial data users require more than just information from metadata to be able to select the data for their purposes. Participants with educational background in spatial science prefer to get access to actual spatial data or sample so they can assess the suitability of the data themselves. Whilst other participants thought that if they could have a look on the data, it would useful to check how the actual data look like. But all of them also think that experiences from other users who worked with the data are useful for the selection process. This is another indicator that the standards or the main guideline for spatial metadata design and production should be improved, and the improvement should help spatial data users with different backgrounds.

The rest of the thesis is designed to address the problems and user expectations of spatial metadata to improve the usability of spatial metadata and recommend how spatial metadata standard ISO 19115 can be extended to provide a solution for better metadata usability.

4.6. Chapter summary

This chapter outlined the analysis method, design, process and results of the established spatial metadata usability evaluation.

The first part explains the analysis method, experiment design as well as step-by-step activities for conducting this evaluation. Protocol analysis and script analysis (content analysis) are implemented for analysing data from TAP and interviews.

The second part of this chapter presents the results and findings related to current spatial metadata usability for spatial data users. The results, findings, and discussions are divided into two groups based on the data collection method; TAP and semi-structured interviews. From the TAP data analysis, effectiveness, efficiency, user satisfaction, usability problems and user expectations are identified. Semi-structured interviews provide additional detail, and more personal user opinions, for the TAP results and findings.

The results indicate that the spatial metadata is neither effective nor efficient for spatial data discovery and selection, and improvements are required of both the metadata records and the user interface. The experiment also revealed some usability problems in both metadata records and user interfaces (websites). Inconsistency and irrelevant information in the metadata records were found in title, keywords, abstracts, data quality and other elements of the metadata. Lack of searching criteria, irrelevant results from the submitted search, and discouraging presentation of the metadata are some prominent problems found in the user interface. Those problems hampered the spatial data users from effectively and efficiently discovering and selecting spatial data. Moreover, the results also revealed that the information presented in metadata records does not suffice the needs of spatial data users to make a selection of spatial data. Previous user reviews and experiences and access to actual data or sample are also required by the users to make their selection.

Accordingly, spatial data users were not satisfied and expected that the websites would provide them with better searching tools and better results in terms of consistency of relevance of information regarding spatial data.

The results and findings from this evaluation chapter are used as the basis for usability improvements and for garnering spatial data users' needs and expectations from the wider community, as explained in the next chapter.

This page is left intentionally blank

5. User-oriented spatial metadata profile

5.1. Introduction

This chapter presents the results of usability evaluations and surveys for identifying user needs and expectations of spatial metadata for data discovery and selection. This chapter then maps the user needs and expectations identified by the surveys with the metadata elements in the ISO 19115:2014—the geospatial metadata standard—to determine the compliance of the standard to the user requirements. Finally, the chapter explains the integration of the user needs and expectations with the ISO 19115:2014, in order to develop the user-oriented spatial metadata profile as an extension of the discovery metadata of the standard.

5.2. Questionnaire structure

In order to address the second objective, an online survey was designed and conducted to gather spatial data users' opinion about spatial data discovery and selection. The questionnaire was developed based on the users' insights from the metadata usability evaluation explained in Chapter 4 and was then circulated internationally via email to the researcher's known associates, via the researcher's LinkedIn account, and via the researcher's Twitter account. Selected social media accounts belonging to spatial communities and organisations were tagged, e.g. CSDILA, CDMPS, etc.

The structure of the questionnaire, as illustrated in Table 5.1, contains five sections: Introduction; Participant's Information; Spatial Data Search and Discovery; Title and Abstract; and Spatial Data Suitability Assessment and Selection. The Title and Abstract sections have been specifically designed to gather information from survey participants relating to their experiences and opinions concerning the contents, number of words, and presentation of the titles and abstracts shown in search results. Both titles and abstracts are key elements of the metadata with regards to the usability of the metadata for both spatial data discovery and selection, as explained in the Results and Interpretation section in

Chapter 4. Other survey questions were developed based on the users' insights, as well as on the ISO 19115:2014 geospatial metadata standards.

Table 5.1: Structure of research questionnaire for user needs and expectations of spatial metadata

Section	Topics	Sub-topics
1	Introduction	✓ Introduction to research and questionnaire ✓ Plain language statement ✓ Consent form
2	Participant's information	✓ Working field ✓ Level of spatial knowledge ✓ Years of experience ✓ Location (country) ✓ Spatial data familiarity ✓ Spatial metadata familiarity
3	Spatial data search and identification	✓ Searching criteria ✓ Presentation style
4	Title and Abstract	✓ Contents ✓ Number of words ✓ Presentation style
5	Spatial data suitability assessment and selection	✓ Critical information ✓ Presentation style

5.3. Results

As mentioned in Chapter 3, Methodology, the questionnaire for the survey was distributed world-widely to get as many participants within the pre-determined timeline. The survey was open for two months, and sixty-one participants from eighteen different countries responded to the questionnaire, as illustrated in Figure 5.1. Since qualitative approach is implemented for this research, the number is suitable for the selected approach and is significant to produce a valid outcome.

The following sections describe the results of the survey, starting with participants' demography.

Figure 5.1: Survey participants' origins

5.3.1. Participants' demography

Information regarding the participants' demography has been extracted from the responses to Section 1 of the questionnaire: Participants' Information. The first set of data extracted from Section 1 is related to the field of work or sectors where the participants spend most of their professional time, as illustrated in Table 5.2.

Table 5.2: Participants' work field

No.	Field of work	Number of participants
1	Public Health	2
2	Auditor	1
3	Biodiversity	1
4	Civil Engineering	2
5	Consumer goods	1
6	Data analytics	1
7	Disaster management	1
8	Environment	2
9	Financial	1
10	Geodesy	1
11	Geospatial data production	1
12	GIS or spatial services	32
13	GNSS positioning	1
14	Higher education	1
15	Land management and cadastre	6
16	Planning	4
17	Psychology	1
18	Remote sensing	1
19	Social science	1
	Total Participants	61

These results show that the survey participants come from diverse professional backgrounds, and most of the participants work in GIS or spatial services where spatial data and information are utilised regularly. The results indicate that there are many spatial data users outside the spatial industry. These users will potentially grow in number over time. Gathering their needs and requirements for spatial data discovery and selection is crucial to increase the discoverability and utilisation of the spatial data, and their responses should be treated accordingly, regardless of their numbers.

As well as gathering data about participants' professional backgrounds, the survey also sought to discover users' level of spatial knowledge and their familiarity with spatial data, as illustrated in Table 5.3.

Table 5.3: Participants' spatial knowledge level

Participants' responses	**Category**	**Number**
I sometimes work with spatial data and have a general understanding of spatial knowledge	Knowledgeable	1
I work with spatial data and have a general understanding of spatial knowledge	Knowledgeable	17
I have never worked with spatial data or information before	Novice	2
I work with spatial data and have extensive understanding of the spatial knowledge at high level	Experts	37
I work with spatial data and I have academic or professional training in spatial science	Experts	1
I work with spatial data, but I do not have academic or professional training in spatial science	Knowledgeable	3

Most of the participants have extensive or knowledgeable spatial knowledge and have worked with spatial data regularly. Two participants do not have any spatial training or education and have never worked with spatial data. Nevertheless, their responses are still important and equally analysed, since to their thoughts about spatial data are useful to identify their potential needs in the event that they decide to find and use spatial data in the future.

These respondents might live and work in countries where there are no operational geoportals nor the National Spatial Data Infrastructure. However, their responses are still valuable to be analysed and included in the profile, so the metadata and the user interface are also useful for them, when they need to work with spatial data and need the geoportals to find and select the required data. By doing so, this research will not only improve the usability of metadata, but might also extend the reach of spatial data outside the spatial communities.

Having similar reasons to the participants' spatial knowledge level, participants' levels of experience in using spatial data portals or websites is also useful information to have when analysing their responses in spatial data discovery and selection. This information can be seen in Table 5.4.

Table 5.4: Participants' experience with spatial data portals or websites

Participant's response	Category	Number
I have never searched for spatial data from web spatial data catalogues or portals	None	5
I have searched for spatial data from web catalogues for some projects in the past	Occasional	30
I regularly search for spatial data from web data catalogues or data portals	Regular	26

A majority of the participants have prior experience with spatial data discovery and selection, which makes their responses particularly valuable for the development of user-oriented metadata profiles. However, responses from those who have less or do not have experience are also important, as they enable us to explore the needs of potential users of the metadata and the user interface. Their responses are valuable to make the profile more acceptable to any spatial data user, regardless of their spatial knowledge or experiences. So, the spatial metadata improvement should also be useful for this group of users.

5.3.2. Spatial data discovery

Close-ended questions for spatial data discovery were developed based on user expectations, identified from the usability testing and interviews as explained in Chapter 4, combined with the discovery metadata for geospatial resources of ISO 19115:2014. Open-ended questions were also given to participants, in order to accommodate additional needs not included in the close-ended questions, as well as to give participants the opportunity to explain their answers. Their responses are categorised based on their level of spatial knowledge, as represented in the following tables.

Table 5.5: Criteria for spatial data discovery for novice participants

No	Searching criteria	Responses	Interface style
1	Data content	2	Dropdown list
2	Geographic coverage	2	Dropdown list
3	Time related content	1	Dropdown list

Table 5.6: Criteria for spatial data discovery for knowledgable participants

No	Searching criteria	Responses	Interface style
1	Data content	21	Free text, Predefined text, Dropdown list
2	Geographic coverage	17	Freehand drawing tool, coordinates, geographic names, buffer tool
3	Time related content	12	Dropdown list, free text
4	Data format	10	Dropdown list, Checklist
5	Spatial scale/resolution	9	Dropdown list, Checklist, Free text
6	Data producer	6	Dropdown list, Checklist

Table 5.7: Criteria for spatial data discovery for expert participants

No	Searching criteria	Responses	Interface style
1	Data content	33	Free text, Predefined text (Catalogue Tree), Dropdown list, Map preview
2	Geographic coverage	35	Freehand drawing tool, coordinates, geographic names, buffer tool, Map preview
3	Time related content	20	Dropdown list, free text
4	Data format	22	Dropdown list, Checklist
5	Spatial scale/resolution	19	Dropdown list, Checklist, Free text
6	Data producer	12	Dropdown list, Checklist

The above tables show that both knowledgeable and expert users require *specific information* as criteria when using spatial data portals and websites to discover and identify potential spatial data. This access to more specific criteria allows knowledgeable and expert users to narrow down their search, tailoring it more efficiently to their specific needs. Based on these participants' responses, criteria for knowledgeable and expert users' searches need to include:

- What the data is about;
- The geographical area that the data covers;
- When the data was created and/or updated;
- The quality of the data; and
- Who the data has been produced by.

As the above tables show, the survey's novice participants gave different answers from the knowledgeable and expert participants. They only require search criteria to cover information relating to what the data is about, the geographical location the data covers, and when the data was created or updated. Unlike the knowledgeable and expert participants' responses, novices' responses omitted the need for search criteria that would enable them to search for information with regard to the quality of the data, or who the data had been produced by. The more interesting information from the novice is their responses regarding spatial data selection, as explained later in another section.

The questionnaire also included questions designed to gather information about users' preferred presentation style for each criterion that could be used for developing the user interface. As can be seen in Table 5.8, participants prefer the dropdown list, check box, or free text for text-based criteria. For the geographic coverage, there are a number of alternatives favoured by the participants such as freehand drawing, coordinates, and geographic names.

The results also indicate more detailed presentation styles preferred by participants, and these preferences can be grouped based on participants' experiences with spatial data searching using web portals.

Table 5.8: Presentation style preferred by participants for search criteria

Frequency/regularity of use of spatial data portals	None	Occasional	Regular
Preferred presentation style for search criteria	✓ Drop down list	✓ Geographic Names ✓ Free text for contents ✓ Bounding box for location search ✓ Bounding box for location search ✓ Dropdown list in range for date	✓ Bounding box for location search ✓ Drop down list ✓ Free hand boundary for location search ✓ Check box ✓ Multiple choice

		✓ Free hand boundary for location search ✓ Drop down list ✓ Filter based presentation as in e-commerce ✓ Filter based presentation as in e-commerce ✓ Free text for location search ✓ Buffer using a coordinate for location search	✓ Catalogue tree for content

As can be seen in Table 5.8, the participants who use spatial data portals occasionally and those who use them regularly share similar preferences as to how search criteria are presented and selected. Drop down list presentation is the most preferred presentation style. Free hand boundary for location-based search criteria is also preferred by most of the participants from the occasional and regular use groups.

5.3.3. Title and Abstract

As mentioned in an earlier section, the questionnaire responses indicate that titles and abstracts displayed in search results are key elements for spatial data discovery and selection. Spatial data users might decide whether they would continue or stop the search process based on the information presented in a title and abstract, as explained in Chapter 4. Inconsistency and irrelevant presentation of information were identified as elements of titles and abstracts of established metadata that hamper data users from efficiently and effectively discovering and selecting spatial data. The questionnaire therefore consists of questions specifically designed to address these problems. Participants' responses to these questions are shown in the following tables.

Table 5.9: Participants' responses on information content for title

Participant category	Novice	Knowledgeable	Expert
Content desired in title	✓ Data content (2) ✓ Geographic coverage (2)	✓ Data content (22) ✓ Geographic coverage (14) ✓ Production date (10) ✓ Spatial scale/resolution (10) ✓ Data format (9) ✓ Last update (4)	✓ Data content (37) ✓ Geographic coverage (32) ✓ Production date (15) ✓ Last update (13) ✓ Data format (23) ✓ Spatial scale/resolution (21)

As illustrated in Table 5.9, the participants' responses indicate a preference for crucial information for identifying spatial data to be presented in the title, so they might discover the potential data for their projects instantly by reading list of titles in the result page. They voted that title should contain the preferred information in no more than 10 words, as illustrated in Table 5.10.

Table 5.10: Participants' responses on number of words for title

Up to 5	**Up to 10**	**Up to 15**	**Up to 20**	**No Limit**
10 responses	27 responses	10 responses	6 responses	8 responses

The participants also gave responses indicating their preferences with regards to the information that should be presented in abstracts. Knowledgeable and expert users gave very similar responses in this regard, as can be seen in Table 5.11. The noticeable difference between the two groups is in their preferences with regards to the production process: only two participants from the knowledgeable user group indicated a preference for this information, compared to fourteen participants from the expert group. The results indicate that expert users prefer more detailed information relating to the spatial data, in order to make their assessment on the data quality and compare it to quality information presented in metadata, as explained in Chapter 4.

Table 5.11: Participants' responses on information content for abstract

Novice	Knowledgeable	Expert
✓ Content description - explanation about the feature and its attributes (1) ✓ Detailed explanation about geographic (location) coverage/extent (1) ✓ Spatial scale/resolution (1) ✓ Production date (2) ✓ Data accuracy or error (1) ✓ Level of detail or generalization of the feature presented in the data, e.g. the lowest road type (1) ✓ Intended use of the data (1) ✓ Production process (1)	✓ Content description - explanation about the feature and its attributes (20) ✓ Spatial scale/resolution (14) ✓ Production date (12) ✓ Detailed explanation about geographic (location) coverage/extent (11) ✓ Data format (10) ✓ Data accuracy or error (10) ✓ Maintenance date (9) ✓ Intended use of the data (8) ✓ Level of detail or generalization of the feature presented in the data, e.g. the lowest road type (6) ✓ Production process (2)	✓ Content description - explanation about the feature and its attributes (34) ✓ Detailed explanation about geographic (location) coverage/extent (29) ✓ Data format (25) ✓ Spatial scale/resolution (23) ✓ Data accuracy or error (22) ✓ Production date (19) ✓ Level of detail or generalization of the feature presented in the data, e.g. the lowest road type (19) ✓ Intended use of the data (15) ✓ Maintenance date (14) ✓ Production process (14)

To maintain the consistency and the readability of abstracts, participants also gave their vote on the maximum number of words presented in an abstract. As shown in Table 5.12, most participants voted for a maximum of 150 words in an abstract. Considering the preferred contents of abstracts as selected by the participants, it is reasonable to expect an abstract of this word count to be able to present all the required information.

Table 5.12: Participants' responses on number of words for abstract

Up to 100	Up to 150	Up to 250	Up to 400	No Limit
8 responses	18 responses	13 responses	7 responses	16 responses

As well as the information provided in response to the close-ended questions, survey participants were also invited to respond to open-ended questions. These asked participants to indicate any preferences for additional information, and to provide any further explanations for their answers. The responses to these open-ended questions were to be presented in the result page for the data discovery.

The participants' responses to these open-ended questions are shown in Table 5.13.

Table 5.13: Additional information for the search result page

No	Item	Description
1.	User rating, data citation	As an online product, users' rating for each dataset might be useful to get insight into the quality or usability of the data
2.	Thumbnail, Pop-up image, Graphical quicklook	A graphical quicklook of the spatial data showing the extent of the data
3.	Keywords	Keywords related to the data
4.	License	Intellectual property of the data that limits its use
5.	Organisation	Data producer or owner
6.	Number of downloads	Number of downloads made by previous users
7.	Location	Geographic location of the data relative to the surrounding area
8.	Examples	Small portion of the data that can be downloaded or accessed
9.	Data preview, graphics preview	Map viewer for spatial data or chart/graph viewer of non-spatial data

The results show that only participants from the expert group chose to give responses to the open-ended questions. Their responses indicate that they prefer to find out certain information related to the data as early as possible, in order to identify the potential and suitable data for their applications. The suggestion that examples and data previews are included in the result list, accompanying the titles and abstracts, also indicates that expert participants prefer to look at and assess the data directly, in order to determine the suitability of the data for their applications and make selection as early as possible.

5.3.4. Spatial data selection

The participants' responses to the spatial data selection questions show the same pattern as their responses to spatial data discovery, as can be seen in Table 5.14. The results indicate that novice users are only interested in what the data is about, as they do not possess sufficient knowledge to assess the suitability of the data. Knowledgeable and expert participants, on the other hand, prefer detailed information about the spatial data. Information about internal characteristics of the data such as the contents, geographic coverage, attributes, accuracy or error remain the major considerations for these participants when it comes to selecting spatial data. Only a few participants considered previous uses (by other users) for the selection. This might indicate that the knowledgeable and expert users would select the spatial data based on their own assessment of the data, rather than on others' experiences.

Table 5.14: Information required by participants for spatial data selection

Novice	Knowledgeable		Expert	
✓ Feature/Theme	✓ Feature/Theme	16	✓ Feature/Theme	25
	✓ Geographic coverage/extent	16	✓ Feature attribute description	25
	✓ Last update (maintenance)	14	✓ Geographic coverage/extent	25
	✓ Data format	14	✓ Resolution/pixel size	25
	✓ Production Date	13	✓ Data format	24
	✓ Spatial scale	13	✓ Spatial scale	23
	✓ Resolution/pixel size	12	✓ Data type	22
	✓ Data producer	12	✓ Use restriction	21
	✓ Use restriction	12	✓ Positional accuracy/error	20
	✓ Data type	11	✓ Price	20
	✓ Price	11	✓ Production Date	17
	✓ Owner contacts	11	✓ Last update (maintenance)	16
	✓ Feature attribute	10	✓ Attribute accuracy/error	16
	✓ Positional accuracy/error	9	✓ Owner contacts	15
	✓ Topology	8	✓ Level of detail/generalization	14
	✓ Production process/history	7	✓ Data producer	13
	✓ Intended use	7	✓ Topology	12
	✓ Level of detail/ generalization	6	✓ Production process/history	12
	✓ Attribute accuracy/error	5	✓ Intended use	11
	✓ Data sample	5	✓ Data sample	11
	✓ Previous uses	3	✓ Previous uses	5

The knowledgeable and expert user participants also required additional information for spatial data selection, which is mainly related to their preference for assessing the data themselves. This additional information includes, for example, access to relevant data, or samples and Web-Map services (WMS) as the data preview. Table 5.15 shows the additional requirements indicated by the participants.

Table 5.15: Additional information required for spatial data selection

Additional information for data selection
✓ Link to relevant data (example: link to the single frame name or ID of images of mosaicked orthoimage) ✓ Graphical preview of the data ✓ Data sample containing complete attributes and domains ✓ Graphical representation of spatial extent or footprint ✓ Licensing (CC, etc.), for commercial use, etc. - perhaps covered by Use Restriction. ✓ A graphical view of the data in a web map ✓ Technical constraints ✓ Frequency of update ✓ Variable/attribute information ✓ Copyright or compliance use of the data ✓ Data assurance in graphical ✓ Future data updates (as planned) ✓ Utilisation and usefulness of the data

The additional requirements for spatial data selection are similar to the additional requirements for data discovery. This suggests that the participants prefer to get the information as early as possible and would assess the data by looking at the sample or preview after they had found the potential data from the titles and abstract.

The results also identify participants' preferred style for the user interface, ie. e-commerce or online shopping style for the portal, and Web-Map service browser for the data preview. One participant from the regular user group specifically mentioned that they were upset with the functionality of the portal they used for searching and selecting spatial data, and indicated that it would be very helpful if the portal could have a window showing relevant data or previous search information, so they would not have to do another search or go back to the result page every time. User rating and user reviews, as identified in the additional information for data discovery, suggest that this response is indicative of a more general user experience. The expectation put forward by this participant is also identified in the metadata usability evaluation explained in Chapter 4.

The above-mentioned information required by participants for data discovery and selection, including the preferred presentation style, are the basis for developing the user-oriented spatial metadata profile, as explained in the following sections.

5.4. User-oriented spatial metadata profile development

Having a list of information required by spatial data users, including the preferred presentation style, for spatial data discovery and selection provides the information required to redesign and redevelop spatial metadata and user interface for spatial data users. However, as the spatial metadata have been created following a standard such as ISO 19115, there is a requirement for developing the user requirements in accordance to the standard. Similarly, it is important to investigate how, or to what extent, the metadata standard meets the user requirements and expectations identified by the results of the questionnaire. This process can be started by mapping the user requirements against the standard.

5.4.1. Mapping the metadata standard with the user requirements and expectations

The latest version of the ISO standard for geospatial metadata is the ISO 19115-1:2014, *Geographic information - Metadata*, Part 1: Fundamentals. According to the standard, a full metadata is an aggregate of 12 metadata classes as illustrated in Figure 5.2, metadata schema.

