AN OPERATION-BASED APPROACH TO THE COMMUNICATION OF SPATIAL DATA QUALITY IN GIS

AMIN ZARGAR

AN OPERATION-BASED APPROACH TO THE COMMUNICATION OF SPATIAL DATA QUALITY IN GIS

by

©Amin Zargar

A thesis submitted to the
School of Graduate Studies
in partial fulfillment of the
requirements for a degree of
Master of Science

Department of Geography
Memorial University of Newfoundland

May 2009

St. John's Newfoundland

Abstract

Spatial data used in Geographic Information Systems (GIS) are prone to uncertainties that can undermine their usability. Improving the GIS users' awareness of these uncertainties requires improvements in the management and communication of the quality information provided with spatial data. Current tools for communicating data quality information in commercial GIS are rudimentary, and alternative tools - mainly developed and applied in academia – remain to be implemented in commercial GIS.

This work develops an alternative operation-based approach to the communication of quality information in GIS. Communicating quality information is studied in the context of GIS operations. A review of GIS operations, one of the main components in GIS applications is performed. Based on the study of GIS operations and quality information, a conceptual link is established between the two components. Using this link, a system is designed to retrieve and communicate applicable quality elements to GIS users. The designed system is then implemented as a prototype in a commercial GIS software, and its usefulness is tested among GIS users.

Acknowledgments

I would like to thank all the people who helped make the production of this thesis possible. These include, first of all, my family, in addition to my supervisor at the Memorial University of Newfoundland, Dr. Rodolphe Devillers. Thesis reviewers and committee members Dr. Alvin Simms (Memorial University of Newfoundland) and Dr. Gary J. Hunter (University of Melbourne), all are gratefully acknowledged. I would also like to thank my colleagues Randal Greene, Jennifer Fowler and Helen Goodchild for their help throughout my MSc.

The provider of the data for this thesis, Dr. Jean Brodeur (Centre for Topographic Information of Sherbrooke, CIT-S, Natural Resources Canada, NRCan) is acknowledged for this important help. Natural Sciences and Engineering Research Council of Canada (NSERC) and Memorial University of Newfoundland (MUN) are acknowledged for providing financial support for this thesis.

Finally, the 25 anonymous participants in the testing of this work's prototype are gratefully acknowledged for their time and effort in answering the questionnaires.

Table of Contents

List of Tables

List of Figures

List of Appendices

1. Introduction

1.1 Introduction

Since their appearance in the early 1980s, commercial Geographic Information Systems (GIS) have evolved into tools used by an increasing number of users to manage, modify, analyze and visualize spatial data. Spatial data, also called *geospatial data* or *geographic information*, are a main component of these systems, representing various Earth phenomena in different types of data structures. Resource managers, environmental scientists and urban planners are only a few examples of the group of people mapping and analyzing geographic phenomena using GIS.

When spatial data are processed by GIS, they can be analyzed and "transformed" through GIS operations to obtain new data (Chrisman, 1999). For instance, an interpolation operation can be used to transform point data representing a phenomenon (e.g., soil pH values at sampling locations) into a continuous representation of that phenomenon.

While being the input to various analyses and a factor to support decision-making in many disciplines, spatial data are subject to quality deficiencies, a concern ultimately shared with *non-spatial* data (Wang and Strong, 1996; Olson, 2003). These deficiencies can undermine the reliability of decisions based on spatial data. While GIS, and information systems in general, provide a high level of precision in numerical modeling and calculation aspects (e.g., ESRI ArcGIS uses double-precision data for location), geographic information is nevertheless understood to be fundamentally "imperfect" in most occasions (Worboys, 1998; Duckham *et al.*, 2001; Zhang and Goodchild, 2002). In other words, there is a risk for misconception that a high level of precision in spatial data (e.g., up to 10 decimal places) would correspond to high level of accuracy in the data (e.g., data being accurate to the order of millimetres). The quality of outputs resulting from GIS analysis is said to be "only as good as" the quality of the input data (Visvalingam, 1991; Herzog, 2000). Deficiencies in spatial data are caused by factors such as the techniques used to acquire the data, and the data models used to conceptualize real-world entities into database objects. One persistent issue has been the challenge concerning the identification and modeling of some real-world features. The definition of such features is said to be bound to our natural language (Mark, 1993; Montello, 2003). Examples include the "vague" definition of the boundary of certain natural features, such as 'mountains' or 'forests', both of which have proven difficult to delineate (Fisher, 1999; Bennett, 2001; Smith and Mark, 2003). The lack of sufficiently detailed and globally accepted concepts used for modeling such features results in an imprecise definition, which in return introduces an "endemic" uncertainty to spatial data (Fisher, 1999; Duckham *et al.*, 2001).

In order to document and share the knowledge about spatial data quality, it is now typical for data producers to document the quality of the spatial data they distribute. This documentation is a part of more general metadata (data about data). For example, if land boundaries have been digitized from paper maps, the information about this process can be recorded within the dataset's lineage information, which records the history of the dataset starting from the data collection. This information can improve users' awareness of possible discrepancies that may result during the process of the dataset's production. In addition to the lineage description, completeness, attribute accuracy, positional accuracy and temporal accuracy are other examples of spatial data quality descriptions that can be documented as part of metadata (Guptill and Morrison, 1995; ISO/TC-211, 2003).

It is generally accepted that increased awareness about spatial data quality can increase the chances of making reliable use of spatial data (Goodchild and Gopal, 1989; Beard *et al.*, 1991; Devillers *et al.* 2007). This awareness should help users to better understand data characteristics, as well as consider the limitations in spatial data. With the advent of the Internet and the increasing availability of online GIS applications (such as web mapping tools), making spatial data quality available has become a priority to help the expanding GIS user community make informed decisions. As uncertainty is inherent to any spatial dataset, communicating information about uncertainty to GIS users has become imperative.

GIS users, meanwhile, have diversified in terms of both their areas of application (e.g., in medical sciences and mobile GIS) and their knowledge about the underlying concepts of geospatial sciences (e.g., a novice and a professional user) (Goodchild, 1995; Elshaw Thrall and Thrall, 1999; Mowrer, 1999; Reinke *et al.*, 2006). When comparing a novice online GIS user to a professional cartographer, there can be a large difference in their understanding of the risks associated with the use of spatial data of poor quality. While the use of spatial data was initially restricted to a smaller number of applications (e.g., land-surveying, military applications, land-use mapping), GIS are now used in a large number of disciplines, ranging from natural sciences and engineering to social sciences. The increasing number of 'new' GIS users, who may lack sufficient knowledge in geography and cartography, can potentially increase the risk of 'misusing' spatial data.

Documenting spatial data quality has been given additional importance as legal cases have emerged, resulting from the use of spatial data of insufficient quality (Goodchild and Kemp, 1990; Onsrud, 1995; Curry, 1998; Agumya and Hunter, 2002, Gervais, 2006). Gervais (2006) argues that having spatial data appropriately described in terms of spatial data quality also helps legally protect data producers in cases of data misuse.

To ensure a consistency in the methods used to describe spatial data quality, and aid in the transfer of metadata together with the data, a number of countries have adopted standards for the transfer of spatial data and documentation of metadata. In the United States, the Spatial Data Transfer Standard (SDTS; NIST, 1992) requires spatial data to be accompanied by descriptive quality information stored in the form of metadata. The inclusion of quality information with other metadata can help the users to "determine the fitness of the set of geospatial data for an intended use" (FGDC, 1998).

Throughout the years, however, in addition to the time-consuming nature of metadata documentation for data producers, establishing effective standards for modeling spatial data quality has proven complex. Often, perspectives differ as to which error and uncertainty factors should constitute quality information and how this information can affect GIS applications' results. This is especially true since spatial data are increasingly used for applications outside their intended design. Based on the description of spatial data quality as documented in the metadata, users should be in a better position to assess if the data would be appropriate for their use, and, ultimately decide on an action. This concept is known as "fitness for use" assessment (see Chrisman, 1983). Assessing the fitness for use is more difficult for users lacking expertise in geographic information as they have little awareness of the process used in the documentation of spatial data quality (e.g., Shortridge and Goodchild, 1999). Such users may also have difficulties understanding the terms used in spatial data quality that are often very technical (Boin and Hunter, 2006). This lack of necessary knowledge may hinder non-expert users' ability to assess the data's fitness for the intended use. Several authors have indicated a communication gap between the mainly "producer-oriented" data quality standards and the users' ability to understand quality information (e.g., Devillers *et al.*, 2005; Boin and Hunter, 2006). The communication of quality information must therefore be made in an understandable and familiar way so that the various GIS users can make use of the documented quality information.

A number of studies have been undertaken to improve the users' awareness of spatial data quality, and these have varied in scope and target. For example, Duckham and McCreadie (2002) discuss a fundamental revision of GIS software by implementing object-oriented approaches. This approach is argued to allow for integrated error-handling in the functioning of GIS software. Others have focused on more immediate issues such as the visualization of the quality information or providing warnings to the users in case of deficient data quality (e.g., Drecki, 2002; MacEachren *et al.*, 2005; Devillers *et al.*, 2007; Levesque *et al.*, 2007). Visualizing quality information (for example, through signs, changing colors and visual aids) has been suggested as an effective and efficient way to warn users of data limitations (MacEachren *et al.*, 2005).

Nevertheless, such approaches and tools remain largely academic with no concrete implementation in commercial GIS software. With an increasing number of GIS users that make use of uncertain spatial data, improving spatial data quality communication remains an ongoing research problem.

1.2 Research Problem

When using GIS software, users are prone to use data and produce results without being aware of the uncertainties that accompany data. The lack of error-awareness among users when using GIS operations can undermine the GIS's outcome by introducing inadvertent errors. When provided, users rarely consult the metadata (Timpf *et al.*, 1996; Van Oort *et al.*, 2005). Metadata are often provided in a flat file or stored separately from the data (such as an XML – eXtensible Markup Language – file in ArcGIS). In this situation, although quality information is provided with the data, users cannot access the relevant spatial data quality information when using various GIS operations for manipulating or analyzing the data. To illustrate, when measuring a distance with GIS software, users will always obtain a very precise measurement, regardless of the positional accuracy stored in the metadata, which is misleading. In this situation, a link between the information about data quality stored in the metadata and the operation used by the GIS user is missing. The absence of such a link can result in the use of spatial data beyond their accuracy.

1.3 Research Hypothesis

The research hypothesis is that GIS users' awareness of spatial data quality can be improved by modifying GIS software to link GIS operations directly with spatial metadata. With the availability of metadata that are now increasingly provided with spatial data, it should be possible to improve the usability of data quality information by associating the quality information to the operations that users perform. Such an association would more closely link GIS applications to the quality information stored in the metadata as suggested by Duckham (2002). This linkage would subsequently make it possible to consider the type of the operation performed by the user as a basis for communicating relevant quality information to the user. The resulting link should allow an improved targeting and usage of the spatial data quality information and subsequently help users to make more informed decisions.

1.4 Objectives

The general objective of this project is to improve the communication of spatial data quality information by providing contextual quality information to GIS users dynamically when using selected GIS operations. The project seeks to narrow the communication gap between the user and the quality information currently available in metadata.

The specific objectives of this project are:

a) To develop a conceptual framework for linking GIS operations with spatial data quality information available from metadata.

b) To develop a framework for applying the conceptual framework to an existing commercial GIS software.

c) To assess the effectiveness of the prototype with users and analyze the results.

1.5 Method

The research method consists of five phases (Figure 1-1).

Phase One involves a review of the literature that helps identifying, first, the problem to be addressed and second, the context in which the research will take place. Relevant literature related to this project is studied. A review of the proposed models and tools for improving the management and visualization of spatial data quality are carried out. This is followed by a study of the GIS applications and operations, and approaches for classifying GIS operations.

