USER PROFILES FOR DATA QUALITY MODELS

MESELE ATSBEHA GEBRESILASSIE
March, 2011

SUPERVISORS:

Ms. Dr. I. Ivana Ivanova
Dr. Javier Morales

USER PROFILES FOR DATA QUALITY MODELS

MESELE ATSBEHA GEBRESILASSIE
Enschede, The Netherlands, March, 2011

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geo-informatics

SUPERVISORS:

Ms. Dr. I. Ivana Ivanova
Dr. Javier Morales

THESIS ASSESSMENT BOARD:

Dr. Ir. de Bey(chair)
External examiner: Dr. Ir. H. J. Uitermark

ABSTRACT

Massive geo-spatial data with heterogeneous characteristics produced from multitude of sources is being shared via SDI. The ever increasing user types coupled with diversified characteristics towards using spatial data and understanding and interpreting its quality information are having easy access to the data. However, users are facing difficulty in determine suitability of the shared data for their purpose. This research aims at devising a mechanism for different users to enable to determine fitness-for-use and serve spatial data based on their quality. User-profiling technique is used to search for spatial data based on fitness-for-use. The ISO 19113 standard quality elements, usage information, and geographic bounding box for spatial extent are used by prioritizing quality elements based on users' preference. Explicit and implicit methods of user profile construction are used to devise a user profiles based system for addressing fitness-for-use for different user types from any data quality model. The system enables spatial data users of different expertise level in GIScience, access spatial data that fits their required quality requirements or applications. It delivers spatial data searching and recommendation services based on its quality. The system is implemented in a prototype for a cadastre domain on data acquired from the Netherlands cadastre of the Overijssel province. The prototype can effectively find spatial data based on specified quality requirements or applications in order of the users' preference. User profiles for spatial data search can provide enhanced means of determining fitness-for-use as it provides a flexible means of searching based on specific quality requirement, applications and specific preferences of users. Based on the prototype implemented, user profile techniques in spatial data quality have good potential in addressing the problems of determining quality of data. By thoroughly assessing users' spatial data use behaviors, a better means of delivering spatial data for users based on its quality can be achieved despite users' expertise level in GIScience for any data quality model. Moreover, the continuously increasing number of user types and quality requirements can be better addressed by maintaining all profiles of a user in using spatial data based on its quality.

Keywords

Spatial data quality, fitness-for-use, user profiles, quality information system for spatial data, data quality model

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all thanks God for every thing, next my deepest gratitude goes to both of my supervisors Ms. Dr. I. Ivana Ivanova and Dr. Javier Morales for their continuous follow up, help and timely guidance for the whole period of this research work. With out their constructive, crucial and timely feedbacks and suggestions this work would have been impossible. My next thanks goes to Prof. Dr. J.A. Jaap Zevenbergen, for his important ideas shared me on my research case study cadastre domain. My thanks is then to my beloved family who shaped my life to where I am today and who are always devoted to see success of my life. My warmest thanks to all my friends here in Enschede and all around the world for their encouragement and day-to-day good wishes which kept me strong and happy during my stay here at ITC. My thanks and appreciations then goes to the sponsors of my study at ITC, Netherlands organization for international cooperation in higher education (nuffic) and Hawassa University of the Federal Democratic Republic of Ethiopia with out their financial help the study was unthinkable.

Thanks to all academics and staff at ITC for their guidance, help and continuous cooperation in realizing my study smoothly. I thank you all, God bless you.

LIST OF ACRONYMS

DBMS Database Management System

EA Enterprise Architect

UML Unified Modeling Language

QIS-SD Quality Information System for Spatial Data

PL/pgSQL Procedural Language/PostgreSQL Structured Query Language

SQL Structured Query Language

PHP Hypertext Preprocessor

HTML Hyper Text Markup Language

SDI Spatial Data Infrastructure

GIS Geographic Information System

GIScience Geo Information Science

ISO International Organization for Standardization

WPS Web Processing Service

TC Technical Committee

FDGC Federal Geographic Data Committee

DDL Data Definition Language

XML Extensible Markup Language

GPS Global Positioning System

GML Geographic Markup Language

PIM Platform Independent Model

PSM Platform Specific Model

DDL Data Definition Language

SDQ Spatial Data Quality

IR Information Retrieval

IF Information Filtering

AQL Acceptable Quality Level

SRID Spatial Reference System Identifier

Chapter 1
Motivation and problem statement

1.1 INTERODUCTION

Massive geospatial data production, with diverse sources shared via spatial data infrastructure (SDI) leads to high availability of spatial data to diverse users. Spatial datasets are increasingly being shared, interchanged and this sharing plays significant role in avoiding duplicate spatial data production. Due to the expansion of World Wide Web (WWW) which enables easy sharing of spatial data reduces the cost of acquisition, processing, managing and maintaining of same data with individual organizations [27]. It helped users in integrating and customizing spatial data to their preferences. Hence many national and regional SDIs are emerging and able to use spatial data in various decision making processes efficiently.

Spatial data are produced from different sources, described using different standards and methods and each of the dataset are produced and processed for specific purpose, and thus they are highly heterogeneous [14]. The heterogeneity of spatial creates difficulty for users to easily use these spatial data. The description is meant for users to understand the behavior of the data before using it. However, the description are not always based on the user needs.

A portion of a spatial data description which can be used to determine if spatial dataset is suitable for other application than intended by producers is the quality information. Spatial data as a product of acquisition and compilation processes possess inherent quality characteristics [11]. Quality information of shared data is expected to tell the potential users the strengths and limitations of the spatial data. Thus, understanding the limitation of the spatial data can contribute its appropriate use. However, quality of spatial dataset cannot be statically defined for every possible application it may be used and needs users to understand the limitations. Quality descriptions of spatial datasets that follow a product specification of producers' predefined quality criteria are not as such usable for other uses. The absence, unclear and cryptic quality descriptions of spatial data are also not easily understandable by all the potential users of it [7]. Thus, leads to misuse of data and causes sever consequences.

Potential users of the dataset are as diverse as the general public. Shared datasets can be easily customized, re-processed or simply used for applications for which they were not originally intended [8]. This depends not only on the type of quality descriptions provided, but also the knowledge of users using the spatial data for their own applications. Therefore, users may misuse spatial data as the all users are not able to understand the data description.

Spatial data user characteristic towards using spatial data and understanding its quality descriptions is a challenge due to the increasing number and type of users. Many users are not aware of the quality aspects of spatial datasets they use; even sometimes they use such spatial data in critical decision making processes [19] of which they have little knowledge about its quality. Some users may know well the pros and cons of using spatial data shared from various unknown sources. These types of users may try to assess the suitability of these spatial data for their intended applications. Moreover, they can try to understand any associated quality descriptions when available. The problem with some other spatial data users however is, they may not know that spatial data

has imperfect characteristics. they do not know if there is quality description of the data they intend to use because it may not be delivered to them in a structure they can understand [40]. Even if quality information is available with the data, they do not have the tools to access and use it to determine if the spatial data they have fits their needs. Therefore, easily understandable and easily discoverable structure and content of quality information is important to find spatial data that best fits one's purpose.

Producers are generally having different perspective on the spatial data quality concept from those of users. By stating some of the intrinsic characteristics of spatial data they tend to provide quality information to users [12]. However; users understood spatial data quality as its capability to fulfill their needs. Hence difference in conceptualizing quality exists between the producers and users of spatial data

For determining fitness of spatial data users need to know the internal characteristics of the spatial data and state their specific quality requirements. Users are therefore expected to specify what they need exactly and to compare their needs with the intrinsic characteristics of potentially relevant spatial data [9, P. 38]. Nevertheless as explained above, it is few users who can even specify and discriminate spatial data as to relevant and non-relevant to their application based on its quality. Therefore; the diverse user needs has to be studied, documented and used to describe spatial data based on these needs and bridge the gap of the user requirement quality specifications and the quality descriptions of spatial data produced and further processing. Thus a means of understanding users' interest for selecting relevant spatial data is needed thereby to develop a means of obtaining relevant spatial data based on users' needs.

SDI involves many types of spatial data users. These users have different varied quality information requirements. In this study the various quality information requirements of users are considered as user profiles towards its quality information. The user profiles are the spatial information requirements of the different user types in terms of the quality information structure, content and its reporting means. User profiling can therefore be used to understand and learn users characteristics and to be used for spatial data delivery that meet users' complex requirements in terms of fitness-for-use. For this research we categorize these users into two broad categories as human and non-human user groups based on their requirements for spatial data and its quality information. We further categorize the human users into two based on how they understand and make use of quality quality information.

Geographic Information system-expert (GIS-expert) users: users who have the analytical understanding and awareness of the principle and applications of geo-information science including its quality. This category includes people who work with map production, visualization, and spatial analysis based on spatial and attributes information.

Non-GIS-expert users: users who do not have knowledge in geo-information but use products of geo-information in their activity. This category may include those who work in decision making processes based on policies and use services provided by geo-information processes; e.g. managers, lawyers, planners and users from the general public.

Non-human users are automated operations that use spatial data in spatial processes. These types of users use spatial data and process for producing services. They consume spatial data in a machine readable format for processing spatial data and communicating it with other automated systems e.g. Web Processing Service.

Users require spatial data quality to be described maintained and managed in ways that enable them to discover and access it easily. This depends on the content of the description, the structure in which it is maintained and managed and in the reporting means to the users. The content, structure and reporting means of spatial data quality have significant influence in understanding, interpreting and using spatial data for decision making process by users.

Standards like the ISO defined a means to describe spatial data quality contents by defining content description parameters [21]. Several studied also define means for maintaining spatial data quality that describes spatial data stored in various forms which we call them here as data quality models. These models maintain spatial data quality information describing the data. These models are meant to cater for quality information discovery and retrieval. Thus, users can easily understand spatial data before use however, not all users are capable of using quality information.

Design of data quality models and data quality communication mechanisms also need to consider the characteristics of users. As explained above, the user groups in SDI, we categorized into requires at least different means of communicating spatial data that fits their needs and its quality information. Techniques of spatial data conversion are also required for some users to understand and use the dataset and the quality information e.g. the non-human users.

Therefore dealing with various spatial data user behaviors, understanding their complex needs of spatial data and profiling these needs for various users in a way that can be used to serve users with their preferred spatial data is the main research interest of this study. The main motivation behind this research work relies on contributing a means of new way of addressing fitness-for-use for users in SDI inspired by the impact of quality in spatial data based decisions of users. This aims at catering for decision making process based on appropriate spatial data which is involved in various levels and types of human life in the current world of information science.

1.2 RESEARCH IDENTIFICATION

Developing a method for communicating spatial data quality information based on user requirements and developing and implementing a system which will serve user with spatial data according to user profiles towards spatial data quality.

1.2.1 Specific objectives

The specific objectives of this research are specifically outlined below

1. To review definitions and concepts in spatial data quality with respect to different users in SDI.

2. To assess techniques of quality information communication to various user in SDI.

    - human expert, human non-expert, and non-human users

3. To develop quality information requirements thereby the user profiles of different user types in SDI.

    - human expert, human non-expert, and non-human users

4. To develop a method for communicating spatial data quality information to users according to their profiles to quality information.

5. To develop a system for serving quality information to the users based on their profiles.

6. To implement the system in a prototype

1.2.2 Research questions

sec:ch1sec22 The research work will focus on answering the following specific questions

1. What are the definitions and concepts of spatial data quality in SDI?
2. What are the techniques for communicating spatial data quality information users in SDI?
3. What are the quality requirements of the specified spatial data user groups in SDI?
4. How to develop a method for communicating quality information to the various user groups based on user profiles?
5. How to develop a system which will serve users of different quality requirements with spatial data based on their profile?
6. How to implement the system for serving quality information to the users based on their profiles to data quality information?

1.2.3 Innovation aimed at

Designing a method for communicating quality information of spatial data to user in SDI and developing a system which will serve specific spatial information requirements of specific users based on user profiles for quality information to determine fitness-for-use.

1.3 METHOD ADOPTED

We studied concepts in spatial data quality, user characteristics towards spatial data use and understanding of spatial data quality, we reviewed the main user expectation of spatial data quality in SDI. Further we reviewed the literature on user profiling and intelligent system construction in information science as technique of information retrieval (IR) and Information filtering (IF). From this concept of IR and IF we tried to consider the method for user profile construction of user quality requirements of spatial data in the SDI context of diversified user behavior towards consuming spatial data. As a result we defined user quality requirements to address fitness-for-use.

Once we anlayse and conceptualized the quality requirement of general users in SDI, we dealt with modeling the requirements based on a use-case driven data modeling. For this we harnessed the UML modeling language using the Enterprise Architect (EA) as a tool for use-case development, realizing use-cases by using activity diagrams we designed a user profile class diagram (conceptual data model). We defined and modeled the class diagrams in EA for user profiles.

For the implementation of our system, we use the PostgreSQL database management system together with PHP for developing the front end as a web application system. The prototype system is developed based on the PHP programming facility, HTML coding and PostgreSQL user defined functions coded by procedural language/Postgresql structured language (plpgsql) database programming languages.

1.4 THESIS OUTLINE

1.4.1 Chapter One: Introduction

An introduction to spatial data quality, spatial data users' characteristics in SDI and motivation points to the research is raised. Further set of research objectives are stated and related questions to meet these objective are outlined, finally the method adopted to address the raised objectives is explained.

1.4.2 Chapter Two: Definitions and concepts in spatial data quality

Definitions and concepts in spatial data quality, spatial data quality use in decision making are explained. The fitness-for-use definition of spatial data quality, and the standards based quality description techniques are assessed. The spatial data quality structures, contents and communicating means to the users are discussed. The gap between the inherent quality characteristics of spatial data and the complex quality requirements of users to determine fitness-for-use of spatial data to reduce data misuse are assessed.

1.4.3 Chapter Three: User profiling in addressing fitness-for-use

User profiling, a technique used in information science for learning users preferences and to meet their expectations is taken to this study as a means of addressing fitness-for-use of spatial data based on its quality. Then the means how fitness-for-use can be addressed by using user profiling is explained. A means of learning users' quality requirements implicitly and explicitly and using this information for determining users' spatial data requirements is discussed. Finally a Quality Information System for Spatial data (QIS-SD) system is proposed to learn spatial data quality requirements of different users and to serve them spatial data based on their profiles.

1.4.4 Chapter Four: System design for QIS-SD

A design of the QIS-SD is presented by using the UML modeling techniques on Enterprise Architect (EA) as UML designing tool. The systems general architecture is presented and the various components and sub-system are discussed. Use-case modeling is used to represent the core functionalities of the system and the use-case realization process is presented by using activity diagrams. Moreover a UML modeling based Conceptual data model of the user profile is designed. Finally by considering an SDI layer, a cadastre as a case study and considering a data quality model for the case study a conceptual framework for QIS-SD is design by combining the user profiles conceptual data model with the case study data quality model. For the system modeling a parcel based table data of the Overijssel region of the Netherlands is used to demonstrate in a prototype in conjunction to arbitrarily defined quality information.

1.4.5 Chapter Five: System implementation in prototype

Chapter five focuses on the system functionality proof, to validate the process of serving users with best possible spatial data base on the fitness-for-use concept. The system prototype developments is based on a locally store cadastral data in the ITC intranet. The prototype development is made using PostgreSQL data base management system, plpgsql database programming language, PHP and HTML dynamic web programming languages. The prototype system is a web based application.

1.4.6 Chapter Six: Discussion conclusion and recommendation

Chapter six deals with study process conclusion on how user-profiling can be effectively used by learning diverse user quality requirements of spatial data is discussed. Moreover, further improved profiling techniques, in terms of quality content and content preference investigation and further researching on spatial data quality requirements of specific application is recommended for further study.

Chapter 2
Definitions and concepts in spatial data quality

2.1 INTRODUCTION

The definition given to quality is subjective to various applications. Quality is a subjective concept and strongly depends on the point of view to individual use of data [3]. Quality can be generally described by the attributes and properties of an object or phenomenon that can be observed and interpreted. Quality is a relative; there is no absolute high or low quality unless it is expressed as a measure against a production specification or a user requirement [4].

In Geographic Information Science (GIScience), quality of spatial data has got slightly different definition [6]being conceptualized differently from the view point of data producers and data consumers. To determine usefulness of a specific spatial dataset for a specific application, we need to have a precisely defined requirements or specifications with which we can compare against the dataset's inherent characteristics. Quality is therefore related to a global perception of users but not an intrinsic characteristic of datasets [5]. The definition of data quality given by the International Organization for Standardization (ISO) is the most commonly used to describe spatial data quality where quality is defined as the totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs" [21]. Therefore quality is viewed as dependent on the capability of the dataset on fitting stated requirements or product specifications.

2.2 SPATIAL DATA QUALITY AND USER CHARACTERISTICS IN SDI

Users of spatial data are continuously increasing with increasingly new and diversified needs of spatial information. Users are frequently becoming producers of spatial data [15]. Some users use various functionalities of spatial data based applications like Google Earth and OpenStreet Map for various purposes. Where as others use spatial data for environmental, economical and risk assessments analysis which needs deeper analytical understanding and skills on spatial data and spatial data processing.

Some users simply use spatial data considering that the spatial data they use does not have discrepancies, others do not know how they can understand if the spatial data they have has imperfect information. Some users also do not understand if spatial data quality information is associated with the data they use. This is usually caused by the spatial data quality description provided by producers which is often cryptic message for them.

Users easily misuse spatial data because of the several reasons.

- Unavailability of spatial data quality information associated with spatial data
- inadequate quality description and summarized quality information provided by producers of spatial data
- lack of knowledge and skill of users to understand, interpret and use spatial data quality information

- lack of tool to help users easily access and use spatial data quality
- inappropriate reporting mechanisms of spatial data quality to users
- misunderstanding of users behaviors towards requirements of spatial data quality formats and structure

These and other factors are challenging users to understand and use spatial data appropriately. They misuse spatial data quality information in their applications and decisions that rely on spatial data. The consequences of these misuses are sever and cause substantial loses and unreliable decisions and polices [8]. To bridge the gap between users understanding of spatial data quality and to enable proper use of spatial data, producers need to understand the users diverse needs, tools required to enable if the spatial data is relevant to them, and methods to specify individual or group user quality information requirements in terms of the content of spatial data content, its structure for easy discovery and delivery to the user in a usable way.