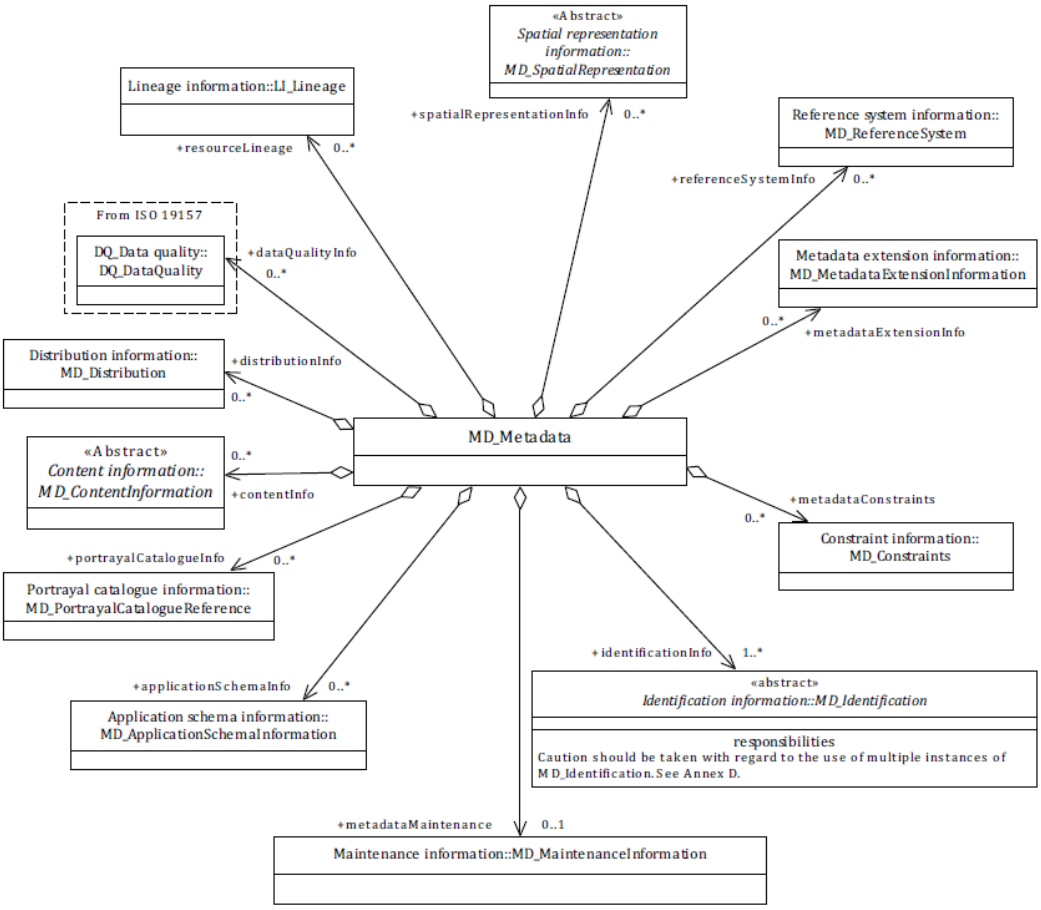


Figure 5.2: Metadata schema

Within the classes, there are metadata elements that contain information related to spatial data characteristics. This information can be used to answer the what, when, where, who, why and how of the spatial data or resources. Table 5.16 shows the user requirements and the relevant metadata elements from ISO 19115-1:2014.

Table 5.16: Map of user requirements with ISO 19115-1:2014

User requirements	ISO 19115-1:2014 Elements	O/M *)
Title	MD_DataIdentification>CI_Citation.title	M/1
Abstract	MD_DataIdentification.abstract	M/1
Content/Theme	MD_DataIndetification>MD_TopicCategoryCode	C/N
Geographic Coverage	MD_DataIdentification>EX_Extent> EX_GeographicExtent. geographicBoundingBox	C/N
Geographic Name	MD_DataIdentification>EX_Extent> EX_GeographicExtent.geographicDescription	C/N

Data format (spatial rep)	MD_SpatialRepresentation> MD_SpatialRepresentationTypeCode	O/N
Data type	MD_Metadata>MD_MetadataScope. resourceScope>MD_ScopeCode	C/1
Scale	MD_DataIdentification.spatialResolution> MD_Resolution.equivalentScale	O/N
Resolution	MD_DataIdentification.spatialResolution> MD_Resolution.distance	O/N
Positional accuracy	<<ISO19157>>DQ_Result>DQ_QuantitativeResult	M/N
	<<ISO19157>>DQ_Result>DQ_DescriptivetiveResult.statement	M/1
Attribute accuracy	<<ISO19157>>DQ_Result>DQ_QuantitativeResult	M/N
	<<ISO19157>>DQ_Result>DQ_DescriptivetiveResult.statement	M/1
Level of detail	MD_Resolution.levelOfDetail	C/1
Data Assurance	<<ISO19157>>DQ_Result>DQ_ConformanceResult.explanation	O/1
Production date	MD_DataIdentification>CI_Citation>CI_Date> DateTypeCode	O/N
Last update	MD_DataIdentification>MD_MaintenanceInformation. maintenanceDate>DateTypeCode	O/N
Maintenance schedule	MD_DataIdentification>MD_MaintenanceInformation. maintenanceAndUpdateFrequency>MD_MaintenaceFrequencyCode	C/1
Producer	MD_DataIdentification>CI_Citation>CI_Responsibility. role>CI_RoleCode	O/N
Intended use	MD_DataIdentification.purpose	O/1
User rating	**No suitable element**	
Number of download (use)	**No suitable element**	
Previous uses	MD_DataIdentification>MD_Usage.specificUsage	M/1
User reviews*	MD_DataIdentification>MD_Usage.identifiedIssues	O/N
	MD_DataIdentification>MD_Usage.spesificUsage	M/1
Utilisation/Usefulness*	MD_DataIdentification>MD_Usage.identifiedIssues	O/N
	MD_DataIdentification>MD_Usage.spesificUsage	M/1
	<<ISO19157>>DQ_Result>DQ_ConformanceResult.explanation	O/1
Thumbnail	MD_DataIdentification>MD_BrowseGraphic.fileName	M/1
	MD_DataIdentification>MD_BrowseGraphic.linkage	O/N
Keywords	MD_DataIdentification>CI_Citation.otherCitationDetails	O/N
	MD_DataIdentification.descriptiveKeywords>MD_Keywords	O/N
Data sample	MD_DataIdentification>CI_Citation. onlineResource	O/N
Data Viewer	MD_DataIdentification>CI_Citation. onlineResource	O/N
Attribute	MD_CoverageDescription>MD_AttributeGroup.contentType	M/N

	MD_Metadata>MD_ContentInformation>MD_Coverage Description	M/1
License (Legal restrictions)	MD_DataIdentification>MD_Constraints.graphic	O/N
	MD_DataIdentification>MD_LegalConstraints	O/N
	MD_DataIdentification>MD_SecurityConstraints	O/N
Use limitation (Technical)	MD_DataIdentification>MD_Constraints.useLimitation	O/N
	MD_DataIdentification>MD_Usage.userDetereminedLimitations	O/N
Price	MD_Distributor>MD_StandardOrderProcess.fees	O/1
Production process	MD_Metadata>LI_Lineage	O/N
Topology	MD_SpatialRepresentation>MD_VectorSpatialRepresentation.topologyLevel>MD_TopologyLevelCode	O/1
Owner/Distributor Contact	MD_Distribution>MD_Distributor.distributorContact	M/1
	MD_DataIdentification>CI_Citation.citedResponsibleParty	O/N

O/M *) Obligation/Maximum Occurrence

* Potentially relevant to the corresponding ISO element

- No corresponding or relevant ISO element

As can be seen from Table 5.16, there are two types of information from the user requirement list that the standard does not currently have elements for accommodating: User data rating and Number of downloads. Two other requirements, User reviews and Data usefulness, do not have exact corresponding elements, but there are established elements that are potentially suitable for them. However, the standard has been prepared to accommodate information related to data use/utilisation by providing MD_Usage sub classes under MD_Identification class. The profile can propose new elements for Rating and Number of downloads under this sub class:

MD_DataIdentification>MD_Usage.numberOfUsage – M/1

MD_DataIdentification>MD_Usage.userDefinedRating>MD_UserRatingCode – M/1

Other requirements frequently mentioned by participants in Chapter 4, and appearing in the survey results, are data sample and data preview. These requirements reflect participants' capacity to make their own assessment of the data suitability by directly exploring and reviewing the data, instead of just relying on the metadata. Both requirements can be accommodated by providing corresponding links to download or access the sample or browse the data WebMap service in the OnlineResource element under the CI_Citation sub class.

Apart from the unavailable elements for the abovementioned user-related information, most of the requirements are accommodated by the standard. It is interesting that the results of usability evaluation of the established metadata reveal missing information, inconsistency and irrelevance problems in the metadata records such as keywords, title and abstract. Missing information is the absence of crucial information in the metadata presented to users. It can be the result of different obligation types of the element in the standard, or of non-existent information. When optional, obligation assigned to an element means that the author may opt to provide the information or not to the element. Assigning the mandatory type for user-required elements can encourage authors or responsible parties for a specific data to make the information available in the first place. The non-existent information might occur due to the implementation of the minimum/core metadata in the standard. The core/minimum metadata is defined as a set of elements that can be used for both metadata management and data discovery. It is designed to help the data producers create metadata for their data inventory. However, the information/element in the minimum metadata may not necessarily suffice for data users to discover and select the data.

The irrelevance and inconsistency problems identified in the evaluation are mainly found in titles and abstracts. Since they are critical information for assessing the spatial data, special efforts were provided for addressing the problems. The potential source of the problems is the fact that both titles and abstracts are assigned with free text domain. Although the standard assigns a definition for title, which is “the name by which the cited resource is known”, it is still too vague to ensure that the author will necessarily provide the relevant information. The same reason applies to abstracts, as the standard defines it as “a brief narrative summary of the resource”. The problems can be addressed by assigning strict rules or guidelines in the standard for authors or machines to abide by when creating titles and abstracts, and other elements when possible. These rules or guidelines will dictate what information should be presented in titles and abstracts, and in what order. It will also dictate the number of maximum words allowed in both titles and abstracts.

ANZLIC, the Spatial Information Council of Australia and New Zealand, developed their metadata profile together with the guidelines for organisations to create their metadata, including detailed explanations for every element. However, the guidelines are vague, and do not prevent the inconsistency problem in the metadata records, as they fail to give clear instructions to maintain the consistency and clarity of the contents, e.g. in title and abstract, as can be seen in Table 5.17 and 5.18.

Table 5.17: Abstract description and guidance in ANZLIC Metadata Profile Guideline Version 1.2 (ANZLIC, 2007)

Name	Abstract
Definition	Brief narrative summary of the content of the resource
Data type	CharacterString
Domain	Free text
Meaning & Purpose	The identification abstract provides additional information about the resource. This may allow users to obtain a better appreciation of the resource and assist them to determine fitness for purpose
Guidance	The abstract should provide sufficient information, such as keywords, to adequately describe the content of the resource. Careful consideration should be given when preparing an abstract as it is an important element for the assessment of a resource

Table 5.18: Title description and guidance in ANZLIC Metadata Profile Guideline Version 1.2 (ANZLIC, 2007)

Name	Title
Definition	Name by which the cited resource is known
Data type	CharacterString
Domain	Free text
Meaning&Purpose	The resource title is the official name for the resource. Where no normal name for the resource, a useful name for the resource should be assigned.
Guidance	If the resource is a text document, use the full title as it appears on the title page; otherwise use a meaningful, plain language phrase for that resource (i.e. do not use the file name) The title naming conventions should be consistently used for related resources (e.g. to facilitate discovery). To discriminate between duplicate titles, a reference to the version should be included in the title.

	For identification purposes, it is important to carefully complete this element. Other users should easily understand the title. If the resource is known by an alternate title, include this in the alternatetitle element.

The rules for maintaining the clarity and consistency of titles and metadata can be inducted in the metadata standards by assigning a special domain for them, or by keeping the free text domain and including the required information in the definition. This domain would be an aggregate of contents of selected metadata elements including their order, as can be seen in Table 5.19.

Table 5.19: Proposed information in domain/definition for title and abstract in user-oriented profile

Element	Data Type	Domain or Definition
Title	CharacterString	<<Content>><<GeographicName>><<Scale/Resolution>> <<Production.Date>><<LastUpdate>> <<SpatialRepresentation>>
Abstract	CharacterString	<<Content>><<GeographicName>><<AttributeDescription>> <<CoverageDescription>><<LevelOfDetail>> <<PositionalandThematicAccuracy>><<Lineage>> <<IntendedUse>><<Maintenance>>

The element mapping results explained above reveal that the ISO 19115-1:2014, as the latest version of the standard for spatial metadata, accommodates most of the user required information, except for user-defined information, e.g. user data rating and user reviews. New elements should be introduced to the profile, to enable the information to be entered into metadata records. The results also reveal that the standard is able to maintain the consistency of the metadata record structure, but that it fails to maintain the consistency and clarity of the information provided in key elements of metadata, i.e. title and abstract, as the effect of free text domain with general definition.

The user-oriented spatial metadata profile is expected to fill the gaps in the ISO 19115-1:2014 by extending the standard, following the extension guidance provided in the standard, as explained in the next section.

5.4.2. Extension of ISO 19115:2014 Discovery metadata for geospatial resources

The user-oriented spatial metadata profile is expected to provide the rules and guidelines to redesign and redevelop metadata records and user interface. The aim of this redesign and redevelopment is to improve the usability of metadata for spatial data discovery and selection.

The profile is essentially an extension of the ISO 19115-1:2014 metadata standard, to incorporate information required by users to discover and select spatial metadata into the standard. The term profile in this thesis does not refer to community profile as described in the ISO 19115-1:2014 document. It is specifically used in this thesis for the extension of the discovery metadata for geographic resources, as in Annex F in the standard document. Therefore, it follows rules for creating extension as given in Annex C in the standard. According to these rules, an extension is allowed if the standard does not accommodate a specific requirement, and one or more of the following extension types shall be allowed:

1. adding a new metadata package;
2. creating a new metadata codelist to replace free text domain of an existing element;
3. creating or expanding a codelist;
4. adding a new metadata element;
5. adding a new metadata class;
6. imposing a more stringent obligation on an existing metadata element; and/or
7. imposing a more restrictive domain on an existing metadata element.

The profile adds new metadata elements, imposes a more stringent obligation to some existing elements, and imposes a more restrictive domain on some existing metadata elements.

Therefore, the following rules for creating the extension are applied:

1. Name, definition or data type of an existing element shall not be changed.
2. Stringent obligation for existing metadata elements is permitted.
3. Restricting the use of domain values from other metadata elements is permitted.

Following the rules, the option for changing the definition of title and abstract is not allowed, and rule number 3 for restricting of the use of domain values from other metadata elements is used instead. Table 5.20 illustrates the profile as an extension of the discovery metadata for geospatial resources of ISO 19115-1:2014.

Table 5.20: User-oriented spatial metadata profile

Metadata element	**Obl.**	**Occ.**	**Description**
Metadata reference*	O	1	Unique identifier for the metadata
Resource identifier*	O	N	Unique identifier for the resource
Resource language*	C	1	The language and character set used in the resource (if a language is used)
Metadata date stamp*	M	N	Reference date(s) for the metadata, especially creation
Resource title*	M	1	Title by which the resource is known
Resource abstract*	M	1	A brief description of the content of the resource
Resource topic category*	M	N	A selection of the 20 elements in the MD_TopicCategory enumeration which describe the topic of the resource
Geographic extent*	M	1	Spatial area of the resource
Geographic location*	M	1	Geographic description or coordinates (latitude/longitude) which describes the location of the resource
Resource spatial format	M	1	A selection of the 5 elements in the MD_SpatialRepresentationType enumeration which describe the method used to represent geographic information in the resource
Resource type*	O	1	A resource code identifying the type of resource e.g. dataset, a collection, an application (see MD_ScopeCode) which the metadata describes
Spatial resolution (equivalentScale/distance)*	M	N	The nominal scale and/or spatial resolution of the resource
Positional accuracy	M	N	Accuracy of the position of features
Attribute accuracy	M	N	Accuracy of quantitative attributes and correctness of non-quantitative attributes and of the classifications of features and their relationships
Level of detail	C	1	Brief textual description of the spatial resolution of the resource

Data Assurance	O	1	Information about the outcome of evaluating the obtained value (or set of process stages) against a specified acceptable conformance level
Reference date (production)*	M	1	A date which is used to help identify the resource (is existed)
Reference date (update)	M	1	A date which is used to help identify the resource (is last updated)
Reference date (next update)	M	1	A date which is used to help identify the resource (will be updated)
Producer	M	N	The responsible organisation for creating the resource
Intended use	O	1	Summary of the intentions with which the resource was developed
User rating	M	N	User vote or rate to the resource based on the conformance of the resource against a set of criteria set by users
Number of download (use)	M	1	The number of occurrences which the data is used
Previous uses	M	N	Brief description of the resource and/or resource series usage (the project)
User reviews	M	N	Brief description of the resource and/or resource series usage (reviews against user specified criteria)
Utilisation/Usefulness	O	N	Textual expression of the descriptive results of conformance against a set of user requirements
Thumbnail	M	1	Graphical presentation of an area covered by the data or an area where the data are not present
Keywords*	M	N	Words or phrases describing the resource to be indexed and searched
Resource on-line Link (Data sample/Viewer)*	M	N	Link (URL) in the metadata for the resource (Data sample or Webmap Service)
Attribute	M	N	Description of the attribute described by the measurement value
Constraints on access and use*	M	N	Restrictions on the access and use of the resource
Use limitation	O	N	Applications, determined by the users for which the resource and/or resource series is not suitable
Price	O	1	Fees and terms for retrieving the resource
Lineage*	O	N	A description of the resource(s) and production process(es) used in the producing the resource
Topology	O	1	Degree of complexity of the spatial relationships
Resource point of contact*	M	1	Name of the person, position, or organisation responsible for the resource
Metadata point of contact*	M	N	The party responsible for the metadata

*Metadata elements from the original discovery metadata for geospatial resource

As can be seen in the profile, several elements of metadata in the discovery metadata are changed from optional (O) into mandatory (M) following user requirements, e.g. online link to resources for accessing data sample and data viewer (Web-Map service). Some elements from the standard are brought into the discovery metadata, e.g. intended use, previous uses, etc. Two new elements are included for user data rating and number of downloads, as well as an element from the standard that could be used for user reviews.

The profile is accompanied by a set of presentation styles and functionalities for presenting the metadata and for developing user interface functionalities. E-commerce or online shopping style, criteria (filter) based search style, and Web-Map service for data preview are the most frequent presentation styles mentioned by the participants. By implementing these presentation styles into the design and development of metadata records and user interface, it is expected that the usability of spatial metadata for data discovery and selection will be greatly improved.

5.5. Chapter summary

This chapter identified user requirements and preferred presentation style for spatial metadata, based on the results of an internationally circulated survey targeting spatial data users with a range of spatial knowledge and relevant expertise levels, and from a range of professional groups.

The results show that the current and mostly used metadata standard, the ISO 19115-1:2014, provides metadata elements to accommodate most of the user required information. However, the standard lacks elements related to user specified information, i.e. elements for user data rating and number of downloads (uses), and an extension is required to accommodate the user requirements for spatial data discovery and selection.

Two new metadata elements are proposed to be added as part of the extension of the standard. The extension is also made to accommodate other required information by changing the obligation type of metadata elements and bringing them into the discovery metadata, to be used as the user-oriented spatial metadata profile. Specific attentions are given to title and abstract elements, where the domains are changed from free text into multiple values from other specified elements, as listed in the title and abstract contents. Finally, the new elements, more strict obligation for some elements and more strict domains for title and abstract, along with other required information for spatial data discovery and selection, are inducted into discovery metadata for geospatial resources in the ISO 19115-1:2014 to develop the user-oriented spatial metadata profile.

The next chapter is dedicated to explaining the development of metadata and user interface prototype based on the user-oriented metadata profile, and to reporting the usability evaluation of the prototype as part of the evaluation stage.

This page is left intentionally blank

6. Validation through prototyping

6.1. Introduction

Having established the user-oriented spatial metadata discovery profile, the final stage of this research aims to achieve the last objective: to test and evaluate the profile's capacity to improve metadata usability. This will answer the last research question regarding the users' acceptance of the newly proposed spatial metadata and the user interface. This chapter reports on the process and discusses the results and findings of the evaluation.

The chapter starts with a brief review of the validation process. Detailed developments of metadata prototypes are provided and explained in section 6.3. Section 6.4 discusses the results and findings, where, to evaluate the users' acceptance of the prototype, the usability of and reactions to the spatial metadata and the user-interface prototype are analysed and interpreted.

The last section discusses the results in accordance to the user-oriented spatial metadata profile, which is a proposed solution for improving metadata usability for spatial data discovery and selection. This is then followed by a summary of the chapter.

6.2. Validation Process

As mentioned in Chapter 3, prototyping and usability evaluation are selected as methods for the validation stage. A prototype is required to implement the profile into functional spatial metadata and to deliver the metadata to users in a technologically familiar environment. Usability evaluation is required to measure the metadata's usability in helping users in spatial data discovery and selection. It is also required to see users' reactions to and their acceptance of the proposed metadata. The validation process was implemented following the UCD approach, as illustrated in Figure 6.1.

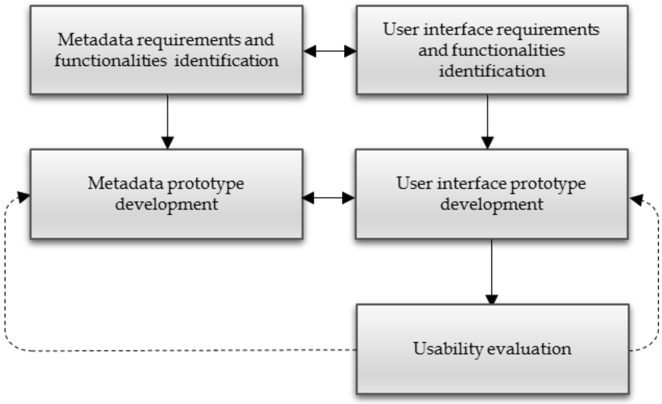


Figure 6.1: Validation process following the UCD approach

The validation process begins with identifying the relationships between the spatial metadata elements from the profile and the user-interface functionalities from Chapter 5. The purpose is to ensure all functionalities are supported by their respective metadata elements and vice versa.

The next step is the prototype development. Here, functional spatial metadata is created by providing the required information for metadata elements of the selected established spatial data, followed by putting them into a dedicated database. The user-interface prototype is developed by designing a spatial data catalogue web page comprising identified functionalities and interfaces to support spatial data discovery and selection. The design is then implemented into a physical web page using selected web technologies (e.g. Apache web server, PHP MySQL database query, ESRI web-map server).

The author makes an initial internal evaluation to ensure all functionalities are covered and functional, e.g. tabs, buttons, windows, etc., and then uses the reviews to refine the system. Once the prototype is ready, the final step of this validation stage is the evaluation of the prototype.

The evaluation was conducted using Think-Aloud Protocol (TAP) and by interviewing a number of selected spatial data users. These users used the web catalogue prototype to discover and pick specific spatial data for be used in a project featuring a simulated real-life scenario.

The following sections describe in detail each step of the validation process: the prototype development, the usability evaluation design, the results, and the interpretations.

6.2. Prototype Development

The prototype is required as a practical implementation of the user-oriented spatial metadata profile so the profile can be used by spatial data users for data discovery and selection. By doing so, the usability of the profile can be evaluated. The prototype development was conducted following a process used for developing real products, a system development framework adapted from Nunamaker Jr et al. (1990), as explained in Chapter 3, The Methodology, and illustrated in Figure 6.2.

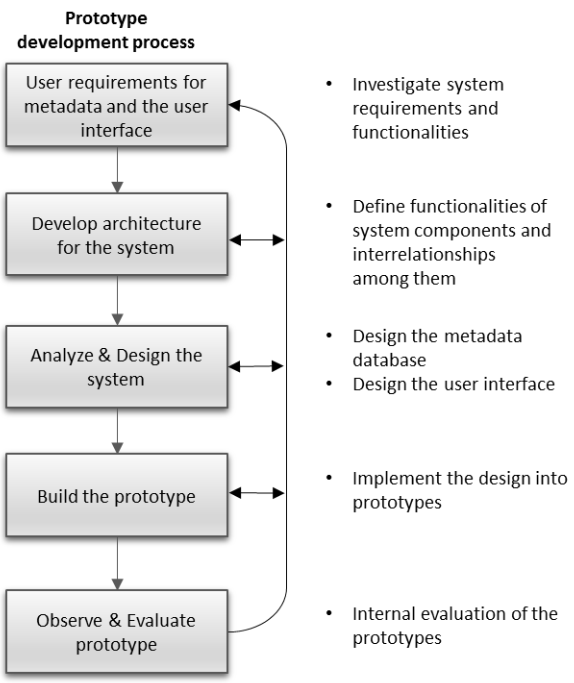


Figure 6.2: Prototype development process (Adapted from Nunamaker Jr et al. 1990)

The process started with identifying the prototype's system functionalities from the profile to be implemented into functional metadata records and the user interface, which was then delivered and displayed to spatial data users in the internet environment. Beside the profile, user expectations and experience, as described in Chapter 5, were considered to help identify the functionalities and presentation design of the interface, e.g. as found in an e-commerce or online shopping design. The result is a set of user-interface functionalities and the relationships between them to the metadata elements, as illustrated in Figure 6.3.