Phase Two of the project presents the development of a conceptual framework to provide a solution to the research problems identified during Phase One. This involves developing a conceptual framework for linking relevant spatial data quality elements to GIS operations. This phase discusses classifications of GIS operations and looks at different models describing data quality and uncertainty. The next task in this phase involves designing an implementation framework which can be used for modifying GIS operations to gather and make use of the available data quality information for data types commonly used in GIS (e.g., typical vector primitives and raster data).

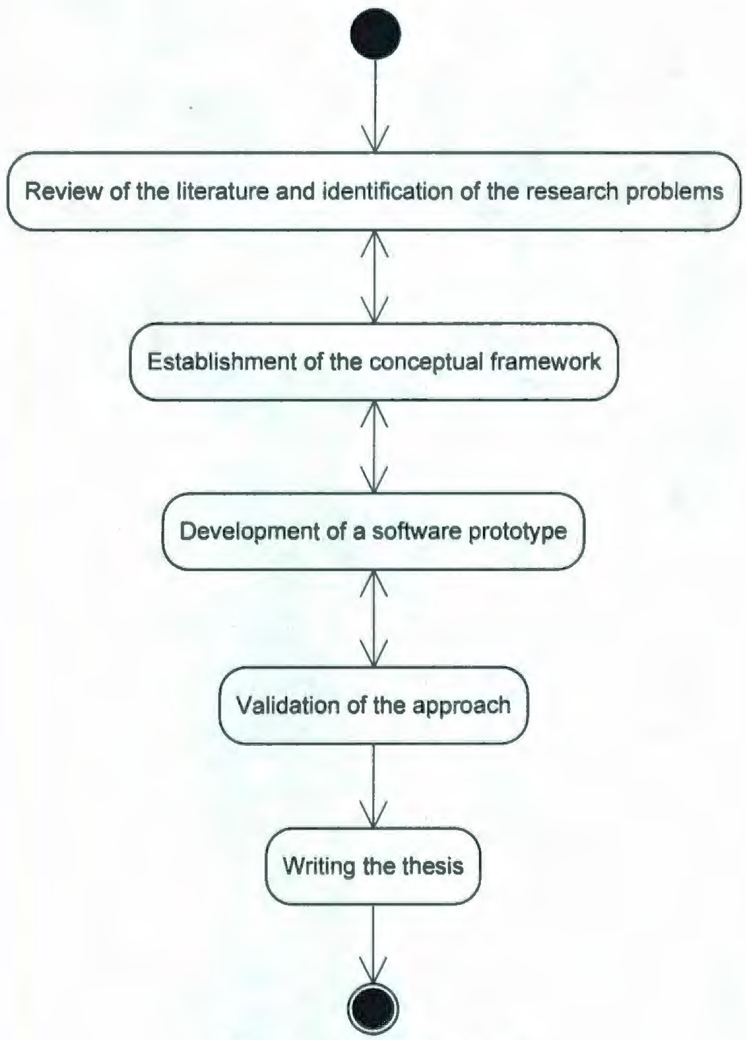


Figure 1-1: Research method

In Phase Three, the concepts developed during Phase Two are implemented into a software prototype. In this phase, two GIS operations ('Measure' and 'Select') are modified and implemented using ArcGIS 9.1 using a subset of a Canadian 1:50,000 topographic map sheet for the city of Sherbrooke (Quebec). The datasets are described by ISO compliant metadata provided in XML format.

The evaluation phase (Phase Four) involves a test of the developed tool or *prototype.* A group of GIS users had to complete a test and fill a questionnaire. The analysis of the responses to the questionnaires, in addition to users' general feedback, enabled an evaluation of the effectiveness of the prototype.

The final phase (Phase Five) is the synthesis of this research and the production of the thesis. This phase included the writing of an article for the Advanced Geographic Information Systems & Web Services conference presented in February 2009 (Zargar and Devillers, 2009).

1.6 Thesis Outline

Following this introduction, Chapter 2 provides a review of the literature pertaining to this research area. General issues, such as the development of quality standards for GIS data, and more specific topics such as the various techniques used for communicating spatial data quality information are discussed. Chapter 3 focuses on GIS operations and presents the development of the conceptual framework for communicating data quality information with GIS operations. The concepts developed in Chapter 3 are formalized into an architecture and then implemented into a software prototype, discussed in Chapter 4. This is followed by an evaluation of the prototype in the final section of Chapter 4. Based on the evaluation, a discussion and a conclusion are given in Chapter 5, and this chapter ends with an opening to possible future research.

2. Background

2.1 Introduction

This chapter provides a review of the literature on different aspects of spatial data uncertainty, quality and quality-awareness. First, spatial data uncertainty is described in relation to the nature and types of uncertainty that can be found. Then, the concept of spatial data quality is presented, as well as how information about uncertainty can be communicated. A number of standards for documenting spatial data quality are presented, followed by a discussion of the usability and effectiveness of these standards. The review then focuses on the issue of improving the usage of spatial data quality information in GIS. The chapter ends with a review of a number of methods proposed for improving the management and communication of quality information in GIS.

It is worth mentioning that while some concepts in this field have established a relatively wide acceptance, others lack a unified theoretical corpus. Terms like uncertainty and quality have for instance a number of definitions, depending on the authors using them. Therefore, the terms discussed in the following sections shall be regarded applicable only within the scope of the references they originate from and are used in this section to provide an overview of the field of study as well as some grounds for the following sections.

2.2 Spatial data uncertainty

The advent of GIS in the 1960s improved the capacities for collecting, storing, analyzing and visualizing geographic information. Similar to other types of information (see for instance Wang, 1996), geographic information is subject to quality deficiencies which had traditionally been a subject of concern among the experts in this field (e.g., cartographers, land surveyors and geographers). However, with the new platform of GIS available, spatial data could be more easily exchanged and manipulated by a larger number of users.

Since the 1980s, with the popularization of GIS and the so-called "democratization" of spatial data (Couclelis, 1992), concerns have been raised about the widespread use of quality-impaired spatial data. The democratization of spatial data resulted in a situation where users could "combine and manipulate information from diverse data sources" to generate a variety of GIS products that only experts could deliver prior to the advent of GIS (Couclelis, 2003, pp. 166-167). As Information Technologies (IT) advanced and the usage of the Internet and other networks became more customary to the public, GIS and spatial data became readily available to a broader user community. This new means of communicating data empowered any interested individual to manipulate spatial data, produce new spatial products and exchange them with other users over the Internet (Goodchild *et al.*, 1997; Couclelis, 2003).

The speed by which spatial data increased in volume and improved in accessibility gathered the attention of researchers in the field of GIS, prompting action to increase user awareness of spatial uncertainty. According to Fisher *et al.* (2006), problems of definition, misunderstanding, doubt, and error can all lead to uncertainty in spatial data. Documenting knowledge about uncertainty in the form of "spatial data quality" has been sought for enabling GIS users to assess the fitness of data for the intended application (Chrisman, 1983; Aalders, 2002).

To define the term uncertainty, Goodchild (1998, p. 50) states that uncertainty is "all that the database does not capture about the real world, or the difference between what the database indicates and what actually exists out there". Theoretically, Fisher *et al.* (2006) categorizes spatial data uncertainty into three types: error, vagueness, and ambiguity.

Error: Error in the context of spatial uncertainty is defined as any deviation from the "true" value (Worboys, 1998). The definition of the "true" value, itself, has been subject of extended discussions. According to Aalders (1996), a possible definition for "true" can be given by the French term "*terrain nominal*", or "nominal ground", a theory for a model, or abstraction, for an "ideal" dataset (for a theoretical discussion of "terrain nominal" see Chrisman, 2006). Therefore, any difference between the data stored in a database and their instance in the ideal dataset can be referred to as error. Examples of error include the wrong number of floors in a building (i.e., error in the theme of the data, usually defined as *thematic accuracy*) or in the accuracy in the horizontal or vertical dimensions of data collected by a GPS (i.e., positional error of data features, usually defined as *spatial accuracy*).

According to Fisher *et al.* (2006), error occurs in cases where geographic objects are inherently "well-defined", but, the recorded attribute, spatial or temporal values are inaccurate. For instance, in the above example, each floor is a well-defined member of the building feature; however, the number of floors is erroneous. This is opposed to cases where the geographic objects are "poorly-defined", in which case the objects can be said to be *vague* or *ambiguous* in description (see below).

Vagueness: Vagueness arises from imprecision in the "concepts used to describe the information" (Worboys, 1998, p. 258). Vague objects are therefore called "poorly-defined" objects. Smith and Mark (2003), in their work: *"Do mountains exist? Towards an ontology of landforms"*, analyze from a number of viewpoints how various landscape features such as cliffs, hills, dales or mountains, inherently lack precise definition. The issue of vagueness can derive from differing human perceptions of the objects (Burrough, 1996). In this regard, specifically within the natural resources disciplines, Fisher *et al.* (2006) argue that "the conceptualization of mappable phenomena and the spaces they occupy is rarely clear-cut" (Fisher *et al.*, 2006, p. 48). An example of vagueness can be portrayed by the minimum physical requirements for a group of trees to be designated as a forest. As an example, Figure 2-1 presents the definition used by different countries to define a forest, based on a combination of canopy cover (in percentage) and tree height (in meters).

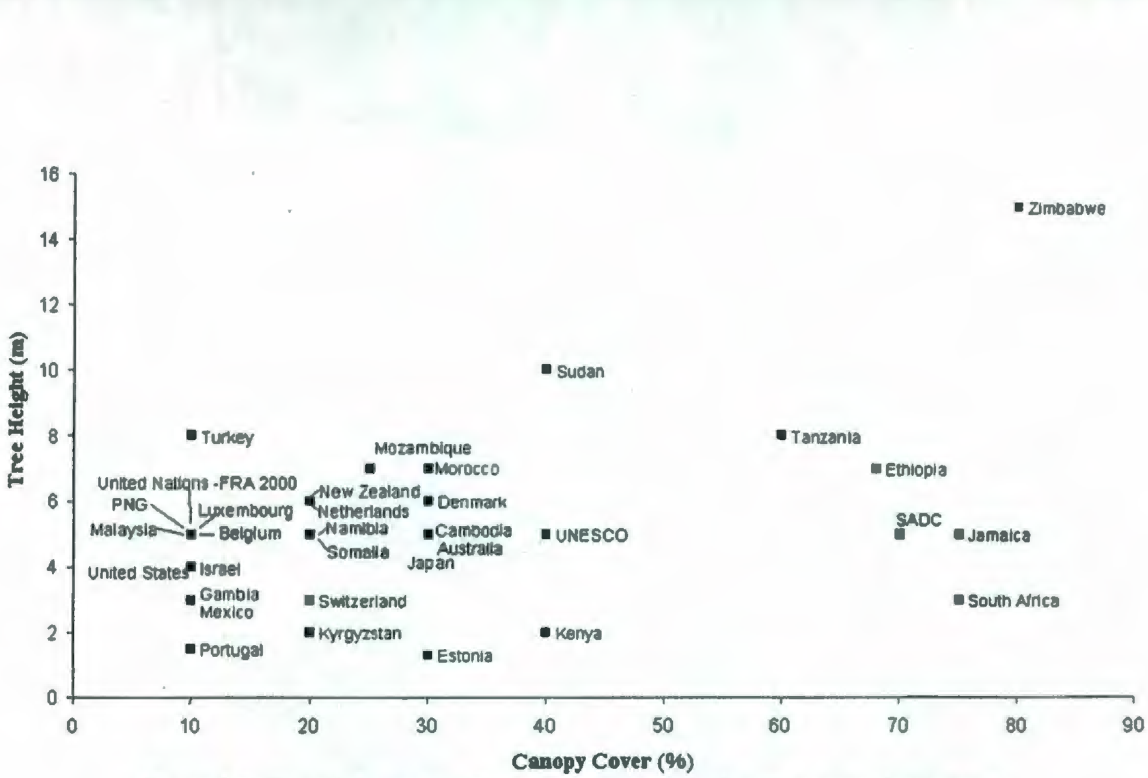


Figure 2-1: The minimum physical requirements of a forest, a case of vagueness (Fisher *et al.*, 2006)

Ambiguity: According to Fisher (1999) ambiguity occurs because of conflicting perceptions of a phenomenon, which results in doubt among users as to how to classify the phenomenon. Ambiguity can have two forms, *discord* and *non-specificity*. Discord occurs when there is difference in the perceptions of the classification schemes that results in the object being assigned to more than one class. An example of discord is a territorial claim between neighbouring countries (Fisher *et al.*, 2006). Non-specificity occurs when the appropriate class for assigning a feature is missing. Classification of soils can be subject to non-specificity as they may not be precise enough. Figure 2-2 presents the uncertainty in spatial data in the form of a hierarchy diagram.