2.3 STANDARDIZATION OF SPATIAL DATA QUALITY

In order to judge suitability of a dataset for a certain application it is important to know the inherent characteristics of the dataset. Standards provide a common method to describe, manage, and present quality information to users [3]. This requires having common parameters to decide if the inherent characteristic of the dataset meets the required quality level set deemed acceptable by the user.

Spatial data produced from different sources using different techniques can have discrepancies in terms of theme space and time. Therefore the description of the data as well as its quality description should explain the limitation of the data for the potential users. This requires a standardized means on how spatial datasets can be explained in terms of the them, space and time to helps users use them in decision making properly.

The dynamism nature of spatial data users and the popularity and wide applicability of spatial information hindered standards fulfilling the users' needs for spatial data that meets users' requirements. Nevertheless the ISO quality standards remained the basis for all other quality standardization organizations, initiatives and quality evaluation activities [36]. Hence the ISO 19113 defined data quality elements are widely used in as quality assessment parameters in spatial data quality both by users and producers of spatial data.

According to the ISO quality standards shall be described using two components.These are data quality overview elements which are Lineage, Usage and Purpose and data quality elements. The ISO 19113 [21] standard takes into account the main data quality quality elements: completeness, logical consistency, attributes accuracy, positional accuracy and temporal accuracy; also known as quantitative spatial data quality elements. Each quality element shall be described by its quality sub-elements.

Quality elements and their sub-elements provide producers with guidelines to describe the internal quality characteristics of datasets. Moreover, the quality information content is meant to help users to determine if spatial datasets fulfill their applications' quality requirements. ISO does not define minimum acceptable levels for the quality elements as they varies with the nature of the users' potential applications. ISO further defined ISO 19114 Geographic information–Quality evaluation procedures [23]. It also defines the ways for reporting the result of quality evaluation procedure, either as evaluation reports or as metadata. For the metadata, ISO further defined ISO 19115 Geographic Information-Metadata[22].

2.4 SPATIAL DATA QUALITY MODELS

Quality information is a metadata that can be stored together with the data it describes or separately in different structure. Spatial data quality model is the means used to manage, organize and structure quality information of spatial data. The quality information structuring and management is directly related to how quality information is conceptualized in terms of the structure of the spatial data it describes. This quality information is important to be accessed and discovered in such a way that users can know the exact quality characteristics of spatial data either in an aggregated way or in different levels of detail of spatial data. The content of spatial data quality model is spatial data and description of the data. There are several implementations approaches.

1. *Metadata catalogue:* is description information of spatial data including the quality description of the data. Catalog record represents a dataset in the context of a specific structure [29]; for spatial data quality information is similarly part of the metadata is the. This type of structure helps to discover the description of spatial data quality but lacks strong connection with the dataset components. The difference in structure between the quality information and dataset makes it difficult to discover spatial data at finer levels.

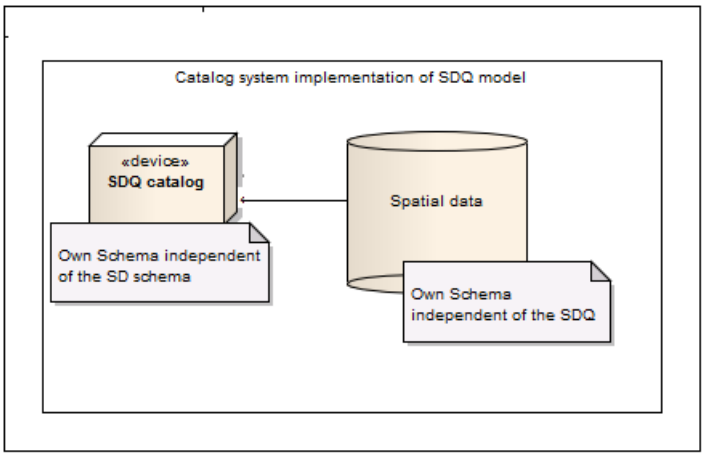

Figure 2.1: Metadata catalogue based spatial data quality model

2. *Database management system:* The data model stores quality information with the same structure and schema of the spatial data it describes. This type of implementation is useful for easy discovery, retrieval and update of spatial data and its quality information; explicitly associating quality information of spatial data in a DBMS can support efficient access to the data at an appropriate level [9, P. 242].

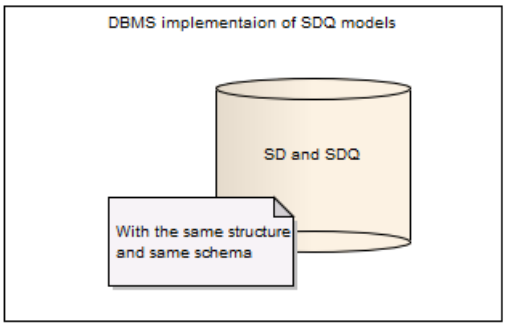

Figure 2.2: DBMS based spatial data quality model

3. *Text report:* Paling text report separated from the data it describes. This type of representation of spatial data quality information organizes quality description of spatial data in a file format. This may help users get summarized knowhow on overall spatial data characteristics but helps little in providing detailed descriptions of spatial datasets and it is difficult to be used with automated processes.for e.g. Positional Accuracy: "Variable", Completeness: "Some features have been eliminated "Street address details partially complete" [19].

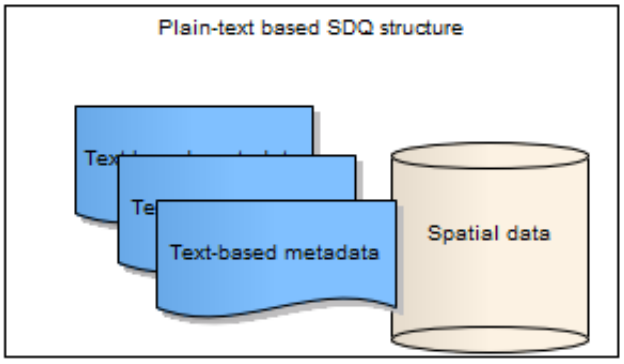

Figure 2.3: Plain-text based Spatial data quality model

These various spatial data quality models in general have strengths and weaknesses in providing sufficient quality description in an easy way for users of different background and different level of understanding. Linking quality information to spatial dataset has immense advantages for users. For example it can reduce the problem of summarized or averaged quality description. It also helps in easy retrieval of both the quality and spatial data to determine the characteristics of each spatial object to an application.

2.5 SPATIAL DATA QUALITY COMMUNICATION

The way quality information is reported to users is a challenge in understanding and use of quality information by users and its applicability in spatial data use. There is a missing link between the spatial data quality industry aim to communicate to users and the way users use information in practice to overcome the consequences of imperfect data [2, 7]. Quality information put in

metadata by producers is largely overlooked by users which leads to a risk of users making poor decisions.

Visualization techniques of spatial data quality used to tell the story about inherent strengths and weaknesses of spatial data are not suitable for users to make informed decisions [2]. The quality statements used to report are vaguely explained [19]. Based on Boin A, et al., [2] survey (emails and interviews) the terminologies used in present-day reporting of spatial data quality are almost absent for the frequent spatial data users. Therefore, Unless a mechanism is devised by which users can be aware of the quality of information in the web environment where data and services of unknown sources and quality can be shared and integrated in a single application, consequences of using them may be costly [44].

2.6 FITNESS-FOR-USE OF SPATIAL DATA

Spatial data producers' perception of spatial data quality mainly depends on the dataset's internal characteristics. These intrinsic characteristics (Internal quality)are resulted from production methods e.g. data acquisition technologies, data model, and storages [9, P. 256]. Internal quality description of spatial dataset is independent of any task [13], unless it is collected and processed for a specific application. There exist no dataset suitable for all potential users nor will its quality meet the needs of all the conceivable uses [32, 10]. Moreover it is impractical if one assumes collecting of spatial data for each and every use and user to fulfill the much complex users' requirements.

The suitability of spatial data for a task can not be determined by the internal quality of the spatial dataset alone. Considering the task, the decisions that should be made based on spatial data, and how decisions are being influenced by the quality of the data [13] are among the few factors. Therefore, studying the behavior of users of the spatial data, their specific quality requirements of the data and specific application requirements are important in determining fitness-for-use.

Internal quality is the production specification of spatial data. For example when spatial dataset is obtained from a 1:1,000,000 scale image digitized 5 years ago, it can be described as a dataset with positional accuracy of 500m, acquisition time of acquisition 2005, and source satellite image. These terms and figures are stating the internal property of the spatial data. They can be specifications of an already produced spatial data acquired by certain assessment and testing by the producer.

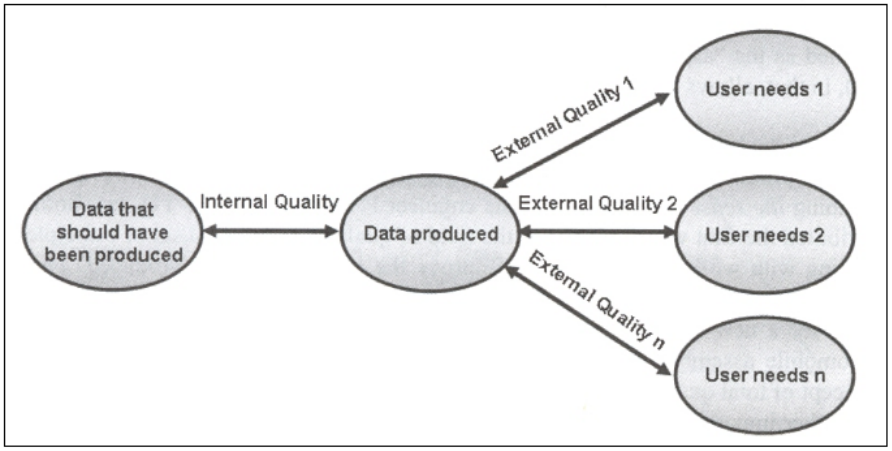

Figure 2.4: Concept of Internal and External quality: taken from [9, P.36]

On the other hand when a user seeks spatial data for certain application, first the user need to specify what quality information she/he needs for the application. For example; positional accuracy of 2.0 meters, spatial data collected before the 2004 devastating tsunami, collected from a field Global Positioning System (GPS) measurement etc. Therefore, these stated requirements are of not spatial data but requirements of the user or a user's application. Quality information requirements stated as per needs of specific users or specific application are called external quality [9, P. 36].

Evaluating fitness-for-use can be an extremely complex task even for GIScience experts due to the heterogeneity of spatial data, the various components of spatial data quality (SDQ) user requirements, and the various reporting approaches [9, p. 243]. Moreover fitness-for-use includes other information beyond the ISO standard quality elements stated in section 2.3.

Generally fitness-for-use of spatial data depends on the behavior of the users and their applications. It requires comparison of the internal and external quality including its content, delivery means, and its structure. To determine fitness-for-use comparision of user specific quality requirements and the spatial data production characteristics is needed. When a spatial data production and processing characteristics (i.e. the internal quality) matches users' specific quality requirements (i.e. the external quality) then the spatial data is said to be suitable for the users' intended application.

Since both external quality and internal quality are descriptions of spatial data quality; similar quality description elements and measurements are usually used. The main distinguishing behavior of these two categories of information is, however, the context they are used for. The internal quality is a product specification provided for potential users to be aware of the intrinsic characteristics of a spatial data. Where as the external quality are the user or application requirements used to determine if a spatial dataset's internal quality meets the stated requirements. Therefore by comparing these two quality specifications, fitness-for-use of spatial dataset can be determined.

The comparison between internal quality of spatial data which is a part of the metadata and the external quality which is the user and application requirements is in terms of both the content and delivery means of spatial data. However, the comparison should be made against more comprehensible and detailed description of the spatial dataset metadata [2]. At the same time; determining fitness-for-use requires learning specific requirements of potential users of the spatial data. This helps in providing spatial data together with its quality description in a format and structure that the prospective users would understand, interpret and make use of it.

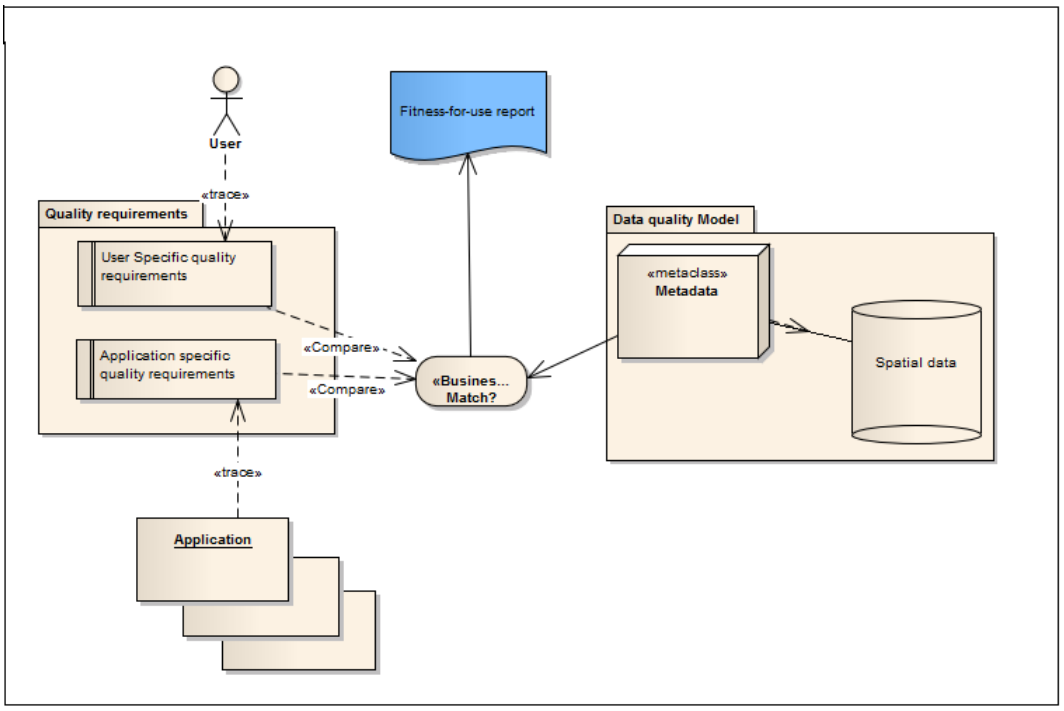

Figure 2.5: Process of determining fitness-for-use

2.7 SUMMARY

In this chapter we tried to build background information on spatial data quality definitions, concepts and users characteristics towards spatial data quality in SDI. Spatial data quality is widely accepted in the producers and users of spatial data as fitness-for-use. To address this various techniques have been developed and being used in terms of the spatial data quality content, reporting means to the user and its structures. These techniques have their own strengths and weakness. The standards based quality descriptions and some of the structures used to store these standards based quality descriptions can be considered as the strengths in addressing fitness for-use. Because techniques are foundations for developing tools for automatic fitness-for-use determining approaches and provides easier access to metadata of spatial data. The weaknesses however remain challenging users in determine fitness-for-use because of various producers use non-consistent standards and all users are not able to understand the standards. Moreover, producers see quality from their own perspective and because user requirements are usually complex, data are becoming heterogeneous but available and shared via SDI from multiple sources. This results in increasing users both in number and type in terms of requirements and skills to determine fitness-for-use based on the internal quality. Therefore; to better address fitness-for-use, a user-profiling technique is introduced in the next chapter of this study to learn the various profiles of users towards spatial data quality requirements and provide them spatial data according to their profiles.

Chapter 3
User profiling for addressing fitness-for-use

3.1 INTRODUCTION

As discussed in chapter 2 of this thesis paper, the definition for spatial data quality is widely accepted as fitness-for-use in the GIScience community. Further, we investigated the literature and found out that there is still misuse of spatial data by users. The misuse of spatial data is because of several reasons. Some of these reasons are absence of quality description of spatial data, inadequate description of spatial data provided by producers, the diversity of spatial data user characteristics and their knowledge of GIScience, heterogeneous and shared spatial data, cryptic and difficult to understand quality information which users usually ignore. Therefore; we introduced a user profiling approach to overcome the misuse of spatial data by users. This technique introduces a new way of looking at the varying and complex users requirements for spatial data in terms of its quality to able to learn user quality requirements and serve them spatial data that fits their needs.

3.2 USER CHARACTERISTICS TOWARDS SPATIAL DATA USE

As stated in section 1.1 we categorized spatial data users into human and non-human users based on their behavior in using spatial data in relation to the fitness-for-use concept.

3.2.1 Human spatial data users

Human users have difference in using spatial data and understanding of spatial data quality information. Therefore, we distinguished them in to two smaller groups based on their expertise level in determining fitness-for-use. However, their difference cannot be easily delineated, rather,we found that providing a possible means of determining fitness-for-use for expert and naïve users in general is important.

1. *Human expert users*: Human expert users include those users who are experts in GIScience are members of this group. These users require wider range of quality parameters to determine if spatial dataset is relevant for their task. These types of users often work with more sophisticated GIScience applications and need detailed characteristics of spatial data. These types of users do not have difficulty to specify their quality requirement based on the ISO 19113 Geographic Information-Quality principles standard quality elements.

2. *Human non-expert users*: Human non-expert user group includes users who are less aware of the spatial data quality impact on the use of spatial data. These users lack understanding of the quality information provided by producers of spatial data. Therefore; the quality information report about whether spatial data of users' interest fits their application or not should be supported by more textual or graphical explanations to reduce the misunderstanding of the "scientific jargon" [20].