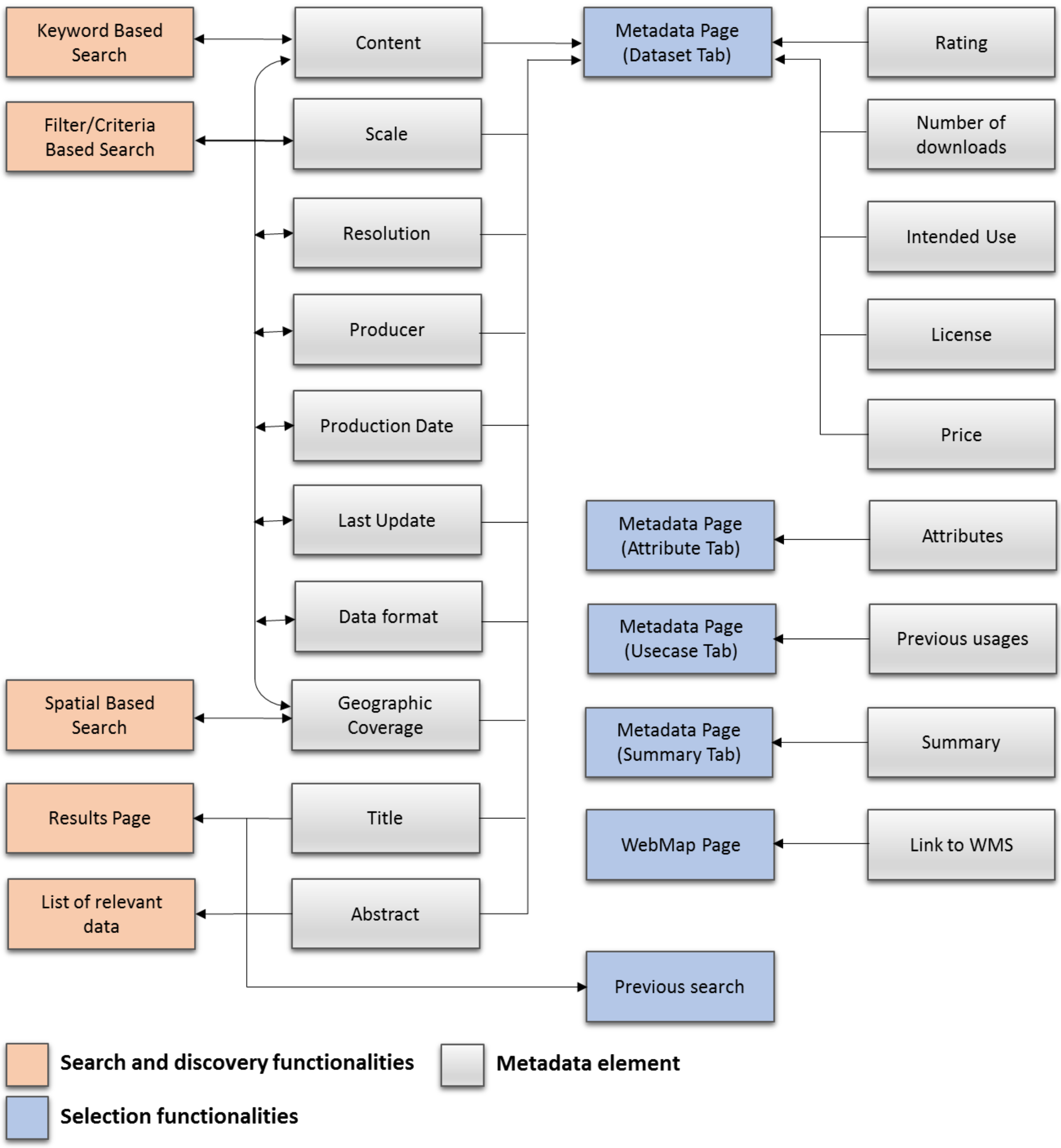


Figure 6.3: The relationship of metadata elements and user-interface functionalities

Two functionalities of the prototype are divided into two groups: search and discovery functionalities and selection functionalities. The search and discovery functionalities are designed to help users find and discover potential spatial data for their projects. The selection functionalities are designed for users to select data based on its suitability to their projects. Each functionality is related to one or more metadata elements and vice versa.

The user interface was developed to accommodate the user-defined functionalities, and followed the general web-page development criteria proposed by ISO 9241 (ISO, 2018):

- *Clarity*. The information content is conveyed quickly and accurately.
- *Discriminability*. The displayed information can be distinguished accurately.
- *Conciseness*. Users are not overloaded with extraneous information. The user interface should allow users to easily find information by not presenting too much information on a single page. Tabs were used to display different information.
- *Consistency*. A unique design that conforms to a user's expectation. The user-interface design should utilise the convention in the web design, especially in e-commerce websites.
- *Detectability*. The user's attention is directed towards the information required.
- *Legibility*. Information is easy to read.
- *Comprehensibility*. The meaning is clearly understandable, unambiguous, interpretable, and recognisable. The user interface should avoid unnecessary information or presentation.
- *Affordances*. The user interface should inform users on where to click or how to perform the right actions to achieve their goals.
- *Minimal Click*. The user interface should allow users to find the required information and to perform their actions using minimum clicks.

The next step is to prepare the system architecture and requirements, hardware and software that will be used to build the spatial metadata database and the website. Figure 6.4. illustrates the system architecture and requirements for the prototypes.

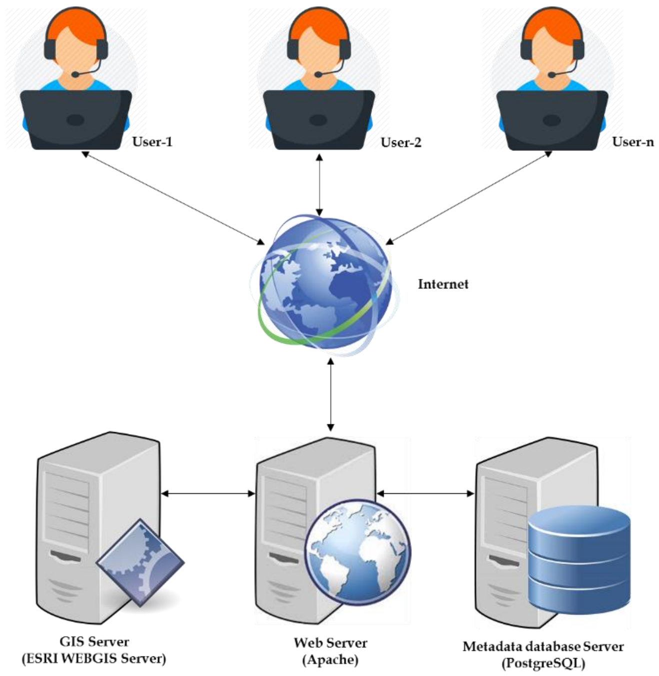


Figure 6.4: The system architecture for the prototypes

The system consists of three parts: (1) web server, (2) metadata database server and (3) GIS server. The web server is designed to store, process and deliver web pages to users using the Hypertext Transfer Protocol (HTTP). The metadata database server is designed for storing and managing spatial metadata records, as well as performing task analyses given by users. The GIS server is designed to process a user request relating to a specific feature, or to map and deliver a web map as per a request.

PostgreSQL is used as the metadata database system. As for the website, Apache is used as the web server with PHP database scripts, and JavaScript is used to create and develop the web pages as well as its functionalities. There are no specific criteria for the selected applications, except they must have the capabilities to create a website that can search and select spatial data from spatial metadata databases, and which can deliver and visualise all the information, including web-map services, from the metadata.

Having defined the system architecture and tools for the metadata database and website development, the next step is the prototype development, both for the metadata and the website (user interface). Spatial data available from the Indonesian Spatial Data Infrastructure was selected for this purpose. The selection was made based on two criteria: available access to spatial data and its services, and a variety of themes. Indonesia was among the first nations that declared their NSDI initiative for spatial data sharing (Masser, 2005). They have set up their National Spatial Data Infrastructure (IDSN) connecting clearing houses from their ministries and local governments that can be accessed by public via its geoportal (https://tanahair.indonesia.go.id). Most of their local governments have also built and utilised SDI for supporting their decision-making process, such as Cianjur Municipal (Kharisma & Sagala, 2018). However, they have been facing difficulties with the improvement of the IDSN due to several problems. One of them is related to metadata development, as can be viewed in the portal. Currently, the Indonesian National Geospatial Agency (BIG) is developing a spatial metadata country profile based on ISO 19115-1:2014 that will be implemented as the national standard (PSKIG-BIG, 2018). This research will be very useful for such initiatives by incorporating the results into the profile development, or even by adopting the functionalities of the prototype.

The prototype itself were implemented as a website that can be used by selected spatial data users during the evaluation stage. Prior to the usability evaluation, an internal assessment made by the author was conducted to ensure that all functionalities are covered and functional, e.g. tabs, buttons, windows, etc.

The website consists of four main pages: the search page (as the homepage); the search-result page covering the data search and discovery functionalities; the dataset metadata page (with tabs); and the web-map service page covering data selection functionalities, as described in the next sections.

Search page

The search page is also the home page of the website, consisting of three different search tools: (1) keyword-based; (2) criteria (filter)-based; and (3) spatial or location-based; as illustrated in Figure 6.5. Users might use their own searching style based on their knowledge and experiences. The keyword-based search is a simple search tool that users can use to type in specific words relevant to the spatial data they are interested in. The criteria-based search provides a set of criteria users can use to narrow down their search. The location-based search allows users to define their own area of interest by drawing a boundary polygon on the map window before submitting their search.

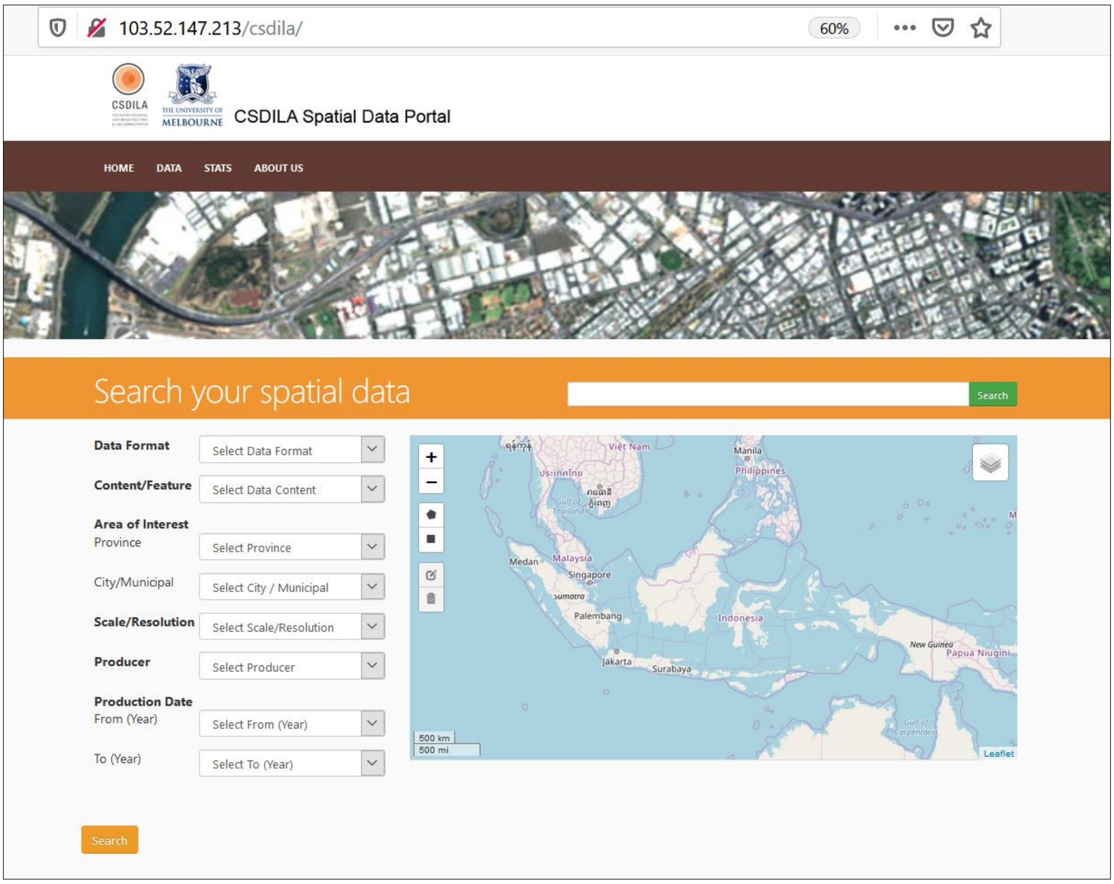


Figure 6.5: Search page (home page)

Search-result page

The search-result page consists of the main window and the relevant data window, as illustrated in Figure 6.6. The main window provides a list of data titles and abstracts responding to their submitted search. The relevant data window presents a list of data titles and abstracts that might relate to the submitted search.

Figure 6.6: Results page

Dataset metadata page

The Dataset tab is the default metadata page presented to users when they click one of the titles on the results page. As illustrated in Figure 6.7, it consists of the main metadata window presenting data characteristics such as the following user-related information: a data rating, the number of downloads, geographic coverage in a graphical form (thumbnail), and the previous-search window providing users results (a list of data titles and abstracts) of their previous (current) searches.

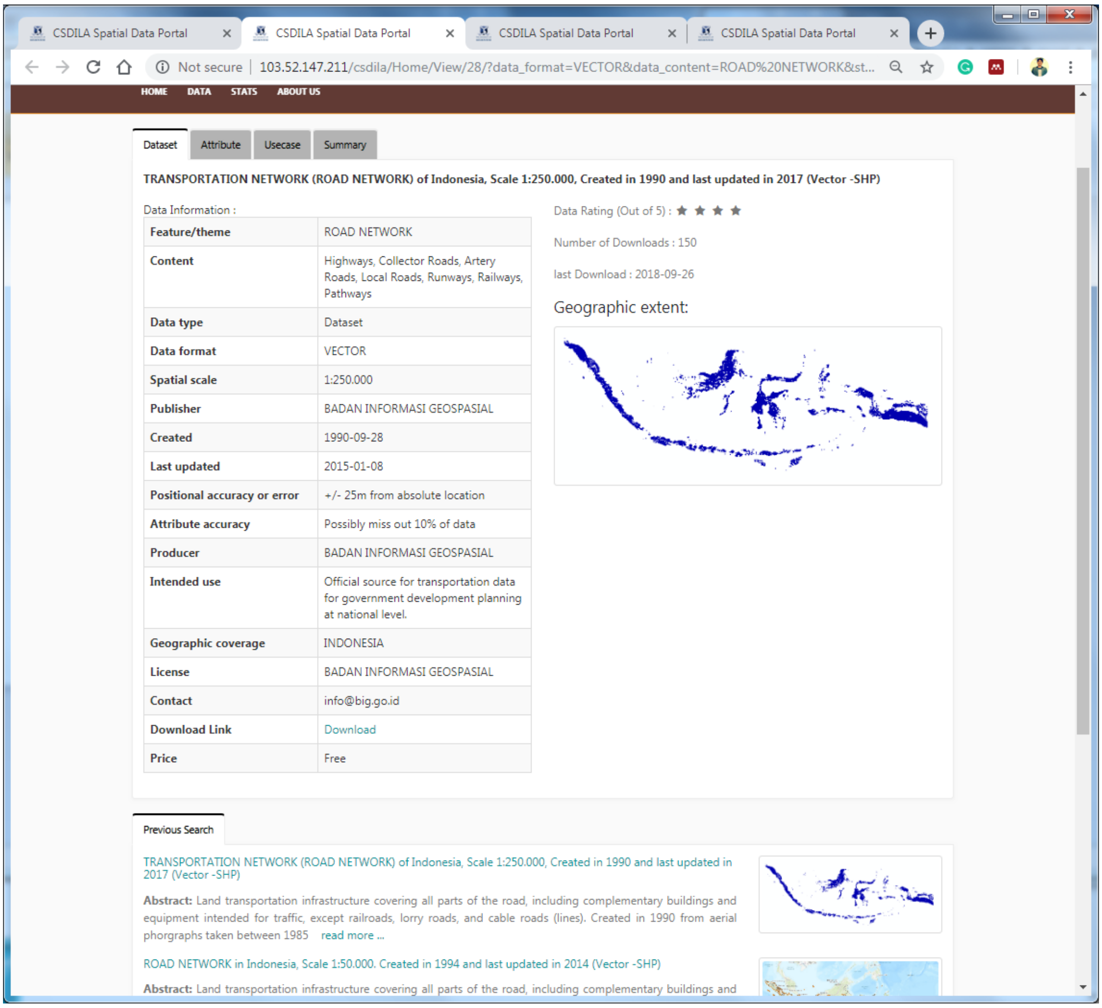


Figure 6.7: Metadata page (Dataset tab)

Attribute metadata page

The next tab provided as part of the metadata page is the Attribute tab. It contains information about the spatial data attributes in a tabular format, which might assist users in getting to know the contents and structure of the data. A previous-search window is also provided below the main attribute window, similar to the Dataset tab, as illustrated in Figure 6.8.

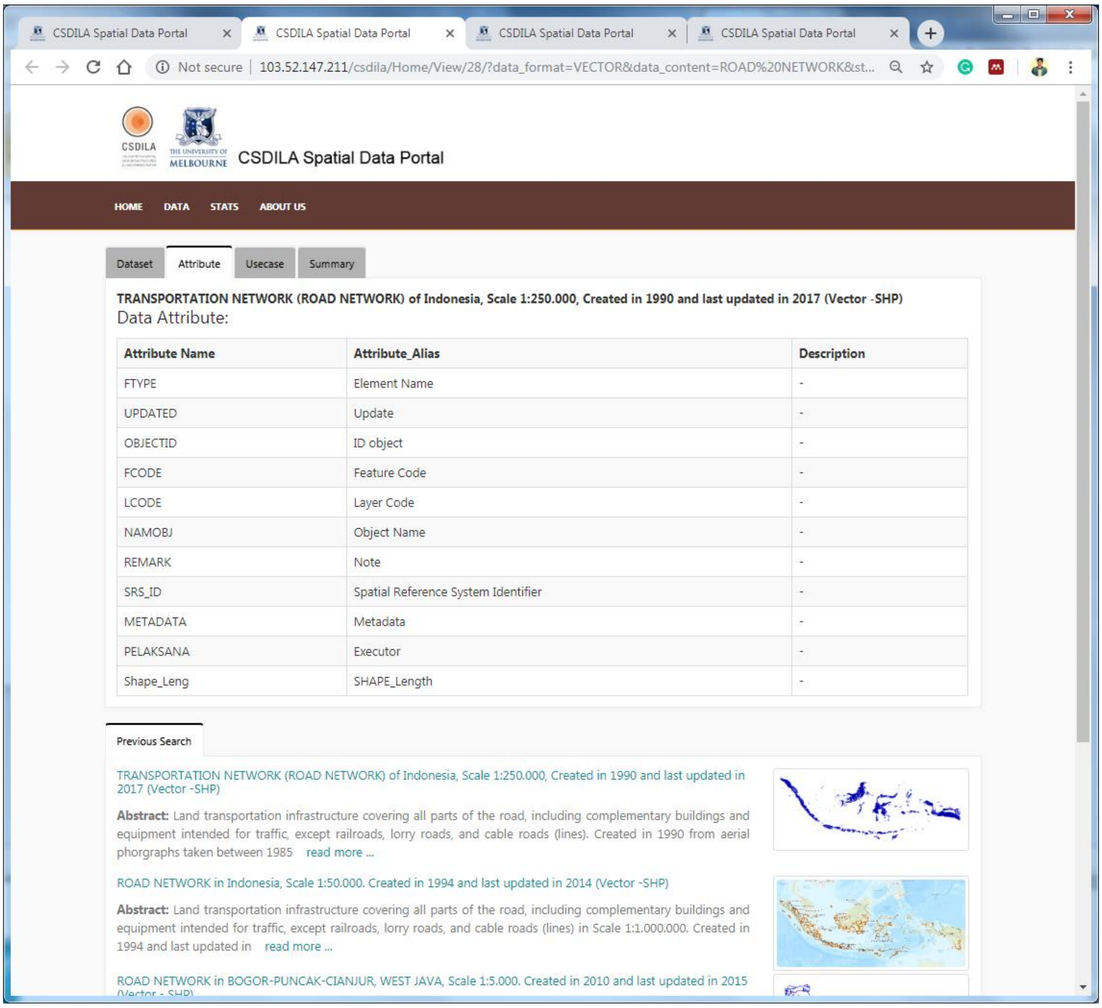


Figure 6.8: Metadata page (Attribute tab)

Usecase metadata page

Figure 6.9 illustrates the Usecase metadata page, a tab that provides users information about the previous usages of the data. In this tab, users are allowed to give their own experiences with the data by submitting reviews into a form. The previous-searches window is also provided on this page.

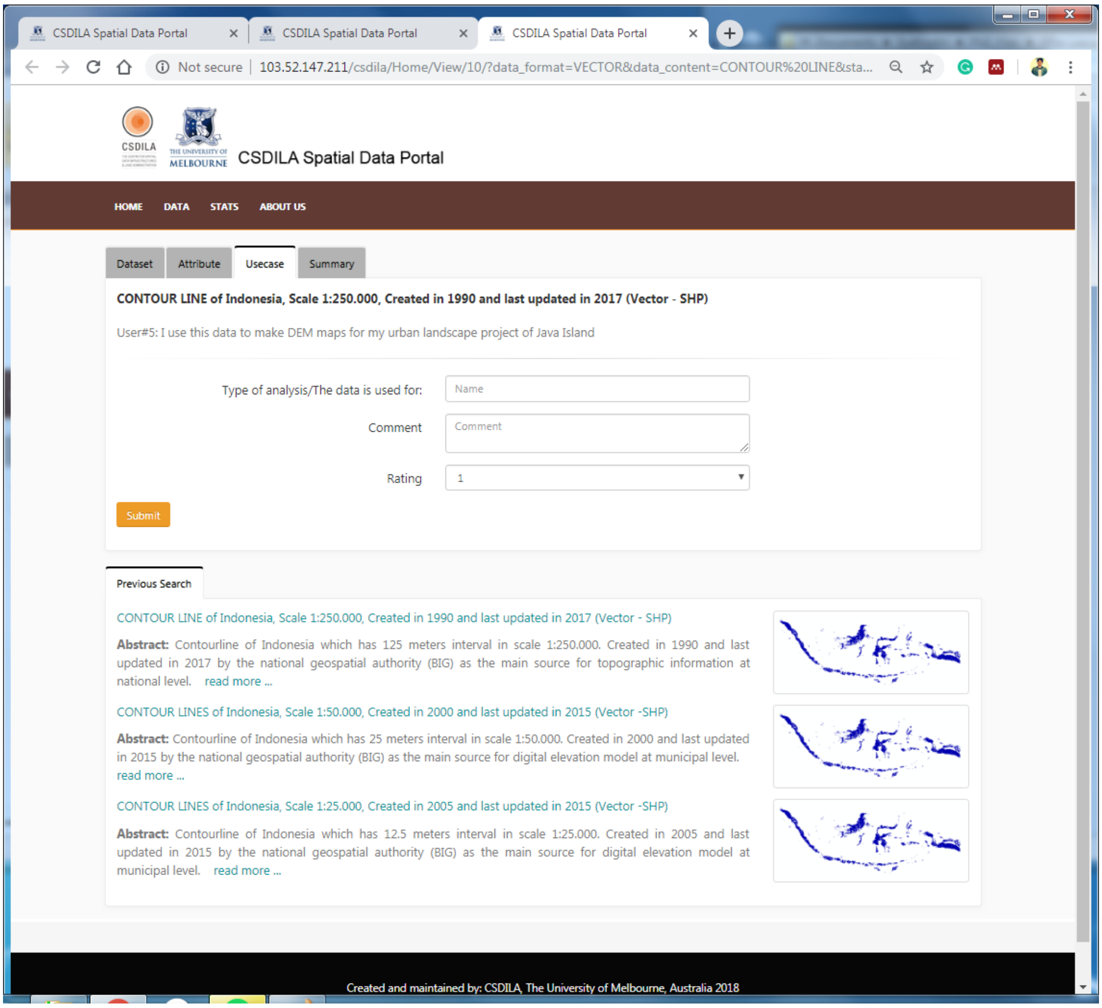


Figure 6.9: Metadata page (Usecase tab)

Summary metadata page

The last tab provided on the metadata page is the Summary tab. It gives users a brief summary of the spatial data, mainly about the production process, but which also includes the data's sources, production method or technology, and the organisations conducting the process, as illustrated in Figure 6.10. The previous-search window is also presented on this page.