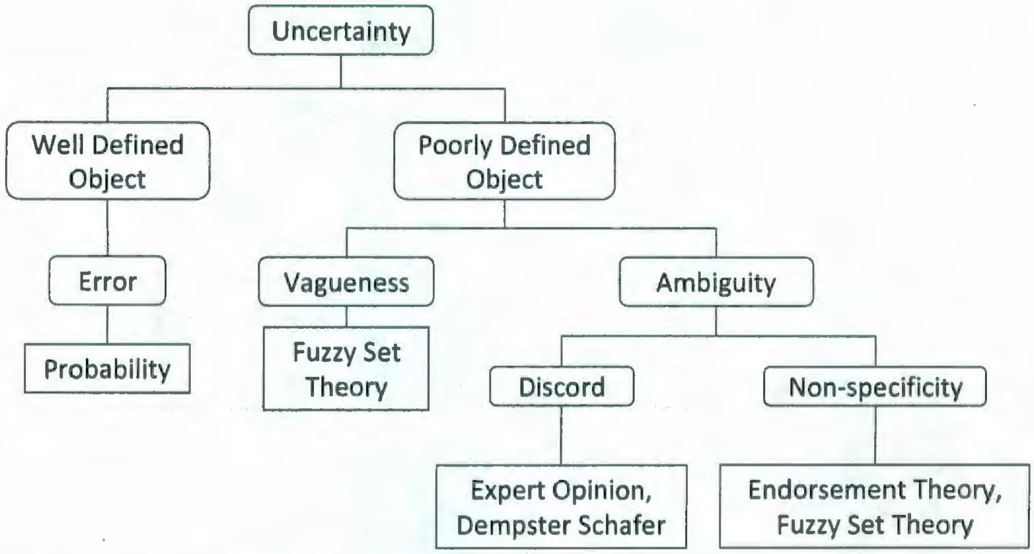


Figure 2-2: A taxonomy of spatial data uncertainty and methods to address them (Fisher *et al.*, 2006; adapted from Klir and Yuan, 1995)

Duckham *et al.* (2001) argues that the uncertainty in spatial data is an "endemic" part of its reality, and cannot always be completely resolved. In the case of error, errors may reoccur in the process of trying to improve the dataset, as "no instrument exists with which error-free measurements can be made" (Van Oort, 2005, p. 4). Both vagueness and ambiguity arise from problems in labelling geographic features. Ambiguity can also persist as data may require greater elaboration in the categorization, for example in pedological mappings, finer categorizations of soil data may be required. Vagueness is also endemic, since the definition of natural phenomena (such as forests) are bound to humans' natural language (Mark, 1993; Montello, 2003), and this language is reflective of the human-environment interaction, which itself is influenced by diverse human localities.

As spatial uncertainty cannot be completely eliminated by only improving data accuracy, researchers proposed solutions which include communicating some of this uncertainty to the GIS user (see for instance, Buttenfield and Beard, 1991; Hunter and Goodchild, 1996; Hunter, 1999). The presence of uncertainty in spatial data further complicates GIS usage, as most GIS operations inevitably introduce further uncertainties to the source data through their transformations. Such uncertainty is understood to undermine the reliability of the GIS products as uncertainty "propagates" from one data source to another (Heuvelink *et al.*, 1989; Burrough and McDonnell, 1998; Heuvelink *et al.*, 2006).

2.3 Spatial data quality

The term uncertainty has been used for "all that the database does not capture about the real world, or the difference between what the database indicates and what actually exists out there" (Goodchild, 1998, p. 50). The term *data quality* generally describes "the degree of excellence in a database" (Veregin, 1998) and has been used by various spatial data quality standards for categorizing some components of uncertainty (King, 2002). The quality of spatial data may be impaired in any of the three dimensions that generally constitute spatial data: spatial, temporal and thematic (Giordano *et al.*, 1994; Beard, 1997). In addition to the types of uncertainty discussed earlier (i.e., error, vagueness and ambiguity), *incompleteness*, *inconsistency* and *imprecision* can impair the quality of GIS data (Worboys, 1998). Worboys defines incompleteness as the "lack of relevant information" (Worboys, 1998), which leaves an ambiguity related to what should be considered as "relevant". ISO/TC 211(2003) defines completeness as "the presence and absence of features, their attributes and their relationships." To illustrate, in the case of missing or superfluous point features in a dataset, inconsistencies are conflicts that arise from the information, for example, when two conflicting objects are simultaneously located at the same location. Imprecision arises from the "limitation on the granularity or resolution at which the observation is made, or the information is represented", for example, objects less than twice the spatial resolution in raster datasets are prone to be omitted (Warren *et al.*, 2004; Worboys, 1998, p. 258).

Since the earlier cautions about spatial data uncertainty given in Chrisman (1983) amongst others, GIS researchers have stressed the need to communicate deficiencies in the quality of spatial data to the users. A number of spatial data quality standards have since been published, such as the international standard ISO 19113, presenting data producers with guidelines for communicating data quality. While these standards may not necessarily be a direct translation of spatial data uncertainty (Fisher *et al.*, 2006), the documentation and the resulting knowledge of quality in a spatial dataset was nonetheless aimed at enabling the users to decide about issues such as the conformance of the dataset to their usage, that is, "fitness for use" assessment (Juran *et al.*, 1974; Chrisman, 1983, 1994; Hunter, 1999, de Bruin *et al.*, 2001).

Enabling fitness for use assessment is one the purposes for publishing metadata. Metadata also provide a number of other pieces of information about spatial data, such as:

1) availability: information used "to determine the sets of data that exist for a geographic location",

2) access: information "needed to acquire an identified set of data", and

3) transfer: information "to process and use a set of data" (Kim, 1999, p. 173).

The first major data quality standard was published in 1994 in the USA, requiring all US federal agencies collecting and producing spatial data to include information about the quality of data with their products (Foresman *et al.*, 1996). The "Content Standard for Digital Geospatial Metadata" published by the Federal Geographic Data Committee (FGDC, 1994) contains guidelines for organizations to include descriptive metadata with their data. Since the first version of FGDC, a second version of the Content Standard for Digital Geospatial Metadata has been published (FGDC, 1998). Another widely-used standard is published by the International Organization for Standardization (*ISO 19115:2003 Geographic information – Metadata*, ISO/TC-211, 2003) and is to be adopted in the USA as a replacement for FGDC (as a profile of the ISO standard) (FGDC, 2006).

The data quality section in these standards is organized into a number of *quality elements*. In combination, quality elements provide the users with a composite picture of quality so they can examine the degree to which spatial data adhere to the intended (or *reference*) quality (Figure 2-3). The quality elements under FGDC (1998) and ISO 19115 (ISO/TC-211, 2003) standards share five elements: lineage, positional accuracy, thematic (attribute) accuracy, logical consistency and completeness. ISO 19115 (ISO/TC-211, 2003) includes an additional element of *temporal accuracy* (Figure 2-4). In Figure 2-3, "*DQ_*", standing for "Data Quality", helps in identifying the spatial data quality elements.

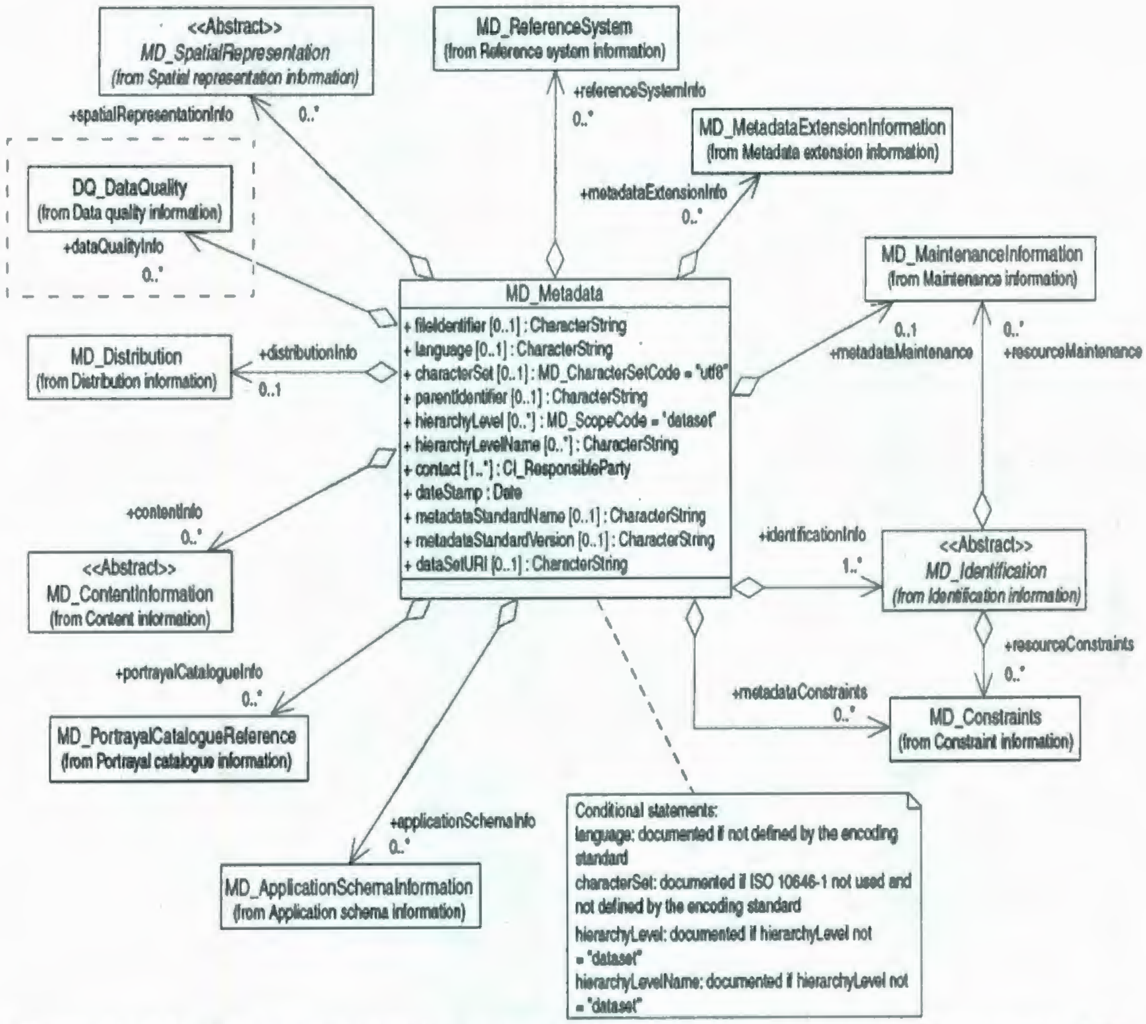


Figure 2-3: Main UML (Unified Modeling Language) class diagram representing the organization of ISO 19115 (ISO/TC 211, 2003) metadata. The dashed rectangle highlights the spatial data quality section.