3.2.2 Non-human spatial data users

Non-human users are automated operations considered as spatial data users. Automated processes can only understand, read and process spatial data and its quality information when encoded into machine readable formats e.g. Extensible Markup Language (XML) encodings. Spatial data together with its quality information having the same structure in a database management system can be used by these users to enable them access spatial data together with its quality information. For this type of data quality model to be used by these types of users there must be a mechanism by which the quality information and spatial data should be converted to other forms. Moreover, users need a machine to machine communication protocols to communicate for requesting and responding information between them. Therefore, when automated services(e.g. quality aware WPS) consumes spatial data for a process the associated quality information of the data is also subjected to the process, as a result a new quality information associated to the output of the process is delivered.

3.2.3 Users' spatial data quality requirements

Users of spatial data may have simple or complex requirements towards spatial data quality information because their needs depend on their applications. The data of users' preferences also depends on the characteristics of the multidisciplinary spatial data available through SDI [3] and users need to select the suitable ones. Users' quality information requirements vary according to each user's characteristics in using spatial data and its quality information characteristics. This quite depends on what type of spatial data does a user requires and the means of spatial data quality description provided to the user. It also depends on the data quality structure used to maintain and manage spatial data and metadata for easy discovery and retrieval.

Users may need quality information based on the different quality elements stated in section 2.3. Some users require based on all the quality element descriptions some other users may need based on few of them. Moreover, users also require the quality elements in various priorities by giving certain weights for each quality element. Sometimes users only want to know quality of spatial data at various levels of data granularity. Users are also interested in accessibility, delivery means and costs of spatial data before they decide to use for their purpose. They are also interested in spatial data of varied locations. Generally this scenario highly depends on individual user, group of users, or specific application requirements. This makes spatial data quality requirements of users and determining fitness-for-use of spatial data very complex.

To determine fitness-for-use of users' purposes, specific quality requirements of users or specific quality requirements of users' applications are required. The specifications need not be limited to several or all of quality elements of ISO 19113 Geographic Information–quality principles but also other factors of fitness-for-use. Therefore, for users to find spatial data that best fits their requirements; a user, group of users or application quality requirements have to be defined .This requires complex scenarios of external user's quality requirements be maintained as users' profiles. The profile information then will help to serve users spatial data of their interest.

3.3 USER PROFILING FOR FITNESS-FOR-USE

User profiling is a common technique used in information retrieval (IR). The rationale behind it is to ease the overload of information generated when users request information from search engines. User profiling is the process of learning user's interests and behaviors [18, 17]. It is a representation of information for individual or group of users that is essential identification of users' behaviors towards certain application. Thus it is used in information retrieval and filtering

for discriminating relevant and irrelevant resources for various users [26]. The user profile,here in our case is exploited to determine what spatial data users potentially require and to serve them with most relevant spatial data for their application. This is accomplished based on the quality characteristics of spatial data and user quality requirements to address fitness-for-use.

As discussed in chapter one and chapter two of this research, the problem of shared and overloaded spatial data, with its heterogeneous characteristics and with users' different quality needs is continuously affecting the proper use of spatial data. Since fitness-for-use requires the user quality requirements or intended applications' requirements; user profiling can be used to effectively learn users' quality needs of spatial data thereby serve most relevant spatial data to them.

In the IR science, explicit feedback of users on their interest resources is considered as effective way of user profile construction. However, the challenge is that not all users are willing to provide feedback [33]. In GIScience the challenge is more difficult. To gather every possible user quality needs is difficult because most users have limited skill to identify and evaluate their application quality requirements. Thus, expecting every profile constituent from the user is not always possible. However, it is possible to learn user's behaviors in terms of the user's spatial data preferences implicitly.

The outcome of the process of user profiling is typically a set of information that reveals knowledge about users' spatial data usage and quality requirement. This typically means capturing the usage history of users to learn their quality information requirements. Thereby to enable automatic selection of a user's requirements when a user is identified to provide spatial data based on the identified requirements from the profile. This spatial data filtering mechanism uses several techniques. These techniques are identifying users, identifying user behaviors in using spatial data, identifying quality requirements of users, and user's preferences on the parameters used to look for spatial data etc. These behaviors depend on spatial extent, on theme and temporal extent of spatial data.

Generating an initial profile and updating an existing profile over time are important aspects of user profiling [26]. For updating, tracking the access history of users as they use spatial data implicitly or new quality requirement capturing from the user explicitly is used. When a new user arrives, where no profile of that particular user is maintained, then by looking for similarity between the user's characteristics and those of similar users' profiles maintained is important. The similarity can be identified by the applications the users use spatial data for, and can be used to maintain their quality requirements. This technique in IR, is named as collaborative filtering [16]; which is based on the assumption that "similar users have similar preferences". Therefore, collaborative filtering technique helps us to generate initial profiles of users for further use in determining fitness-for-use of spatial data. Here the assumption is similar applications have same quality requirements.

In this research, we propose a user profiling based system called Quality Information System for Spatial Data (QIS-SD) to cater for enhanced means of addressing fitness-for-use in SDI. QIS-SD follows the information retrieval (IR) and information filtering (IF) techniques and principles in information science as mentioned previously. In IR and IF user profile is usually constructed either directly (explicitly) by users supplying their interest or automatically (implicit) methods [38]. We use both these two methods to learn the quality requirements of user profile. It helps to determine quality requirements of newly arriving users and to keep updating quality requirements of existing users. Thus, the user profile is used as a source of metadata about user's quality requirements by which the system serves spatial data to users based on their profiles.

3.3.1 Explicit profiling for QIS-SD user profile construction

The simplest way of obtaining information about users' needs is through data input via user interfaces. This is called explicit (static) profiling and it is used to analyze a user's static and predictable characteristics [33]. QIS-SD will accept the following information from users explicitly via a web based interface for human users. For non-human users, the interaction is based on machine to machine communication protocol and XML/GML standard data format [30] as web services use these protocols to communicate with other service.

- *Basic user information:* The user profile is information about the user to identify who accessed what type of data (in terms of the spatial data quality characteristics) during registration. This information is useful in identifying datasets and application that a particular user is interested in. When a user is correctly identified, the corresponding quality requirement of the user is identified in the profile for further use and it helps for creating individual user sessions in the system.

- *User quality information requirements:* When a user want to find out spatial data that fits a specified set of quality requirements the user provided, these set of quality requirements are stored as the user's profile. These sets of quality requirements are in fact the basis for determining fitness-for-use of spatial data to the user. Moreover, in case where the request returned no result, the requirements initially entered are stored in the profile of the user. This helps find out newly available spatial data in later use of the system.

- *Users' intended application:* User application for which users' access data is important constituent of the user profile. This can provide information about the possible application area of users, and their preference spatial data for that application. Moreover, it provides quality information requirements for those who know what data they want but do not exactly know how to state the quality specifications of the data they have to use for an application. Therefore, by using the user intended application the QIS-SD can determine the dataset that fits users' applications. Users' intended application also caters for means to group users who share similar applications this means similar quality requirements of users. This application description is also part of the ISO 19113 Geographic Information–Quality Principle quality overview elements of spatial data. Usage describes uses of a dataset by the data producer or other, distinct, data users [21] which can provide with the notion of similarity among users in user profiling.

- *Spatial extent:* users need to specify the area of their interest from which they could search for spatial data that fits their specific quality requirements. This can be done in two ways. By specifically stating the Longitude and Latitude of area of users interests or by using a geographic bounding box on a map. While the former can be used by expert users who can easily identify the coordinate points and the zoom level of their interest, the later is easy for non-expert users to use it intuitively. The bounding box can have various shapes to include various shaped features. The bounding box shape should consider a point, a line string, and a polygon types of features. The spatial extent defined by a geographic bounding box can be used by considering the type of the spatial features in which the user is interested in. When the spatial extent can be defined dynamically in different shapes spatial features ca be checked if they lay with in the extent of the bounding box. This approach may exclude features which lies around the bounding box. However, considering the centroid of the features to be with in the specified extent can include these objects if most of the features lie inside defined extent. Defining a bounding box on maps is also

not easy for every user. Another method of specifying spatial extent is using a geographic boundaries. This can be based on counties or provinces or other administrative boundaries. Therefore, users can select the spatial object they are looking for a specified location based on these administrative boundaries. This approach is specially important for non-expert users, as they can simply choose an administrative boundary by the name of that country or province, district.

3.3.2 Implicit profiling for QIS-SD user profile construction

Implicit/dynamic profiling is the process of capturing and analyzing user's activity or actions to determine their preferred resources [41]. Therefore, in our system by using this technique users' quality requirements are capture. Accordingly the captured quality information is used to update the particular user's profiles. This is important for identifying each user's quality requirement history thereby it is possible to identify what type of spatial data often a user is interested in terms of its quality. To construct and continuously update the user profile information, QIS-SD captures the following information implicitly

- *User identification:* a user who had registered and with profile information is identified during login. The user information is used to monitor what dataset the user has accessed following each login time. The dynamic web application based system uses a web session for monitor each users interaction as used in clustering of web users [43]. This is important to update the profile of the user's interaction with the system.
- *counter:* which tracks the number of times quality requirement is used to search spatial datasets, by a user. The counter informs how often a particular data have been used instead of storing the same information as a new profile. This is important in avoiding any repetitive information in the profile.
- *Weight:* When users search spatial data based on quality elements, they usually prefer to consider some elements more relevant than others. This is a concept of users' preferences ranking for advanced search mechanisms based on multiple requirements with various relevancies to a user [34]. QIS-SD uses a weighting technique by averaging a previous weight of each data quality element with a new weight used by the user when searching for spatial data.This keeps track of order of data quality elements relevance. We considered the highest weight as most relevant and the smallest weight as least relevant quality element during the system search.

The profile keeps changing when user continuously interact with the system. Profile update takes place when user makes use of spatial data that fits their purpose both statically and dynamically. Unlike many profiling techniques in IR, in this study limited information about the user, only the name and identifier of the user are required. Both the identifier and the name will help in identifying user profile to be used in determining fitness-for-use of spatial data.

3.4 DETERMINING FITNESS FOR USE OF SPATIAL DATA BASED ON USER PROFILE

QIS-SD is meant to serve geospatial users spatial data based on their quality requirements. It is meant to find the best possible spatial data that fits a particular user needs based on the fitness-for-use concept. As explained earlier in 2.6 fitness-for-use is determined by comparing the internal quality of spatial data and external quality of user requirements. Therefore, the QIS-SD system searching mechanism is based on user's behaviors towards spatial data quality need. These needs

are mainly specified user quality requirements (external quality) or users' stated application. However these comparisons include other factors. These factors are prioritizing requirements, searching spatial data on a specific area of interest (i.e. geographic extent) and based on all or some of the data quality elements.

To perform the comparison:

- Users could specify and tell the system their explicit requirements, at least one quality element should be used
- They can state their applications and the system can find spatial data suitable for similar application from the usage. As defined in ISO 19115 overview quality elements are defined by the data producer or from user defined default application and their stated quality requirements from expert users [22].
- The system can also suggest spatial data to users based on their past data access history tracked by the system in the profile of users. To accomplish this, the system uses the information stored in the system mainly those types of information stated in section 3.3.1 and 3.3.2 above.

Users find spatial data based on various quality information requirements for their multitude of applications. In the profile, multiple numbers of quality elements can be used to search spatial data that fits the quality elements values. Therefore; the quality information used to search spatial data will always be based on fitness-for-use.

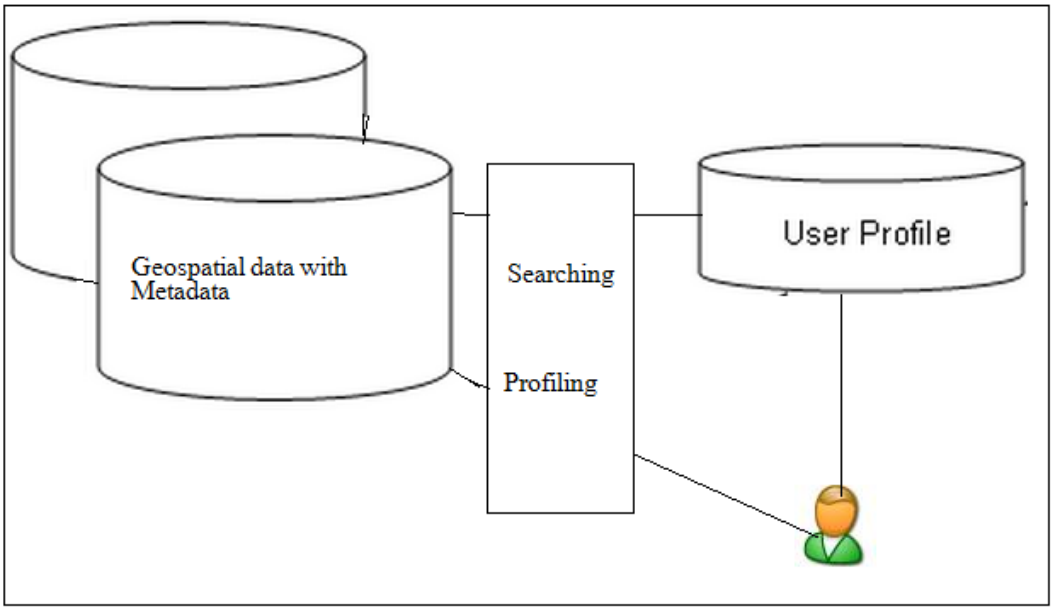

Figure 3.1: Spatial data access based on fitness-for-use

3.4.1 Spatial data retrieval functions

Not all users can specify the exact description of their quality requirements in accordance to the quality elements used because specification of external quality is very complex and depends on expert knowledge of spatial data users. Average users, and non-expert users can, however, specify

the purpose they are looking spatial data for. This typically is important for non-expert users to find applicable spatial data for their intended application. For example municipality authorities can use the applicable information of certain spatial data at a managerial decision making levels. Data quality overview elements are critical for assessing the quality of a dataset for a particular application [21]. Moreover users can easily specify the spatial extent before the actual search of spatial objects. Therefore finding spatial data that fits one's purpose is made by explicitly stating quality requirements or by stating the intended application in the specified geographic extent.

3.4.2 Spatial data recommendation functionality

To recommend a user with spatial data relevant to the user's application, we can consider the previous data access history by the user and retrieve these data. This could create an overload of output, especially for frequently frequent users. Therefore to solve this possible overload of spatial data that could be retrieved, an initial weighting means of quality information can be considered to allow users find spatial data with highest weight quality elements. This is also used to select those quality elements with highest quality values based on the choice of the user of which quality element the user wants to stress more.

Another alternative for recommending a user spatial data relevant to the user's application is based on quality information requirements of the user to retrieve spatial data. This can be found from the user external quality requirements tracked when user request spatial data each time by entering set of quality requirements. Since a user could probably use several quality requirements to access spatial data, based on the weights the user gave to each of the quality elements is selected. This quality information, once retrieved from the user requirements, is stored in the system and can be used to search spatial data that meets the requirements.

These two approaches have their own weakness and strengths. The first one while fast, it does not suggest newly available spatial data with similar quality information. In fact with new spatial data some quality differences may be available especially in terms of temporal aspects of quality. However, temporal aspect of data may not always be required. Therefore, newly available spatial data could be important. The strength of this approach is that it does not need more computation and does not create access delay. Where as the second approach can include newly available spatial data, it needs more computation and causes access delay.

3.4.3 User weighting of quality elements based on preferences

The quality elements a user may use to search spatial data that fits a purpose can be weighted to prioritize some quality elements over others. This will help identify the best fitting spatial data in case where multiple candidates spatial data exist. The weights given to each quality element is based on the users weighting preferences. The weighting of these quality elements is based on the most preferred quality element to overtake the priority of others while the other elements are also important to search data. When a user has quality requirement with same or similar preference for a user or use, based on individual quality elements or group of quality elements, the weight of quality elements is also used prioritizes the quality elements. The weighting mechanism we used is based on a value of five for the highest weight and a value of one for the least weight. The weight stored in the profile information is using a simple averaging of weight each time a user enters weight for each quality element.

3.5 SUMMARY

After developing the conceptual understanding of spatial data quality, users and user characteristics in using spatial data quality in chapter two of this thesis work. This chapter aimed at analyzing the existing user profiling techniques in IR and IF and exploited the technique in similar way to device a mechanism through which to address fitness-for-use in SDI. Using this method in spatial data retrieval specifically for addressing fitness-for-purpose of users' applications, enables users find out only relevant spatial data that best fits their requirements.

In this chapter we proposed a QIS-SD system for addressing fitness-for-use of spatial data for users by profiling their quality requirements or their intended application requirements and deliver spatial data based on their preference of delivery. The system initially learns user characteristics explicitly and keeps updating the profiles through time to provide the best fitting spatial data according to users profile. The next chapter deals with the design of the QIS-SD system and a case study background based on which a prototype implementation is made.

Chapter 4
System Design for QIS-SD

4.1 INTRODUCTION

This chapter focuses on the detailed design of QIS-SD, which is presented based on a use-case based system design. We used a use case diagram to illustrate the core functional requirements of the system. The use-case is then elaborated based on activity diagram as a means to explain detail system and user interaction. Based on the activity diagram the information that has to be documented in the profile is explained and a conceptual data model for the user profile is presented. The user profile data model is then used with a quality model for a cadastral domain as a case study to develop a prototype system demonstration. Moreover, background information for the case study is presented to reflect the relevance of the profiling technique in the selected domain.

4.2 SYSTEM REQUIREMENTS

The proposed system QIS-SD is aimed at delivering spatial data based on fitness-for-use. To accomplish this, several services need to be designed. Generally the core service of the system are specified below as functional requirements. Moreover, non-functional requirements are also stated

4.2.1 Functional requirements

To deliverer spatial data based on user profile information for spatial data quality to the user, the system need to be equipped with application logic to address the following requirements.