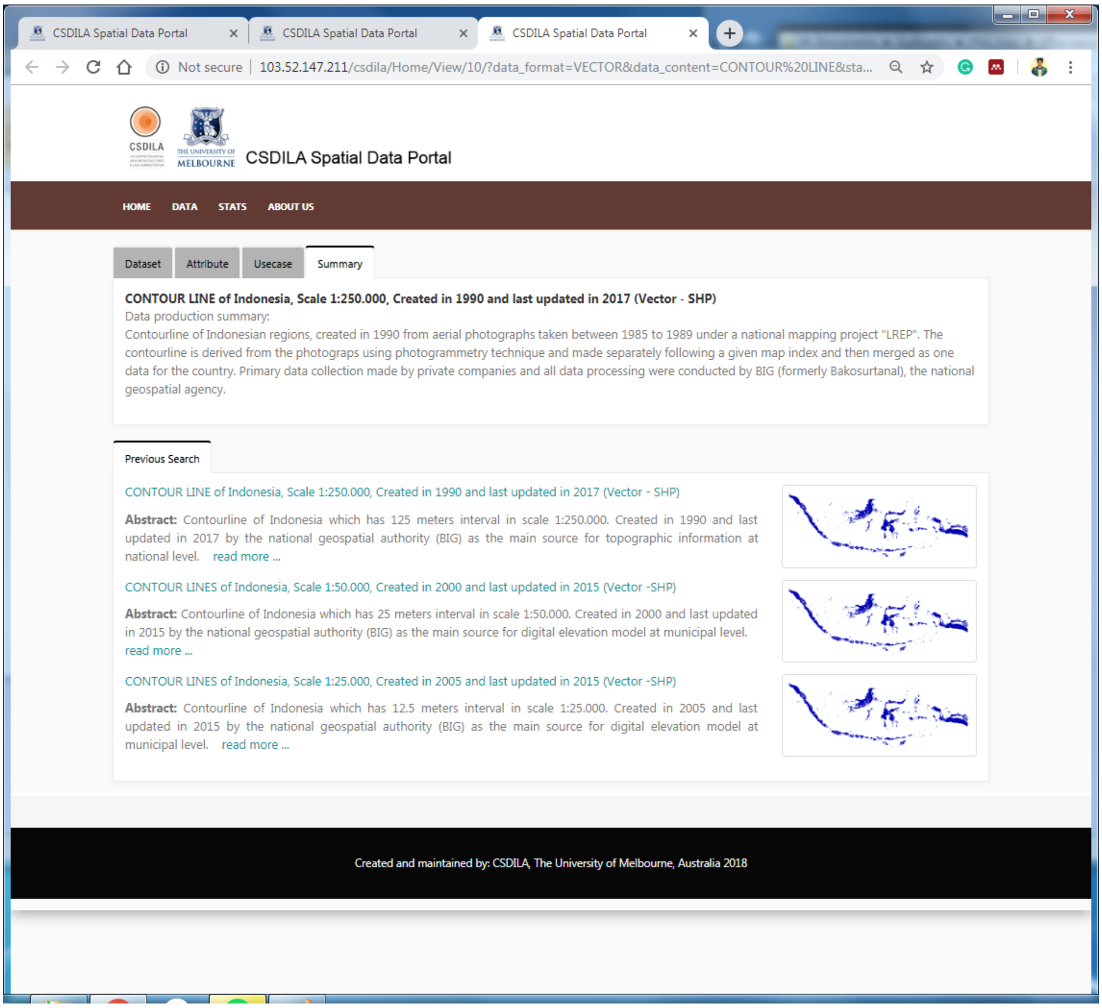


Figure 6.10: Metadata page (Summary tab)

Web-map service (WMS) page

The web-map service (map preview) page is designed to show a preview of the selected spatial data in an OGC web-map service (WMS) format. Users can activate the page by clicking the thumbnail (graphical information showing the geographic extent of the data) either on the result page or on the dataset metadata page. The page contains a map window and a legend window that allow users to play with and familiarise themselves with the data, as illustrated in Figure 6.11.

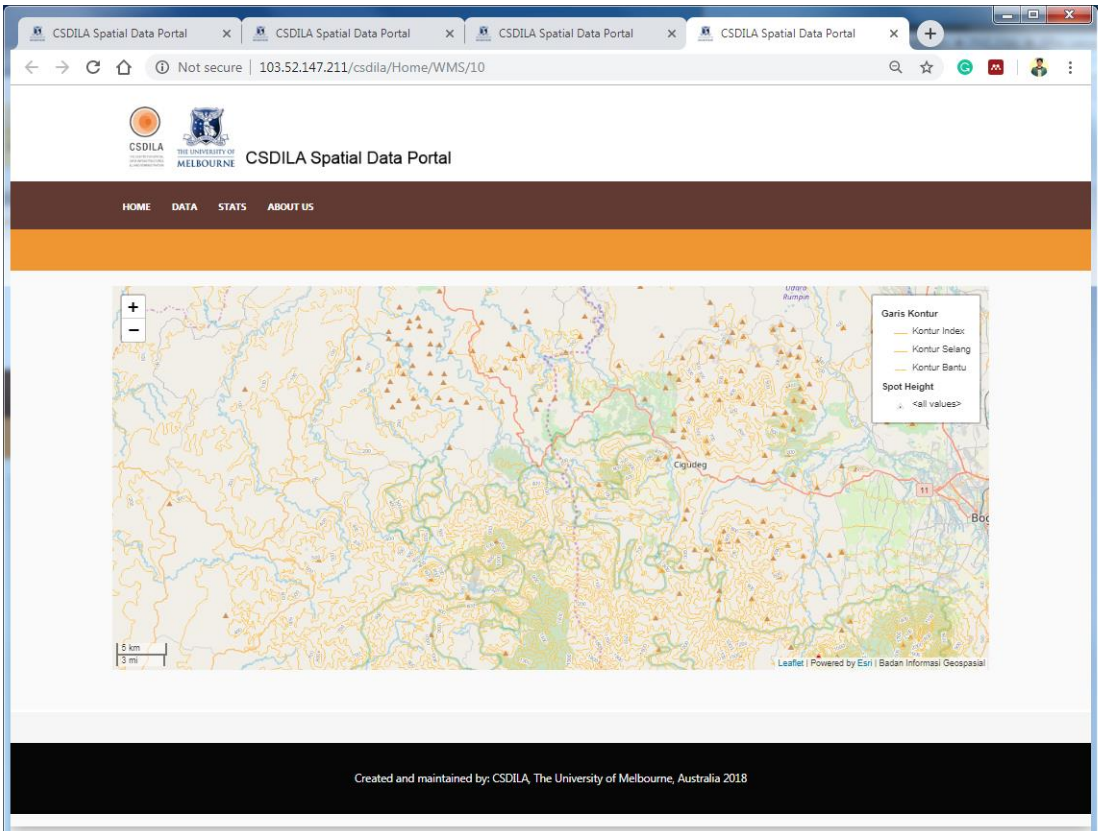


Figure 6.11: Web-map service (WMS) page (Map Preview)

6.3. Usability Evaluation Design

Since the usability evaluation is designed to detect any improvement of metadata usability, the evaluation stage was implemented by conducting a usability evaluation process similar to that used in the first stage, as explained in chapter 4. The purpose is to maintain the consistency of the methods and indicators so the results can be compared.

As in the previous usability evaluation, the data-collection methods for this evaluation purpose is Think-Aloud Protocol (TAP) usability testing combined with semi-structured interviews. Eight spatial data users from various professional backgrounds participated in the data collections: a planner, three spatial data professionals, and four professionals with computer science backgrounds. They all work closely with spatial data in different areas and in various projects in different organisations.

The data was analysed using a combination of protocol analyses, and transcript analysis was implemented for data analysis. Here, any comments or expressions are considered to have the same weighting potential, even though any comment is only submitted once by a user. The level of importance of a comment is not based solely on how often it is revealed but also by how much it is discussed during the interaction between the user, the prototype, and the interviewer. The results were interpreted to determine the usability of the proposed prototype based on following criteria, which are similar to the previous usability evaluation:

1. *Effectiveness*. The extent to which spatial data users successfully discover and select spatial data.
2. *Efficiency*. The resource (time) spent by spatial data users to successfully discover and select spatial data.
3. *User satisfaction*. Emotions, feelings, and experiences of spatial data users during the data-discovery and selection process.

The evaluation was also designed to identify usability problems, as well as user expectations, which will be useful for future development following the iterative nature of UCD approach.

6.4. Results and Interpretation

The data from the TAP and the interviews were analysed accordingly to extract the information required for evaluating the usability criteria of the prototype: its effectiveness, its efficiency, and its user satisfaction. It is was also required for identifying problems and gathering user expectations.

By conducting protocol analysis for the TAP data and by conducting script analysis for interview data, the results indicated that the usability of the prototype for both spatial data discovery and selection is higher compared to the usability of the established metadata, as explained in Chapter 4. However, the interpretations of the results should be carefully discussed and explained before it can be concluded.

Detailed explanations about the results and the interpretations of the usability evaluation for the prototype are explained in the following sections.

6.4.1. Effectiveness

Effectiveness of the prototype is measured by the successfulness of the participants in achieving the objective of the following tasks: (1) to discover the potential data; (2) to select the data to use or to not use for the given scenario, as explained in the following sections.

Spatial data discovery

The usability of the evaluation results in the identification stage suggested that the effectiveness of the spatial metadata is a combination of the usability of the user interface and the metadata itself. Therefore, it's essential to address the issues and problems existing in both of them. The results from the usability evaluation conducted in this stage confirm the previous results and allow us to have a deeper discussion about this.

The following phrases indicate that the participants were able to discover the required spatial data from using the prototype.

> "It's good, the searching is easy and straightforward, with some options of searching. I can see the results are relevant."
> "The titles in the result page (are) really helpful to quickly find the most relevant data."
> "The titles in the result list. They give me more information than any other portal I've used before."

Most participants had a clear understanding about the discovery tools, especially the filter-based one.

> "The information (used) in the filters in (the) homepage are important."

Their searches were straightforward. They used the required filters with an expectation that the system would respond by giving them the spatial data matching their submitted criteria.

> "The *search* was *done in no time*, (as) it can be *easily filtered based on user interest* and the *results (titles) are quite clear and consistent*."

They were impressed with the results, especially the titles, as they found that the titles contained information that they could use to quickly find the most relevant data for the project. Information about what the data is about, who created it, when and where it was created, and how it is stored are all there. The information was useful for them, enabling them to quickly recognise the potential of the data for the given project, and it can be used as the basis for data selection if they have no other information or if they must make an instant selection.

> "(The) *titles* are very *helpful* since they *contain all the required information* (what, where, when, who)."

Compared to the previous evaluation results, this spatial metadata for data discovery is more effective, since the metadata from the two Australian spatial data portals do not clearly indicate whether participants successfully discover the required data or not.

Due to irrelevant results to submitted search criteria and the absence of any filters, the scale of information presented to participants, and thus the inconsistency of the information presented in the titles, meant that hardly any could discover all the data they required within a given time. If a user had more time and decided to continue the search, they might be able to discover the data they needed. However, this was quite a frustrating process, a result of the aforementioned issues with both the metadata records and the user interface.

The effectiveness of the prototypes is as expected, since they were designed and developed to address the issues. However, there is also another fact that can be discussed regarding the effectiveness of the prototype when compared to the existing metadata and data portal, in regard to their production and development.

The metadata records presented in the two established portals were created and maintained by different organisations, and the portal is maintained by another different organisation. Those organisations might have both used the same metadata standards, ISO 19115:2003, when they designed their metadata, as well as when they designed the portal. However, from the evaluation, it was found that metadata records vary from one to another, and the information does not necessarily support the discovery system or vice versa. Meanwhile, the prototypes, both the metadata records and the discovery system, were designed and developed simultaneously and then synchronised. Therefore, the information contained in metadata records support the functionalities of the discovery system, and vice versa.

This fact indicates that the effectiveness of the spatial metadata for spatial data discovery can be improved greatly by addressing both metadata records and the user-interface (discovery tools) issues simultaneously and by synchronising the information in the metadata records with the functionalities in the discovery system, as based on user requirements and expectations.

Spatial data selection

Participants found that the prototypes give them the information needed to decide whether to use the data for a selected project or not. However, these results should be interpreted carefully to determine the effectiveness of the prototypes. The following phrases might be useful to begin with:

> "*Information* presented *in the metadata* page *are important and relevant*, but some key information for a specific type of data, e.g. interval for contour data, should be presented in the table (view)."
> "All and all, *the information presented in metadata is clear and necessary*."
> "*Usecase* gives me an idea what *dataset might be suitable* and what other people said about it."

The above phrases indicate that the information presented in the prototypes was useful and understandable for the participants when selecting spatial data. However, the phrases do not allow us to clearly determine the effectiveness of the prototypes for spatial data selection. In this case, effectiveness is defined as a participant making clear and confident decisions, rather than vague ones. The following phrases contain participants' thoughts about their decision-making process that could be useful to detect the effectiveness of the prototypes for spatial data selection:

"*I am confident* with the information *to make a decision* to use the data or not."

"But *I need the reference to the original data source* (such as USGS or other institution) to *make sure that the data* is legit and comes from a valid source."

"If I *only* have the *metadata information*, I am *not confident* to make a *decision* and I *definitely need to look the data*."

The above phrases about the participants' decisions can be divided into two groups: (1) confident and clear; (2) conditional. The former group is clear, as participants would decide based on the information in metadata. The latter group indicates that participants might have made any decision, but that they might also need other information potentially beyond the presented information. For instance, if we read these phrases carefully:

"This *authoritative information* (publisher and license) *is important* for the *validity* of the *data*."

"*I know* that the *data is produced* and published by a *valid organisation*, so there is no reason I cannot make a decision whether to use or not the data for my project."

The phrases "authoritative information", "validity", and "valid organisation" refer to the participants' specific knowledge of the organisations presented in the metadata. The phrases give an indication that they can only make a clear decision if they know the validity of the data's producer. In other words, they are not able to decide if they do not know the information.

If we read another phrase:

"But *I also need to check the data* once it's downloaded to *make sure* that the *data* would be *fit to this project*."

"If *I only* have the *metadata information*, I am *not confident* to make a *decision* and I *definitely need to look the data*."

It gives an indication that the information presented in the metadata is still insufficient for this participant to make a strong decision. Instead, they insist that a direct experience with the actual data, or with information from other sources, is necessary as part of their decision-making process. Similar issues, among others, were also found in the previous usability evaluation for the established spatial metadata.

The aforementioned phrases were stated by the participant before they found other capabilities provided by the prototypes, those specifically designed to address the issues. Presented to the participants was additional information, such as Usecase tab, data ratings, or new capabilities – like those appearing on online shops (e.g. a list of relevant data, previous searches, as well as web-map services for a data preview). When they found and explored the web-map service as a map preview, the following phrases were mentioned by the participants:

> "The information in this page (*web-map*) *is* somehow *more important* to me because no matter the quality of the metadata, *I* (would*) need to check the data.*"

> "*This (webmap service) is a must*, because I don't have to download and check the data to play (with) and look at the data. It saves time because I don't have to *download a lot of data* and contact a lot of persons *to know about the content of the data*."

> "*It (webmap service) gives me* a *complete sense of the data*. For example, if I need a parcel data, and if *I can have a look into that specific* area without buying or downloading it first, *I would be very happy*, and I think it (*the webmap service*) *should be a metadata element*."

The above phrases indicate that the participants still need to see and check the data to get the information that cannot be provided textually in nature metadata. To some extent, the spatial data web-map service can fill this need by providing a live preview of the spatial data without the need for downloading or buying the data. To some extent, the map preview is more important than other information presented in metadata regardless of its quality.

However, not all the participants agreed with this. One participant had a different opinion about their decision process, saying that they still might have to look at other additional information regarding the data from other sources, as can be seen from the following phrase:

> "*I* am *okay* with the *information* in the *metadata* and the presentation, but I might *have to look* on *other information* from *other source* as *additional information*."

This participant mentioned specific requirements regarding the spatial data attributes such as the attribute type, statistics, and other quality standards, e.g. quality evaluation methods and the results.

The next phrases were obtained from participants' verbal comments when they used the Usecase tab and saw the rating.

> "*Usecase and rating help me to have a peace of mind* about the *fitness of the data* for my purpose."

> "*The usecase is important*. *If* I, let's say, find out that *the data had been used by* figures or *organisations that are reliable*, then *it will give me the confidence to use the data*. *It enhances the quality* in a sense of *reliability.*"

> "It's complementary, but it *(usecase) is important* because *it gives me objective reviews* based on *user actual experience*, (whilst) the information in the *metadata* page *presents the facts* about spatial data. *I will consider both* information *to make a decision.*"

> "I like the rating and number of downloads (as it may suggest that) *the higher the rating and number of downloads* might be *more relevant or reliable.*"

These phrases indicate how the Usecase tab and the user reviews on spatial data could be important to the participants' decision-making process. The confidence level of the participants improved when they had this information. However, the participants also mentioned the dangerous potential of the rating and the user reviews if they were not presented properly. As the planner mentioned in a phrase below:

> "The persons who give *a rating and download* the data can be *from different backgrounds* with *different interests*. It would be better and *make more sense if* the rating and download are *presented* in a graphics *showing groups of users* and numbers of download and *uses for each group.*"

According to the planner, and the others, the user review, the rating, and the number of downloads should be presented in detail and should include the type of project and the group of users. Otherwise, the information can be useless at minimum and misleading at maximum, as they might get the wrong impression about the suitability of the spatial data for their own projects.

Other additional information providing insights about the prototype is the spatial data relating to the participants' searches. It is one of the implementations of online shopping features, as requested by users in the user requirements:

> "That would give users useful information about the possibility and availability of other data that might be required for their projects. It will save time and provide insights to users in data searching process."

The phrase was confirmed by the participants and their explanations are interesting. The related-data feature potentially saves time for them and provides other data that might be useful for their project. It might also provide the participant insight about the data content by looking at other relevant data in the list. Again, this information should be carefully prepared as it can be misleading if the additional data is not relevant or wrongly determined.

Having the results and interpretation of the TAP and the interviews data, featured above, we can identify several findings that should open a discussion relating to the effectiveness of the spatial metadata for spatial data selection.

Most of the negative comments in the first stage were due to the data's inaccessibility or the lack of a sample or preview, as well as the absence of additional information on the data as evidence of its reliability. By providing access to the web-map service as a data preview and to user-related information (e.g. reviews, a data rating), more positive phrases were found in the data-selection activities conducted in this usability evaluation. This indicates that the effectiveness of the prototypes for spatial data selection is higher than when compared to the established metadata.

However, apart from the quality of the metadata records and the user interface, the effectiveness of the spatial metadata for data selection is subject to individual knowledge (i.e. spatial concepts, the reputation of data producers or distributors), experiences with the spatial data (so they can use this when they check the data), the availability of spatial data, and also a participant's trust of the metadata, as following phrases reflect:

> "What if it's a rare data that only a few people know about it?"

> "What if it's the only available data for that specific theme?"
>
> "I think I might need additional information from other sources."

Nevertheless, the validation stage provides the required data and information for evaluating the effectiveness of the user-oriented metadata prototypes for spatial data selection. They indicate that the effectiveness, when compared to previous evaluation of the established metadata, is improved. Additional information about or capabilities of the prototypes – such as user-related information (the Usecase tab, a data rating, the number of downloads), related data to the current search, and the web-map service (WMS) as the data preview – proved useful in increasing the participants' confidence in deciding whether to select the data for their given project or not.

6.4.2. Efficiency

The results from the usability evaluation indicate that the efficiency of the prototypes for spatial data discovery and selection is higher than the efficiency of the established metadata, as explained in Chapter 4. The following sections describe the results and interpretations of the TAP data and the interviews data from the usability evaluation of the prototypes.

Spatial data discovery

Efficiency of the prototypes for spatial data discovery is determined by observing the time spent by participants on the discovery task. The participants comments are then analysed during the data discovery process.

The researchers' observations during the evaluation show that all participants were successful in immediately discovering the required spatial data for the given scenario. This was expected, as the metadata records and the user interface were specifically prepared following the user's needs and expectations, as presented in the user-oriented metadata specification from chapter 5. The content of each required element, used by the participants as the search criteria, was prepared, as well as the functionalities of the user interface to support the discovery process.

As explained in the prototype development section, both the metadata and the user interface were designed, developed, and synchronised simultaneously.

The following phrases indicate the highly efficient discovery process conducted by the participants.

> "I can easily filter the search and the results (titles) are quite clear and consistent."
>
> "It's good, the searching is easy and straightforward, with some options of searching (keywords, filters). I can see the results are relevant."

As mentioned by the participants, they found it easy to use the discovery tools. The criteria are straightforward, and they can submit all the criteria required for their search to find the given spatial data. The results were even better, especially the titles, when they contained the basic requirements for spatial data discovery and when they were presented consistently, as can be identified from the following phrases:

> "Titles contain all the required information (what, where, when, who) and the consistency is important."
>
> "The titles saved my time to get basic information about the data without the need to open the metadata page."

According to the participants, the titles contain information to answer all questions regarding the who, what, why, when, and how related to spatial data. What the data is about; where is covered by the data; when the data is produced or last updated; who the producer is; and how the data is conditioned (represented by the scale and data format), and all presented in titles with a reasonable length of words.

The titles alone are sufficient for the participants to make a quick selection on spatial data, and the abstract provides them with additional information useful for more detailed descriptions about the data. All the information is presented consistently, which makes it easy to follow and to read, even if the results consist of a lot of data.

However, for similar reasons, the participants agreed that the abstract should not be presented in the result, since the titles already provide them with the required information. They instead prefer succinct, yet useful information to be presented alongside the titles, e.g. keywords, to get a glimpse of the context of the spatial data. They also prefer the abstracts to be moved into the metadata page, so the page could contain more results.