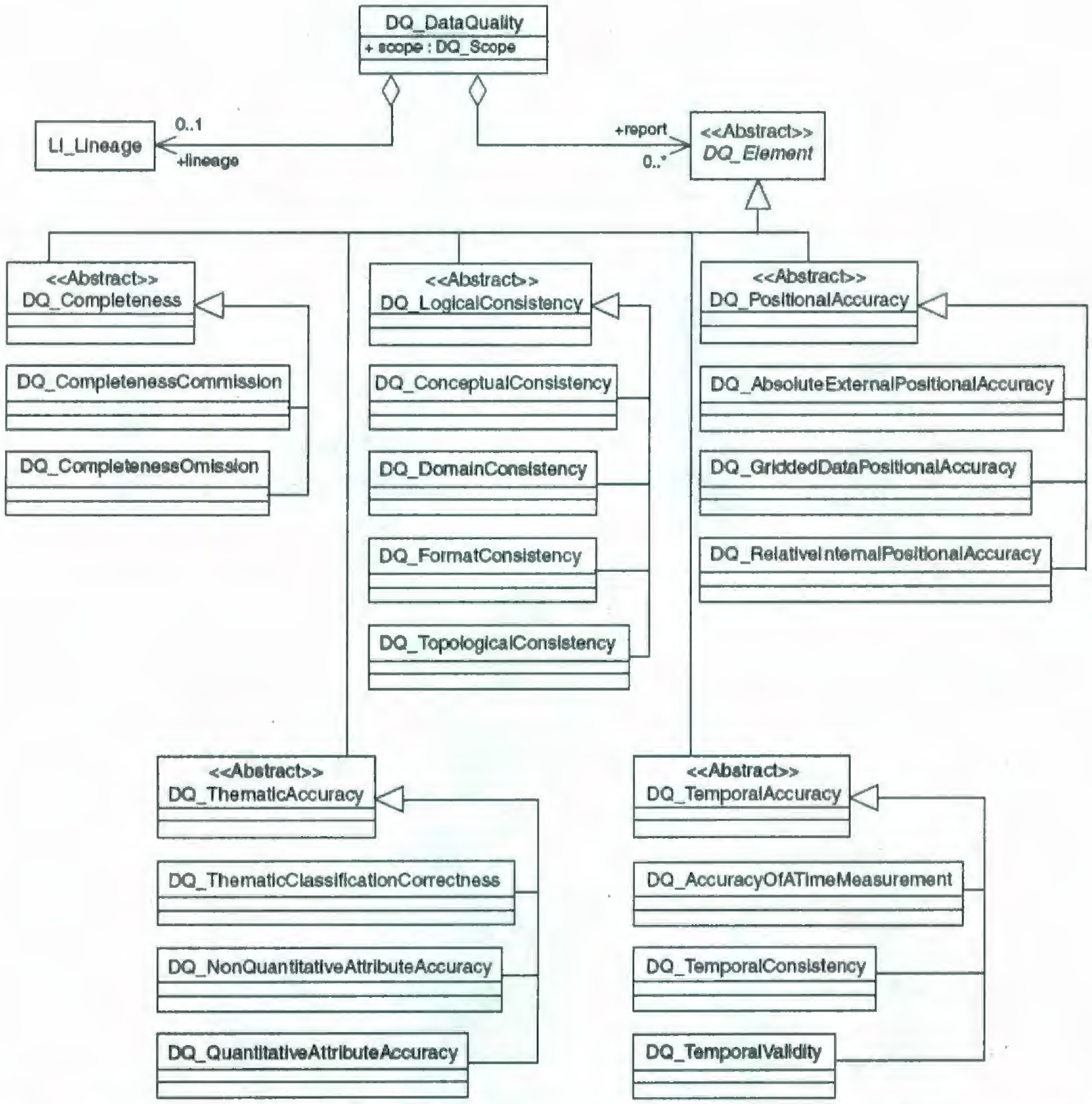


Figure 2-4: UML class diagram representing the elements composing data quality (ISO/TC 211, 2003)

These spatial data quality elements are described below:

1. Lineage: "information about the events or source data used in constructing the data specified by the scope or lack of knowledge about lineage" (ISO/TC 211, 2003, p. 48).

2. Positional Accuracy: "accuracy of the position of features" (ISO/TC 211, 2003, p. 53).

3. Attribute Accuracy: "accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships" (ISO/TC 211, 2003, p. 54).

4. Completeness: "presence and absence of features, their attributes and their relationships" (ISO/TC 211, 2003, p. 51).

5. Logical Consistency: "degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical)" (ISO/TC 211, 2003, p. 52).

6. Semantic Accuracy: "the quality with which geographical objects are described in accordance with the selected model. It is related to the meanings of 'things' of the universe of discourse. Refers to the pertinence of the meaning of the geographical object rather than to the geometrical representation" (Salgé, 1995; from Guptill and Morisson, p. 139). This element is not part of the ISO standard but has been used by other standards and hence mentioned in this list.

7. Temporal Accuracy: "accuracy of the temporal attributes and temporal relationships of features" (ISO/TC 211, 2003, p. 53).

With the introduction of spatial metadata standards, an increasing number of spatial data producers have since used these standards to document metadata. The data quality section in metadata provides information for quality elements for which information could be obtained. Completion of data quality information is however optional in a number of these standards (e.g., ISO 19115) which may limit the documentation of data quality in practice.

Metadata vary in configuration; they are provided in a number of formats such as unstructured text files (e.g., Microsoft Word or basic text files), structured text files (e.g., HTML and XML) and databases (Najar, 2006). Structured XML files are becoming increasingly popular for metadata storage and they have proven to provide an "extremely powerful and flexible" standard for formatting information (Green and Bossomaier, 2002, p. 95; Kazakos *et al.*, 2005). In this regard, ISO has additionally recently released the ISO 19139 (ISO/TS, 2007) standard for the XML implementation of the metadata standard ISO 19115.

2.4 Usability of spatial data quality information

Since the introduction of spatial data quality standards and the publication of quality information with spatial data, the availability of such information does not seem to have significantly helped the GIS users in determining data's fitness for use. Quality information has been ignored in most GIS applications (Openshaw, 1989; Agumya and Hunter, 1999; Couclelis, 2003; Van Oort and Bregt, 2005).

According to Qiu and Hunter (2002), two problems undermine the usefulness of spatial data quality: "(1) that the current method of reporting data quality is poorly structured and remains physically separated from the actual data, and (2) that current ... [GIS] packages continue to lack the tools to effectively manage (that is, to store, manipulate, query, update and display) any embedded data quality information." (Qiu and Hunter, 2002, p. 230). Boin and Hunter (2006; 2009) also argue that the terms used by the standards to describe data quality are too technical to be understood by most GIS users. More fundamental concepts in GIS have also been scrutinized. From a data uncertainty point of view, GIS's fundamental approach for modeling geographic phenomena as GIS data have been deemed poorly constructed (Goodchild *et al.*, 2007). For example, Kemp (1997) earlier questioned the appropriateness of raster datasets for modeling continuous phenomena such as terrain surface elevation or temperature maps. Since such phenomena are continuous in nature, the discrete primitives (i.e., pixels, or, cells) in a raster fail to include the inherent continuity in such phenomena (see also, Shi, 2002; Goodchild and Haining, 2004). Similarly, others such as Burrough and Frank (1995), Fisher (1998) and Fisher *et al.* (2006) have called for changes in the methods used for modeling geographic phenomena as GIS data. Since geographic data include cases where features are fuzzy in boundary or vague in the nature of definition, GIS must accommodate for vagueness and ambiguity to accompany geometric data at the lowest level (i.e., at the level of primitives, such as points, lines and polygons). According to Schneider (1997), the treatment of spatial objects with indeterminate boundaries is especially problematic for computer scientists who are confronted with difficulties such as:

- how to model such objects in a database system so that they correspond to the user's intuition,
- how to finitely represent them in a digital format,
- how to develop spatial index structures for them,
- how to draw them.

Computer scientists are thus accustomed to an *abstraction process* of simplifying spatial phenomena of the real world "through the concepts of conventional binary logic, reduction of dimension, and cartographic generalization to precisely defined, simply structured, and sharply bounded objects of Euclidean geometry like points, lines, and regions" (Schneider, 1997, p. 241, see also Couclelis, 2003 for a discussion of these issues). Increasing awareness about ontological issues related to geographic objects has also been discussed. Associating ontological meanings to both spatial data (Smith and Mark, 2001; Comber *et al.*, 2007) and data quality (Vasseur *et al.*, 2003; Mostafavi *et al.*, 2004; Frank, 2007), can lead to a more appropriate representation of spatial data.

Regardless of the deficiencies in the methods used for modeling geographic features, a number of problems can be directly attributed to spatial data quality standards. Fundamentally, the derivation of any spatial data quality standard is subjective to user needs since uncertainty is perceived as an *objective* issue while quality is more *subjective* (Worboys, 1998). While the uncertainty arising from both error and lack of definition (i.e., vagueness and ambiguity) can be studied from a standalone point of view, the transformation of uncertainty into quality and its implications for users may vary as user requirements and application cases differ. Couclelis (1992) describes this scenario: "[while] data error can be discussed in absolute terms, data quality is a function of context and purpose." Hence, the demands for spatial data quality information are said to be varying or "heterogeneous" among the various operators of spatial data (Duckham, 2002; Devillers *et al.*, 2005). As a consequence, the objective set earlier of enabling users to assess data's fitness for use (the capability to make an evaluation of the appropriateness of a dataset for a specific need or application) remains to be materialized (Boin and Hunter, 2009). Beard and Buttenfield (1999) discussed how different types of GIS users, for example, a database designer versus a decision-maker, require different approaches to data quality modeling. Users may also be disparate in the level of knowledge and may lack experience working with quality information. A professional user may be better placed to make use of quality information compared to a novice user without formal education and training (see, for example, Hunter *et al.*, 2003). Devillers and Jeansoulin (2006) discuss how fitness for use relates to users' individual cases when using spatial data.

Furthermore, quality standards have been labelled producer-oriented instead of being user-oriented by some authors. For instance, Comber *et al.* (2006, p. 281) argue that "the specification of quality standards continues to reflect data production interests, reporting the easily measurable and showing that the data producer can follow a recipe rather than more fully communicating the producer's knowledge of the data." Accordingly, they are more "a formalization of production procedures and tests understandable by data acquisition specialists, than meaningful information for a general audience useful for decision-making processes" (Devillers *et al.*, 2005, p. 206).

A number of modifications have been proposed for spatial data quality standards in order to increase their usability for users. A user-oriented approach to the modeling of data quality must be flexible in allowing different formats for reporting data quality, such as maps and graphics (Timpf *et al.*, 1996; Beard, 1997), and must accommodate methods to allow the inclusion of additional quality elements (Morrison, 1995; Aalders, 1996). Changes in the terminology of quality standards have also been proposed, for example, replacing the term 'metadata' with 'product description', and 'lineage' with 'history' (Hunter *et al.*, 2005).

2.5 Usability of quality information

Along with improvements in spatial data quality models, researchers have concurrently aimed at enhancing GIS software's capabilities for making better use of the data quality information (Agumya and Hunter, 2002). Research in this direction essentially aims at improving the handling of available quality information, ideally leading to a better usability of quality information in GIS software. Unwin (1995) describes the concept of "error-sensitive" GIS, which is capable of handling both geographic information and uncertainty that surrounds it. According to Devillers (2004), in order to establish an error-sensitive GIS, improvements in the management and communication of spatial data quality are required. The two issues will be discussed in the following section.

2.5.1 Managing spatial data quality

According to Duckham (1999), despite the long history of research in spatial data quality, commercial GIS include little or no data quality management capabilities. One of the most advanced commercial GIS software supporting some elements about data quality is probably IDRISI, which uses the Root Mean Square Error (RMSE) stored in a raster's attribute values for some of its probability analysis operations (Clark Labs, 2006). Qiu and Hunter (2002) argue that on-the-shelf GIS tools "lack the tools to manage (that is, to record, manipulate, query, update and display) data quality information" (Qiu and Hunter, 2002, p. 230). Such capabilities, as argued by Duckham and McCreadie (2002, p. 62), are a "rarity outside the research laboratory". An important concern in managing spatial data quality has been to enable the usage of quality information at different levels of detail (Hunter, 2001; Qiu and Hunter, 2002; Devillers, 2004; Devillers *et al.*, 2007). Lack of varying levels of detail in the description of spatial data, or "granularity", undermines their usefulness (Hunter, 2001). A number of approaches have been discussed and tested for improving spatial data quality management in GIS software that are described below.