- Intelligent system that find spatial data based on user quality requirements. Users may specify few or several quality requirements, so the system need to provide provides a flexible searching mechanism based on specified quality requirements.
  - A system that searches spatial data based on flexible number of quality requirements specified by the user.
  - A system that searches spatial data of user interests based on user specific applications
- The system should allow users to prioritize their searching parameters every time they request spatial data if necessary.
- A system that delivers spatial data to users in a format of users' preference. Users may sometimes want spatial data description (the metadata)of certain spatial data to use the details of spatial data in decision making process.
- A system should have the capability for searching spatial data in a user defined spatial extents.

- A system that suggests users spatial data based on their past spatial data request and access history. This is helpful for frequent users of the system to ease routine search criteria specification.

4.2.2 Non-functional requirements

- The system should have a web based interface for ease of use for all users
- Secured system based on login name password credentials for identifying users' preferences user session and user profiles.

4.3 THE SYSTEM COMPONENTS AND ARCHITECTURE

The QIS-SD has several components to accomplish its main aim of addressing fitness-for-use efficiently. These components are: user login, user registration, data retrieval and profile update sub-systems. The system general architecture is based on the Service oriented architecture in which services are posted on to the registry.

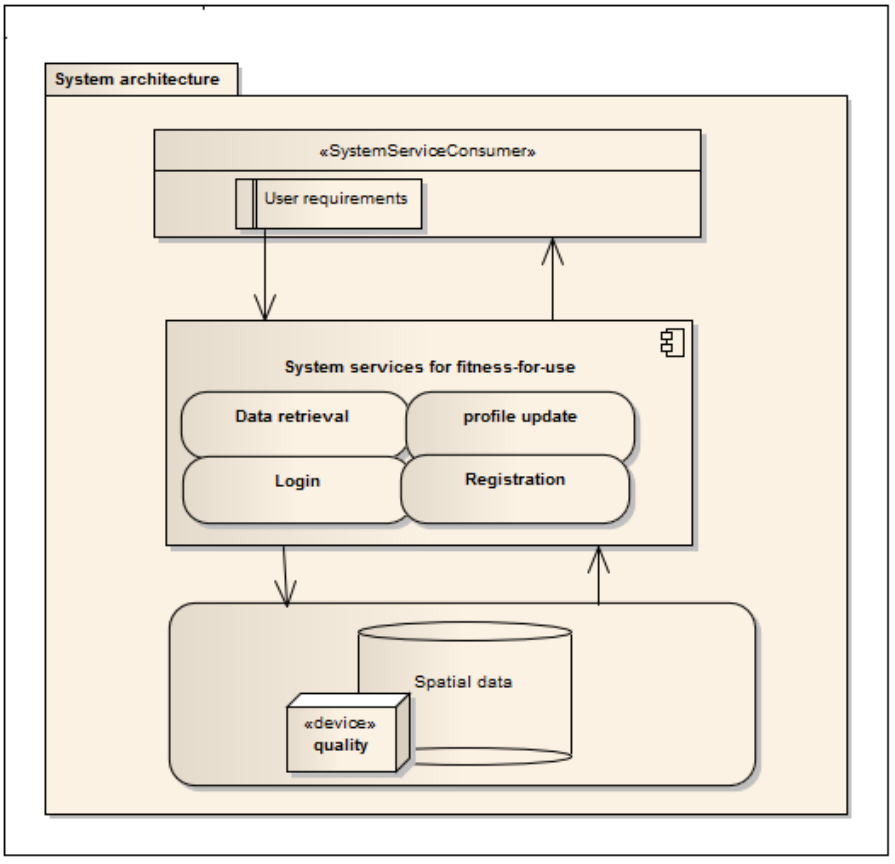

Figure 4.1: System architecture and components

4.3.1 Login service

As shown on figure 4.1 *Login service* is used to identify whether a user who attempt access exists in the system or not if existed to determine its profiles. It is also used for protecting the system from

unauthorized users. The login service is also important for creating user sessions. The session is an important system component for user profile construction.

4.3.2 Registration Service

Registration is required for users of the system by which users can provide their identification and requirements to the system. During registration, user credentials and requirements of each user are recoded. This information will be the initial profile of users. Registration users can provide other information related to their spatial data preferences.

4.3.3 Data retrieval Service

Retrieval system service is the core of the QIS-SD for used to searching for spatial data and deliver it to users based on fitness-for-use concept. The service is used to search spatial data directly based on user's specified quality requirements. Moreover, it retrieves quality requirements of users from their profiles maintained in the system then determine what spatial data users need. As discussed earlier in section 3.4.1. The retrieval service for spatial data is also based on users' applications.

4.3.4 Profile update Service

This system service is meant to continuously update the user profile maintained in the system. The update information is typically the one provided in sections 3.3.1 and 3.3.2

QIS-SD profile information is a collection of information about users behaviors for using spatial data based on its quality. It mainly consists of the following information

- user identification information
- user quality requirements
- user applications
- user's preference information on each of the quality elements the user used to search spatial data i.e. weight
- the spatial extent users are interested in
- information about users' data access history

The user profile information collected implicitly or explicitly in the user profiles based system is used for recommender service and for spatial data producers as a knowledge base about the potential user requirements. This information the can be input for producers to know the possible user requirements for quality of spatial data.

4.4 USE-CASE DEFINITIONS FOR QIS-SD

Use-cases focus on system users, user actions, and system processes thereby shows an abstracted view of what a system can do to a user [24, P. 60]. It is a way of using a system and understanding the system in a context. When users come to know what exactly a system should do for them, use-cases demonstrate the interaction between the user and system. Use cases also represent the units of functionalities or services of a system, or sub-system.

For our system QIS-SD, we define use-cases by looking at how the different users (actors of the system) need to use the system to meet fitness-for-use. It has several functionalities for constructing and using user profile information to use in addressing fitness-for-use by retrieving relevant

spatial data to users based on their profiles. These functionalities are mainly for searching the profiles to find out user's requirements and for retrieving spatial data based on user requirements to address fitness-for-use. Even though there is no standard way of designing use-cases for our purposes, we follow the specifications and guidelines for designing use-case diagrams as explained in[1, P. 82]. Thus we define three main use-cases for the QIS-SD:

- Retrieve relevant spatial data: This aims at finding out spatial data that fulfills user's quality requirements or user applications' quality requirements
- Recommend relevant spatial data: this aims at identifying a user's profile and delivers the user with spatial data that meets the user's requirements based on the profiles already stored in the system. The system also lets the user to choose a delivery means.
- User registration: Users need to be registered users when they access any spatial data. Therefore, the system provides them with registration facility prior to any spatial data access

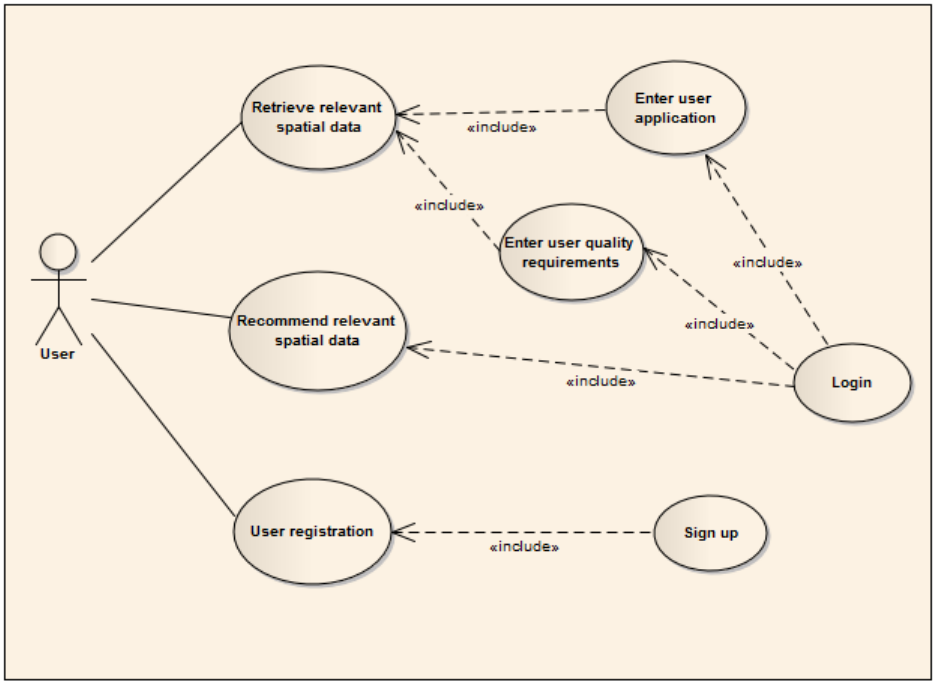

Figure 4.2: System use case diagram

To explain the above use-cases figure 4.2, we define activity diagrams which shows how all the use-case processes accomplish their set of activities to meet the desired goal. The activity diagrams are presented in a simplified and smaller functionalities for easier understanding of the activities.

4.5 ACTIVITY DIAGRAM FOR QIS-SD

Activity diagram help system development to visually illustrate sequence of activities in a system. It contains action states, which are the finest granularity building block of activity diagrams, and represents activities and sub activities [1, P. 232]. The following activity diagrams states the main functionalities and activity sequences of the QIS-SD to elaborate each of the use-cases defined above.

Relevant spatial data retrieval service: User specific quality requirements or user intended application are accepted from the respected user as input to the system. The quality requirements are compared against the metadata of spatial data to select the best fitting spatial data to the user. The quality requirements are also weighted according to the user's preferences of each quality element used to search spatial data. Moreover, these quality requirements are used to update the profile of a user either by updating a counter or tracking as a record of profile information for the respected user. These quality elements are also weighted each time the user want prefer to give more relevant quality element and less relevant quality element to filter the best available spatial data.

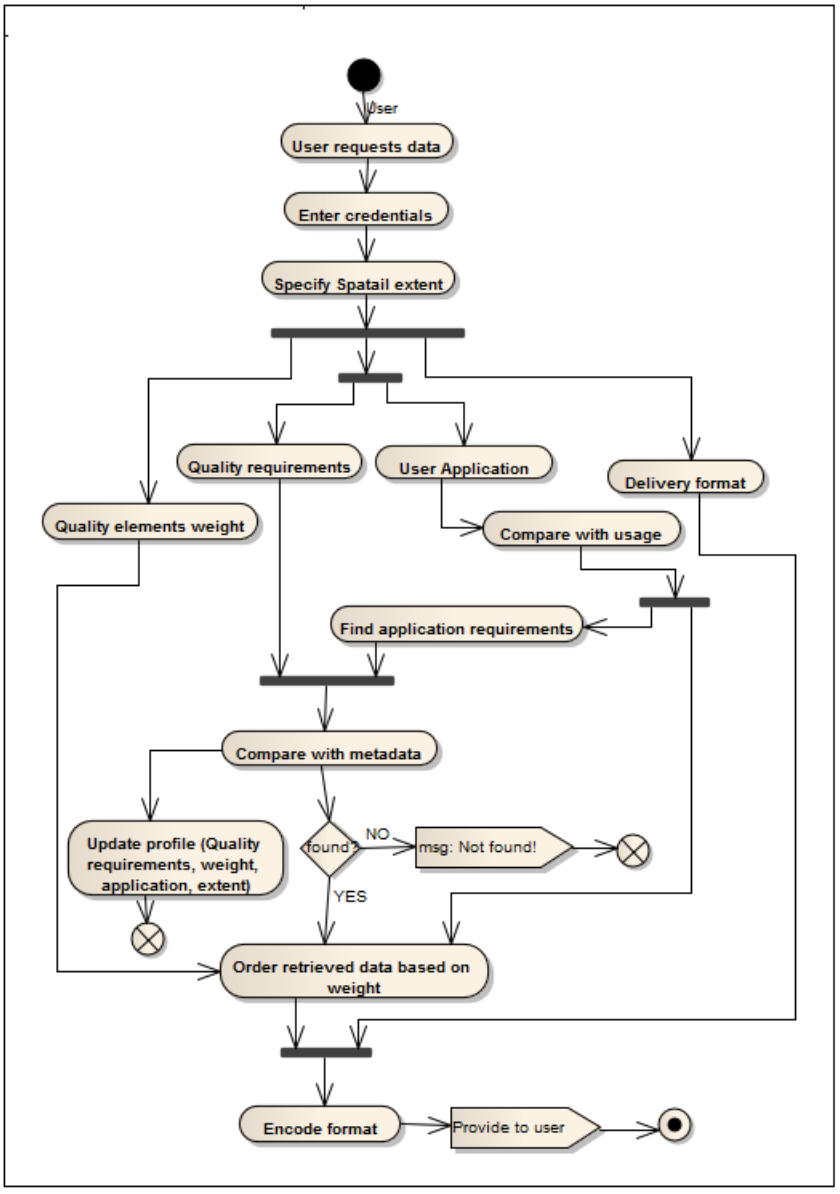

Figure 4.3: Relevant spatial Data retrieval service

Recommender service: retrieves spatial data based on user preference quality requirements of an identified user. The service also recommends based on access history and frequently used quality requirements. The quality requirements of the user or the application are determined from the user profile. As explained on section 3.3.2 a counter is used to determine the quality requirements of a user based on use frequency and weight of quality elements the user has been used.

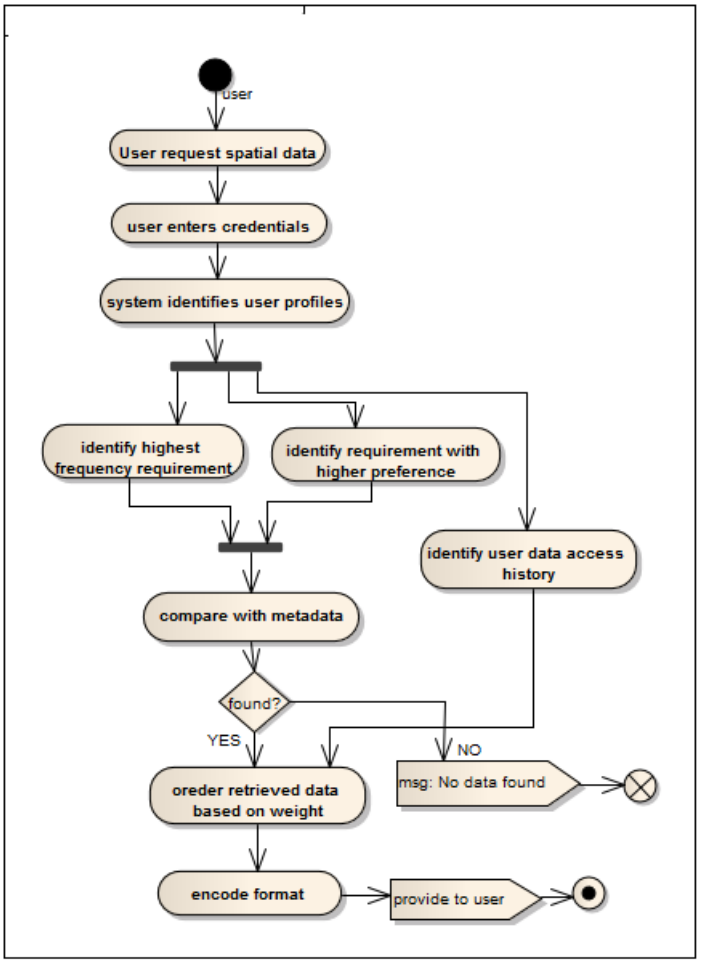

Figure 4.4: Recommender service activity diagram

The above figures 4.3 and 4.4 show retrieving spatial data based on specific quality requirements and the requirements are from the profile. the profile is therefore continously updated when users interact with the system.

4.6 CONCEPTUAL DATA MODEL DESIGN FOR USER PROFILES

The user profile data model mainly consists of information about users, user intended applications, external quality of users and applications, weights of each external quality and the geographic extent users may be interested to look for spatial data. User profiles keep track of users' behaviors when using spatial data by keeping users interests in terms of spatial data that fits their needs. This is used to personalize the application specific quality information to users who use these applications. Extent information from which users seek spatial data, the individual data

quality element's weight users give for prioritizing are also tracked each time when users access spatial data. The weights are averaged and stored in the profile and are used for potentially relevant spatial data to users.

Through time, users will have several user specific quality requirement, several applications, several areas of interest in terms of spatial extent but always keeps one weight for each data quality element for each user. The weight kept in the profile is the average of all weights a user used for a data quality element. Thus, the tracked information for each application a user uses spatial data is ranked based on the weights. These requirements may be user specifically entered or learnt from application requirements. The set of quality requirements of a user with higher count (frequently used to search spatial data), and the highest weighted quality requirement is used to determine fitness of spatial data for recommendation.