Regarding these evaluation results, the efficiency of the prototypes for spatial data discovery is greater than when compared to the established metadata. The reason is clear: the spatial metadata records were prepared to provide all the information (element) required by spatial data users to discover spatial data, and the user interface provides the required functionalities to support the discovery, e.g. filters (or searching criteria). Another reason is that the results (title and abstract) were consistently prepared, following the user requirements in the specification, to make it easier for the participants to recognise critical information when identifying potential data for the given project. The established metadata, however, did not provide similar information and did not have the consistency seen in the titles of the spatial data discovery prototypes.

Spatial data selection

The participants found that the required information for selecting spatial data was easily identifiable and understandable. They were instantly able to find critical information and to make their decision on the spatial data. The participants found it easy to decide whether they wanted to select and use the data for the given project or not based on the presented information, as reflected in the following phrases:

> "I'm quite impressed with the website, as it's clear and informative, straightforward, and the information is very useful and enough for me to make a decision whether to use or not the data."

> "I think the system is easy to use. The information is understandable."

However, there were some negative comments from the participants, mainly about the user interface, as can be identified from the following phrases:

> "I think the information is useful to know about the data, but I didn't know that the titles (in the result list) can be clicked to find metadata. I think you should give a clear instruction on the result page."

> "I didn't notice about that (the web-map service) at the first place, so it requires a sign or instructions to make users notice it in the first place."

There were similar phrases in the data that allow us to identify the efficiency of the prototype for spatial data selection.

As can be seen from the phrases, efficiency of the metadata was considerably high since the participants mostly gave positive statements regarding the importance of the presented information as well as the relevance of the information. There were some negative comments regarding the information (or metadata contents), but such comments are more likely to be used for improvements, as explained later in the User satisfaction and Expectations section. However, there were comments from the planner worth noticing, as they questioned the relevance of some information to their search:

> "I don't know this (contour type: index contour, additional contour). For me, contour line is just one, contour."
>
> "I don't know this dataset type: dataset."

It does not mean that the information is not there, but that the additional contour types were confusing for him, as he would normally only look for the main contour data. As for the dataset type, when it was explained by the researcher during the interview, they agreed that the information was useful. However, they requested a different description to be displayed.

In terms of the user interface, the design and functionalities of the prototype in presenting the metadata make it easier for the participants to quickly identify the information. Credit can be given to the table format for succinctly presenting information, i.e. using a limited number of words while still being informative.

A special credit can also be given to the web-map preview (WMS). As most participants required seeing and checking the data before making any decision, most stated that the preview would save them time when deciding whether certain data could be used for the project or not:

> "I think it is good to have it (the web-map service) because it will save my time to download and check the data (also the money if it's not free) to check whether the data is valuable for my work or not."

> "Web-map service is a must, because I don't have to download and check the data and can play and have a look at the data and save time since I don't have to download a lot of data and contact a lot of persons to know about the content of the data."

The data, along with the user review, allows participants to identify the web-map service as making the data-selection process more efficient (regarding both time and money), as they could play and check the data first without having to download and buy it. They could also save time by not having to get the information from other sources.

Participants found that the previous-search window was helpful since it provided them with the result or their current search at the bottom of main metadata page. They could jump to other results from the window without needing to perform another search. This saved time for the selection process.

From the above descriptions about the evaluation results, the efficiency of the prototype for spatial data selection has improved when compared to the established metadata evaluation, where the participants had difficulties identifying and understanding the information presented in metadata. It is not that the participants found no difficulties with the prototype, but that the prototype allowed them to make a quicker decision, and to some extent a cheaper one.

The results show that the efficiency of spatial metadata for spatial data selection is determined by: (1) the completeness of the metadata against the users' selection criteria; (2) the relevance of the content (information) to users; (3) the presentation style, as well as the functionalities, which should make it easier for users to recognise the information.

6.4.3. User Satisfactions and Expectations

User satisfaction is related to the emotions, feelings, and experiences of the participants when they worked and utilised the prototypes during the usability evaluation process. It was not easy to measure and determine them. However, we can use the phrases to detect the participants' satisfaction and their expectations of the prototype, as exemplified below.

> "This is a good way of (metadata presentation), based on my previous experience with metadata with other software like arcgis, mapinfo and other types of (system). This approach gives more comprehensive way of presenting metadata."

Similar phrases can be found in the data, which indicate that the participants were satisfied with the information presented in the prototype, as well as with the user interface (presentation and functionalities). A participant who happened to be an ex-national open data manager mentioned that he was impressed with the prototype as it was informative, clear, and easy to use. He also suggested that the prototype should be commercialised or implemented as a product.

The participants also gave a lot of comments about the prototypes that can be identified as their expectations to improve the prototypes. Table 6.2 summarises the user's expectations of the prototype to better help them in spatial data discovery and selection process.

Table 6.1: User expectations of the prototypes

User interface	**Metadata records**
✓ Default searching tool is the simple one (keywords) and can be extended to the advance (filter and spatial based) ✓ Detailed data format for both vector and raster data, e.g. shp., cad., cov., etc. for raster or tiff., jpeg., png., etc. for raster ✓ Clear instructions for the spatial based search tool ✓ Clear instructions for opening the map preview (web-map service) ✓ Clear instructions for opening the metadata page from the search-result page ✓ Alternative presentation of the titles ✓ Detailed data rating based on pre-defined criteria in accordance to spatial data quality or user-specified criteria ✓ Detailed user review based on previous experiences ✓ More functionalities in map preview, e.g. object identification and statistics for the attribute information	✓ Additional information i.e. the validity of the data (valid, used, obsolete) ✓ Detailed information about previous users and the number of downloads i.e. type of project, field of work ✓ Detailed information for data rating based on pre-determined criteria as on e-commerce web pages ✓ Additional information in the attribute data i.e. statistics, data attribute type

From Table 6.2, we can identify that the user's expectations were mainly related to the user interface, as the participants considered the information sufficient, and that fine tuning should only be made to the functionalities and to the presentation. As for the metadata records, most of the participants agreed that the detailed user reviews, the number of downloads, and the user data ratings are necessary. It was mentioned by a participant in an earlier section that the user-related information was important in assisting their decision, but only when this detailed information related to the user and to the project given.

Compared to the user satisfaction of the established metadata, the participants were more satisfied with the prototype. Two participants who performed the evaluation for both metadata searches comments on the prototype more positively than those compared to the previous ones.

The result is as expected, since the prototype was designed and developed based on the previous usability evaluation results and was catered to meet the user's requirements and expectations of the spatial metadata and the user interface. However, users will never be fully satisfied, and that is normal, as they will always find new needs and new expectations. That is why usability is something that should be evaluated regularly to ensure that the metadata, as well as other products, meet at least most of its user's expectations.

6.4.4. Usability Problems

Similar to the user's expectations, usability problems were identified because participants will always have new needs and expectations of the metadata. However, unlike the problems identified in the previous evaluation, these found by the participant during the prototype evaluation did not prevent them from discovering and selecting the spatial data.

The most noted problem during the evaluation process was the unclear instruction on how to click or activate important features for the spatial data selection, such as the map preview (web-map service). Most of the participants did not know about the preview as it was not presented altogether with the metadata. Participants who found the preview, either gave more attention to the the mouse-over style link or accidentally clicked the link, which is located in the thumbnail picture (geographic coverage). A similar problem was also found regarding the spatial-based search tool, as illustrated in Figure 6.12. This problem is considered a high risk, as the map preview has proven to be a crucial tool and this issue could jeopardise the participants' decision-making process.

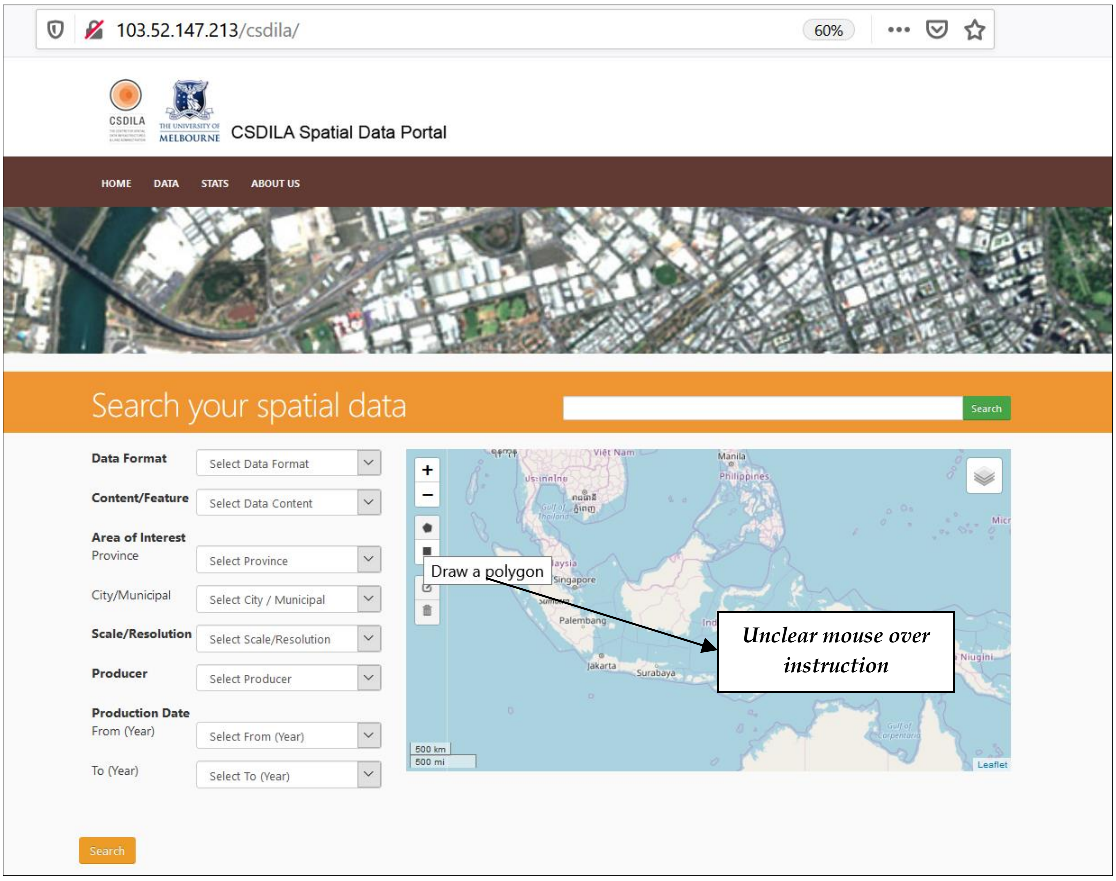


Figure 6.12: Unclear instruction for spatial based search tool

A similar instruction problem was also found on the search-result page, where participants did not know that the titles could be clicked to open the metadata page. This is, again, considered a high-risk problem as users with no experience with the website, or with spatial data searching tools, would stop there and fail to continue the process. This is a typical usability problem with a user interface or a website (Ferreira & Pithan, 2005)

In terms of the information from the metadata records, participants found some of the information irrelevant, as happened to the planner in an earlier section when he looked at the content of the contour data. It did not necessarily prevent them from getting the relevant information, i.e. the main contour line, but it did take some time for them to identify the important information.

Other problems identified during the evaluation process were mainly related to the expectations of the participants, such as regarding the user data rating and the user reviews, where the participants requested detailed information rather than just a general statement such as good or bad, as mentioned in section 6.4.3. All the identified problems were important, and they should be addressed in the next development stage of the spatial metadata and the user interface.

Having explored and interpreted the usability evaluation results of the newly proposed metadata prototype and user-interface prototype, the next section discusses how these results have addressed the last research question and how the overall research has contributed to the spatial community.

6.5. Discussion

As identified in Chapter 2, current spatial metadata is less helpful and less usable for data users to discover and select spatial data from distributed geospatial data portals or clearing houses. This is due to the problems with the metadata records, as they are created following a standard originally developed for data-inventory management purposes. It was also identified that the usability of spatial metadata is a combination of the metadata records and the user interface, as it is the medium for users to interact and work with metadata. From a research perspective, this study proposes a hypothesis that the usability of spatial metadata will improve if the metadata and the user interface is designed and developed in a synchronised manner following user-oriented metadata profile.

This validation stage is designed to evaluate the usability of the user-oriented spatial metadata prototype created following the profile from Chapter 5. The objective is to identify any usability improvement of the metadata when compared to the usability of the existing metadata, as explained in Chapter 4. This will result in evidence supporting whether the hypothesis is proven or not.

Using the three attributes of usability as adapted from the ISO criteria – effectiveness, efficiency and user satisfaction – the results indicate significant improvements in usability across all criteria. The hypothesis is proven, and, as a response to the last question, the participants (spatial data users) reacted positively and were satisfied with the metadata and the user interface for helping them discover and select spatial data.

In addition to that, there are some areas for further discussion, such as the usability of the metadata and the metadata profile as indicated from the results of the validation stage and the previous stages, which can be the contributions of this research to the geospatial community.

Firstly, from metadata design and development point of view, it is believed that the implementation of the metadata standard is a source of less usable metadata for users to discover and select. A close look at the standard in the geospatial community, ISO 19115-1:2014, the latest edition and the most widely known and used as explained in Chapter 5, revealed that it has accommodated almost every aspect of spatial data characteristic used for data discovery and selection, except for user-related elements, i.e. user data ratings, numbers of downloads, and user reviews. Another identified issue with the standard is the free text domain in the characterString data type, which might generate inconsistencies and irrelevant metadata. Obligation type is another issue found in the standard that might cause an insufficiency of information in the metadata. Geospatial authority, such as ANZLIC, has created and published metadata guidelines within their jurisdiction to be used for spatial metadata creation (ANZLIC, 2003). However, they still use the free text domain and a similar obligation type, and the result, metadata that is less usable for users, is explained in Chapter 4.

The results from the validation stage indicate that maintaining the consistency and relevance of the information presented in metadata, can be achieved by proposing more stringent obligations to the elements required by users, e.g. intended use, data samples, and replacing the free text domain of crucial elements, such as title and abstract, with a compound domain comprising contents from other selected elements following the user requirements. In the meantime, the relevance of spatial data to, as well the confident level of, the users can be improved by adding new user-based metadata elements, i.e. a user data rating, and a number of downloads.

However, the data rating and number of downloads have to be presented carefully and in detailed. Otherwise, it can be misleading. As mentioned by a participant, the producer rating on specific data quality attributes, e.g. geometric and attribute accuracy, are facts that should highly be considered in spatial data selection. Whilst data rating and number of downloads made by previous users are helpful and will add his confident about the validity of the data. But the information is only valid if the rating and number of downloads are accompanied by detailed descriptions of the users who gave the rating or download the data, e.g. type of application or industry sector that used the spatial data and type of analysis as well as the professional background of the user. The information would be helpful for users to determine the relevance of the spatial data to their own purposes of using the data. For instance, if the rating was given by a hundred of users but most of them are from the industry sectors different from his/hers with different type of analysis then it might not be useful for them. Or, if the spatial data is the only available data that can be downloaded and made by a sole organisation, it cannot adequately represent the quality of the data suitable for users' projects. In that case, users might need and look into, and make their decisions based on, producer ratings and detailed spatial data quality attributes provided by the producers. Therefore, although the results from the evaluation stage gave the evidence about the improvement on the metadata usability by proposing user rating and number or downloads, there is still an urgency to put producer rating and detailed data quality information in the user-oriented profile.

The results validate the need for an improvement for the spatial metadata standard to be more usable to spatial data discovery and selection for data users. The improvement is not necessarily fundamental, i.e. adding or removing a class. It can be done by proposing an extension, such as developing the user-oriented spatial metadata profile, which has been proven sufficient.

Secondly, the results validate the proposed synchronised development approach for the spatial metadata and the user interface in increasing and maintaining the usability of the metadata for data discovery and selection. In the distributed geospatial data catalogue platform, as explained in Chapter 2, the spatial metadata records and the spatial metadata catalogue are not necessarily created and maintained by the same person or organisation, and nor is the catalogue gateway, where the user interface and its functionalities are defined and developed. As the usability of the metadata is a combination of the metadata and the user interface, a synchronised development approach, with the users in the centre, is required to maintain the usability of the metadata, as illustrated in Figure 6.13.

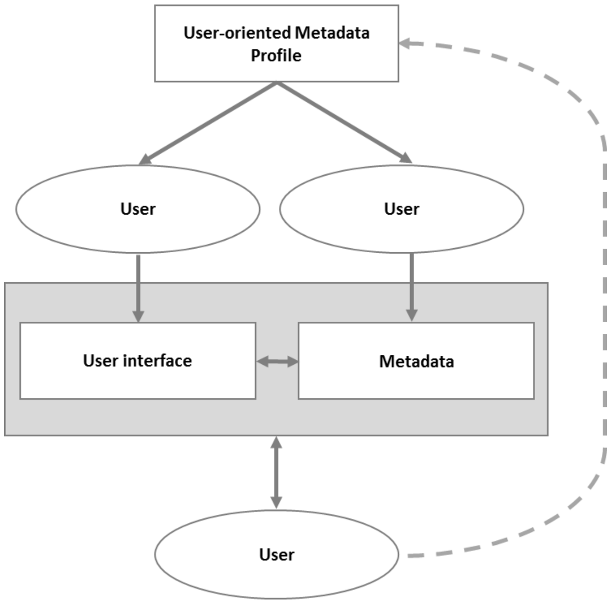


Figure 6.13: Synchronised development approach for spatial metadata

The next section summarises the chapter with brief explanations about the results and how these results address the fourth and fifth objectives and the last research question.

6.6. Chapter Summary

Now that we have established the user-oriented spatial metadata discovery profile from Chapter 5, this chapter covers the design, development, and usability evaluation of practical metadata prototypes, both for the metadata and the user interface, in spatial data discovery and selection. The goal was to implement the user-oriented metadata specification into practical metadata and to deliver the metadata to users to help them find and select the potential spatial data for a given application.

A set of metadata prototypes for selected spatial data from the Indonesian Geospatial Agency (BIG) were accordingly created, e.g. Road Networks, Contour

Lines, Administrative Boundaries, etc., along with a website for the spatial data discovery tool as a user interface. The prototypes were built on open-source platforms, i.e. using PostgreSQL for the spatial metadata database and Apache for the web server. As for the spatial data service, an OGC-compliant WMS was created using ArcGIS Server.

The prototypes were then evaluated by conducting TAP usability testing and semi-structured interviews with eight spatial data users with varying backgrounds. They were asked to complete a series of tasks in a given scenario within a given timeframe and to think aloud, giving verbal expressions during the TAP.

The results showed that participants were really interested in the spatial data search tool, especially the criteria (filter)-based option, as it allowed them to narrow down their search as early as possible. Participants also found that the result page, which listed data matching the search criteria, was very pleasing and easy to understand. The consistency, length, and content of information presented in the title and the abstracts were helpful for users to easily recognise what the data was about. Accordingly, they could quickly preview the potential data for the given application before clicking them to check the detailed information. The list of relevant data to the search was also interesting to participants, as they can use the information to get to know other data beyond their knowledge that might be useful for their projects. In terms of efficiency, this feature saved time while searching for other required data. This finding shows the potential of linked metadata as the next metadata development.

Moving on to the main metadata page, the results found users agreed that most of the information presented in the Dataset tab was useful for them to assess the quality of the data. The tabular presentation of the metadata made it easier for them to instantly identify information most relevant to their interests, e.g. data contents with the level of detail, positional and attribute accuracy, contact detail, etc. Information about user reviews, data ratings, and number of downloads were also interesting and important to participants as they can use them as alternative measures of a data's quality and fitness for use. Such information was unavailable in the previous usability testing. However, participants requested that the information be presented in detail for them to be able to get the relevant information. Accordingly, the form provided on the Usecase tab should have the same requirements so users might provide their experience with the data properly.

Another noteworthy finding from the results was that participants were really interested in the web-map service (WMS) feature of the data, as it gave them the chance to access the data, thereby giving them the confidence to assess its suitability. They could check the data, something missing in the previous usability testing of the established metadata.

The users' reactions and comments indicate that they were satisfied with the prototypes, apart from the several requests for other improvements, which are expected from a usability study. This answers the last research question about the users' responses and acceptances of the proposed spatial metadata and user-interface prototype as the manifestation of the user-oriented spatial metadata profile. Thus, the results also proved the hypothesis.

The results also validate the contributions of this research to the geospatial community. Firstly, UCD is evidently a useful method to be implemented in a research context. Secondly, this research suggests that the spatial metadata standard, ISO 19115-1:2014, should be improved by extending it with additional metadata elements, by changing the more stringent obligation type, and by replacing the free text domain with the contents of other relevant elements, as required by spatial data users. The user-oriented spatial metadata discovery profile, as an extension of the ISO19115 spatial metadata discovery for geospatial resource, is evidently able to increase the usability of spatial metadata.

Lastly, the results validate the proposed synchronised development approach for the spatial metadata and the user interface in increasing and maintaining the usability of the metadata for data discovery and selection. As the usability of the metadata is a combination of the metadata and the user interface, a synchronised development approach, with users at its centre, is required to improve and maintain the usability of the metadata.

The next chapter of this thesis will present the final conclusions of this research by revisiting the research questions and objectives and by examining the results and findings in response to those questions and objectives. It also provides possible future research and key recommendations resulting from this study.

This page is left intentionally blank

7. Conclusions and future directions

7.1. Introduction

Despite the considerable interests and applications of the spatial metadata in the geospatial community, we know little about the usability of spatial metadata for spatial data users (Bulens et al., 2013; Boin & Hunter, 2008; Boin, 2008). There have been criticisms that established spatial metadata are not necessarily useful for spatial data users as they have been created following producer-oriented standards, which were originally designed for data inventory management purposes with limited involvement of spatial data users (Goodchild, 2007; Goodchild, 2009; Fisher et al., 2010). Against this background, this research aimed to investigate the usability of spatial metadata for spatial data users and how improvements could be made. To achieve the aim, this thesis investigated how spatial data user involvement in the design and development of spatial metadata could improve the usability of metadata for spatial data discovery and selection.

This concluding chapter examines the outcomes of the study, discusses the key contributions to the field of knowledge, and suggests areas for future research.

7.2. Research Aim and Objectives

Chapter 1 detailed the backgrounds to this research, the importance of spatial data and metadata for geospatial community and the structure of this research. As mentioned in Chapter 1, the main aim of this research is to:

> Improve the usability of spatial metadata for spatial data discovery and selection by proposing user-oriented spatial metadata profile as the basis for synchronised development approach for the metadata and the user interface.

The research was originally constructed following a hypothesis as follows:

> Implementing a user-oriented spatial metadata profile as the basis for n synchronised development approach for metadata and user interface will increase the usability of the spatial metadata for data users.

The hypothesis was then translated into a research question as follows:

> How can user-oriented spatial metadata profile be developed and implemented in synchronised spatial metadata development to improve the usability of spatial metadata?

To answer the research question, and test the research hypothesis, the following research questions were defined accordingly:

1. What are the current spatial metadata usability and problems?
2. What do users need and expect from spatial metadata?
3. How can the usability problems, user needs and user expectations be used to develop the user-oriented metadata profile?
4. How can the profile be used to develop user-oriented metadata and user interface?
5. How do users perceive and react to the newly proposed metadata and user interface?

Before setting up how the research would be carried out to achieve all the objectives and answer these research questions, Chapter 2 reviewed the principal concepts and theories related to spatial metadata, as well as recent developments and research. Chapter 2 concluded that spatial metadata usability depends on the relevance of the information presented in the metadata to their users and the usability of the user interface as the media for delivering and presenting the metadata. Therefore, usability improvement should involve spatial data users in real-life activities, and address actual usability problems, as well as user expectations emerging from actual interactions between spatial data users and spatial metadata. The outcomes of Chapter 2 were used as the basis for formulating the research structure.