One of the proposed approaches is to use indicators to communicate spatial data quality to GIS users. Quality indicators provide users with a quick insight on quality information which can help prevent unwanted risks (Devillers *et al.*, 2002). The Quality Information Management Model (QIMM) developed by Devillers *et al.* (2005) is capable of managing spatial quality information at different levels of data detail (from the *geometric primitive* level to the *dataset* level), and for spatially-varying quality information using indicators. The incorporation of an object-oriented (OO) design has also been proposed as a framework to manage data quality in GIS (Anderson and Stonebraker, 1994; Duckham and Drummond, 1999; Duckham *et al.*, 2001; Qiu and Hunter, 2002). This approach allows for the management of spatial data quality at different levels, and more closely associates data quality with spatial data information using the "objects" model (Duckham and McCreadie, 2002).

2.5.2 Communicating spatial data quality

According to Paradis and Beard (1994, p. 25), "lack of effective communication of data quality ... is an essential issue hampering the successful utilization of GIS." Although, 15 years later, GIS are widely used without having an effective way to communicate data quality, efforts to improve this communication increased in importance due to the nature of today's GIS users (that may not have a formal experience in geographic information anymore) (Devillers *et al.*, 2005; 2007). Except for a few academic tools that can be added as extensions to GIS, current GIS do not allow one to communicate spatial data quality (Boin and Hunter, 2006). Moreover, an improved communication of spatial data quality is closely related to other developments and improvements in the field of data quality. Among the literature, methods developed for spatial data quality handling and management are argued to simultaneously support data quality communication. The QIMM prototype by Devillers *et al.* (2005) and the OO approach by Duckham and McCreadie (2002) both demonstrate improvements in the communication of spatial data quality as a result from their respective research. A number of communication approaches have been described to help the user decide on the fitness for use of some data.

Communicating metadata statements directly to the users has proven problematic as they may be too technical for users to understand (Hunter *et al.*, 2005). Instead, the use of indexes (Veregin and Hargitai, 1995; Cheng, 2001) and indicators (Devillers *et al.*, 2002; 2005) of measurable data quality elements has been discussed as an alternative (see also Agumya and Hunter, 1999). According to Devillers *et al.* (2005) quality indicators can represent data quality in a hierarchy of 'data quality indicators' and 'data' dimensions. For example, in the data quality indicators dimension, the five data quality elements of ISO 19115 represent the second level of the hierarchy (Figure 2-4). These indicators are themselves composed of lower level indicators in the third level (e.g., commission and omission for the completeness element). Indicators can provide a quick insight into data quality, and thus help in preventing potential risks of data misuse. This approach, however, is limited in some aspects: it is not applicable to all data quality elements, such as lineage (Agumya and Hunter, 1999; Devillers *et al.*, 2005). Also, this approach requires many assumptions and decisions about user needs, and there is an inherent loss of information: for instance, the decision-maker may still require access to the original data quality (Devillers *et al.*, 2007).

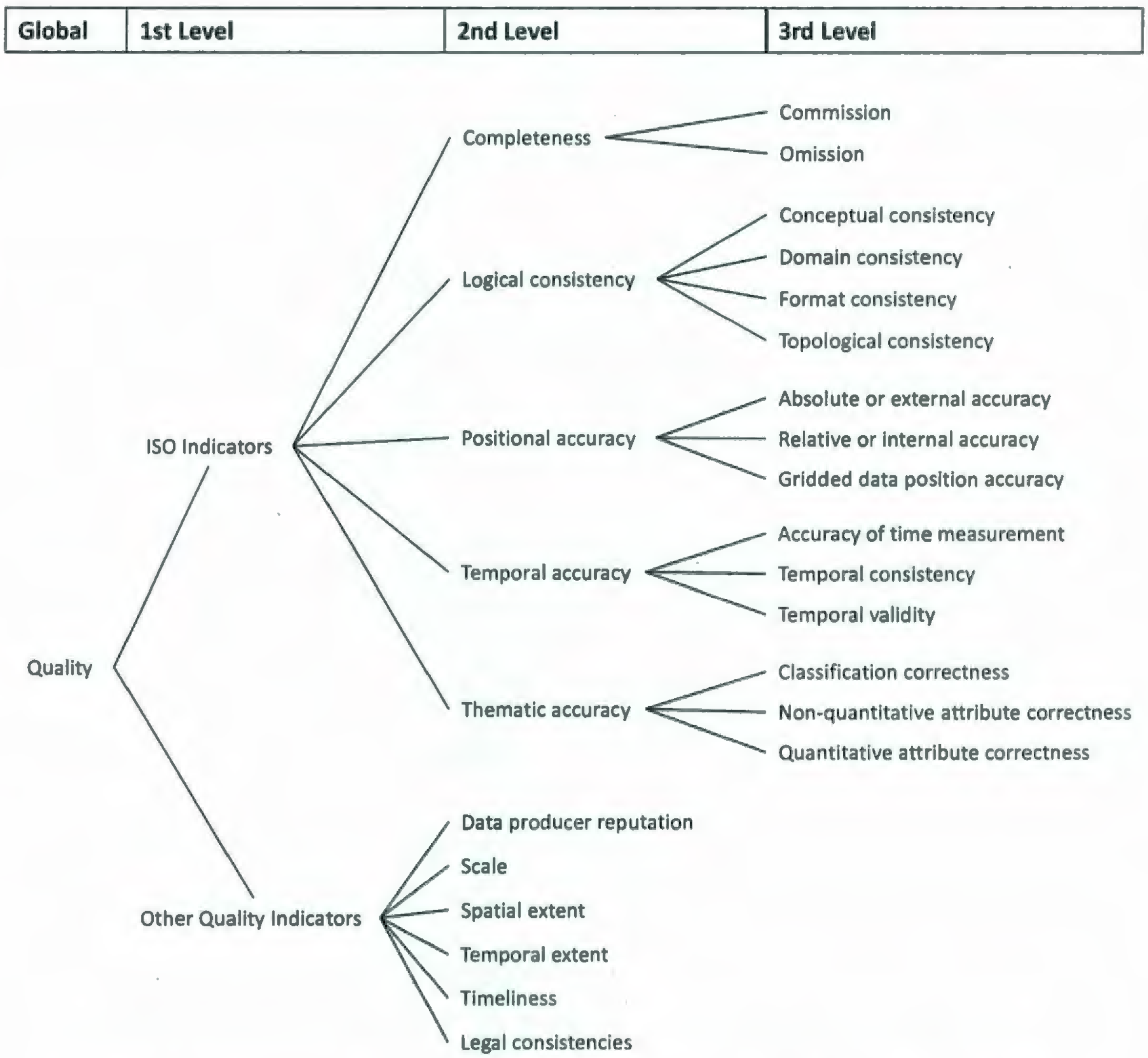


Figure 2-5: Hierarchy of indicators in the 'Quality Indicator' Dimension (Devillers *et al.*, 2005)

Determining the "risk" of using erroneous datasets has also been considered for assessing the data's fitness (Agumya and Hunter, 1999; Van Oort and Bregt, 2005). This approach focuses on the undesired consequences that can arise from using uncertain data to support a decision as well as the different strategies that can be used to cope with the risk (e.g., reduction, absorption). The risk-based approach, accordingly, computes an estimate of the risk exposure by summing risks due to the varying graveness of an adverse event, such as the inundation depth in the case of a flood

(Agumya and Hunter, 1999). Based on the quantified risk in terms of, for instance, *dollars* or *lives lost*, the decision maker can respond to the risk by choosing a number of approaches that have been discussed in Agumya and Hunter (1999, 2002) and Levesque *et al.* (2007). Possible responses include avoiding carrying out the tasks, trying to find alternative ways to minimize the risk (i.e., risk reduction), transferring the risk to another party (e.g., insurance company), or accepting the liability of using erroneous data (i.e., risk absorption). According to Agumya and Hunter (1999) the risk-based approach, however, is complex and requires radical simplifications of its concepts for risk quantification. Additionally, risk estimation is subjective as the parties involved may have conflicting perceptions of the risk (Agumya and Hunter, 1999). Van Oort and Bregt (2005) cite problems such as the lack of tools, theory, and poor documentation of spatial data quality to further complicate the risk analysis process for the users.

Visualizing quality information is an approach that is largely considered to be advantageous for communicating spatial data quality (MacEachren *et al.*, 2005). Earlier works such as the NCGIA visualization initiative (Buttenfield and Beard, 1991; Beard and Mackaness, 1993; McGranaghan, 1993) have established the concepts for a visual quality representation system. Visual representations of data quality, accompanied by multimedia methods such as animation (e.g., Fisher, 1993; Ehlschlaeger *et al.*, 1997) and sound (e.g., Fisher, 1994; Krygier, 1994) have also been proposed. Visual interface components such as help balloons, pop-up boxes and "wizards" (Hunter and Reinke, 2000; Levesque *et al.*, 2007), dashboards and signs such as traffic lights, have also been used in research (Devillers *et al.*, 2007; Huth *et al.*, 2009). These components were found to be intuitive for the users and allowed the delivery of the intended warning message in a more efficient manner.

3. A Conceptual Framework for an Operation-based Communication of Spatial Data Quality

3.1 Introduction

Improving the usability of spatial data quality information is a research challenge that can benefit from advances in the communication of spatial data quality in GIS (Duckham and McCreadie, 2002; Devillers *et al.*, 2005). As suggested by these works, spatial data quality communication can improve users' awareness of the limitations related to spatial data and hence reduce the potential risks of misusing spatial data.

As discussed in Chapter 2, earlier research proposed a number of approaches and led to the development of a variety of tools for communicating spatial data quality. Nevertheless, such academic approaches have not yet led to implementations in commercial GIS. This chapter sets the basis for an approach that associates spatial data quality information and GIS operations.

3.2 An overview of GIS operations

GIS operations are one of the main components of GIS software, providing a wide range of geospatial analytical capabilities (Albrecht, 1996). Together with spatial data, they constitute the "kernel" of GIS (Yuan and Albrecht, 1995). GIS operations have also been called GIS *functions* (for example, Goodchild, 1987; Burrough *et al.*, 1996). The more recent term, GIS *services,* refers to all types of functionalities that are used in a distributed environment, such as the basic tasks of viewing, querying data and discovering services (Gomez Benitez, 2002).

GIS operations perform a wide range of user tasks in GIS: from simpler ones, such as distance measurement, to more complex tasks, such as overlay. GIS operations are defined in the context of this thesis as any tool available in a GIS environment that allows a certain task to be performed. It is hence not restricted to spatial analysis tools but also includes other tools used for instance for data management, computation of basic statistics, or querying, that are not concerned with the spatial component of spatial data.

Since the advent of the first GIS, these systems have found applications in new fields. Meanwhile, the analytical capability of GIS has increased resulting in an extension of the number of GIS operations (Albrecht, 1997). Today, it is usual to expect GIS packages to include more than 200 operations. For example, Grass 6.0 GIS includes 300 operations (Ganguly *et al.*, 2005), there are 250 "modules" in IDRISI Andes (Clark Labs, 2006) and ArcGIS 9.X[1] includes 440 operations (ESRI, 2004). In addition to the available operations, more specialized tools and extensions become available by developers using programming platforms such as ArcScripts[2].