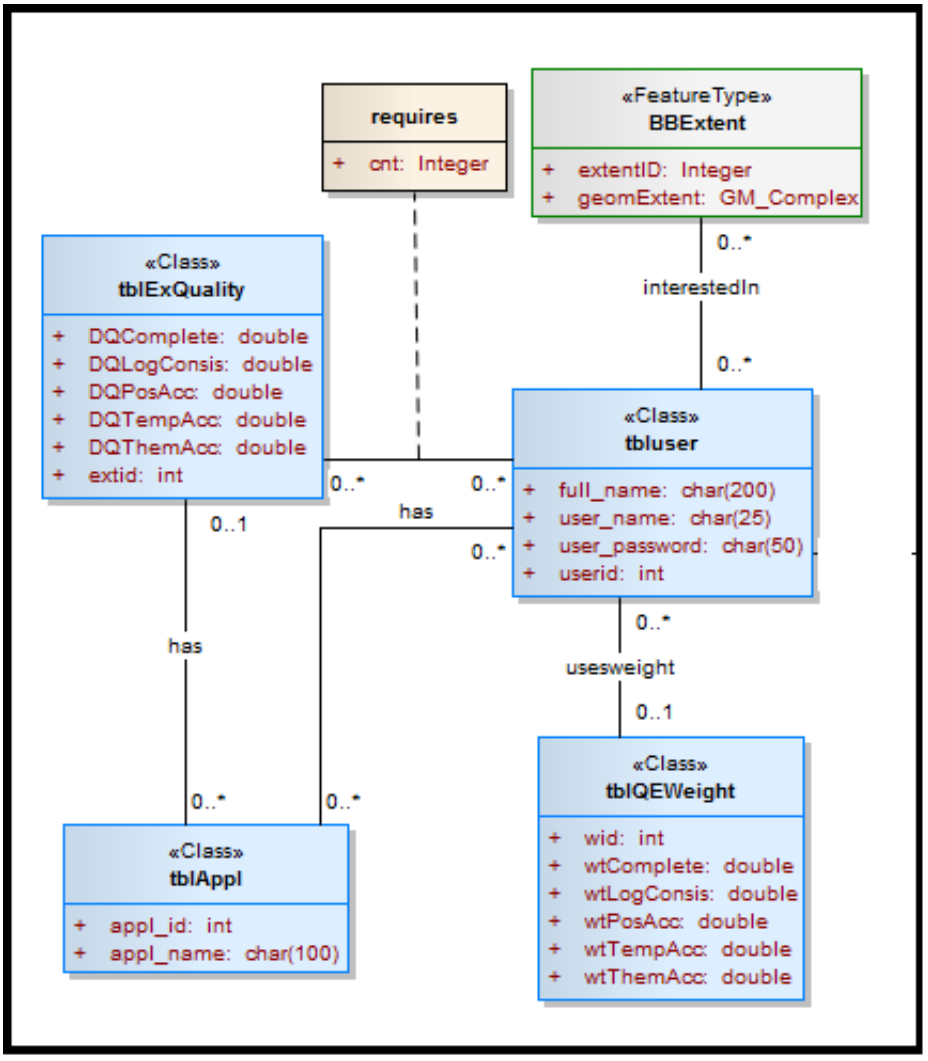

Figure 4.5: Conceptual data model for User profile

4.7 CASE STUDY CADASTRAL SYSTEMS IN SDI

The case study we selected focuses on the process of selecting best fitting spatial data for use in cadastral system. We select cadastre as our case study because we believe that cadastral systems involve quality sensitive data as well as laws that involve acceptable quality levels for decisions to rely on.

As one of the base layers of a Spatial Data Infrastructure (SDI), cadastre system provide users data for mapping locations and extents of parcels, establishing ownership of rights, and determine the value of rights [25]. It contributes support in land and fiscal policies for policy makers. Cadastral system supports governments to recognize the benefits being returned to them from exchange of data via the SDI [42] and provides up-to-date spatial and non-spatial information for market places, economic developments and social stability. Cadastral systems help to reduces land disputes, to improved environmental management, to avoid land stealing and corruption in taxation by creating transparency in land administration and promotes good governance. Cadastral systems envisaged to benefit many potential users who use land information for public and commercial applications [28] because it holds data on geodetic, topographic, road network, parcel based land holding etc. It provide services and data for users interested in infrastructure planning and social services delivery for urban and rural community, land use and resource management. In the modern multipurpose cadastre the parcel remains the dominantly important constituent for these and other processes in cadastre.

In general cadastral systems should mainly include the following information [31, 39].

- A piece of land in the real world: The geometry (spatial component) of the land consisting of the location and its spatial extent.
- An unambiguous identifier for each piece of land: used to avoid ambiguities in the spatial reference and identification
- A description of the spatial location (i.e. the boundary): determined by the point measurements this depends on the precision of the point measurement and point definition itself
- Attributes of the piece of land: the description including text records of attributes of the spatial object

4.7.1 Spatial data quality for Cadastral Processes

Some of the main processes in cadastre are recording and indexing of a new land survey, tax administration which includes assigning Parcel Identification Number (PIN), updates and establishes relationships of different parcels and updating values of records. Producing and updating maps i.e. mapping of newly surveyed parcels, road networks public utilities, constructions and archiving maps for publication. Creating maps for variety of uses e.g. tourist route maps, land use maps, land cover maps, administrative boundaries, public facility centers etc.

Quality of spatial objects in cadastre should be maintained and evaluated carefully. Unless, it may cause significant financial and legal problems related to ownership rights of parcels, planning and maintenance of utilities and facilities. It also affects decisions made related to social service.

- Up-to-date spatial and non-spatial information should be maintain especially in fast growing areas massive changes exist within a short period of time
- Correct and optimal accuracy in geometrical measurements and attribute information observations thereby correct area estimations are important for using in fair taxation and estimation of financial planning.

- Full and timely description of rights of subjects on objects is important for taxation, court, financial institution processes, planning etc.

In general quality information of cadastral data helps various users to make appropriate decisions. These decisions may range from financial, environmental, and planning to dispute resolution in courts and fair tax by taxation authorities.

4.7.2 Users of spatial data in cadastral

Cartographer: Spatial data use in cadastre helps in creating cadastral maps used to map parcels, buildings, roads, waterways landscape elements and other utilities [37]. It also shows specific landmarks which people can use for navigation, including natural features like lakes and streams and other natural features. For this purpose complete data of features in the area of interest and measurement of point objects up to acceptable level of accuracy, scale, spatial reference system (SRS) and time related information and other attribute information are important. Moreover, assessment of the topological aspects of objects is required for creating maps that can be used by various users.

Planners: To making appropriate decisions for infrastructure maintenance, constructions, upgrading and expansion of utilities planners use spatial data frequently. When planner involve in planning in certain area; extensive spatial analysis is made. The spatial analysis may include the closeness of the utility to buildings, electric transmission line and closeness of data network etc. Moreover correctly identifying the location, intersections and under/over passes of these networks is crucial. Therefore complete spatial and non-spatial data of the area of interest accuracy of object measurements time related information and other related spatial and non-spatial information is needed.

Municipal authorities: Authorities in a municipality are responsible to the facilities in their locality. These facilities include local roads, parks, public centers etc. During maintenances and construction of facilities, municipalities require the cadastre for a data of specific location. They need accurate measurements of features for example; dataset digitized from a lower scale map cannot be used for calculating road segment size because it affects the area size of the segment. Temporal accuracy of the data may also help in determining which road segments should be maintained when and for prioritizing. In terms of the data requirement of users, correct information on each feature road segments, local and high way roads, public owned and privately owned facilities, commercial, residential and recreational area, service centers of districts is very important. Therefore Attribute information should be accurate and should be well documented and easily accessible by the municipality authority as they frequently rely for multitude of decisions.

Courts: Courts usually deal with legal issues related to locations based disputes and passes decisions using both spatial and non-spatial information evidences. Courts may involve in resolving disputes, or on violation of laws of land management, on disputes related to land value compensation and sales related disputes. For these and related issues courts need correct, accurate and up-to-date spatial and non-spatial information; thus, they can make reliable decisions.

Financial institutions: Financial institutions are those working in banking and insurance. These institution officials usually make sure that a person who has relation with these institutions has the appropriate mortgage information. In such cases complete and up-to-date information on the customer's holdings is only sourced from the cadastre. Usually the area of parcels, its time of registration and validity of the registration period, and correctness of the associated information are important for them in finance related decisions.

Taxation: Taxation is divided into two types; property tax and land tax (van Oosterom, et al., 2006). In most countries Land tax or property tax is performed annually. The land tax

Table 4.1: sample quality information for application

ApplicationName	postionalAccuracy	AttributeAccuracy	LogicalConsistency	Completeness	TemporalAccuracy
taxation	0.2	100	80	95	99
urban planning	0.5		100	90	85
courts decision	1	100	78	70	
real estate	0.2			100	90

is completely dependent on the area size of the parcel owned by a person or organization and associated tax rate per unit square size is given. Accurate measurement of area of parcels should be with in the acceptable level. Moreover temporal information associated with: changes made through in the object should be available as the change in the property though time affects the tax. Land use and location related attribute information and market value of the holdings are also factors in taxing amount thus, they need to be accurate.

4.7.3 Data Used for prototype implementation

For implementing our prototype system, we used a parcel based Cadastral objects that have registered rights taken from the Dutch Kadaster (esd00-object). Arbitrary quality information corresponding to the quality elements in a tuple based spatial or attribute information is used in a PostgreSQL database management system. The main aim of this system is to show how to identify specific quality information of spatial data that users can make use while looking for data that fits-user-specific or application specific quality requirements. Therefore; the quality information used in this prototype is by no means reflection of the actual quality information the esd00-object dataset stored in our system in tblobject-data table.

For our prototype system implementation we use some arbitrarily assigned quality information for applications by which users can look fitting data for similar application. These data shown in the table below does not represent the actual quality information requirements fort the sated application. The actual data need to be gathered and maintained by experts in the field.

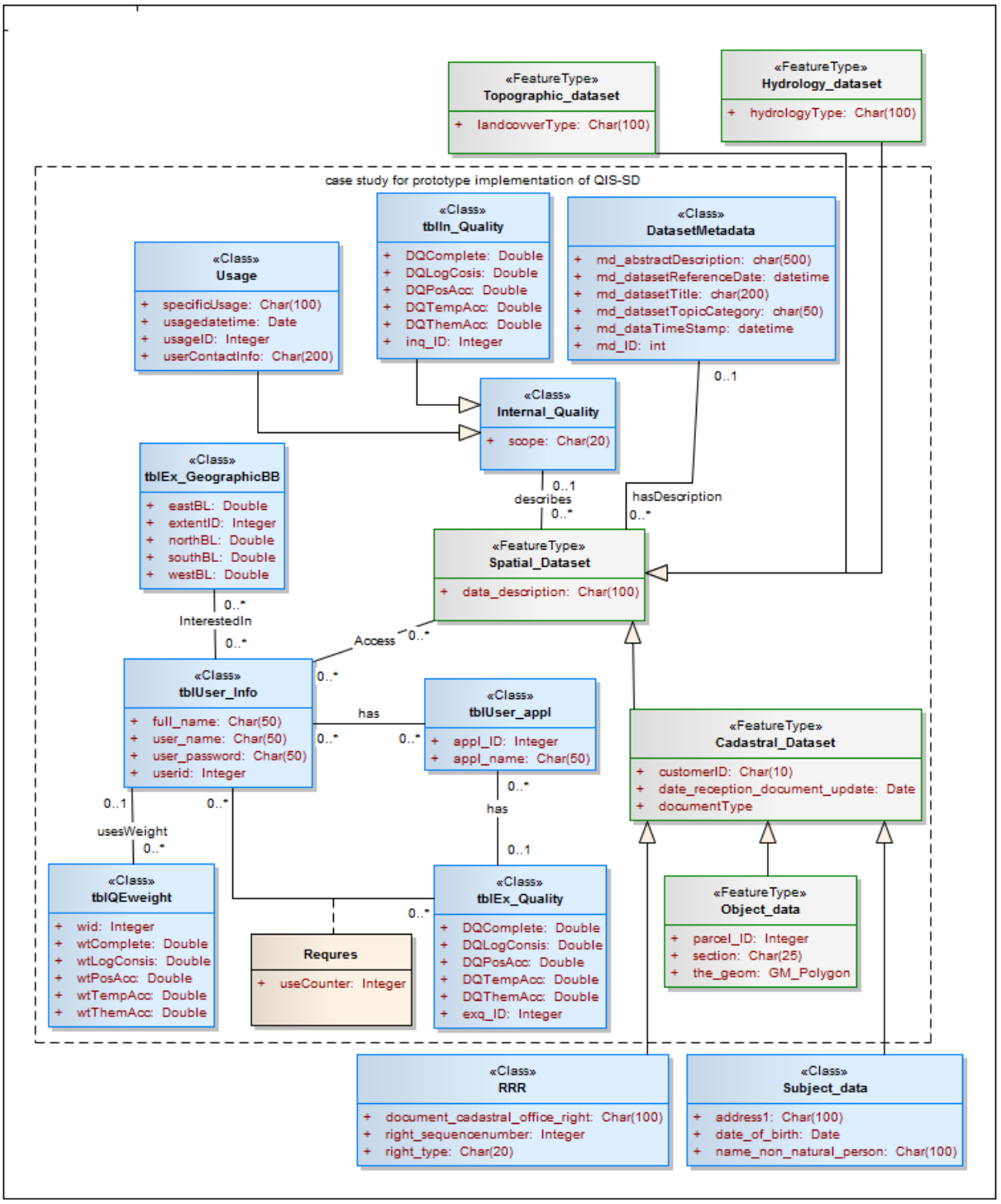

Figure 4.6: Conceptual data model for QIS-SD for the prototype

4.8 SUMMARY

This chapter presented the QIS-SD system design by defining the system requirements to the data model design for user profiles base quality information system. different services provided by QIS-SD and detailed description of service for meeting the objective of this research as stated in section 1.2. It stated the interaction between the user and the system while the user retrieving spatial data that fits the users needs either based on the requirements of his application or specified quality requirements the user may enter any time. The system interaction illustrated by use-case which shows the system and user interaction is presented. The use-case is elaborated by activity diagram and a conceptual data model for the user profile is designed. The conceptual data model is meant for the profile information which is dynamic that gradually updates to learn the behavior of the system users. This behavior is in terms of user spatial data consumption based on quality requirements of users. Moreover; background information about the case study selected for the system prototype implementation. By introducing why we select cadastral system as our case study we proceeded to explain the associated data and user characteristics in cadastral and the possible spatial data use an SDI domain. Moreover; the chapter highlighted the need for quality information and the potential impact when cadastral data is misused. Finally the data that we considered for the prototype implementation is introduced which is taken from Netherlands cadastre. The prototype implementation of the system based on the design given in section 4.6 is given in the next chapter for prototype.

Chapter 5
System Implementation in a prototype

5.1 INTRODUCTION:

This chapter focuses on the implementation of the proposed QIS-SD system in a prototype for demonstration. As can be seen on section 4.6, we designed the conceptual platform independent data model (PIM) for user profile. In this chapter as part of the implementation of the system in prototype

We used PostgreSQL based Procedural language/Postgresql Structured Query Language (PLPGSQL) to create our functions to accomplish the tasks in side the database. The remaining code is programmed in Hypertext Preprocessor (PHP) Programming language. The prototype system is evaluated based on parcel information and with arbitrary quality information assigned for the five geographic quantitative quality elements as described in ISO 19113. From the overview quality elements, we included usage information for user application based search.

5.2 TRANSFORMATIONS

Transformations from the conceptual data model which a platform independent model (PIM)to logical data model platform specific Model (PSM) is made. Then by transforming the PSM to physical model we generate PostgreSQL based data definition language (DDL) script. The transformation of the prototype system implementation is from the PIM given in section 4.6 is transformed to PSM (the PostgreSQL/PostGIS). Then the PSM is transformed to DDL to create actual database of our system. The prototype conceptual model consists of cadastral objects from the Netherlands cadastre and we took the whole attributes used in that database.

The prototype system is developed using PostgreSQL database management system as back end using PLPGSQL database programming used to develop some function in the side the database. The front end and some functionalities of the system are also developed using PHP as a dynamic web page.

5.3 SYSTEM SERVICES

This prototype system implementation shows how the main functionalities of the system work to accomplish the objectives. The requests of users and the response of the system are shown below for each system service we implemented in the prototype.

5.3.1 System login service

Users need to provide user credentials each time they request data. These credentials also help for creating a session for the users.

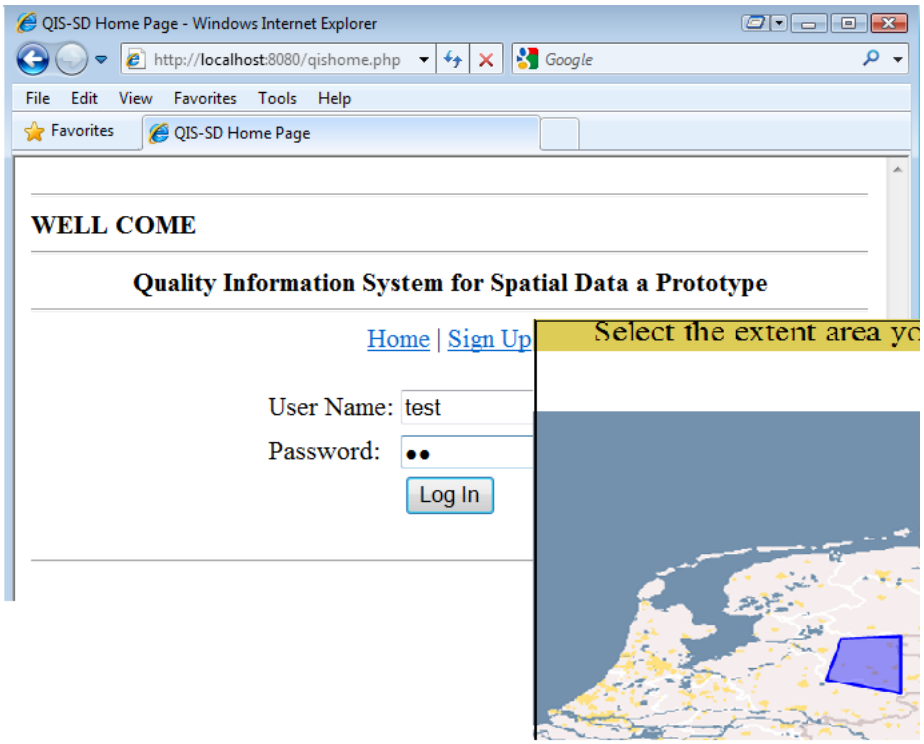

Figure 5.1: System Login service

5.3.2 User specific quality requirements spatial data search

This service aims at satisfying users spatial data request as explained on figure 4.3. The request is based on quality requirements for spatial data. the system will take user quality information requirements or users' application for which they request spatial data. Based on their requirements the system retrieves spatial data that best fits the user's needs.

When the spatial objects inside the user defines bounding box are compared if their centroid lies With in the bounding box, transformation of Spatial Reference System Identifier(SRID) conversion is required. Because the SRID of the bounding box as extracted from the openlayer may not always be the same to the spatial data users would be looking. Therefore, generally the following SRID conversion is used.

When search for spatial data with in a specific area is located by bounding box, the centroid of the object is checked if it is inside the bounding box. The POSTGIS "ST_WITHIN function" is used to check if an object's centroid is inside the desired location then to compare the quality information associated to the object. To accomplish this, an ST_WITHIN function takes other POSTGIS functions to convert the SRID of the Coordinate of the bounding box to that of the data. generally the following functions are used as discussed in [35].