Having gathered information on how spatial metadata is created, delivered and presented to spatial data users, Chapter 3 explained the process for selecting the research methodology in order to achieve the research objectives and answer the research questions. The process concluded that a user-centred design (UCD) should be adopted as the methodological approach to investigate and explore user experiences with spatial metadata. In doing so, the research design was developed accordingly utilising three UCD methods: Think-Aloud Protocol Usability testing (TAP), surveys and (semi-structured) interviews. The methods were mixed qualitative and quantitative, with more emphasis put on qualitative methods, consistent with the UCD approach. The survey as a quantitative method was also analysed using a qualitative approach. Implementation of the methods was uniquely designed to achieve the research objectives and answer the research questions, as reviewed and discussed in the following sections.

7.2.1. Objective 1: *To investigate spatial metadata usability, identify problems and user experience of spatial metadata*

To improve spatial metadata usability for spatial data users – in accordance with the aims of this research – the study first aimed to investigate the usability of existing spatial metadata. The purpose of this investigation was to determine usability and to identify key problems that might hamper spatial data users in finding and selecting spatial data, as well as user experience with metadata.

Chapter 4 explained this process and discussed the results of usability evaluations of spatial metadata from two available spatial data portals in Australia. Employing Think-Aloud Protocol (TAP) and semi-structured interviews, the evaluation was conducted by five spatial data users with varying educational and professional backgrounds against three usability attributes as the criteria: effectiveness, efficiency and user satisfaction.

The outcomes of Objective 1 validated the defined research problem – that spatial metadata is not necessarily usable for spatial data users to discover and select spatial data from the geoportals. More importantly, the results contributed to a detailed understanding of usability problems and their origins. The outcomes also revealed that spatial data users were not satisfied as they faced difficulties with both metadata and the user interface. Unclear and irrelevant information as found in the title and abstract, incomplete or missing information as found in the attributes and data quality, are some of the problems in the metadata that prevent users from effectively and efficiently discovering and selecting spatial data. A lack of filters for searching, inconsistency in titles, abstracts, keywords and results (as presented in the list, titles or abstracts) and inconsistency of results in different modes (i.e. newest and most relevant) as well as static web page design were some of the problems found in the user interface. These problems hampered the users' ability to effectively discover and select spatial data.

The outcomes also revealed user insights about the spatial data selection process, in particular where they expected to get access to data or sample data to check and assess the suitability of the data for their project. Users also considered previous users' experiences with the data as a crucial element in data selection.

To help meet research objectives 2 and 3, the identified problems and users' expectations were used as the basis to design the questionnaire for gathering user needs and expectations of spatial metadata from broader spatial data users.

7.2.2. Objective 2: *To identify the user needs and expectations of spatial metadata and compliance of the metadata standard, the ISO 19115-1:2014, to these needs.*

Understanding the usability status of existing spatial metadata and the related problems is only half of the information required to create the baseline for the metadata usability improvement. The other half is understanding what spatial data users need and expect when they try to discover and select spatial data from geospatial portals.

The outcomes for Objective 2 (i.e. problems and user expectations) were used as the basis to design the questionnaire for a survey to identify user needs and to assess compliance of the ISO 19115-1:2014 to these needs. Consequently, as explained in Chapter 5, the outcomes of Objective 2 revealed, first, a list of requirements for both spatial metadata and the user interface proposed by spatial data users to help them discover and select spatial data. Second, compliance of the ISO 19115-1:2014 with these needs and metadata standards.

It was revealed that the ISO 19115-1:2014 accommodates most of the users' needs, except for user data ratings and user reviews. Therefore, new elements should be proposed to add information about the metadata records to the profile. The standard maintains the consistency of the metadata structure, but it fails to maintain the consistency and clarity of the information provided in key elements of metadata, namely the title and abstract resulting from free text domains with broad definitions.

Therefore, there is no need to replace or change the standard fundamentally. Instead, the profile can be developed as an extension of the ISO standard by: (1) adding two new elements, (2) giving more stringent obligations to the required elements, and (3) replacing the free text domain for title and abstract with a list of specified contents forming selected elements.

7.2.3. Objective 3: *To propose a synchronised metadata development based on the user-oriented metadata profile*

Development of the spatial metadata records and the user interface synchronously contributes to addressing the metadata usability problems; absence/missing information, irrelevance and inconsistent information. Both the metadata records and the user interface were developed following the same standard – the user-oriented metadata profile. Metadata elements, structure of the metadata database, and content information for the metadata elements along with functionalities for the user interface are defined based on the profile.

The outcomes revealed that the synchronised approach was able to ensure that the metadata records provide the required information to support the functionalities of the user interface, and vice versa. After discussing spatial metadata for spatial data sharing, the development approach for metadata and user interface was highlighted as being essential to the usability of metadata. The synchronised development approach, based on the user-oriented metadata profile, ensure that the required information for users to discover and select spatial data was provided in the metadata records and presented in the user interface.

7.2.4. Objective 4: *To test and evaluate the profile*

The outcome of Objective 4 illustrated a successful implementation of the user-oriented spatial metadata profile in a synchronised development approach for metadata and user interface. A prototype system was developed following the synchronised approach as a manifestation of the profile with a practical metadata database and a geospatial portal accommodating the user requirements for spatial data discovery and selection.

Employing eight spatial data users from various educational and professional backgrounds, the prototype was tested and evaluated against the same usability attributes as the criteria. It was highlighted that key features for the users to discover spatial data were the filter-based search tool with selected criteria, the result page with title and abstract consisting of required information and presented consistently with a reasonable length (number or words) – all in line with their expectations. For spatial data selection, the highlighted features were the web-map spatial data viewer and user related information, such as data rating and user reviews, which help users to explore and check the data based on their own interests.

The outcomes validated that using the user-oriented spatial metadata profile, combined with a synchronised development approach for spatial metadata and user interface, improved the usability of spatial metadata for data discovery and selection.

7.3. Conclusions on research questions

To conclude, the research question to be answered was:

> How can user-oriented spatial metadata profile be developed and implemented in synchronised spatial metadata development to improve the usability of spatial metadata?

The research question was answered by:

- Reviewing the spatial metadata system to identify any possible issues that could affect the usability of spatial metadata for spatial data users. The outcomes highlighted two areas: spatial metadata creation and spatial metadata presentation. With respect to spatial metadata creation, spatial metadata is created following a standard, i.e. ISO 19115, that is considered producer-oriented and originally designed for spatial data inventory management. Further, regarding in spatial data presentation, spatial metadata is highly depended on the user interface. Therefore, the usability of spatial metadata for users is determined by the usability of the spatial metadata records and user interface, and any attempt to improve spatial metadata usability should include both.
- Selecting the methodology to be implemented for this research. Since usability focuses on the users and the aim is to improve the usability, the User-Centred Design (UCD) approach was selected as the most appropriate research methodology. Beside its focus on user-oriented product design and development, UCD also has various of methods which specifically focus on usability, i.e. usability testing, and other methods that can be used to gather information from spatial data users.
- Evaluating the existing spatial metadata usability using usability attributes: effectiveness, efficiency and user satisfaction as the criteria. Further, identifying the problems and user requirements and expectations of spatial metadata that potentially hamper spatial data users to discover and select spatial data by implementing selected UCD methods. The outcomes were used as the basis for usability improvements and to develop the user-oriented spatial metadata profile.

- Proposing a synchronised development approach as the implementation of the user-oriented profile. The approach was implemented synchronously to design and develop spatial metadata and the user interface to ensure that important information for the profile was accommodated in the metadata records and the functionalities in the user interface accommodated this information, and vice versa.
- Evaluating the user-oriented profile by first designing and developing a prototype system following the proposed synchronised approach, consisting of spatial metadata and user interface prototypes, followed by a usability evaluation implementing similar UCD methods with comparable evaluation criteria as the benchmark.

7.4. Main findings and contributions to knowledge

The research has produced several findings that can contribute to the current knowledge in academic discourse of spatial metadata as follows:

1. The User-Centred Design (UCD) approach with various implementing methods has proven suitable as a methodology for investigating spatial metadata usability as an information product. Understanding the nature of the metadata, including the metadata system, is essential to selecting the appropriate UCD methods – in order to meet the requirements of the research and the constrains (e.g. time and financial resources). The approach can be implemented to investigate the usability of spatial metadata for other users based on their interest in the metadata, e.g. the usability of spatial metadata for data analysis.
2. Spatial metadata is an information product that needs mediums to utilise it. Therefore, the usability of spatial metadata is a combination of the metadata (information in the metadata) and the user interface as the medium. Failing to maintain usability of any of the two will affect the usability of the metadata. Therefore, any attempt to evaluate or improve the usability of spatial metadata should include both factors.

3. Usability attributes, as adopted by ISO 9241-11:2018: effectiveness, efficiency, and user satisfaction, have been proven suitable for evaluating the usability of spatial metadata and can be used for future research regarding spatial metadata.
4. Usability of spatial metadata, as with other products, should be subject to evaluation on a regular basis, as over time the users and applications of spatial metadata might change as well as the technology for creating and disseminating spatial metadata. As implemented in this research, the "list of relevant data" feature adopted from e-commerce proved to be helpful and liked by the users. This is an example of something that was not known 20 years ago and might evolve in the next 20 years.
5. ISO 19115-1:2014 – the metadata standard for geospatial resources is useful for maintaining the structure of spatial metadata and the interoperability of the metadata across metadata catalogues. However, it is not able to maintain the consistency and relevance of information presented in the metadata, for example the results of free text domains for many of descriptive elements. Title, abstract and descriptive data quality elements are some of the essential elements for users to discover and select spatial data that have free text domains. This research found that this reduced usability in terms of spatial metadata, but successfully addressed the problem by assigning a list of specified information from selected elements to replace the free text.

7.5. Future research directions

It is expected that this research will provide useful outcomes for improving spatial metadata usability to help spatial data users discover and select spatial data for their purposes. However, further investigations could be made in the future as directed by this thesis' outcomes.

Firstly, as highlighted, the effectiveness of spatial data selection can be increased by giving detailed user-related information, such as in user reviews and ratings. Investigations of the criteria for spatial data rating will be useful for improving the usability of spatial metadata in future. Other investigations as to how user reviews should be procured and the validation process for the reviews, as well as the presentation styles, are also important.

Secondly, following the results regarding the titles and abstract, a further investigation could be made into other descriptive elements in the ISO 19115 standard with the free text domain to maintain the consistency and relevance of the information presented in spatial metadata. This direction of research will be useful for spatial metadata automation.

Finally, as mentioned in the Chapter 6, the "list of relevant data" feature was liked by users as it assisted them in identifying other data potentially useful for their project, and users' proposed some characteristics for constructing relevance. A further investigation on linked spatial data could be made to investigate relevance.

This page is left intentionally blank

References

Aalders, H. J. G. L. (2005). An Introduction to Metadata for Geographic Information. *World Spatial Metadata Standards*, 3–27. https://doi.org/10.1016/B978-008043949-5/50004-8

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, *37*(4), 445–456.

Ahonen-rainio, P. (2005). Visualization of Geospatial Metadata for Selecting Geographic Datasets. In *Technology*.

Ahonen-Rainio, P., & Kraak, M. J. (2005). Deciding on fitness for use: Evaluating the utility of sample maps as an element of geospatial metadata. *Cartography and Geographic Information Science*, *32*(2), 101–112. https://doi.org/10.1559/1523040053722114

Alamdar, F., Kalantari, M., & Rajabifard, A. (2016). Towards multi-agency sensor information integration for disaster management. *Computers, Environment and Urban Systems*, *56*, 68–85. https://doi.org/10.1016/j.compenvurbsys.2015.11.005

Albertoni, R., De Martino, M., Podestà, P., Abecker, A., Wössner, R., & Schnitter, K. (2018). LusTRE: a framework of linked environmental thesauri for metadata management. *Earth Science Informatics*, *11*(4), 525–544.

ANZLIC. (2007). *ANZLIC Metadata Profile Guidelines*.

Arenas, H., Harbelot, B., & Cruz, C. (2014). A semantic web approach for geodata discovery. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8697*, 117–126.

Asseng, B., Rahman, Z. A., & Said, M. N. (2018). Determinant Factors Of Spatial Data Sharing Of Local Spatial Data Infrastructure In Indonesia. *Geomatika*, *24*(1), 1–12.

Athanasiou, S., Georgomanolis, N., Patroumpas, K., Alexakis, M., & Stratiotis, T. (2015). TripleGeo-CSW: A middleware for exposing geospatial catalogue services on the semantic web. *CEUR Workshop Proceedings*, *1330*, 229–236.

Austin, J., & Delaney, P. F. (1998). Protocol Analysis as a Tool for Behavior Analysis. *The Analysis of Verbal Behavior*, *15*(1), 41–56. https://doi.org/10.1007/bf03392922

Batcheller, J., Gittings, B., & Dunfey, R. (2009). A Method for Automating Geospatial Dataset Metadata. *Future Internet*, *1*(1), 28–46. https://doi.org/10.3390/fi1010028

Batcheller, J. K. (2008). Automating geospatial metadata generation-An integrated data management and documentation approach. *Computers and Geosciences*, *34*(4), 387–398. https://doi.org/10.1016/j.cageo.2007.04.001

Belward, A. S., & Skøien, J. O. (2015). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*, *103*, 115–128.

Berg, B. L., & Lune, H. (2012). Qualitative Research Methods for the Social Sciences p. 408. *Harlow, Essex, England: Pearson*.

Bevan, N. (2009a). International standards for usability should be more widely used. *Journal of Usability Studies*, *4*(3), 106–113.

Bevan, N. (2009b). What is the difference between the purpose of usability and user experience evaluation methods. *Proceedings of the Workshop UXEM*, *9*, 1–4.

Bielecka, Elżbieta. (2015). Ocena primernosti uporabe prostorskih podatkovnih nizov. *Geodetski Vestnik*, *59*(2), 335–348. https://doi.org/10.15292/geodetski-vestnik.2015.02.335-348

Bielecka, Elzbieta, Leszczyńska, M., & Halls, P. J. (2014). User perspective on geospatial data quality. Case study of the Polish Topographic Database. *9th International Conference on Environmental Engineering, ICEE 2014*, *May*. https://doi.org/10.3846/enviro.2014.193

Bishop, W., & Grubesic, T. H. (2016). Geographic information, maps, and GIS. In *Geographic information* (pp. 11–25). Springer .

Boin, A., & Hunter, G. (2008). What Communicates Quality to the Spatial Data Consumer? *Quality Aspects in Spatial Data Mining*, *34*, 285–296. https://doi.org/10.1201/9781420069273.sec5

Brovelli, M. A., Boccardo, P., Bordogna, G., Pepe, A., Crespi, M., Munafò, M., & Pirotti, F. (2019). Urban Geo Big Data. *2019 Free and Open Source Software for Geospatial, FOSS4G 2019*, *42*(4), 23–30.

Brown, M., Sharples, S., Harding, J., Parker, C. J., Bearman, N., Maguire, M., Forrest, D., Haklay, M., & Jackson, M. (2013). Usability of Geographic Information: Current challenges and future directions. *Applied Ergonomics*, *44*(6), 855–865. https://doi.org/10.1016/j.apergo.2012.10.013

Bryman, A., & Cramer, D. (2004). *Quantitative data analysis with SPSS 12 and 13: A guide for social scientists*. Routledge.

Bulens, J., Vullings, W., Houtkamp, J., & Vanmeulebrouk, B. (2013). Usability of Discovery Portals. *16th AGILE Conference on Geographic Information Science*, 1–5. https://agile-online.org/index.php/conference/proceedings/proceedings-2013

Caulton, D. A. (2001). Relaxing the homogeneity assumption in usability testing. *Behaviour and Information Technology*, *20*(1), 1–7. https://doi.org/10.1080/01449290010020648

Chafiq, T., Ouadoud, M., Oulidi, H. J., Fekri, A., & Elaloui, A. (2018). Towards SDI Services for Geological Maps Data. *International Conference on Advanced Intelligent Systems for Sustainable Development*, 34–44.

Chan, T. O., Feeney, M. E., Rajabifard, A., & Williamson, I. (2001). The dynamic nature of spatial data infrastructures: A method of descriptive classification. *Geomatica*, *55*(1), 65–72.

Chen, Z., & Yang, Y. (2019). Semantic relatedness algorithm for keyword sets of geographic metadata. *Cartography and Geographic Information Science*, 1–16.

Choy, L. T. (2014). The strengths and weaknesses of research methodology: Comparison and complimentary between qualitative and quantitative approaches. *IOSR Journal of Humanities and Social Science*, *19*(4), 99–104.

Comber, A. J., Fisher, P. F., & Wadsworth, R. A. (2007). User-focused metadata for spatial data, geographical information and data quality assessments. *10th AGILE International Conference on Geographic Information Science*, 1–13. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.7278&rep=rep1&type=pdf

Committee, F. G. D. (1998). Content Standard for Digital Geospatial Metadata (CSDGM) Version 2-FGDC-STD-001-1998. *FGDC, USGS*, *590*.

Cooper, A. K., Moellering, H., Hjelmager, J., Rapant, P., Delgado, T., Laurent, D., Coetzee, S., Danko, D. M., Düren, U., Iwaniak, A., Brodeur, J., Abad, P., Huet, M., & Rajabifard, A. (2013). A spatial data infrastructure model from the computational viewpoint. *International Journal of Geographical Information Science*, *27*(6), 1133–1151. https://doi.org/10.1080/13658816.2012.741239

Corti, P, & Lewis, B. (2017). Making temporal search more central in spatial data infrastructures. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, 93.

Corti, Paolo, Kralidis, A. T., & Lewis, B. (2018). Enhancing discovery in spatial data infrastructures using a search engine. *PeerJ Computer Science*, *4*, e152.

Craglia, M., & Shanley, L. (2015a). Data democracy–increased supply of geospatial information and expanded participatory processes in the production of data. *International Journal of Digital Earth*, *8*(9), 679–693.

Craglia, M., & Shanley, L. (2015b). Data democracy – increased supply of geospatial information and expanded participatory processes in the production of data. *International Journal of Digital Earth*, *8*(9), 679–693. https://doi.org/10.1080/17538947.2015.1008214

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. SAGE publications.

Danko, D. M. (2011). Geospatial metadata. In *Springer handbook of geographic information* (pp. 191–244). Springer.

De Martino, M., & Albertoni, R. (2011). A multilingual / multicultural semantic-based approach to improve Data Sharing in an SDI for Nature Conservation. *International Journal of Spatial Data Infrastructures Research*, *6*, 206–233. https://doi.org/10.2902/1725-0463.2011.06.art10

de Vries, W. (2006). Why Local Spatial Data Infrastructures (SDI's) are not Just Mirror Reflections of National SDI Objectives--Case Study of Bekasi, Indonesia. *The Electronic Journal of Information Systems in Developing Countries*, *27*(1), 1–28.

Devillers, R., Bédard, Y., Jeansoulin, R., & Moulin, B. (2007). Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science*, *21*(3), 261–282. https://doi.org/10.1080/13658810600911879

Devillers, Rodolphe, Gervais, M., Bédard, Y., & Jeansoulin, R. (2002). Spatial Data Quality : From Metadata to Quality Indicators and Contextual End - User Manual. *OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management*, *March 2002*, 45–55. https://doi.org/10.1.1.460.4972

Dhanda, A., Remondino, F., & Santana Quintero, M. (2018). *A metadata based approach for analyzing UAV datasets for photogrammetric applications*.

Di, L., Moe, K. L., & Yu, G. (Eugene). (2009). Metadata requirements analysis for the emerging sensor web. *International Journal of Digital Earth*, *2*(April 2008), 3–17. https://doi.org/10.1080/17538940902866195

Download GeoBuiz 2018 Report - Global Geospatial Outlook 2018 is here. (2018). https://geobuiz.com/geobuiz-2018-report.html

Drafting Team Metadata and European Commission Joint Research Centre. (2013). *INSPIRE Metadata Implementing Rules : Technical Guidelines based on EN ISO 19115 and EN ISO 19119 - V.1.3.* 99. https://doi.org/MD_IR_and_ISO_20090218

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* the MIT Press.

Evans, B. J. (1997). Dynamic display of spatial data-reliability: Does it benefit the map user? *Computers and Geosciences*, *23*(4), 409–422. https://doi.org/10.1016/S0098-3004(97)00011-3

Fan, J., Yan, J., Ma, Y., & Wang, L. (2018). Big data integration in remote sensing across a distributed metadata-based spatial infrastructure. *Remote Sensing*, *10*(1), 7.

Ferreira, S. M., & Pithan, D. N. (2005). Usability of digital libraries: A study based on the areas of information science and human-computer-interaction. *OCLC Systems and Services*, *21*(4), 311–323. https://doi.org/10.1108/10650750510631695

FGDC (1998). *Content Standard for Digital Geospatial Metadata*.

Fisher, P. F., Goodchild, M. F., & Jakobsson, A. (2010). Data Quality and Quality Management -Examples of Quality Evaluation Procedures and Quality Management in European National Mapping Agencies. *Spatial Data Quality*, 216–229. https://doi.org/10.4324/9780203303245_chapter_fifteen

Fonteyn, M. E., Kuipers, B., & Grobe, S. J. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research*, *3*(4), 430–441.

Fraser, B., & Gluck, M. (1999). Usability of geospatial metadata or space-time matters. In *Bulletin of the American Society for Information Science* (Vol. 25, Issue 6, pp. 24–28).

Fugazza, C., Pepe, M., Oggioni, A., Tagliolato, P., & Carrara, P. (2016). Streamlining geospatial metadata in the Semantic Web. *IOP Conference Series: Earth and Environmental Science*, *34*(1), 0–8. https://doi.org/10.1088/1755-1315/34/1/012009

Gilliland, A. J. (2008). Setting the stage. *Introduction to Metadata*, *2*, 1–19.

Göbel, S. (2001). Visualization techniques in metadata information systems for geospatial data. *Advances in Environmental Research*, *5*(4), 415–424. https://doi.org/10.1016/S1093-0191(01)00093-4

Göbel, S., & Lutze, K. (1998). Development of meta databases for geospatial data in the WWW. *Proceedings of the 6th ACM International Symposium on Advances in Geographic Information Systems*, 94–99.

Goertzen, M. J. (2017). Applying quantitative methods to research and data. *Library Technology Reports*, *53*(4), 12–18. https://journals.ala.org/index.php/ltr/article/view/6325

Goodchild, M. (2008). *Spatial Accuracy 2.0*. http://www.geog.ucsb.edu/~good/papers/453.pdf

Goodchild, M. F. (2007). Beyond Metadata: Towards User-Centric Description Of Data Quality Michael F. Goodchild 1. *Quality*, 1–11. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.162.1404&rep=rep1&type=pdf

Gray, D. E. (2014). Theoretical perspectives and research methodologies. *Doing Research in the Real World*, *752*.

Greenberg, J. (2003). Metadata and the world wide web. *Encyclopedia of Library and Information Science*, *3*, 1876–1888.

GSDICookbook. (2004). Developing Spatial Data Infrastructures: The SDI Cookbook. *GSDI/Nebert, January*.

Guarda, T., & Augusto, M. F. (2019). *Geographic Marketing Intelligence: GMI Model*.

Guarda, T., Augusto, M. F., & Lopes, I. (2019). Geographic market intelligence as a competitive advantage. *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–5.

Guptill, S. C. (1999). Metadata and data catalogues. *Geographical Information Systems: Principles and Technical Issues*, *1*(2), 677–692.

Hackos, J. T., & Redish, J. (1998). *User and task analysis for interface design*.