GIS operations have been studied using different criteria to produce classifications of these operations. These criteria correspond to the different views individual researchers

[1] With the inclusion of two extensions ArcView, 40; ArcInfo, 200; Spatial Analyst and 3D Analyst, 200 (ESRI, 2004)
[2] ESRI developers and users exchange their tools and scripts using ArcScripts (http://arcscripts.esri.com/).

have of GIS operations. For example, Albrecht (1996) uses the GIS users' viewpoint to find the most important operations for the users and uses a task-oriented approach to group them. Chrisman (1999) classifies operations based on the level to which data are transformed by the operation (see Table 3-1). These different perspectives led to different classifications of GIS operations (Table 3-1). As such, a variety of approaches for classifying GIS operations exists and, depending on the context in which the research takes place, any classification can be seen as valid and applicable (Bose and Reitsma, 2005).

Studying the classifications in Table 3-1 allows a better understanding of the nature of the large number of existing GIS operations, and potentially of similarities, existing within each group of operations that could be used for this project. After being studied from various viewpoints, the proposed classifications provided a base for further exploring the functioning of GIS operations. These classifications were therefore studied and summarized. The aim was to find similar behaviour among operations within different groups when data uncertainty is considered along with the normal functioning of operations.

Table 3-1: Summary of approaches used for classifying GIS operations

Proposed by	Criteria for categorization	No. of classes	Classes
Dangermond (1983)	Various forms of input and editing considered. Operations in a similar "manipulation technique" category are grouped together.	**10**	1) Data retrieval 2) map generalization 3) map abstraction 4) map sheet manipulation 5) buffer generation 6) polygon overlay and dissolve 7) measurement 8) grid cell analysis 9) digital terrain analysis 10) output techniques
Tomlin (1983); **Berry** (1987); **Tomlin** (1990)	Sequence from the most simple to the most complex	**4**	1) Reclassify map categories 2) overlay maps on a point-by-point or region-wide basis 3) measure simple or weighted distance connectivity 4) characterize cartographic neighbourhoods
Goodchild (1987)	The involvement of different spatial data object classes and object-pairs relationships (e.g., the distance between any pair of objects)	**6** (3 groups of 2)	1) Require access: a) only to the attributes of one class of objects, b) to both attributes and locational information for a single class of objects 2) a) create object-pairs from one or more classes of objects, b) analyze attributes of object-pairs 3) a) require access to attributes and locational information for more than one class of objects or object-pairs, b) create a new class of objects from an existing class
Rhind and Green (1988)	A study based on user needs and available operations	**6** headings (19 groups)	Data: 1) input and encoding 2) manipulation 3) management 4) retrieval 5) analysis 6) display
Giordano *et al.* (1994)	The effect of the function on an uncertainty ellipsoid	**12** (in 3 stages)	**Input**: restructuring, compilation, editing; **Analysis operations**: logical, arithmetic, overlay, geometric property, geometric transformation, geometric derivation; **Output**: restructuring, reporting, visualization
Albrecht (1996; 1997)	Derived from users' perspectives, basic analytical functionality	**6** (20 operations)	Search, locational analysis, terrain analysis, distribution neighbourhood, spatial analysis and measurements
Chrisman (1999)	Level of transformation, "the degree to which the information is inherent in the data model or must be inferred through other relationships"	**4**	Transformations: 1) by extraction 2) based on attribute rules 3) with geometric processing only 4) complete transformation

3.3 GIS operations and uncertainty

GIS operations have been programmed to operate with perfect data, thereby assuming that data are exempt from error and quality issues (Burrough, 1999; Yang *et al.*, 2006). Hence, GIS lack procedures that would account for deficiencies in the quality of the input data or resulting from GIS processes. For example, as Meaden and Kapetsky (1991) illustrate, different GIS operations engage at producing output data regardless of data quality (Figure 3-1).

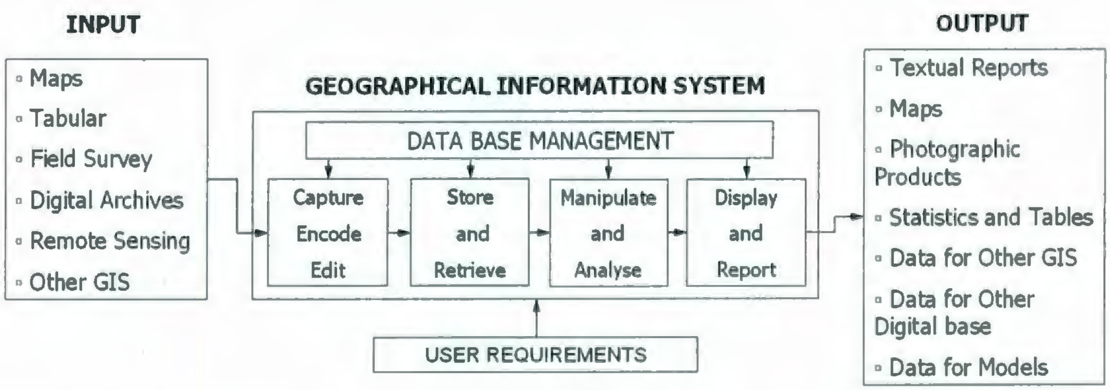


Figure 3-1: The derivation of various outputs by performing GIS operations (Meaden and Kapetsky, 1991)

Given this simplified view of GIS operations, it is nevertheless imperative to account for the deficiencies in the data quality which undermine the accuracy and reliability of GIS analyses (Heuvelink, 1998; Agumya and Hunter, 2002; Gervais, 2006). Since GIS applications are constituted from a number of operations (Timpf, 2003), the entirety of the user application is prone to error when erroneous data are used (Ganguly *et al.*, 2005). In order to avoid undesired consequences of using erroneous data, researchers have suggested providing the users with information about the resulting uncertainty.

Various researchers have since studied the effects of errors on the accuracy of the output from GIS operations (such as Fisher, 1995; Heuvelink, 1998). According to Heuvelink (1998), error in the outcome of an operation originates from two sources: *input error* and *model error*[3]. The *input* error results from error in the data used by the operation. *Model* error is the error caused by the operation itself. For instance, when creating a new dataset from an existing feature dataset using a "Select" operation (either by attribute or location), the error attached to the created dataset originates from both the error in the original data, and the operation and its configuration (here, the particular query used). Figure 3-1 can be re-drawn to include these two sources of error (Figure 3-2).

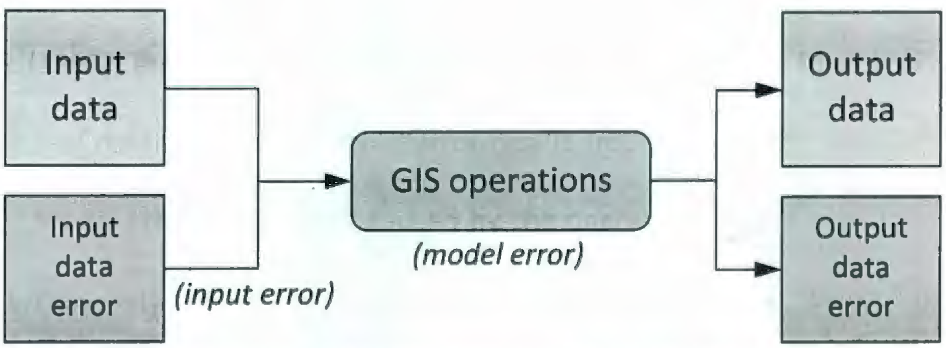


Figure 3-2: Simplified view of data transformation in GIS with the inclusion of error

The "error-aware" GIS design proposed by Duckham and McCreadie (2002) seeks to address the problem of erroneous data used in GIS analysis by integrating error handling methods with GIS applications. The aim is to create an environment that improves, firstly, awareness of the quality of the spatial data used by a GIS operation, "helping

[3] Beard (1989) identified "user error" (e.g., errors in projection or misinterpretation errors) as an additional source of error.

users to understand and make better use of spatial data quality" (Duckham and McCreadie, 2002, p. 62), and secondly, awareness of the quality of the outcome from an operation by "enabling developers to make a range of error handling tools available for users to access" (Duckham and McCreadie, 2002, p. 62, see also Openshaw, 1989).

Among the proposed approaches for an improved error-awareness in GIS is the object-oriented (OO) approach discussed by Duckham *et al.* (2001) and Duckham (2002), for improved handling of data and accompanying error. Current GIS are however based on a relational database model. In an OO GIS model, *objects* consist of both data and methods for handling data. By applying an OO approach, it is possible to include instructions with the data for handling errors that accompany the data. Handling data error can therefore be more closely integrated with GIS data.

Pending the implementation of an OO GIS design (Egenhofer and Frank, 1992; Duckham and McCreadie, 2002), this approach is yet to be provided with error handling techniques that would be able to provide a notion about the resulting error once data are transformed. Although the effects of input and model error on spatial data quality have been studied under the topics of error modeling, error propagation and sensitivity analysis (Veregin, 1989, 1995; Heuvelink, 1998; Crosetto *et al.*, 2001), this study has proven complex. Complex scenarios can involve complicated GIS operations (Keukelaar *et al.*, 2000; see also Couclelis, 1992), also involving several datasets (Östman, 1996), at times resulting in a significant accumulated error (Keukelaar *et al.*, 2000). The study of error propagation through GIS operations is hence complex and requires further research (Couclelis, 1992; Shi *et al.*, 2003; Heuvelink *et al.*, 2006).

With the expansion of available error models created for individual operations (such as those created for a few of the *interpolation* operations; Krivoruchko and Crawford, 2003; Kyriakidis and Goodchild, 2006), such models can be included in the estimation of error in an OO GIS.

3.4 Improving spatial data quality communication in GIS operations

While models for computing errors that result from GIS operations are yet to be incorporated in GIS software, other approaches have been suggested that can similarly be used to improve awareness of data quality. Some of these approaches attempt to improve awareness of data quality prior to GIS analysis, such as visualizing data quality for better awareness (e.g., Huth *et al.*, 2009). Others, such as the risk-based approach (see Agumya and Hunter, 1999) use the adverse effects of a post-usage scenario to communicate quality to the users.

3.4.1 Approaches to the communication of spatial data quality

A variety of approaches have been explored for improving the communication of spatial data quality information. According to Agumya and Hunter (2002) the capability for communicating such information in commercial GIS is limited to metadata management systems associated with some GIS (such as ArcCatalog for ESRI ArcGIS). Such capability has not largely improved.

The method presented in this thesis considers different techniques for delivering data quality information to the users. These include:

1) The visualization of metadata in various forms as viable means to improve awareness of spatial data quality (MacEachren *et al.*, 2005). Interfaces, often using dashboards and signs such as traffic lights, have been particularly popular in research for visualizing data quality information (Duckham and McCreadie 2002; Devillers *et al.*, 2005; Huth *et al.*, 2009).

2) Providing familiar visual warnings in a GIS environment, such as help balloons and pop-up boxes, to deter users from performing illogical operations (Duckham, 2000; Hunter and Reinke, 2000; Levesque *et al.*, 2007). Using audio and animation to provide warnings have also been studied (Fisher, 1993, 1994; Fisher, 1994; Krygier, 1994; Ehlschlaeger *et al.*, 1997).

3) Communicating spatial data quality in terms of the risk associated with using spatial data (Agumya and Hunter, 1999; Van Oort and Bregt, 2005; Levesque *et al.*, 2007). Risk analysis (RA) provides the users with different scenarios resulting from using erroneous data (Agumya and Hunter, 1999). Essentially, data quality is translated into the "risk" of using the data. The feasibility of any action taken once data are used in an analysis (e.g., discard the data or accept the risk) can be justified by considering the acceptability of the risk involved for the user. Although RA has been argued to be suitable and pursued in a number of studies, it has been deemed limited because of several constraints (Van Oort and Bregt, 2005).

Such approaches can be incorporated in GIS to provide warnings when GIS operations are executed. Using these approaches, users could be provided with warnings about quality issues associated with the input data.