- ST_WITHIN completely with inside the bounding box, it takes two geometry objects, here we are using with the centroid of the feature objects to be retrieved

- $ST\_centroid$ Used to compare an object by its geometric center
- $ST\_GeomFromEWKT$ Return a specified ST_Geometry value from Extended Well-Known Text representation from the object to make the two geometry representations the in the same SRID and enable the ST_WITHIN function work.
- $ST\_transform$ to transform the coordinates to similar geometry with the object to be compared.
- $ST\_GeomFromText$ Return a specified ST_Geometry value from Well-Known Text representation for the ST_Transform function work. for transforming to similar SRID.
- $GETSRID$ to find the SRID of the object in order to match with the bounding box SRID

$ST\_WITHIN(ST\_Centroid((ST\_GeomFromEWKT(od.the\_geom))),$
$ST\_transform(ST\_GeomFromText(BB, 4326), getsrid(od.the\_geom)))$

The above statement is used to compare if the centroid of spatial objects users are interested are inside the bounding box i.e. BB

The following algorithm is meant for searching spatial data based on user specific quality requirements. The associated PHP and PLPGSQL based program is provided on the appendix B.

Algorithm 1: Spatial data retrieval

S_D is set of spatial data
S_{IQ} is set of internal quality of data
S_{wQ} is set of weights of quality elements
U_{BB} is user bounding box selection
S_r user preference data delivery
if $ST\_centroid(S_D)$ IS INSIDE U_{BB} **then**
  if $S_{IQ} = S_{QR}$ OR S_{QR} IS NULL **then**
    SELECT $S_D(S_{IQ})$
    if $S_D(S_{IQ}) > 1$ **then**
      identify S'_{wQ}
      identify S''_{wQ}
      identify S'''_{wQ}
      SELECT S_D
      WHERE $max(S_{IQ}(S'_{wQ}))$ OR
      $max(S_{IQ}(S''_{wQ}))$ OR
      $max(S_{IQ}(S'''_{wQ}))$
      ORDER BY S'_{wQ}
    end if;
  end if;
end if;

During search, filtering of spatial data whose either of quality element non-exists is considered as unfit data even if the rest of the parameters are fulfilled. This is accomplished by the following filtering query as part of the $WHERE\ clause$ of the select statement.

- WHERE $S_{IQ} = S_{QR}$ OR S_{QR} IS NULL.

This automatically ignores the unspecified quality element in the external quality and ignores spatial data if a parameter which specified in the external quality does not have internal quality.

The GML and metadata output snapshot of these searching algorithm are shown in the following picture

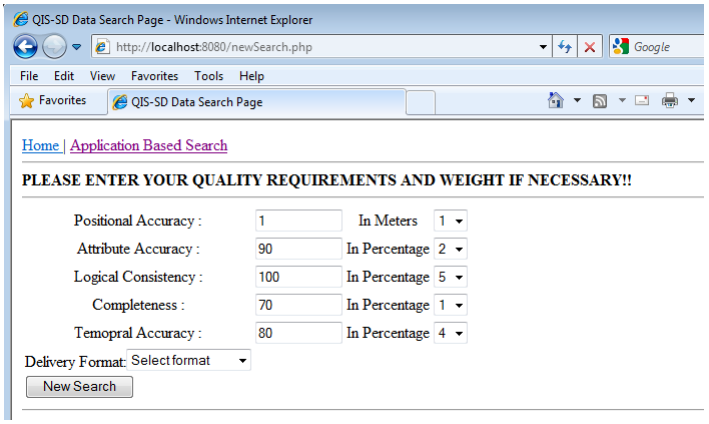

Figure 5.2: System searching service

5.3.3 Application based spatial data search

The application based quality information of spatial data can be obtained from predefined spatial data quality requirements or from the data quality overview elements usage information. In this study usage information is only considered. This prototype implement is for spatial data search based on user specific application searches spatial data from both the usage information and user defined applications. The corresponding code for this algorithm is on appendix C.

The following algorithm shows the user specified application based search

Algorithm 2: Retrieve spatial data based on user application

A_u is user application
D_u is data usage information (usage quality overview element)
S_D set of spatial data
A_{QR} Application quality requirements
S_{wQ} Set of quality element weights
S_r set of spatial data delivery means as chosen by the user
if $centroid(S_D)$ IS INSIDE U_{BB} **then**
  if $A_u = D_u$ **then**
    SELECT $S_D(D_u)$
  else
    Identify A_{QR}
    Identify S'_{wQ}
    if $A_{QR} = S_D(S_{QR})$ **then**
      SELECT $S_D(S_{QR})$
      ORDER BY S_{wQ}
    end if;
  end if;
end if;
S_D encode to S_r

The searching process starts from the overview quality elements specifically by comparing the usage information of spatial data (if any). If the usage information can show the usefulness of spatial data for that of the user, those data are retrieved to the user. The next means of finding data relevant to a user application is by searching the quality requirements of an application stored by default in the system. The quality requirement then is used to search a new data. The later approach may cause system performance effect, however for our prototype system our data is stored locally and we are using limited spatial data and the system performance factor is small.

The following snapshot shows the application based search with some results

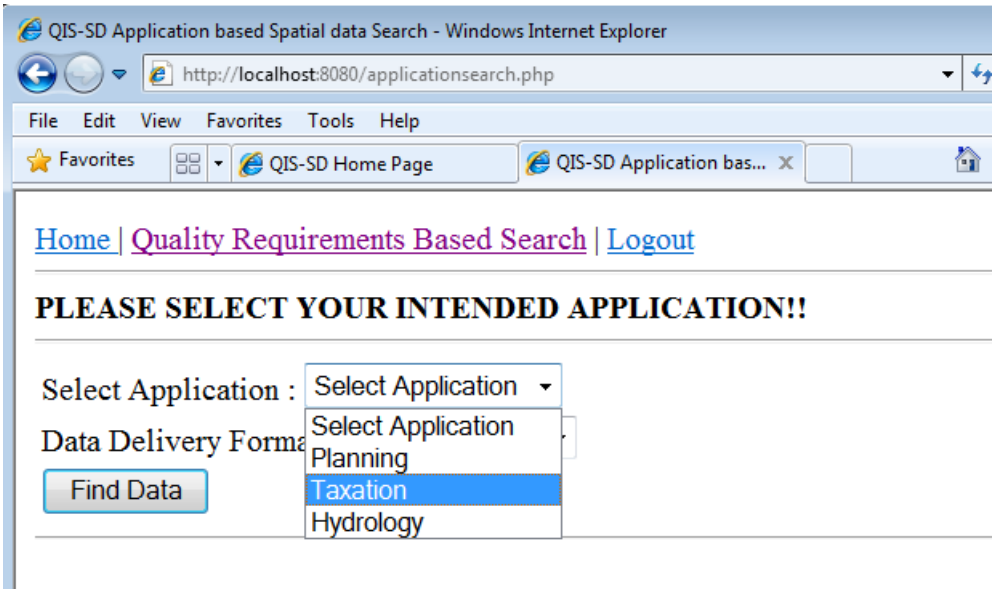

Figure 5.3: Application based search system service

Users are expected to select an application they want data for. The dropdown box is populated from the applications defined by default in the database. When users' intended application is not exactly the same to what is available in the dropdown box, they can select similar applications. Usually the application based search is important for average users who do not require spatial data for sophisticated spatial analysis. Therefore; quality information for general applications is easy to define and maintain to use as default requirements of these users.

5.3.4 Profile information update

When users interact with the system, for retrieving relevant spatial data, the system builds each user's profile. The profile information holds information that is relevant to the user's further interaction with the system.

The profile information is tracked when users enter their requirements to search data. Therefore the update information that take place in the profile is illustrated below.

Assume the following for all the update algorithms below

- S_{QR} is set of user quality requirements
- U_c current user
- S_{wQ} set of quality elements weights
- S_{wQnew} is the new weight entered
- S_D is set of spatial data
- S_{IQ} is set of spatial data internal quality

The following algorithm is meant for recording user spatial data quality requirements when the user searches spatial data. All new set of quality requirements a user used are recorded in the

profile. When an existing set of quality requirements is used a counter is updated to show how often it is used.

Algorithm 3: Profile information update

1: **if** S_{QR} exists in EX-QUALITY with U_c **then**
2: UPDATE REQUIRES SET counter +1
3: EXIT
4: **else**
5: INSERT INTO EX-QUALITY with S_{QR}
6: INSERT INTO REQUIRES S_{QR} with U_c AND SET counter to 1
7: **end if**;

The following algorithm is meant for updating weight of quality elements a user uses during searching spatial data. This works by simple averaging previous weight with new weight for a quality element. The update is made implicitly while user searches for spatial data. The wright stored is then used for later use during recommender service.

Algorithm 4: Update quality elements weight by averaging

1: **if** S_{wQ} AND U_c Exist **then**
2: $S_{wQ} = (S_{wQ} + S_{wQnew})/2$
3: UPDATE WEIGHT with S_{wQ}
4: **else**
5: INSERT WEIGHT (S_{wQnew} AND U_c)
6: **end if**;

The following algorithm is meant for recording a user's spatial data access-history. The user data access history is made while a user is accessing spatial data based on the retrieving service. Based on the user preference expressed by the weights, the most preferred data is recorded as a user preference.

Algorithm 5: update user data access history

1: **if** $S_{QR} = S_D(S_{IQ})$ **then**
2: **if** S_D AND U_c exists **then**
3: EXIT
4: **else**
5: identify the first output S_D
6: INSERT INTO access-history VALUES U_c, S_D
7: **end if**;
8: **end if**;

Assume that a single spatial data will be retrieved if it's more than one it is providing in order of relevance form most relevant to least relevant and the former is tracked. This information once explicitly learn by the system, it is used by the system to suggest spatial information relevant to users up on login.

5.3.5 System recommender service

The profile information is also used for suggesting users with spatial data that users are interested in based on their previous quality information preferences and previous access of data. When a user logs into the system, links are provided to select any of the recommended data. Therefore; based on the weights given to each of the quality elements the spatial data with highest quality information for the most preferred quality element is retrieved.

Assume the following:

- U_c is current user
- S'_{wQ} is the highest weight element in the profile
- S_r is a data delivery means user chosen
- Desc Descending

The user spatial data access history can be recommended to the user with the order of the the user's preference in terms of the data quality preference of the user from the profile. The following algorithm recommends spatial data based on user data access history.

Algorithm 6: Recommending data based on user data access-history

1: SELECT S_D FROM access-history
2: WHERE U_c
3: ORDER BY S'_{wQ} Desc
4: S_D encode to S_r

The following algorithm is for recommending a user spatial data that fits the frequently used set of quality requirement by the user from the user's profile information.

Algorithm 7: Quality requirements with frequently used

1: **if** $ST\_centroid(S_D)$ IS INSIDE U_{BB} **then**
2: **if** $S_{QR} = S_D(S_{IQ})$ **then**
3: SELECT S_D
4: WHERE S'_{wQ}
5: ORDER BY S'_{wQ} Desc
6: **end if**;
7: **end if**;
8: S_D encode to S_r

The following algorithm recommends spatial data based on the higher weight set of requirements from the profile. First it finds higher weight attribute from the weights given used by the user. Then selects the set of requirements of the user from the profile with higher weight attribute, then that set of quality requirement is used to search data.

Algorithm 8: Quality requirements with frequently used

1: identify S'_{wQ}
2: identify $S_{QR}(S'_{wQ})$
3: **if** $S_{IQ} = S_{QR}(S'_{wQ})$ **then**
4: SELECT $S_D(S_{IQ})$
5: **end if;**
6: S_D encode to S_r

5.4 SUMMARY

In this chapter a prototype implementation procedures of the QIS-SD is described shortly. It includes the functionalities of the system that are tracking user access history, user quality requirements, and the spatial extent users are interested in with user assigned weights of user quality requirements for constructing user profiles in using spatial data. Initially the spatial data retrieval functionality algorithm is presented. This algorithm is meant for retrieving the best possible available spatial data based on user explicit requirements input searched and provided to the user based on the weighted quality elements. For users who can not specify their quality requirements, functionality is presented to specify their required application on certain location of interest which is identified by spatial extent. The system also suggests spatial data for users based on their previous quality requirements or applications use history.

Generally the prototype implementation shows all the user profiling based techniques to address fitness-for-use for a data quality model we discussed in chapter three and designed in chapter can be fully implemented. The full implementation of these methods with some possible enhancements that we recommend in the next chapter can develop a full fledged system for searching spatial data based on the ever increasingly diversified user requirements for data quality.

The introduction of user profiles method for data quality models in SDI is helpful not only for users, but also the spatial data producers community can make benefit of it. Because the user profiles can be used as a knowledge base for the producers to enable them predict the possible spatial data requirements in terms of its quality. The knowledge base is therefore usable for producers in providing metadata of their products based on user profiles. A general discussion and further research to enhance the system and related research issues are discussed and stated out for recommendation in the following chapter.

Chapter 6
Conclusion and recommendation

6.1 INTRODUCTION

This chapter presents the summarized view of the thesis work by focusing on the specified objectives and the questions raised to meet these objectives. Further issues for future research are also recommended.

6.2 CONCLUSION

The objective of this research work was to devise a method for constructing user profiles for spatial data quality users, thereby design a system and implement it in a prototype for serving users with spatial data based on their profile. In chapter 1 having introduced the problem statement and motivation for the research; we specified several research objectives and research questions to answer these questions thereby to meet the objects.

The concept of spatial data quality has been perceived differently in the producers and users of spatial data community. While producers look for standards and guidelines for describing, maintaining and communicating spatial data to users, the users on the other hand look for spatial data from the view point of meeting their purposes. Even though producers may understand the users' perspectives, producers can not produce spatial data that satisfy every potential user requirements. Moreover, as massive spatial data becomes easily available and shared via SDI, the user community gets more diversified and their characteristics towards spatial data use get wider. Hence, producer-centric way of describing spatial data does not fulfill all potential user requirements. This creates difficulty in determining fitness-for-use as a result spatial data is being misuse.

The web technology brought a great opportunity for many users of spatial data for accessing, customizing, and using spatial data easily. However, the problem of cryptic description of spatial data quality and its means of communicating to prospective users is hampering its proper use. Communicating of spatial data quality to users of different expertise level in understanding, interpreting and using spatial data quality is affecting users' ability of determining fitness-for-use. While GIScience expert users can easily understand quality of spatial data when provided, non-experts are facing difficulties to understand metadata which is commonly used to communicate quality of spatial data. The high availability and easy to use and customize spatial data influences the distinguishing line between expert and non-expert users to be fuzzy that ultimately result in metadata frequently being overlooked by many users. Therefore, spatial data description that considers diversified user requirements for quality content, structure and reporting means is necessary to ease the users' problems of searching for spatial data based on required quality.

Automated users such as web services with different ways of using spatial data need different approaches in communicating quality. They consuming spatial data and quality are not by specifying requirements. Communicating them with machine readable format spatial data and quality, needs different approach. Therefore tools for automatic extraction of spatial data based on profiles

of users are more usable for users to automatically determine fitness of data for application and use quality in automated processing services.

Users are always keen to find spatial data that meets their requirements. However their requirements are not easy to state. User quality requirements can be very complicated, and continuously vary from application to application and from user to user. Users may use various quality criteria for searching spatial data that fits their applications. This scenario is difficult for producers of spatial data to predict all the divergent requirements. Therefore, producers can only describe the internal quality of spatial data and it is up to the users to determine suitability of the data. Mostly non-expert users are unable to specify their quality requirements in terms of values for standard quality elements and cannot easily determine best fitting data for their applications. User profile techniques we introduced in this research are effective to capture users' behaviors in using spatial data and quality. Thoroughly analyzing users' behaviors towards using spatial data quality and construct users profiles based on their characteristics is potentially promising approach for addressing fitness-for-use effectively. For users of various expertise levels with varying requirements one can maintain all possible application of spatial data based on the profile. This enables users easily determine fitness-for-use of spatial data with out detail investigating the contents of the quality information as it is not easy for every user. Moreover, similarity among users who share characteristics based on their profile can be used for suggesting spatial data.

Therefore, categorizing users to expert and non-expert may help in constructing initial user requirements for fitness-for-use. However, categorizing users initially for profiling contributes little because user profiles are constructed on individual user basis and the profile can only help in formulating default quality requirements for possible application domains for spatial data.

Techniques of communicating spatial data quality to users as described above are very user specific as the human understanding of the quality information is very diverse. Quality information communication separated from the data it describes for users to understand it and use it for searching spatial data based on fitness-for-use can only work for highly skilled users as they can understand the quality content and its structure. However, for many users communicating spatial data based on metadata reporting are not to understand and use not only for the naïve users but also for average and some expert users. Therefore, methods that serve spatial data based on quality requirements maintained as user profiles are more effective in addressing fitness-for-use than communicating quality information that few users can easily understand. This is because the ultimate goal of quality communication to users is looking for data that fits a specific quality level required for applications.