Han, H., Giles, C. L., Manavoglu, E., Zha, H., Zhang, Z., & Fox, E. A. (2003). Automatic document metadata extraction using support vector machines. *2003 Joint Conference on Digital Libraries, 2003. Proceedings.*, 37–48.

Hee Kim, H., & Ho Kim, Y. (2008). Usability study of digital institutional repositories. *Electronic Library*, *26*(6), 863–881. https://doi.org/10.1108/02640470810921637

Henry, S. B., Lebreck, D. B., & Holzemer, W. L. (1989). The effect of verbalization of cognitive processes on clinical decision making. *Research in Nursing & Health*, *12*(3), 187–193.

Hill, L. L. (1999). Collection metadata solutions for digital library applications. *Journal of the American Society for Information Science*, *50*(13), 1169–1181. https://doi.org/10.1002/(SICI)1097-4571(1999)50:13<1169::AID-ASI3>3.0.CO;2-3

Howe, M. A., McGreevy, R. L., & Zetterström, P. (1996). CORRECT: A correction program for neutron diffraction data. In *NFL Studsvik internal report*. NFL, Uppsala University.

Imran, M. (2019). Enabling Crowdsourcing in the Framework of User-centred SDIs for Information Management of Geographical Volunteer Content. *2019 5th International Conference on Information Management (ICIM)*, 7–12.

Ivánová, I., Morales, J., de By, R. A., Beshe, T. S., & Gebresilassie, M. A. (2013). Searching for spatial data resources by fitness for use. *Journal of Spatial Science*, *58*(1), 15–28. https://doi.org/10.1080/14498596.2012.759087

Iwaniak, A., Kaczmarek, I., Łukowicz, J., Strzelecki, M., Coetzee, S., & Paluszyński, W. (2016). Semantic Metadata For Heterogeneous Spatial Planning Documents. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *3*(1).

Jiang, Y., Li, Y., Yang, C., Hu, F., Armstrong, E. M., Huang, T., Moroni, D., McGibbney, L. J., & Finch, C. J. (2018). Towards intelligent geospatial data discovery: a machine learning framework for search ranking. *International Journal of Digital Earth*, *11*(9), 956–971.

Jokela, S. (2001). Jokela S. Metadata Enhanced Content Management in Media Companies. *Acta Polytechnica Scandinavica, Mathematics and Computing Series*, *155*(114), pp. https://aaltodoc.aalto.fi/bitstream/handle/123456789/2361/isbn9512256932.pdf?sequence=1

Jung, T., Do, E. Y.-L., & Gross, M. D. (1999). Immersive redlining and annotation of 3D design models on the web. In *Computers in Building* (pp. 81–98). Springer.

Kaláb, Z., Čápová, D., Moravcová, O., & Kondrová, L. (2019). Metadata of the geophysical data sources in the Czech Republic: Case study in the IGN. *18th International Conference on Geoinformatics-Theoretical and Applied Aspects*, *2019*(1), 1–5.

Kalantari, M., Rajabifard, A., & Olfat, H. (2009). Spatial metadata automation: a new approach. *Spatial Science Conference*, 629–635.

Kalantari, M., Rajabifard, A., Olfat, H., Pettit, C., & Keshtiarast, A. (2017). Automatic spatial metadata systems–the case of Australian urban research infrastructure network. *Cartography and Geographic Information Science*, *44*(4), 327–337. https://doi.org/10.1080/15230406.2016.1154805

Kalantari, M., Rajabifard, A., Olfat, H., & Williamson, I. (2014). Geospatial Metadata 2.0 - An approach for Volunteered Geographic Information. *Computers, Environment and Urban Systems*, *48*, 35–48. https://doi.org/10.1016/j.compenvurbsys.2014.06.005

Kardos, J., Moore, A., & Benwell, G. (2005). *Visualising attribute and spatial uncertainty in choropleth maps using hierarchical spatial data models*.

Kharisma, I., & Sagala, S. (2018). Implementation of Spatial Data Infrastructure in Cianjur District for Disaster Risk Management Purposes. *IOP Conference Series: Earth and Environmental Science*, *145*(1), 12013.

Kowal, K. C., & Prbreveidal, P. (2012). Online georeferencing for libraries: The british library implementation of georeferencer for spatial metadata enhancement and public engagement. *Journal of Map and Geography Libraries*, *8*(3), 276–289. https://doi.org/10.1080/15420353.2012.700914

Kyzirakos, K., Savva, D., Vlachopoulos, I., Vasileiou, A., Karalis, N., Koubarakis, M., & Manegold, S. (2018). GeoTriples: Transforming geospatial data into RDF graphs using R2RML and RML mappings. *Journal of Web Semantics*, *52*, 16–32.

Lacasta, J., Nogueras-Iso, J., Béjar, R., Muro-Medrano, P. R., & Zarazaga-Soria, F. J. (2007). A Web Ontology Service to facilitate interoperability within a Spatial Data Infrastructure: Applicability to discovery. *Data and Knowledge Engineering*, *63*(3), 947–971. https://doi.org/10.1016/j.datak.2007.06.002

Lafia, S., Xiao, J., Hervey, T., & Kuhn, W. (2019). Talk of the Town: Discovering Open Public Data via Voice Assistants (Short Paper). *14th International Conference on Spatial Information Theory (COSIT 2019)*.

Lee, S. (2005). Jenny Preece, Yvonne Rogers, and Helen Sharp (Eds.). *Information Design Journal*, *13*(3), 264–266.

Li, M., Guo, W., Duan, L., & Zhu, X. (2017). A case-based reasoning approach for task-driven spatial--temporally aware geospatial data discovery through geoportals. *International Journal of Digital Earth*, *10*(11), 1146–1165.

Limbach, T., Reiterer, H., Klein, P., & Müller, F. (2003). Visualizing Metadata: LevelTable vs. GranularityTable in the SuperTable/Scatterplot Framework. *Human-Computer Interaction: Theory and Practice (Part II)*, *2*, 1106–1110.

Litwin, L., & Rossa, M. (2011). Geoinformation metadata in INSPIRE and SDI: Understanding. Editing. Publishing. *Lecture Notes in Geoinformation and Cartography*, *9783642158612*, iii–iv. https://doi.org/10.1007/978-3-642-15862-9

Mao, B. J., Vredenburg, K., Smith, P. W., & Carey, T. (n.d.). *User-centered design practice*. *48*(3), 105–110.

Mao, J. Y., Vredenburg, K., Smith, P. W., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*, *48*(3), 105–109. https://doi.org/10.1145/1047671.1047677

Masser, I. (2005). *GIS worlds: creating spatial data infrastructures* (Vol. 338). Esri Press Redlands, CA.

Mathys, T. (2005). The GO-GEO! Spatial Data Portal: a Data Discovery and Research Tool for the Social Sciences in UK Academia. *ASSIGNATION-LONDON-*, *22*(4), 15.

Maxwell, J. A. (2004). Using qualitative methods for causal explanation. *Field Methods*, *16*(3), 243–264.

Najar, C., Rajabifard, A., Williamson, I., & Giger, C. (2007). A Framework for Comparing Spatial Data Infrastructures: An Australian-Swiss Case Study. *Research and Theory in Advancing Spatial Data Infrastructure Concepts*, *5*, 202–213. https://minerva-access.unimelb.edu.au/bitstream/handle/11343/26675/117527_framework.pdf?sequence=1&isAllowed=y

Neumaier, S., Savenkov, V., & Polleres, A. (2018). Geo-semantic labelling of open data. *Procedia Computer Science*, *137*, 9–20.

Nielsen, J. (1993). *Usability Engineering*. Academic Press Limited.

Nielsen, J. (2010). What Is Usability? *User Experience Re-Mastered*, *September*, 3–22. https://doi.org/10.1016/b978-0-12-375114-0.00004-9

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231.

NISO. (2014). *Understanding M Etadata What Is Metadata ,*. http://www.niso.org/apps/group_public/download.php/17446/Understanding Metadata.pdf

Norlund, P. (2010). *Automatic and semi-automatic methods for metadata creation and maintenance: long term implementation of the INSPIRE directive*. *August*. http://www.diva-portal.org/smash/record.jsf?pid=diva2:383048

Nunamaker Jr, J. F., Chen, M., & Purdin, T. D. M. (1990). Systems development in information systems research. *Journal of Management Information Systems*, *7*(3), 89–106.

OGC. (n.d.). *No Title*. Retrieved August 27, 2019, from https://www.opengeospatial.org/ogc/glossary/g

Olfat, H. (2013). Automatic Spatial Metadata Updating and Enrichment. *PhD Thesis*, *January*. http://www.csdila.unimelb.edu.au/publication/theses/

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H., & Williamson, I. (2012). Spatial Metadata Automation: A Key to Spatially Enabling Platform. *International Journal of Spatial Data Infrastructures Research*, *7*, 173–195. https://doi.org/10.2902/ijsdir.v7i0.279

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H., & Williamson, I. P. (2013). A GML-based approach to automate spatial metadata updating. *International Journal of Geographical Information Science*, *27*(2), 231–250. https://doi.org/10.1080/13658816.2012.678853

Olfat, H., Kalantari, M., Rajabifard, A., Williamson, I. P., Pettit, C., & Williams, S. (2010). *Exploring the key areas of spatial metadata automation research in Australia*. https://ezp.lib.unimelb.edu.au/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=ir00004a&AN=irunimelb.265613&site=eds-live&scope=site%5Cnhttp://repository.unimelb.edu.au/10187/9210

Patroumpas, K., Georgomanolis, N., Stratiotis, T., Alexakis, M., & Athanasiou, S. (2015). Exposing INSPIRE on the Semantic Web. *Journal of Web Semantics*, *35*, 53–62. https://doi.org/10.1016/j.websem.2015.09.003

Patroumpas, Kostas, Alexakis, M., Giannopoulos, G., & Athanasiou, S. (2014). TripleGeo: An ETL tool for transforming geospatial data into RDF triples. *CEUR Workshop Proceedings*, *1133*(c), 275–278.

Phillips, A., Williamson, I., & Ezigbalike, C. (1999). Spatial data infrastructure concepts. *Australian Surveyor*, *44*(1), 20–28. https://doi.org/10.1080/00050351.1999.10558768

Pierce, L. (2019). Progress in SAR Metadata Standards. *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, 4435–4437.

Potter, J. (1997). Discourse analysis as a way of analysing naturally occurring talk. *Qualitative Research: Theory, Method and Practice*, *2*, 200–222.

PSKIG-BIG (2018). *Outlook 2018*.

Rautenbach, V., Coetzee, S., Çöltekin, A., Pettit, C., Pijper, L., Madden, M., Christophe, S., & Lkhamjav, O. (2019). TOWARDS ESTABLISHING AN OPEN CATALOGUE FOR GEOSPATIAL EDUCATIONAL RESOURCES. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Revi, A. (2014). *The post-2015 development agenda and the importance of the urban SDG: expert meeting on regional preparations for Habitat III. Jakarta: UNGGIM; 25 November 2014*.

Salk, C. F., Sturn, T., See, L., Fritz, S., & Perger, C. (2016). Assessing quality of volunteer crowdsourcing contributions: lessons from the Cropland Capture game. *International Journal of Digital Earth*, *9*(4), 410–426.

Schauppenlehner, T., & Muhar, A. (2018). Theoretical Availability versus Practical Accessibility: The Critical Role of Metadata Management in Open Data Portals. *Sustainability*, *10*(2), 545.

Seale, C. (2004). *Social research methods: A reader*. Psychology Press.

Smits, J. (2009). *Spatial Metadata : An International Survey on Clearinghouses and Infrastructures Spatial Metadata : An International Survey on Clearinghouses and Infrastructures*. *9374*(December 2014), 37–41. https://doi.org/10.1300/J104v27n03

Soltanieh, K., Entry, M., Rajabifard, A., Kalantari, M., & Binns, A. (2009). *Minerva Access is the Institutional Repository of The University of Melbourne Author / s : Title : SDI and Metadata Entry and Updating Tools Date : Citation : Publication Status : Persistent Link : File Description : SDI and metadata entry and updating to*.

Specka, X., Gärtner, P., Hoffmann, C., Svoboda, N., Stecker, M., Einspanier, U., Senkler, K., Zoarder, M. A. M., & Heinrich, U. (2019). The BonaRes metadata schema for geospatial soil-agricultural research data--Merging INSPIRE and DataCite metadata schemes. *Computers & Geosciences*, *132*, 33–41.

ISO (1998). *ISO 9241-11. 1998 Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability.*

ISO (2014). *ISO 19115-1:2014 Geographic Information - Metadata*.

Sutanta, H., Aditya, T., Budi, P., Diyono, D. L., & Farida, A. (2014). I-SRI, an SDI Readiness Index for Local Government in Indonesia. *Paper Presented at FIG Congress, Kuala Lumpur*, 15–21.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Sage.

Timpf, S., Raubal, M. M., & Kuhn, W. (1997). Experiences with metadata. *Proceedings of the 7th International Symposium on Spatial Data Handling*.

Toro Herrera, J. F., Carrion, D., Albertella, A., & Brovelli, M. A. (2019). Cross-Border Open Data Sharing: Gioconda Project. *2019 Free and Open Source Software for Geospatial, FOSS4G 2019*, *42*(4), 233–238.

Tsou, M. H. (2002). An operational metadata framework for searching, indexing, and retrieving distributed geographic information services on the internet. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2478*, 313–332.

Ureña-Cámara, M. A., Nogueras-Iso, J., Lacasta, J., & Ariza-López, F. J. (2019). A method for checking the quality of geographic metadata based on ISO 19157. *International Journal of Geographical Information Science*, *33*(1), 1–27.

Vredenburg, K., Smith, P. W., & Carey, T. (2002). Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, Changing Ourselves, CHI '02. *Conference on Human Factors in Computing Systems - Proceedings*, *1*, 471–478.

Ware, A. (1988). *Citizens, Parties and the State: a reappraisal. Princeton*.

West, L. A., & Hess, T. J. (2002). Metadata as a knowledge management tool: Supporting intelligent agent and end user access to spatial data. *Decision Support Systems*, *32*(3), 247–264. https://doi.org/10.1016/S0167-9236(01)00102-6

Wigton, R. S., & McGaghie, W. C. (2001). The effect of obesity on medical students' approach to patients with abdominal pain. *Journal of General Internal Medicine*, *16*(4), 262–265.

Woodcock, J., Franombe, J., Franck, A., Coleman, P., Hughes, R., Kim, H., Liu, Q., Menzies, D., Simón Gálvez, M. F., Tang, Y., & others. (2018). A framework for intelligent metadata adaptation in object-based audio. *Audio Engineering Society Conference: 2018 AES International Conference on Spatial Reproduction-Aesthetics and Science*.

Yong, W., Shuaishuai, L., Li, L., Minzan, L., Ming, L., Arvanitis, K. G., Georgieva, C., & Sigrimis, N. (2018). Smart sensors from ground to cloud and web intelligence. *IFAC-PapersOnLine*, *51*(17), 31–38.

Yuliadi, T., Aditya, T., & Vries, W. (2011). A local spatial data infrastructure to support the merapi volcanic risk managment: A case study at Selman regency. *Indonezia. Indonesian Journal of Geography*, *43*(1), 25–48.

Zenus, M. (2018). *Is IoT data taking geospatial analytics to the next level?* https://www.geospatialworld.net/article/is-iot-data-taking-geospatial-analytics-to-the-next-level/

This page is left intentionally blank

Appendix 1 – The Metadata usability Survey

Metadata usability survey

Thank you for your interest in participating in this survey. This survey is part of PhD study at the Centre for Spatial Data Infrastructure and Land Administration (CSDILA), the University of Melbourne which is seeking to improve the usability of metadata for spatial data through a more user-centric approach. This survey intends to identify user requirements and expectations in spatial data discovery and selection processes. The collected data will be used as the basis for creating user-oriented metadata specification. Such specification will help to create more user-friendly metadata that will help spatial data users effectively and efficiently discover and select spatial data from web catalogs or data portals.

This survey should not take longer than 25 minutes to complete. The questions are of a purely informational/factual nature and there will be no direct or indirect financial interest for the researchers in this study. The information you provide in the survey will remain confidential and anonymous, the analysis outputs will be used in my PhD research. The information will be safeguarded, subject to any legal requirements. Should you wish to withdraw at any stage, or to withdraw any unprocessed data you have supplied, you are free to do so without prejudice.

Should you require any further information, or have any concerns, please do not hesitate to contact Syahrudin, at: ssyahrudin@student.unimelb.edu.au or Dr. Mohsen Kalantari, at: mohsen.kalantari@unimelb.edu.au.

This research has been granted clearance by Human Research Ethics Committee with Ethics No. 1545774.3. Should you have any concerns about the conduct of this research, you can contact the Executive Officer for Human Research Ethics at The University of Melbourne by phone +61 3 8344 2073 or fax +61 3 9347 6739.

Please note that this survey will be closed on Thursday, November 15, 2018.

Thank you for your support and contribution.

Kind Regards,
Syahrudin

PhD Candidate
The Centre for Spatial Data Infrastructure and Land Administration
Melbourne School of Engineering
The University of Melbourne
Australia

* Required

Consent Form

I consent to voluntarily participate in this survey, the details of which have been explained to me, and I have been provided with a written plain language statement.

I understand that if I agree to this consent form it will be retained by the researcher.

I understand that my participation will involve in a questionnaire, and I agree that the researcher may use the results as described in the statement.

I acknowledge that the possible effects of participating in the testing and the questionnaire have been explained to my satisfaction.

I have been informed that I am free to withdraw from the project at any time without explanation or prejudice and to withdraw any unprocessed data I have provided.

I have been informed that the tasks and questions are of a purely informational/factual nature, and there will be no direct and indirect financial interest for the researchers in this study.

I have been informed that the confidentiality of the information I provide will be safeguarded, subject to any legal requirements.

I have been informed that my personal details will be kept confidential.

1. **I have read and agreed to the consent information above.** *

Check all that apply.

- [ ] Yes

Participant Information

2. **1. Which of the following is your current field of work? Please specify if you select other.** *

Mark only one oval.

- Planning (Urban, Landscape Architecture)
- Land management and Cadastre
- GIS or spatial services
- Civil engineering
- Transportation
- (Public) health
- Environment
- Disaster management
- Oil and gas or mining
- Consumer goods
- Other: ____________________

3. **2. Please indicate which of the following statements describes your level of spatial knowledge. Please specify if you select other.** *

Mark only one oval.

- I have never worked with spatial data or information before
- I work with spatial data, but I do not have academic or professional training in spatial science
- I work with spatial data and have a general understanding of spatial knowledge
- I work with spatial data and have extensive understanding of the spatial knowledge at high level
- Other: ____________________

4. **3. Please indicate your experience in your current field.** *

Mark only one oval.

- 0-3 years
- 3-5 years
- 5-10 years
- > 10 years

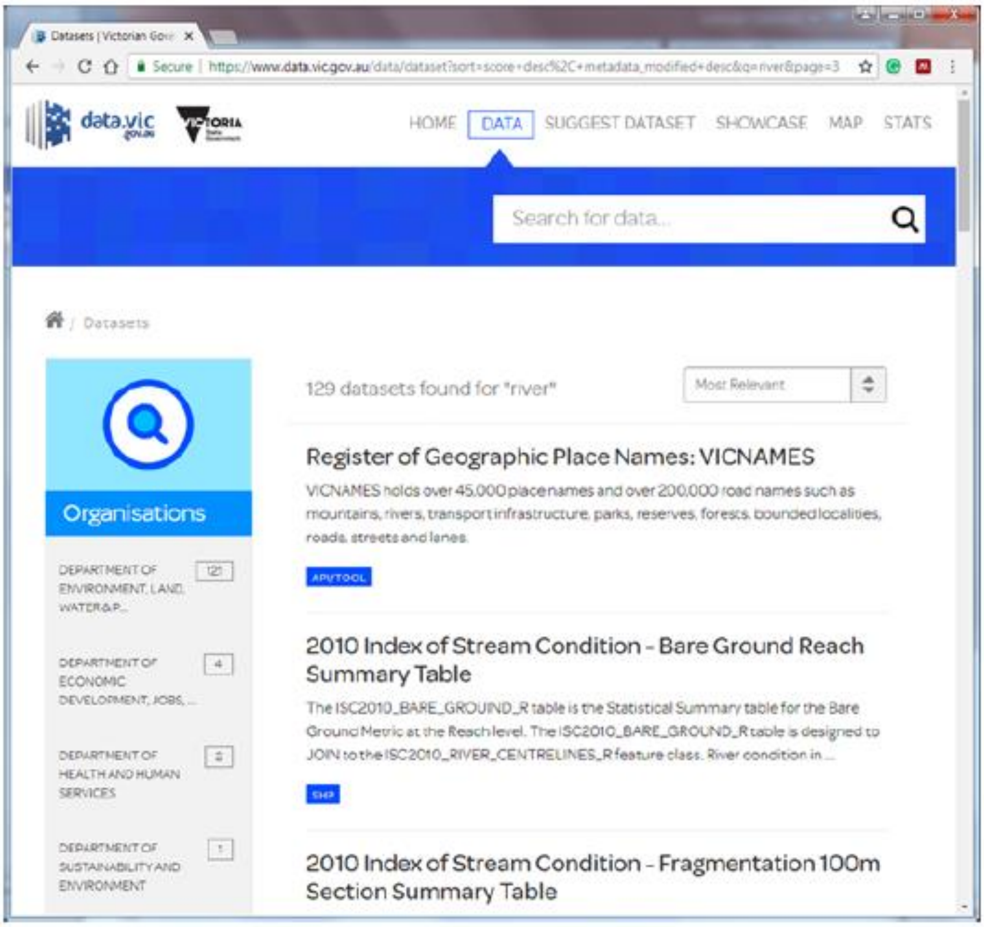
5. **4. In what country(s) you mostly work?** *

6. **5. Which type of spatial data do you work with?** *

Mark only one oval.

- Vector
- Raster/Image
- Both
- Neither

Figure 1. Example of a data portal (http://data.vic.go.au/data)



7. **6. Spatial data portals or web catalogs are internet-based applications (websites) that have capabilities to help their users search and find spatial data using metadata, as shown in Figure 1. Do you have any experience in searching for spatial data from web (internet-based) spatial data catalogs or portals? Please explain if you select other.** *

Mark only one oval.

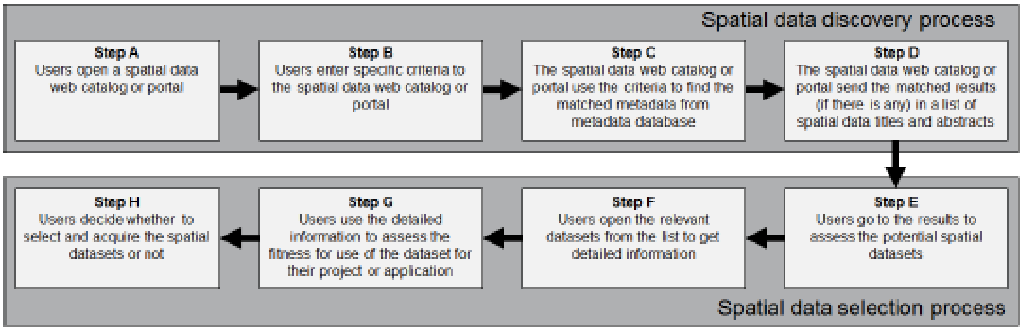
- No, I have never searched for spatial data from web spatial data catalogs or portals
- Yes, I have searched for spatial data from web catalogs for some projects in the past
- Yes, I regularly search for spatial data from web data catalogs or data portals
- Other: ______________________

Spatial data search

Spatial data discovery and selection are important activities for searching and acquiring spatial data based on fitness for purpose of the spatial data for different projects or applications. Figure 2 describes a generic process of spatial data discovery and selection using spatial data web catalogs or portals. Step A to Step D are activities conducted by spatial data users for spatial data discovery. Spatial data selection is carried out by conducting Step E to Step H.