3.4.2 Approaches to classification of GIS operations

Since there are more than 200 operations in some important GIS software (see Section 3.2), establishing a mechanism to communicate the error for individual GIS operations can be challenging as well as time-consuming. This study therefore looks into the different classifications of GIS operations to examine if groups of operations could be considered instead of individual operations for any approach for handling uncertainty. The review of the literature (Section 3.2) identified a number of approaches for classifying GIS operations. Unlike the more straightforward *fields* and *objects* categories for GIS data structures (Schuurman, 2005), GIS operations have not been categorized with such a widely accepted classification framework[4]. From the number of classifications proposed for GIS operations, each of them approach the classification of operations from a different perspective (Dangermond, 1983; Tomlin, 1983, 1990; Goodchild, 1987; Rhind and Green, 1988; Giordano *et al.*, 1994; Albrecht, 1996, 1997; Chrisman, 1999).

From the classifications presented in Table 3-1 (page 37), two were considered to be applicable and will help in defining the framework of this research: Giordano *et al.* (1994) and Albrecht (1996).

Giordano *et al.* (1994): This approach is the only one to explicitly relate the classification of GIS operations to data uncertainty. This classification divides GIS operations into

---

[4] Goodchild (1987) characterized GIS functionality as vendor-based and application-driven. GIS in this regard still lacks what was earlier described as "formalized definitions, categorizations, [and] terminologies" (Clarke, 1986, p.175).

three groups based on the stage at which the operation is called during the use of the GIS. These stages are: input, analysis and output. For the purpose of modeling uncertainty, an "abstract model" of geographical data is described, structured along three dimensions: space, theme and time. The three dimensions are illustrated as a cube (Figure 3-3). Since the *true* location of data along each dimension is subject to a certain degree of uncertainty (as spatial data are never perfect), an ellipsoid depicting uncertainty encircles the information (Figure 3-3a).

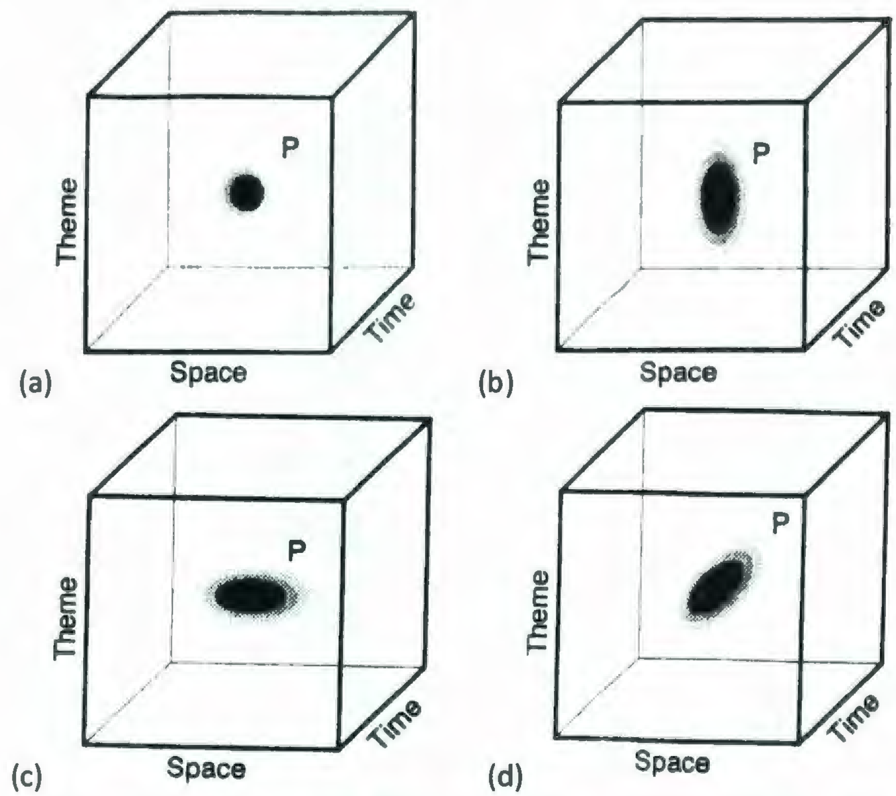


Figure 3-3: A geographical data cube representing (a) an object with initial uncertainty and then, with uncertainty elongated in the three axes representing spatial data's (b) thematic, (c) spatial or (d) temporal dimensions (Giordano *et al.*, 1994)

During an operation, depending on the type of operation performed, the magnitude of the uncertainty can increase in the given directions. For example, the ellipsoid will elongate along the *space* dimension if spatial uncertainty is larger, such as presented on Figure 3-3c.

This "effect" of the operation on the size of the uncertainty ellipsoid is accordingly used as a criterion for further dividing the operations within each stage into a number of groups (12 in total). These groups are ordered based on the magnitude of the uncertainty resulting from their application.

Despite this taxonomy, Giordano *et al.* (1994, p. 49) maintain that, in practice, it is difficult to construct a classification of operations that is "completely consistent with the concept of increasing uncertainty". This is because "[the] effects of a given set of functions [operations] on uncertainty depend on the nature of the source data, including their accuracy characteristics and the way in which functions are applied to these data" (Giordano *et al.*, 1994, p. 49). Additionally, the fact that GIS operations tend to affect the uncertainty ellipsoid with unequal magnitudes in the three dimensions, makes an exact classification a more difficult task.

The Giordano *et al.* (1994) approach can nevertheless provide a notion for the transformations performed by GIS operations that affect the accuracy of data (see also Lanter and Veregin, 1992). The fundamental notion of increase in uncertainty in the spatial, thematic and temporal dimensions can provide the users with information about the relative magnitude of the uncertainty resulting from the data transformation. This information may be incorporated into a system that communicates data quality in GIS to improve users' awareness of uncertainty.

Albrecht (1996): The classification of GIS operations proposed by Albrecht (1996) was also considered helpful from a number of viewpoints. Albrecht's approach can improve error-awareness by simplifying the user's understanding of the process of GIS applications. This approach presents 20 universal analytical GIS operations capable of building all types of GIS applications. The 20 operations are classified under six main groups (Table 3-2).

Table 3-2: Universal analytical GIS operations (Albrecht, 1996)

Group				
Search	Interpolation	Thematic Search	Spatial Search	(Re)classification
Locational Analysis	Buffer	Corridor	Overlay	Thiessen/ Voronoi
Terrain Analysis	Slope/Aspect	Catchment/ Basins	Drainage/ Network	Viewshed Analysis
Distribution/ Neighborhood	Cost/ Diffusion/ Spread	Proximity	Nearest Neighbor	
Spatial Analysis	Multivariate Analysis	Pattern/ Dispersion	Centrality/ Connectedness	Shape
Measurements	Measurements			

The 20 elementary operations can be compounded to produce environmental models performing various GIS applications. An example of such models is given in Figure 3-5.

The idea of combining, or "chaining", GIS operations and services to perform applications has been a recurring theme within the GIS literature (see Giordano *et al.*, 1994; Albrecht, 1996; Alameh, 2003). According to Giordano *et al.* (1994, p. 47), in a "closed system" of analysis functions, "the output of any function can serve as the input to another function." Research by Lemmens and De By (2002, p. 3) suggests that "GIS operations can exist as basic, fine grained functions, such as select or overlay, or as compound operations that consist of finer grained operations." These operations can hence chain to compositely perform larger tasks and applications such as more complex GIS operations or environmental models used in hydrology and landscape ecosystem research (Albrecht, 1996; Lemmens and De By, 2002; Alameh, 2003; Bernard *et al.*, 2003; Einspanier *et al.*, 2003; Timpf, 2003; Yue *et al.*, 2006). Figure 3-4 illustrates the concept whereby a GIS operation can itself be composed of a number of GIS operations.

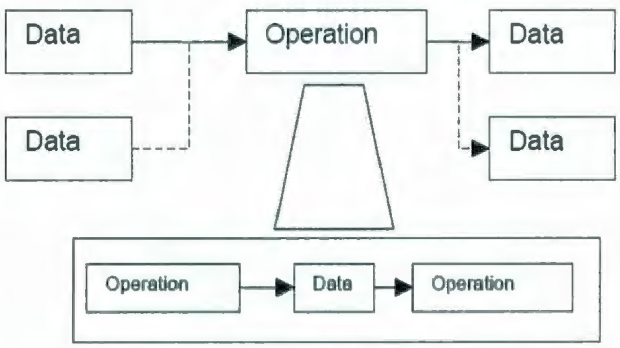


Figure 3-4: Data-operation chain (Lemmens and De By, 2002). The lower section illustrates the constitution of larger operations from more fine-grained operations

This approach has already been applied in different projects, such as interoperability of GIS services (Aditya and Lemmens, 2003; Lemmens *et al.*, 2006), Spatial Data Infrastructures (SDI) (Aditya and Lemmens, 2003; Wytzisk and Sliwinski, 2004; Kiehle *et al.*, 2006) and Web GIS (Tsou, 2001; Peng and Tsou, 2003; Lemmens *et al.*, 2007).

Albrecht's (1996) approach is used by Ganguly *et al.* (2005) for modeling errors in GIS operations. It proposes using a "chain of geospatial operations" (Ganguly *et al.*, 2005, p. 2) for modeling error in GIS operations by replacing "complex geo-computational models" (Ganguly *et al.*, 2005, p. 3) with simpler GIS operations. Once larger models are broken down into smaller operations, methods such as Artificial Neural Networks (ANN) are used to model errors in individual operations. By estimating the error resulting from individual operations in the sequence of GIS operations, this approach can therefore model the error associated with the entire process. Figure 3-5 presents Ganguly *et al.*'s (2005) approach to error modeling. The dashed line represents the chain for the accompanying error terms.

Using Albrecht's (1996) approach, research on error propagation through GIS operations can be restricted to a limited number of simple operations (Ganguly *et al.*, 2005). As a result, while being able to cover all types of GIS functionality, this approach has the advantage of providing the capability of examining the transformation of error at each sub-element of a GIS operation; a step-wise approach to GIS operations and models. Error-modeling research has already been carried out on a number of the operations in the Albrecht classification (1996). For example, Veregin (1989; 1995) studies the effects of error on the accuracy of an overlay operation and Shi *et al.* (2003) studies error effects on vector-based buffer analysis. For the purpose of increasing error-awareness, GIS software can use the models produced from such studies to provide users with an interface displaying information on the effects of input error on the output data.

In conclusion, by applying a combination of approaches in Giordano *et al.* (1994) and Albrecht (1996), respectively, it is possible to assess the main resultant uncertainty in the outcome of a given analysis, and to restrict the number of GIS operations under study. While error-modeling research strives to cover all GIS operations, which is a challenging task, the combination of these two approaches can benefit from a granular approach to GIS operations and services (Lemmens and De By, 2002). It is possible to increase error-awareness in smaller and simpler steps of operations with accompanying information describing error effects from literature such as Giordano *et al.* (1994).