Tools like geoportals for searching spatial data are available on SDI; however, the search lacks to deal with the behaviors of its users specifically in the quality aspect. A user profile captured from thoroughly assessed user characteristics in using quality information can be used to search spatial data effectively. The profile information has individual user quality requirements and from these requirements similarity among user behaviors can be identified. Therefore, profile information helps to identify user preferences on spatial data and a flexible and more robust searching and recommendation methods of spatial data are achievable in both group and individual user basis

From this perspective, the system we designed maintains information about user information, user quality requirements with preferences on each requirement and on an individual user basis. This information is important in determining fitness-for-use of spatial data based on individual user profiles. This information can also contribute to user profiles but the group is not necessarily based on the categories (as expert and non-expert) we defined in the introduction. One typical example of this is users with similar application have similar requirements for that specific application not necessarily for other applications. Moreover, all applications can not always be designated as expert user application or non-expert user application as stated previously as a user

do not have distinguishing line between the expert and non-experts. The prototype system implementation reveals this idea. Therefore, the user categorizing can help to assess several commonly used applications by a certain group but the profile information is individually maintained and its use in searching spatial data based on individual requirements. However, the communication of spatial data quality for non-human user is an important categorizing of users. This is because not only the communication but also the result of communicated spatial data for these users is to find a new spatial data with new quality information and not to determine fitness-for-use as the way human users do

Generally the user profiles techniques we introduced in this research are based on the implicit and explicit ways of learning users' behaviors. Through these methods of profiling, specific quality requirements of users, their intended applications, the spatial extent they and user specified weights of quality elements are stored. The searching method we used in this system uses filtering technique to find out the best available spatial data based on users' quality requirements. The prototype implementation of the system is presented with description and pseudo-code algorithms of the main functionalities. We then attached the PHP and PL/PGSQL based script of the prototype implementation in the appendix.

In conclusion user types are increasing in number and their characteristics in using spatial data and quality is getting more and more diversified. Automatic quality aware computer processes are also increasing. Categorizing human users to expert and non-expert users' groups provides little support in using group based user profiles. However, it can overcome certain difficulty in determining fitness-for-use by identifying commonly used applications quality requirements. Therefore constructing user profiles in terms of users' behaviors in using spatial data and quality will help to learn all types of users despite their expertise level.

Finally we conclude that all the objectives set out for this research are well met with the exception of demonstrating the non-human user use of our system. This is because, it needs a machine to machine communication which is different from the login process our system provides via web based interface. As a result we indicated it for further research below. The result from the prototype implemented indicates us that user profile techniques in addressing fitness-for-use is promising approach for the ever increasing user types. When quality models are designed based on standards, user profiles that consider certain standard of implementing these data quality models can greatly improve the search for the best data by all types of users. This is proved in our design of the prototype based on the ISO 19113 quality elements. However it needs a more robust and more sophisticated profile information analysis for determining users' similarity in terms of their spatial data use. Therefore, as a continuation work of this study several issues are pointed out below for further research.

6.3 RECOMMENDATION

For future research in enhancing the services provided in this study to cater for more information profiling and more flexible services and searching methods we recommend the following research ideas.

1. Time constraint limited us to deeply assess the communication techniques quality aware WPSs use when they process spatial data together with its quality information. Therefore, For the system to work with automated system not by typed login but by service to service communication of systems, we recommend the design to be modified in such a way that service to service communication and XML based data transfer capabilities can be incorporated. Detailed exploration of the ISO 19139 Geographic Information - Metadata - XML schema implementation be made, and quality awareWeb Processing Services like the

UncertWeb [5] be well assessed for understanding the inputs, outputs and the processing behaviors of awareWeb Processing Service in terms of quality use.

2. We used requirements based on specified list of standard quality elements(only quantitative elements) and usage information in this study. However, including lineage and purpose information requirements to determine fitness-for-use would also contribute to enhance the result of this system in both specific data retrieval and recommender services. So for further study we recommend the profiling to be more inclusive in terms of requirements in determining fitness-for-use. Moreover, supporting other more sophisticated weighting techniques can enhance the recommender service of our system. Thus, we recommend for further evaluation of other weighting techniques to be used with this system as a user preference representation and to include other spatial data delivery mechanisms as part of the user profile.

3. The system developed as a prototype in this study is only as an initial proof of concept for usefulness of user profiles. It needs to be set into SDI real-case like geo-portals. This can make SDI based web service aware of the quality aspect of all their service and data understandable by all potential users. So we recommend a means of connecting this user profiles method design in this study to the actual SDI.

Bibliography

[1] J. Arlow and I. Neustadt. *UML and the unified process : practical object oriented analysis and design*. Addison Wesley, Boston etc, 2001.

[2] A.T. Boin and G.J. Hunter. What communicates quality to the spatial data consumer. In *Proceedings 5th International Symposium on Spatial Data Quality*. ITC, 2007.

[3] M. Caprioli, A. Scognamiglio, G. Strisciuglio, and E. Tarantino. Rules and Standards for Spatial Data Quality in GIS Environments. In *International Cartographic Conference–ICC*, pages 10–16. (ICA, Durban, South Africa, 2003.

[4] A. Coote and L. Rackham. Neogeographic data quality–is it an issue. In *AGI Geocommunity conference, ConsultingWhere Ltd*. AGI, ConsultingWhere Ltd., 2008.

[5] L. Dassonville. Quality Management, Data quality and users, Metadata for geographical information. In *Proceedings of the International Symposium on Spatial Data Quality '99*. OEEPE/ISPRS, 1999.

[6] R. Devillers, Y. Bédard, and R. Jeansoulin. Multidimensional management of geospatial data quality information for its dynamic use within GIS. *Photogrammetric Engineering & Remote Sensing*, 71:205–215, 2005.

[7] R. Devillers, Y. Bedard, R. Jeansoulin, and B. Moulin. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science*, 21:261–282, 2007.

[8] R. Devillers, M. Gervais, Y. Bédard, and R. Jeansoulin. Spatial data quality: from metadata to quality indicators and contextual end-user manual. In *OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management*, pages 21–22. OEEPE/ISPRS, 2002.

[9] R. Devillers and R. Jeansoulin. *Fundamentals of Spatial Data Quality (Geographical Information Systems series)*. ISTE London etc., 2006.

[10] R. Devillers, A. Stein, Y. Bédard, N. Chrisman, P. Fisher, and W. Shi. Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities. *Transactions in GIS*, 14(4):387–400, 2010.

[11] C.R. Ehlschlaeger, A.M. Shortridge, and M.F. Goodchild. Visualizing spatial data uncertainty using animation. *Computers & Geosciences*, 23(4):387–395, 1997.

[12] A.U. Frank. Metamodels for data quality description. *Data quality in Geographic Information: From error to uncertainty*, pages 15–29, 1998.

[13] A.U. Frank, E. Grum, and B. Vasseur. Procedure to select the best dataset for a task. *Geographic Information Science*, 3234:81–93, 2004.

[14] J. Gao, W. Zhang, and L. Meng. A method for heterogeneous spatial data integration with storage agent in grid. In *Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on*, pages 1230–1233. IEEE, 2005.

[15] M.F. Goodchild. Beyond Metadata: Towards User-Centric Description of Data Quality, Keynote paper. *5th Int. Symposium Spatial Data Quality, ITC,Netherlands*, pages 13–15, 2007.

[16] M. Grčar. User profiling: Collaborative filtering. In *Proceedings of the conference on data mining and warehouses (SIKDD 2004). Ljubljana, Slovenia*. Citeseer, 2004.

[17] Y. Hijikata. Implicit user profiling for on demand relevance feedback. In *Proceedings of the 9th international conference on Intelligent User Interfaces*, pages 198–205. ACM, 2004.

[18] F. Hopfgartner and J.M. Jose. Semantic user profiling techniques for personalised multimedia recommendation. *Multimedia Systems*, 16:255–274, 2010.

[19] G.J. Hunter, A.K. Bregt, G.B.M. Heuvelink, S. Bruin, and K. Virrantaus. Spatial Data Quality: Problems and Prospects. *Research Trends in Geographic Information Science*, pages 101–121, 2009.

[20] GJ Hunter, S. Hope, Z. Sadiq, A. Boin, M. Marinelli, AN Kealy, M. Duckham, and RJ Corner. Next-Generation Research Issues in Spatial Data Quality. In *Proceedings of SSC*, pages 865–872. Citeseer, 2005.

[21] ISO/TC211. *ISO 19113:2002 Geographic Information Quality Principles*. International Organization for Standardization, Geneva, 2002.

[22] ISO/TC211. *19115:2003 Geographic Information Metadata*. International Organization for Standardization, Geneva, 2003.

[23] ISO/TC211. *ISO 19114:2003 Geographic information – Quality evaluation procedures*. International Organization for Standardization, Geneva, 2003.

[24] Ivar Jacobson, Grady Booch, and James Rumbaugh. *The unified software development process*. Addison-Wesley Longman Publishing Co. Inc, Boston., 1999.

[25] J. Kaufmann, D. Steudler, and Working Group 1 of FIG Commision 7. *Cadastre 2014: A vision for a future cadastral system*. International Federation of Surveyors, 2001.

[26] T. Kuflik and P. Shoval. Generation of user profiles for information filteringâĂŤresearch agenda (poster session). In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information trieval*, pages 313–315. ACM, 2000.

[27] D.J. Maguire and P.A. Longley. The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems*, 29(1):3–14, 2005.

[28] SIA Majid. Benefits and Issues of Developing a Multi-Purpose Cadastre. *International Archives of Photogrammetry and Remote Sensing*, 33(B4/1; PART 4):22–29, 2000.

[29] J. McHugh, S. Abiteboul, R. Goldman, D. Quass, and J. Widom. Lore: A database management system for semistructured data. *ACM Sigmod Record*, 26(3):54–66, 1997.

[30] X. Meng, Y. Xie, and F. Bian. Distributed Geospatial Analysis through Web Processing Service: A Case Study of Earthquake Disaster Assessment. *Journal of Software*, 5(6):671, 2010.

[31] G. Navratil and A.U. Frank. Processes in a cadastre. *Computers, Environment and Urban Systems*, 28:471–486, 2004.

[32] H.J. Onsrud. Liability in the use of geographic information systems and geographic datasets. *Geographical information systems: Management issues and applications*, pages 643–652, 1999.

[33] D. Poo, B. Chng, and J.M. Goh. A hybrid approach for user profiling. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, page 9. IEEE, 2003.

[34] S. Pooja, D. TYAGI, and P. BHADANA. Weighted Page Content Rank for Ordering Web Search Result. *International Journal of Engineering Science and Technology*, 12(2):7301–7310, 2010.

[35] P. Ramsey. PostGIS manual. *Refractions Research Inc*, 2005.

[36] M.Z. Sadiq and M. Duckham. Integrated Storage and Querying of Spatially Varying Data Quality Information in a Relational Spatial Database. *Transactions in GIS*, 13(1):30–42, 2009.

[37] M. Salzmann, A. Hoekstra, and T. Schut. Cadastral Map Renovation–a Dutch Perspective. In *xxx FIG Congress. Brighton*. FIG, 1998.

[38] S. Schiaffino and A. Amandi. Intelligent user profiling. In *Artificial intelligence*, pages 193–216. Springer-Verlag, 2009.

[39] M.A. Silva and E. Stubkjær. A review of methodologies used in research on cadastral development. *Computers, Environment and Urban Systems*, 26(5):403–423, 2002.

[40] C. Traynor and M.G. Williams. Why are geographic information systems hard to use? In *Conference companion on Human factors in computing systems*, pages 288–289. ACM, 1995.

[41] J. Weakliam, D. Wilson, and M. Bertolotto. Personalising map feature content for mobile map users. *Map-based Mobile Services*, pages 125–145, 2008.

[42] I.P. Williamson. Best practices for land administration systems in developing countries. In *International Conference on Land Policy Reform, July*, pages 25–27. LAP-C, Jakarta, 2000.

[43] J. Xiao. Mining Evolving Web Sessions and Clustering Dynamic Web Documents for Similarity-Aware Web Content Management. *Advanced Data Mining and Applications*, pages 99–110, 2008.

[44] A. Zargar and R. Devillers. An Operation-Based Communication of Spatial Data Quality. In *Advanced Geographic Information Systems & Web Services, 2009. GEOWS'09. International Conference on*, pages 140–145. IEEE, 2009.

Appendix A: Quality requirements based Spatial data retrieval

The function defined here are used in other programs for searching spatial data based on quality for a specified user.

The following function retrieves spatial data based on quality requirements retrieved from user profiles. The user profile's quality requirements can be fetched based on frequently used requirements of higher weight requirements.

```
--for retrieving data that meets the
--user specified quality requiremenets
-- Function: retrieverelevantdata()

-- DROP FUNCTION retrieverelevantdata();

CREATE OR REPLACE FUNCTION retrieverelevantdata()
  RETURNS void AS
$BODY$
DECLARE

BEGIN

INSERT INTO tempdata
(SELECT tblObject_data.parcel_id,
tblObject_data.x_coordinate,
tblObject_data.y_coordinate,
tblObject_data.building_code,
tblObject_data.land_use_type_non_built,
tblObject_data.year_of_purchase,
tblObject_data.indication_more_cadastral_objects,
tblObject_data.description_parcel_part,
tblObject_data.cadastral_municipality_code,
tblObject_data.conveyor,
tblObject_data.date_of_agreement,
tblObject_data.area,
tblObject_data.indication_estimated_area,
tblObject_data.the_geom

--when comparing internal and external quality
--check for null requirements
FROM (tbltemp_qlty INNER JOIN tblin_quality
ON tbltemp_qlty.DQPosAcc >= tblin_quality.DQPosAcc
```

```
OR tbltemp_qlty.DQPosAcc IS NULL
AND tbltemp_qlty.DQThemAcc <= tblin_quality.DQThemAcc
OR tbltemp_qlty.DQThemAcc IS NULL
AND tbltemp_qlty.DQLogCosis <= tblin_quality.DQLogCosis
OR tbltemp_qlty.DQLogCosis IS NULL
AND tbltemp_qlty.DQComplete <= tblin_quality.DQComplete
OR tbltemp_qlty.DQComplete IS NULL
AND tbltemp_qlty.DQTempAcc <= tblin_quality.DQTempAcc
OR tbltemp_qlty.DQTempAcc IS NULL)
  INNER JOIN tblObject_data ON tblObject_data.inq_id= tblin_quality.inq_id);
--DELETE FROM tbltemp_qlty;  --empty the temporary table
END;
$BODY$
  LANGUAGE plpgsql VOLATILE
  COST 100;
ALTER FUNCTION retrieverelevantdata() OWNER TO gebresilas23986;
```

The following function takes the user specified application, the retrieves the default user defined quality requirements of the application compares it with the internal quality to and retrieves the data

```
-- Function: retrievedatatoappl(text)

-- DROP FUNCTION retrievedatatoappl(text);

CREATE OR REPLACE FUNCTION retrievedatatoappl(text)
  RETURNS void AS
$BODY$
DECLARE
pol geometry;
BEGIN
INSERT INTO tempData (SELECT parcel_id,
x_coordinate,
y_coordinate,
building_code,
land_use_type_non_built,
year_of_purchase,
indication_more_cadastral_objects,
description_parcel_part,
cadastral_municipality_code,
conveyor,
date_of_agreement,
area,
indication_estimated_area,
the_geom
FROM (tbltemp_qlty INNER JOIN tblin_quality
ON tbltemp_qlty.DQPosAcc = tblin_quality.DQPosAcc
OR tbltemp_qlty.DQPosAcc IS NULL
AND tbltemp_qlty.DQThemAcc = tblin_quality.DQThemAcc
```

```
OR tbltempquality.DQThemAcc IS NULL
AND tbltemp_qlty.DQLogCosis = tblin_quality.DQLogCosis
OR tbltemp_qlty.DQLogCosis IS NULL
AND tbltemp_qlty.DQComplete = tblin_quality.DQComplete
OR tbltemp_qlty.DQComplete IS NULL
AND tbltemp_qlty.DQTempAcc = tblin_quality.DQTempAcc
OR tbltemp_qlty.DQTempAcc IS NULL)
  INNER JOIN tblObject_data ON tblObject_data.inq_id= tblin_quality.inq_id AND
  ST_WITHIN (ST_Centroid((ST_GeomFromEWKT(od.the_geom))),
  ST_transform(ST_GeomFromText (pol,4326),
  getsrid(od.the_geom)))=true);
END;
$BODY$
  LANGUAGE plpgsql VOLATILE
  COST 100;
ALTER FUNCTION retrievedatatoappl(text) OWNER TO gebresilas23986;
```

The following function retrieves spatial data quality requiremets of user identified by a username. The quality requirements are retrieved by the highest weight element

```
CREATE OR REPLACE FUNCTION finduserqualityrequirements(integer, text)
  RETURNS void AS
$BODY$
DECLARE

BEGIN
INSERT INTO tbltemp_qlty (SELECT
    exq.DQPosAcc,
    exq.DQThemAcc,
    exq.DQLogConsis,
    exq.DQComplete,
    exq.DQTempAcc,
    exq.exq_id

    FROM tblUser_Info AS u,
 Requires AS urq,
 tblex_quality AS exq

   WHERE  u.userid=$1 AND
  u.userid=urq.userid AND
  urq.exq_id=exq.exq_id AND
  useCount IN
(SELECT max($2) AS wt
   FROM tblUser_Info AS u,
Requires AS urq,
tblex_quality AS exq
   WHERE u.userid=$1 AND
 u.userid=urq.userid AND
```

```
       urq.exq_id=exq.exq_id));
END; --it returns nothing simply loads the info to tmp table
$BODY$
  LANGUAGE plpgsql VOLATILE
  COST 100;
ALTER FUNCTION finduserqualityrequirements(integer) OWNER TO gebresilas23986;
```