Please answer the following questions in the context of Figure 2.

Figure 2. Generic spatial data searching process using spatial data web catalogs or portals



8. **7. When you search for specific spatial data (Step B in the Figure 2), which of the following do you use as your search criteria? Please select all that apply and please specify if you select other.** *

Check all that apply.

- [ ] Content of dataset (e.g. roads, rivers, buildings, etc.)
- [ ] Geographic (location) coverage/extent
- [ ] Production date
- [ ] Maintenance date (last update)
- [ ] Spatial scale
- [ ] Spatial resolution
- [ ] Data format
- [ ] Data producer
- [ ] Other: ______________________

9. **8. If you have any preferred presentation style for each of the criteria (e.g. drop-down list, free-hand area boundary, etc.), please specify and explain.**

10. **9. Are these criteria different for vector and raster (image) datasets?** *

Mark only one oval.

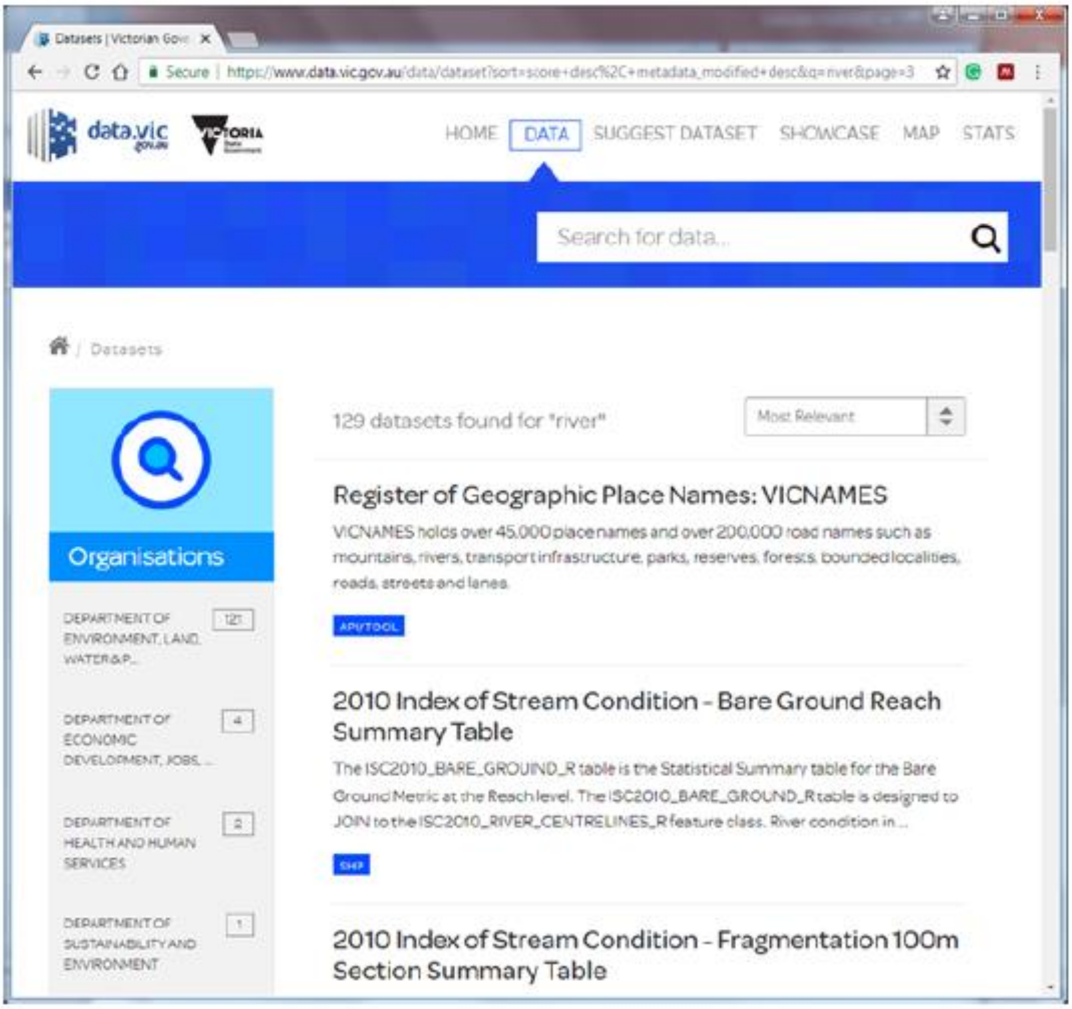
- Yes
- No
- I am not sure

11. **10. Please explain your answer in the question 9.** *

When you search for specific spatial datasets using your own criteria, the web data catalogs or portals would normally response by giving your searching results in a list of titles and a few lines of abstracts, as shown in Figure 3.

Please answer the following questions in the context of Figure 3.

Figure 3. An example of a search result from Victoria data portal using the word "River" as the criteria



12. **11. If you want to make a quick assessment of fitness-for-purpose (suitability) of the provided datasets in the list based on their titles, as can be seen in Figure 3, what information would you prefer to see in the titles? Please select all that apply and please specify if you select other.** *

Check all that apply.

- [ ] Content of dataset (e.g. roads, rivers, buildings, etc.)
- [ ] Geographic (location) coverage/extent
- [ ] Production date
- [ ] Maintenance date (last update)
- [ ] Spatial scale
- [ ] Spatial resolution
- [ ] Data format
- [ ] Other: ____________________

13. **12. How many words do you think is necessary for a title?** *

Mark only one oval.

- Up to 5 words
- Up to 10 words
- Up to 15 words
- Up to 20 words
- No limit

14. **13. Should the information on the title be different between vector and raster (image) datasets?** *

Mark only one oval.

- Yes
- No
- I am not sure

15. **14. Please explain your answer for the question 13.** *

If you need to read the full abstract of the data from the list, you need to click on a title, or in some portals you can click on the "show more" button at the end of the presented abstract. The portal will provide you with the full abstract and some information from the metadata that might help you assess the fitness for use of the data. Figure 4 and Figure 5 are examples of full abstracts and some elements of metadata provided in spatial data portals when you click on a title from the list.

Please answer the following questions in the context of Figure 4 and Figure 5.

Figure 4. A full abstract of a spatial dataset from Victoria data portal

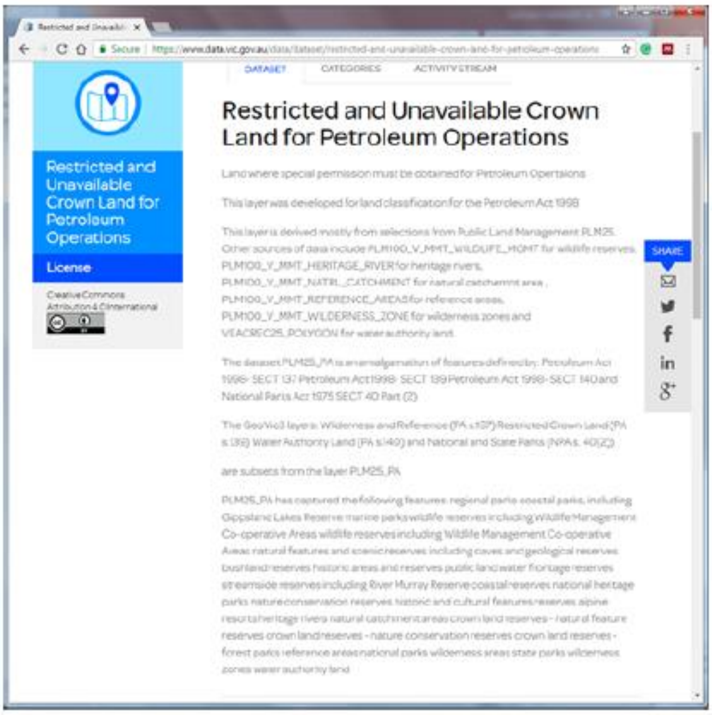
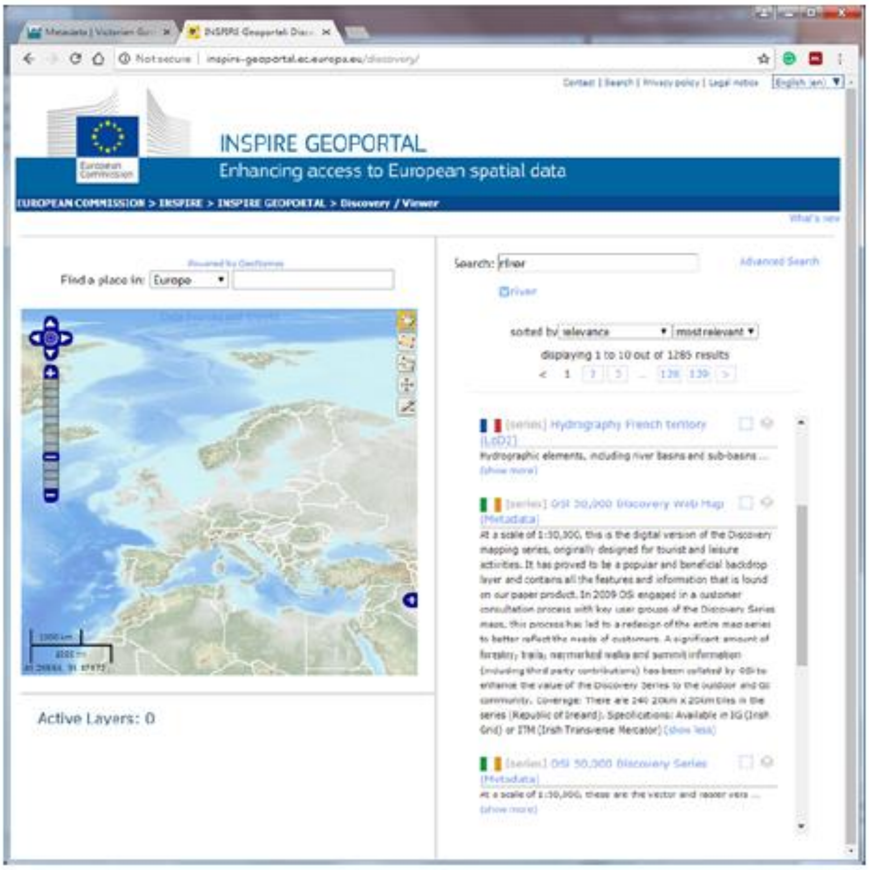


Figure 5. A full abstract of a spatial dataset from INSPIRE discovery portal



16. **15. If you want to make a quick assessment of the fitness-for-purpose (suitability) of the datasets based on their abstracts, which of the followings would you prefer ? Please select all that apply and please specify if you select other.** *

Check all that apply.

- [ ] Content description (explanation about the feature and its attributes)
- [ ] Detailed explanation about geographic (location) coverage/extent
- [ ] Production date
- [ ] Maintenance date
- [ ] Spatial scale
- [ ] Spatial resolution
- [ ] Data accuracy or error
- [ ] Level of detail or generalization of the feature presented in the data (e.g. the lowest road type)
- [ ] Data format
- [ ] Production process
- [ ] Intended use of the data
- [ ] Other: ____________________

17. **16. How many words do you think an ideal abstract would contain?** *
Mark only one oval.

- Up to 50 words
- Up to 100 words
- Up to 150 words
- Up to 200 words
- Up to 250 words
- Up to 400 words
- No limit

18. **17. Should the information on the abstract be different between vector and raster (image) datasets?** *
Mark only one oval.

- Yes
- No
- I am not sure

19. **18. Please explain your answer for the question 17.** *

20. **19. In addition to above details about each dataset in the result list (i.e. title and abstract with occasional indication of the format of the dataset, as shown in Figure 3), is there any additional information that can help you quickly identify most relevant datasets from the list without clicking the link and going to the detailed information about the spatial datasets?** *
Mark only one oval.

- Yes
- No

21. **20. If your answer to the question 19 is "yes", please specify and provide a preferred presentation method (e.g. textual or graphical) for each piece of information.**

Figure 6. A simplified viewer metadata in Victoria data portal from its respective metadata record

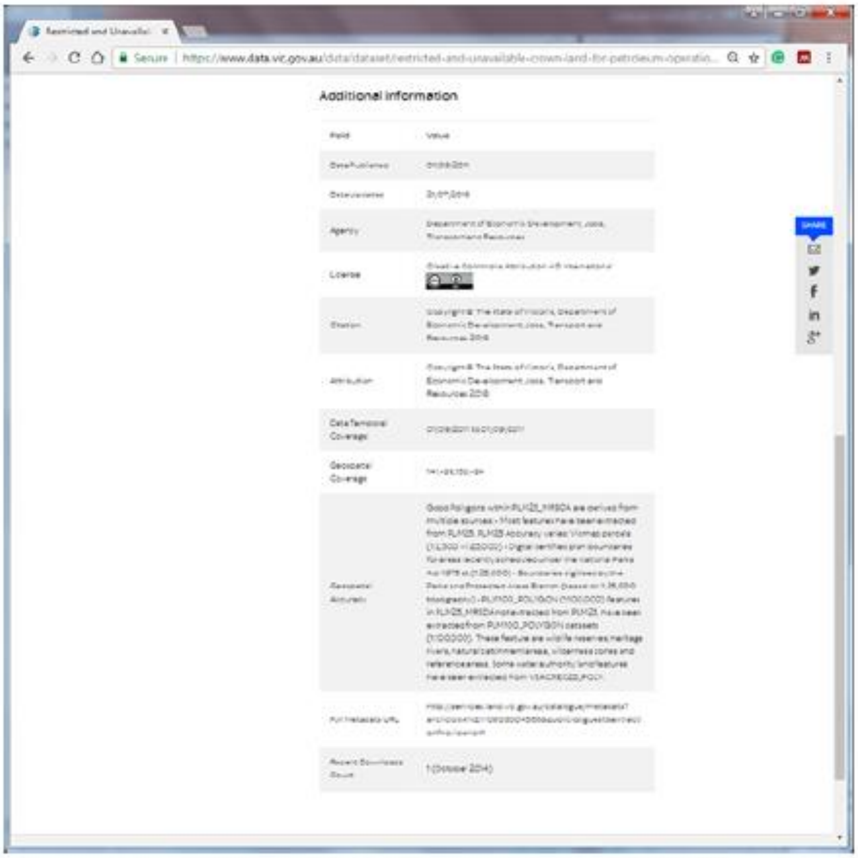
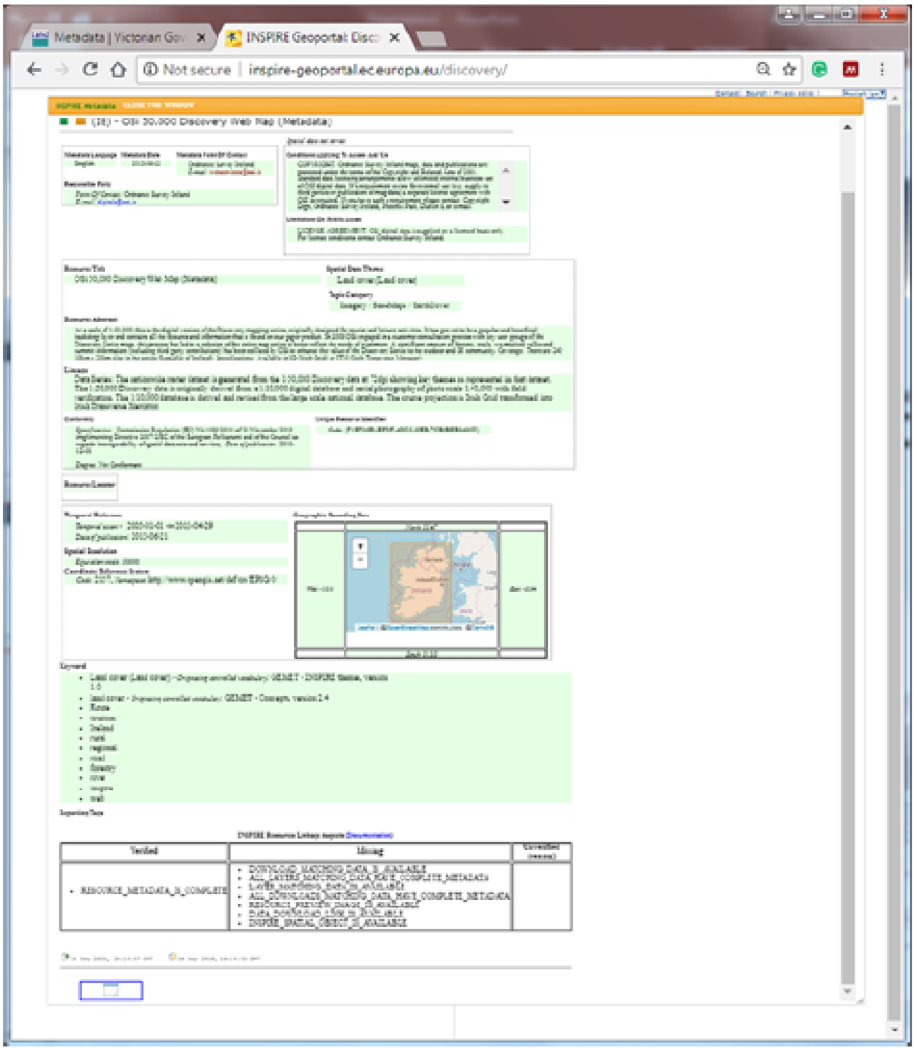


Figure 7. A detailed viewer version of metadata in INSPIRE data portal from a metadata record



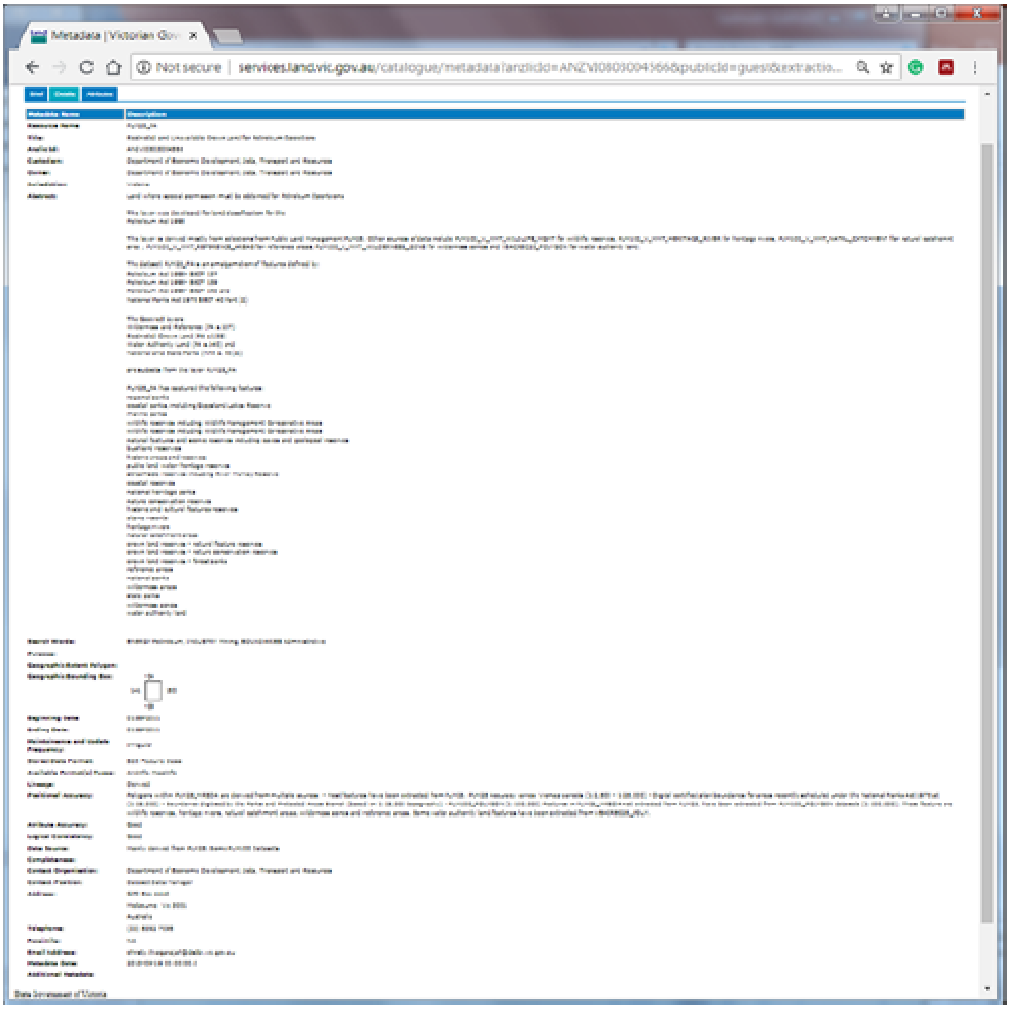
22. 21. In step G of spatial data searching process (Figure 2), when user clicks on a title of dataset in the search results, he/she will be taken to another page with more detailed information about the dataset. Sometimes this page presents a detailed information about the spatial datasets to the user (see Figure 7). In other instances, a simplified information with a link to more detailed information is presented (see Figure 6). In your opinion, which of these two methods are more suitable for you? Please explain if you select other. *

Mark only one oval.

- ◯ Simplified (As in Figure 6)
- ◯ Detailed (As in Figure 7)
- ◯ Other: ____________________

23. 22. If you selected the simplified version, what information do you prefer to be presented in the page? And please provide your preferred presentation method for each of them.

Figure 8. A detailed presentation of a metadata record for a spatial dataset from Victoria data portal



24. **23. If you selected the detailed metadata presentation, the following information about a spatial dataset might be presented to you. Which ones do you prefer to see to help you describe spatial data and assess the suitability of the data for your specific project (Step G in spatial data searching process)? Please select all that apply.**

Check all that apply.

- [ ] Feature/Theme (content of the data, e.g. roads, rivers, building, etc.)
- [ ] Feature attributes description
- [ ] Geographic (location) coverage/extent
- [ ] Production date
- [ ] Maintenance date (last update)
- [ ] Spatial scale (for vector data)
- [ ] Resolution/pixel size (for raster data)
- [ ] Positional (horizontal and/or vertical) accuracy or error
- [ ] Attribute (feature description) accuracy or error
- [ ] Level of detail or generalization of the feature presented in the data (e.g. the lowest road type)
- [ ] Data type (vector, raster, etc.)
- [ ] Data format (shapefile, cad file, etc.)
- [ ] Topology
- [ ] Production process (data history)
- [ ] Intended use of the data
- [ ] Data producer
- [ ] Previous uses
- [ ] Data sample
- [ ] Price
- [ ] Use restriction (copyright)
- [ ] Owner contacts information

25. **24. If there is any, please provide your preferred presentation method (e.g. textual or graphical) for any specific information that you think requires a specific presentation method for your answer in question 23.**

26. **25. In addition to the given list in question 23, is there any additional information that is crucial for you to identify, assess and select appropriate datasets for your projects or applications?**

Mark only one oval.

- ( ) Yes
- ( ) No

27. **26. If your answer to question 25 is "yes", please specify and explain, including your preferred presentation style.**

28. **Should you wish to participate in the later stage of this study or have any other inquiry regarding your participation, please provide your email address.**

Thank you for your participation.

I am expecting to present the results of my research in my thesis and if you are interested to learn more you can contact me at ssyahrudin@student.unimelb.edu.au.

This page is left intentionally blank