Appendix B: Spatial data retrieval

This program is the whole of spatial data searching based on user specific quality requirements, this program also updates the weights, the access history is tracked and the quality requirements are tracked. This program makes call to the plpgsql functions defined in appendix A.

```
<!--Here the PHP code continues -->
<?php
//============accespts and Updates spatail extent====================
$conn = pg_connect("host=itcnt07 port=5432 dbname=esdkad08
user=gebresilas23986 password=23986" );
$d = $_POST['output'];
//echo $d;
echo "<a href ='newsearch.php'>CLick here</a>
TO search data inside the bounding box";
//$pupex = pg_query($conn, "SELECT POPULATEBB('$d')");
//===========Accepts user quality requirements======================


//accepting new quality information from the interface
$pa = $_POST['txtpa'];
$aa = $_POST['txtaa'];
$lc = $_POST['txtlc'];
$cm = $_POST['txtcm'];
$ta = $_POST['txtta'];
//=============================================
$search = $_POST['btnSearch'];  //accepts new search button event
$user = $_SESSION['username']; //current user name
//data delivery format of user preference
$dataformat = $_POST['delivery'];


//================================================================
if($search)
{
$newpa = $_POST['wpa']; //accepting new weight
$newaa = $_POST['waa'];
$newlc = $_POST['wlc'];
$newcm = $_POST['wcm'];
$newta = $_POST['wta'];

$conn = pg_connect("host=itcnt07 port=5432 dbname=esdkad08
user=gebresilas23986 password=23986");
```

```
 //finds the id of the current user
$sql= pg_query($conn, "SELECT finduserid('$user')");
$id = pg_fetch_array($sql);
$uid = $id[0];
//existing weights of a user
$qry = "SELECT w.wtDQPosAcc,
   w.wtDQThemAcc,
   w.wtDQLogConsis,
   w.wtDQComplete,
   w.wtDQTempAcc,
   w.wid
FROM tblweight as w, tbluser_info as u
WHERE u.userid ='".$uid."' AND u.wid = w.wid";
$qryex = pg_query($conn, $qry); //excutes result of query
$wght = pg_fetch_array($qryex); //accepts the result to array
//counts the number of items returned from query
$rows = pg_num_rows($qryex);
//if the weight exists updates it else inserts in to the weight table
if($rows != 0)
{
//fetch the available weights of the user
 $owpa = $wght[0];
 $owaa = $wght[1];
 $owlc = $wght[2];
 $owcm = $wght[3];
 $owta = $wght[4];
 $id = $wght[5];

//updates each of the weights of the elements
// by averaging to keep stay b/n 1 to 5
 $owpa = (($owpa+$newpa)/2);
 $owaa = (($owaa+$newaa)/2);
 $owlc = (($owlc+$newlc)/2);
 $owcm = (($owcm+$newcm)/2);
 $owta = (($owta+$newta)/2);
//update weights by averaging
 $qry = "UPDATE tblweight SET wtDQPosAcc = '".$owpa."',
  wtDQThemAcc='".$owaa."',
  wtDQLogConsis = '".$owlc."',
  wtDQComplete = '".$owcm."',
  wtDQTempAcc ='".$owta."'
  WHERE wid ='".$id."'";

 $upres =pg_query($conn, $qry); //excutes update query
}
else
{
```

```
 $sqlins = "INSERT INTO tblweight VALUES($newpa,$newaa,$newlc,$newcm,$newta)";
 $res = pg_query($conn,$sqlins); //excutes the insert query
 if(!$sqlins)
 {
  die("Error: No weight is stored"); //Error of insertion report
 }
//find the newly inserted weight id and assing the
// current user with the new weight
 $sqlid = pg_query($conn,"SELECT findmaxweightid()");
 $maxid = pg_fetch_array($sqlid);
 $newID = $maxid[0];
 $upsql = pg_query("UPDATE tbluser_Info SET wid = $newID WHERE userid =$uid");
}
//============================================================================

//=======// tracking the maximum weight given its its attribute name===============
$maxW = array($newpa,$newaa,$newlc,$newcm,$newta);
$maxVal = MAX($maxW); // finds maximum weited values
////finds maximum weited quality element its index and its attribute name
$maxattr = array_search($maxVal,$maxW);
switch ($maxattr)
{case 0:
$attrname = "DQPosAcc";
break;
case 1:
$attrname = "DQThemAcc";
break;
case 2:
$attrname = "DQLogConsis ";
break;
case 3:
$attrname = "DQComplete";
break;
case 4:
$attrname = "DQTempAcc";
break;
default:
die ("Error: OUT OF ARRAY INDEX!!");
}
//=========================================================
//check if not all the fields are emptry atleast one has to be specified
if((!$pa)&&(!$aa)&&(!$lc)&&(!$cm)&&(!$ta))
{
 echo "YOU NEED TO SPECIFY AT LEAST ONE QUALITY REQUIREMENT!";
}
else
 {
//==============================Delivery means selection==============
```

```
switch ($dataformat)
{
//====The search result is delivered as a data description format============
  case "Description":
  //===================attributes to be retrieved============================
  //the most prefered quality element is used
//   to sort, from higher quality to lower
  if($attrname == "DQPosAcc")
  {
          $st.="SELECT DISTINCT OD.parcel_id,
OD.x_coordinate,
OD.y_coordinate,
OD.year_of_purchase,
OD.cadastral_municipality_code,
OD.conveyor,
OD.date_of_agreement,
OD.area,
OD.the_geom,
INQ.*,
dmd. *
FROM tblin_quality as INQ, tblobject_data AS OD, datasetMetadata as dmd
WHERE OD.inq_id = INQ.inq_id AND ST_WITHIN(ST_Centroid((ST_GeomFromEWKT(od.the_geom))),
ST_transform(ST_GeomFromText ('.$d.',4326), 28992))= '".true."'";
//Check if each field if not empty, if empty ignores
//if not added to the query string
            if($pa)
               $st =$st." AND INQ.DQPosAcc <= $pa";
            if($aa)
 $st= $st." AND INQ.DQThemAcc >= $aa";
            if ($lc)
                 $st= $st." AND INQ.DQLogConsis>= $lc";
            if ($cm)
                 $st = $st." AND INQ.DQComplete >= $cm";
if($ta)
 $st = $st." AND INQ.DQTempAcc >= $ta";
 // to order and enable to select the highest one
$st = $st . " ORDER BY $attrname ASC";
echo $st;
$result = pg_query($conn, $st);  //excutes the SQL query of the $st
$rowsnum=pg_fetch_array($result); //fetches the retrieved record to array
$numbOfrecords= pg_num_rows($result); //counts the number of rows retrieved

if($numbOfrecords == 0) //when there is no data returned
{
Die ("<br>\n NO DATA MEETS THE ENTERD QUALITY SPECIFICATION!!");
}
//=====================display data to the user================================
 //display output to the user
```

```
 while($n = pg_fetch_assoc($result))
 {
  echo "<br>\n";
  foreach ($n as $key=>$value)
 {
echo strtoupper($key);
echo "==>".$value."<br>\n ";
 }
  }
}
 else
 {
 $st.="SELECT DISTINCT OD.parcel_id,
OD.x_coordinate,
OD.y_coordinate,
OD.building_code,
OD.land_use_type_non_built,
OD.year_of_purchase,
OD.indication_more_cadastral_objects,
OD.description_parcel_part,
OD.cadastral_municipality_code,
OD.conveyor,
OD.date_of_agreement,
OD.area,
OD.indication_estimated_area,
INQ.$attrname,
OD.the_geom

FROM tblin_quality as INQ, tblobject_data AS OD
WHERE OD.inq_id = INQ.inq_id AND
AND ST_WITHIN(ST_Centroid((ST_GeomFromEWKT(od.the_geom))),
ST_transform(ST_GeomFromText ('.$d.',4326), 28992))= '".true."'" ;
//Check if each field if not empty,
//if empty ignores if not added to the query string
            if($pa)
 $st =$st." AND INQ.DQPosAcc <= $pa";
            if($aa)
 $st= $st." AND INQ.DQThemAcc >= $aa";
            if ($lc)
                 $st= $st." AND INQ.DQLogConsis>= $lc";
            if ($cm)
                 $st = $st." AND INQ.DQComplete >= $cm";
if($ta)
 $st = $st." AND INQ.DQTempAcc >= $ta";
//helps to order and enable to select the highest one
$st = $st . " ORDER BY $attrname DESC";

$result = pg_query($conn, $st);  //excutes the SQL query of the $st
```

```
$rowsnum=pg_fetch_array($result); //fetches the retrieved record to array
$numbOfrecords= pg_num_rows($result); //counts the number of rows retrieved

if($numbOfrecords == 0) //when there is no data returned
{
Die ("<br>\n NO DATA MEETS THE ENTERD QUALITY SPECIFICATION!!");
}
//====================display data to the user==============================
 while($n = pg_fetch_assoc($result)) //display output to the user
 {
  echo "<br>\n";
  foreach ($n as $key=>$value)
 {//for($i = 0;$i<=$numbOfrecords;$i++)

echo strtoupper($key);
echo "==>".$value."<br>\n ";
 }
}
}
break;
//====================The search result is delivered as a GML data format===
case "GML":
if($attrname == "DQPosAcc")
{
$st = "SELECT DISTINCT asgml(the_geom),
OD.parcel_id,
OD.x_coordinate,
OD.y_coordinate,
OD.building_code,
OD.land_use_type_non_built,
OD.year_of_purchase,
OD.indication_more_cadastral_objects,
OD.description_parcel_part,
OD.cadastral_municipality_code,
OD.conveyor,
OD.date_of_agreement,
OD.area,
OD.indication_estimated_area,
INQ.$attrname,

FROM tblin_quality as INQ, tblobject_data AS OD
WHERE OD.inq_id = INQ.inq_id AND
ST_WITHIN(ST_Centroid((ST_GeomFromEWKT(od.the_geom))),
ST_transform(ST_GeomFromText ('.$d.',4326), 28992))= '".true."'";
//Check if each field if not empty,
//if empty ignores if not added to the query string
            if($pa)
                $st =$st." AND INQ.DQPosAcc <= $pa";
```

```
if($aa)
                $st= $st." AND INQ.DQThemAcc >= $aa";
            if ($lc)
    $st= $st." AND INQ.DQLogConsis>= $lc";
            if ($cm)
                 $st = $st." AND INQ.DQComplete >= $cm";
if($ta)
 $st = $st." AND INQ.DQTempAcc >= $ta";
 //helps to order and enable to select the highest one;
$st = $st . " ORDER BY $attrname ASC";
$result = pg_query($conn, $st);
$doc = new DomDocument("1.0");
$root = $doc->createElement('data');
$root = $doc->appendChild($root);
while($row = pg_fetch_assoc($result)) {
$node = $doc->createElement('spatialdata');
$node = $root->appendChild($node);
foreach($row as $fieldname => $fieldvalue)
{
if ($fieldname != 'asgml')
{
$node->appendChild($doc->createElement($fieldname, $fieldvalue));
}
else
{ $fragment = $doc->createDocumentFragment();
$fragment->appendXML($fieldvalue);
$node->appendChild($fragment);
}
}
    }
$doc->save("D:\www\GMLSearchResult\SearchResult.xml");
if($doc)
echo "Successfully saved to file <br>\n";
}
else
{ //attributes to be selected
 $st = "SELECT DISTINCT asgml(OD.the_geom)
OD.parcel_id,
OD.x_coordinate,
OD.y_coordinate,
OD.building_code,
OD.land_use_type_non_built,
OD.year_of_purchase,
OD.indication_more_cadastral_objects,
OD.description_parcel_part,
OD.cadastral_municipality_code,
OD.conveyor,
OD.date_of_agreement,
```

```
OD.area,
OD.indication_estimated_area,
INQ.$attrname,

FROM tblin_quality as INQ, tblobject_data AS OD
WHERE OD.inq_id = INQ.inq_id AND
ST_WITHIN(ST_Centroid((ST_GeomFromEWKT(od.the_geom))),
ST_transform(ST_GeomFromText ('.$d.',4326), 28992))= '".true."'" ;
//Check if each field is not empty,
// if empty ignores if not added to the query string
            if($pa)
                $st =$st." AND INQ.DQPosAcc <= $pa";
if($aa)
                $st= $st." AND INQ.DQThemAcc >= $aa";
if ($lc)
                 $st= $st." AND INQ.DQLogConsis>= $lc";
            if ($cm)
                 $st = $st." AND INQ.DQComplete >= $cm";
if($ta)
 $st = $st." AND INQ.DQTempAcc >= $ta";
//to order and enable to put the highest quality one on the top;
$st = $st . " ORDER BY $attrname DESC";
$result = pg_query($conn, $st);
$doc = new DomDocument("1.0");
$root = $doc->createElement('data');
$root = $doc->appendChild($root);
while($row = pg_fetch_assoc($result)) {
     $node = $doc->createElement('spatialdata');
     $node = $root->appendChild($node);
     foreach($row as $fieldname => $fieldvalue) {
if ($fieldname != 'asgml') {
$node->appendChild($doc->createElement($fieldname, $fieldvalue));
} else {
$fragment = $doc->createDocumentFragment();
$fragment->appendXML($fieldvalue);
$node->appendChild($fragment);
}
}
}
$doc->save("D:\SearchResult.XML");
if($doc)
echo "Successfully saved  on D:\SearchResult.xml<br>\n";
}
break;

default:
echo "<br>\n Please select format of delivery!";
break;
```

```
} //end of search start
//=======================================update quality requirements==========
//checks if the newly used id is not in the requirements list
$qrchk =pg_query($conn,"SELECT *
FROM tblex_quality
WHERE DQPosAcc = $pa,
  DQThemAcc = $aa,
  DQLogConsis =$lc,
  DQComplete = $cm,
  DQTempAcc =$ta");
$n = pg_num_rows($qrchk);
$k = pg_fetch_array($qrchk);
if($n==0)
{
$sqlin ="INSERT INTO tblex_quality values('$pa','$aa','$lc','$cm','$ta')";
$qryupdate = pg_query($conn,$sqlin);

//finds the newly entered quality requirement id
 $qid=pg_query($conn,"SELECT findmaxexqid()");
 $mqid = pg_fetch_array($qid);
 $id = $mqid[0];

  $sqlins = "INSERT INTO Requires values ('$uid',1,'$id')"
 $updatereq=pg_query($conn,$sqlins); //updates requirements
        }
else
{
 //update only the counter
 $id = $k[5];
 $updateR = ($conn, "UPDATE Requires
SET usecount = 1
WHERE userid ='$uid' AND exq_id = '$id'");
}
//=======================================update accesshistory==========
 }
} //end of searchbutton
?>
```

Appendix C: Application based search

This function retrieves spatial data based on user specific applications displays the data as GML format or metadata together with the spatial data it self depending based on the choice of the user. This program makes use of the plpgsql functions stated in Appendix A.

```
<?php

$conn = pg_connect("host=itcnt07 port=5432 dbname=esdkad08
user=gebresilas23986 password=23986" );
$BB = $_POST['output'];

$user = $_SESSION['username']; //current user
$find = $_POST['btnfind']; //accept button even
$dataformat = $_POST['delivery']; //data delivery format of user preference
if($find )
{
$app = $_POST['appl']; //need to be compared in caps and lower cases
$conn = pg_connect("host=itcnt07 port=5432 dbname=esdkad08
user=gebresilas23986 password=23986" );
//retrievens Full name of the user
$name = pg_query($conn, "SELECT findname('$user')");
$nam = pg_fetch_array($name);
switch($dataformat)
{
 case "Description":
//attributes to be retrieved
$sql = pg_query($conn, "SELECT DISTINCT parcel_id,
OD.x_coordinate,
OD.y_coordinate,
OD.building_code,
OD.land_use_type_non_built,
OD.year_of_purchase,
OD.indication_more_cadastral_objects,
OD.description_parcel_part,
OD.cadastral_municipality_code,
OD.conveyor,
OD.date_of_agreement,
OD.area,
OD.indication_estimated_area,
OD.the_geom,
ovq. *
FROM tblObject_data as OD, tbloverviewquality as ovq,
tbldatauser as du
```

```
WHERE ovq.usage = '".$app."' AND
du.ovqid = ovq.ovqid AND du.gid = OD.gid AND
ST_WITHIN(ST_Centroid((ST_GeomFromEWKT(od.the_geom))),
ST_transform(ST_GeomFromText ('.$BB.',4326), 28992))= '".true."'");
$result = pg_fetch_array($sql);
$numbOfrecords= pg_num_rows($sql);

if ($numbOfrecords !=0)
{//display data to the user
echo "<u>THE FOLLOWING DATA HAS BEEN USED FOR SIMILAR APPLICATION!</u><br>\n";
 while($n = pg_fetch_assoc($result)) //display output to the user
 {
  echo "<br>\n";
  foreach ($n as $key=>$value)
 {//for($i = 0;$i<=$numbOfrecords;$i++)

echo strtoupper($key);
echo "==>".$value."<br>\n ";
 }
}
}
else
{
 //plpgsql function call for quality requirements search
$Sqlr = pg_query($conn, "SELECT findapplqualityrequirements('$app', '$user')");
//plpgsql function call for retrieving data
$sqld = pg_query($conn, "SELECT RETRIEVEDATATOAPPL()");
//display to the user
$data = pg_query($conn, "SELECT DISTINCT * FROM tempData");
$res = pg_fetch_array($data); //fetch result to array
$numReco = pg_num_rows($data); //count returns rows


if($numReco !=0)
{
echo "<u>THE FOLLOWING DATA MEETS THE APPLICATION REQUIREMENTS!</u><br>\n";
 while($n = pg_fetch_assoc($result)) //display output to the user
 {
  echo "<br>\n";
  foreach ($n as $key=>$value)
 {//for($i = 0;$i<=$numbOfrecords;$i++)

echo strtoupper($key);
echo "==>".$value."<br>\n ";
 }
}
}
else
```

```
{
echo "NO DATA IS AVAILABLE FOR THIS APPLICATION!!";
}
}
 break;
 case "GML":
$sql = pg_query($conn, "SELECT DISTINCT
asgml(OD.the_geom),
parcel_id,
OD.x_coordinate,
OD.y_coordinate,
OD.building_code,
OD.land_use_type_non_built,
OD.year_of_purchase,
OD.indication_more_cadastral_objects,
OD.description_parcel_part,
OD.cadastral_municipality_code,
OD.conveyor,
OD.date_of_agreement,
OD.area,
OD.indication_estimated_area,
OD.the_geom,
ovq. *
FROM tblObject_data as OD, tbloverviewquality as ovq,
tbldatauser as du
WHERE ovq.usage = '".$app."' AND
du.ovqid = ovq.ovqid AND du.gid = OD.gid AND
ST_WITHIN(ST_Centroid((ST_GeomFromEWKT(od.the_geom))),
ST_transform(ST_GeomFromText ('.$BB.',4326), 28992))= '".true."'")
break;
 default:
echo "No Delivery format selected!";
break;
  }
}

?>
```