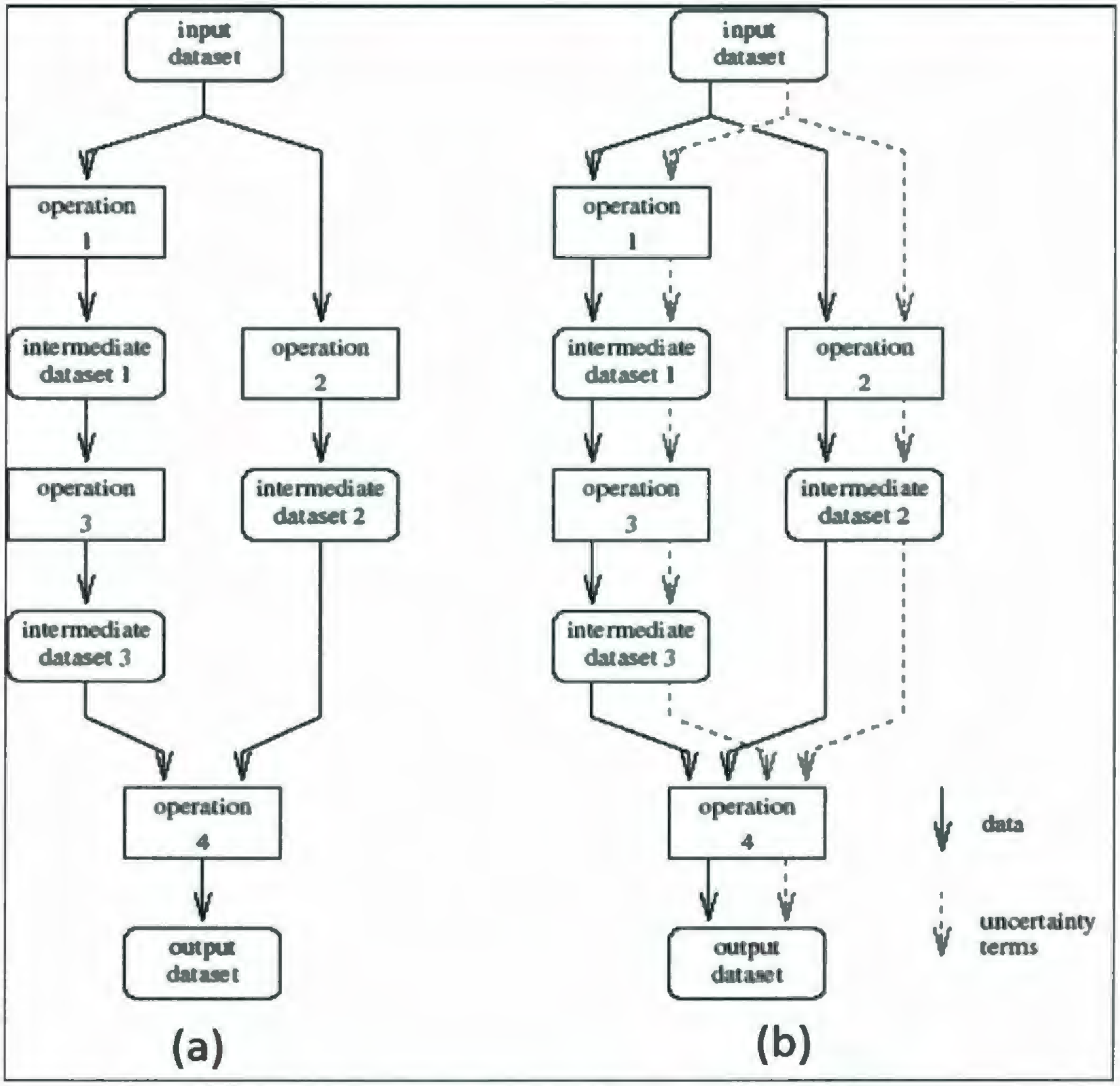


Figure 3-5: A chain of geospatial operations (a), with the presence of uncertainty terms (b) (Ganguly *et al.*, 2005).

3.5 A conceptual framework for linking GIS operations and data quality information

In this section, the mechanisms by which GIS software carry out user operations and the interaction of the components involved in this process are studied. This setup will then be analyzed for establishing of a system where communicating input data quality is linked to GIS operations. As described earlier, such a link should enable an interaction between the two components, allowing the communication of relevant quality information during a GIS operation. A number of the "universal" operations defined by the Albrecht (1996) classification will later be used as examples for implementing this link.

The proposal for linking spatial data quality information with GIS operations has been mentioned by Bédard *et al.* (2004). This work raised the idea of relating spatial data quality and GIS operations to improve GIS capabilities to handle error, calling for the identification of a set of rules for linking data quality information to GIS operations (Bédard *et al.*, 2004).

When using the current setup of GIS software, a user that wants to use an operation while consulting the metadata[5] will interact with three components: spatial data, operations and metadata. For consulting the quality of the data used in an operation, the users must first locate the input datasets and then locate the associated metadata,

[5] As per the guidelines in metadata standards, spatial data quality information is stored in a section within metadata and not as individual accompanying data; hence *metadata* has been used hereafter as the component when discussing quality information in a GIS environment.

which is typically managed and accessible by a separate system (for instance ArcCatalog in ArcGIS). Therefore, the current connection between operations and spatial data quality information can be regarded as indirect, involving at least two steps illustrated in Figure 3-6.

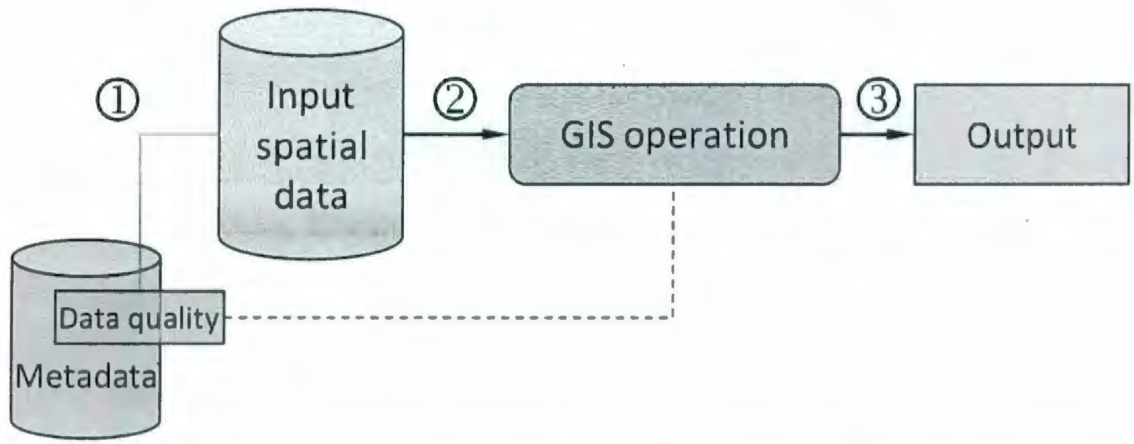


Figure 3-6: A simplified view of the steps involved for performing a GIS operation. Link ① depicts the setting in current GIS where ideally, users consult metadata prior to using GIS operations (link ②) and producing results (link ③). The dashed line depicts an envisioned connection between GIS operations and quality information, where quality information is directly provided with GIS operations

By applying the appropriate techniques, the awareness of quality information could be improved by communicating relevant input data quality information to the users when GIS operations are used. If a mechanism could allow direct access to data quality information during the usage of GIS operations, users could simultaneously be provided with relevant quality information that would describe the quality of input data. The path the users take to access quality information could therefore be shortened. It should be noted that this mechanism will communicate the quality of the input data, and not the quality from the outcome of the operation (which would then require modeling input and model error).

In order to create this link, the components in Figure 3-6 have been re-arranged to have metadata and GIS operations adjacent (Figure 3-7a). As a result, metadata (and thus data quality) are now adjacent to both spatial data and GIS operations and metadata can be schematically linked to both. This new configuration can provide a step towards studying the association between GIS operations and metadata. A number of modifications may be applied to GIS in order to establish a link between GIS operations and data quality information (depicted in Figure 3-7b).

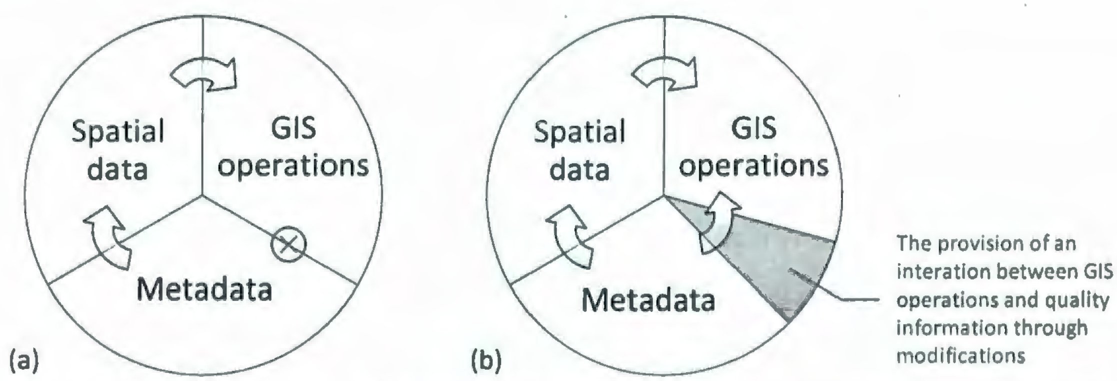


Figure 3-7: A conceptual view of GIS functioning. (a) Figure 3-6 redrawn, current GIS functioning: the three components are represented side by side with no direct link between GIS operations and metadata, (b) the proposed approach for including a mechanism to link GIS operations and quality information

In order to establish this link, earlier research can be consulted to find a suitable platform for its implementation. "Service metadata" provide information that describes services and their properties, including their input requirements, computational requirements, and output requirements (Tsou, 2002; Radwan *et al.*, 2003; Kuhn, 2005; Aktas *et al.*, 2005, 2007). Such metadata can be considered as a method for describing GIS services or *operations* when the implementation is local (i.e., not distributed). For the purpose of establishing a link between GIS operations and quality information, the concept of service metadata may be extended to include information about the applicability of quality elements in the operation. Such metadata can describe GIS operations within the scope of the association of quality information with GIS operations. The service metadata can thus be extended to facilitate an exchange between the two components (GIS operations and quality information). In this work, this extension to service metadata has been termed *Operation Quality Metadata* (OQM). OQM may include information describing individual GIS operations with regard to their link with quality information.

As it has been illustrated in Figure 3-7 (above), OQM would be located between the metadata and the GIS operations components, linking these two components (marked by the symbol ⊗ in Figure 3-7a and the grey area in Figure 3-7b). Based on existing literature, it is possible to identify a number of roles that OQM may perform. The potential involvement of OQM can be discussed in the two distinct stages of the functioning of GIS operations mentioned in Figure 3-2 (page 39) and Figure 3-4: data input and data output.

1. Input stage: At the input stage of a GIS operation, users specify the input datasets and set the required parameters for the operation (such as setting a search radius value in an *Inverse Distance Weighted (IDW) interpolation*). At this stage, GIS users may be provided with warnings and information about data quality. Once the input data are specified, GIS operations will be instructed to perform a number of quality-awareness tasks along with their normal procedures. Modifying GIS operations will permit the retrieval of the appropriate data quality information from the metadata, during the input stage. Giordano *et al.* (1994) study the effects of GIS operations on uncertainty in the spatial, temporal and thematic dimension of spatial data (also in Lanter and Veregin, 1992). Amongst other outcomes from this research it is argued that first, GIS operations affect uncertainty in either dimensions asymmetrically, and second, the amount of overall uncertainty introduced by each operation varies. The spatial data quality section in metadata provides information about some aspects of uncertainty in each of these dimensions, that is thematic, temporal and positional accuracy. OQM can accordingly use the Giordano *et al.* (1994) approach to indicate which data quality elements from the metadata are relevant for a given operation. This will extend some aspects of expert systems and knowledge-based rules described by Fischer and Nijkamp (1992) and Duckham (2002), which provide rules relating data quality to various types of spatial data. An example of such a rule is "positional accuracy can only apply to geometric information" (Duckham, 2002, pp. 185-186). These rules can be extended to add a 'GIS operations dimension' that would describe the applicability of spatial quality elements to GIS operations. For example, for the operation "Select by Attribute", the elements of *attribute accuracy, completeness* and *logical consistency* are immediately applicable, while *lineage, positional accuracy* and *temporal accuracy* may be hidden or excluded from the applicable elements (see Figure 3-8).

This research also notes that the selection of applicable quality elements for a given operation can be complex. This selection has been made a-priori, by a GIS expert, based on a most likely uses. For instance, in "Select by Attribute", this operation would always be affected by *attribute accuracy*, however, *temporal accuracy* or *lineage* may be relevant in some less common situations. For example, *lineage* may become important if the spatial coverage of data or their specifications have evolved over the lifetime of the database. Similarly, *temporal accuracy* may also become relevant when using spatio-temporal data, e.g., position of moving objects can be influenced by time accuracy. However, such cases are fairly rare in real-world usages of GIS.

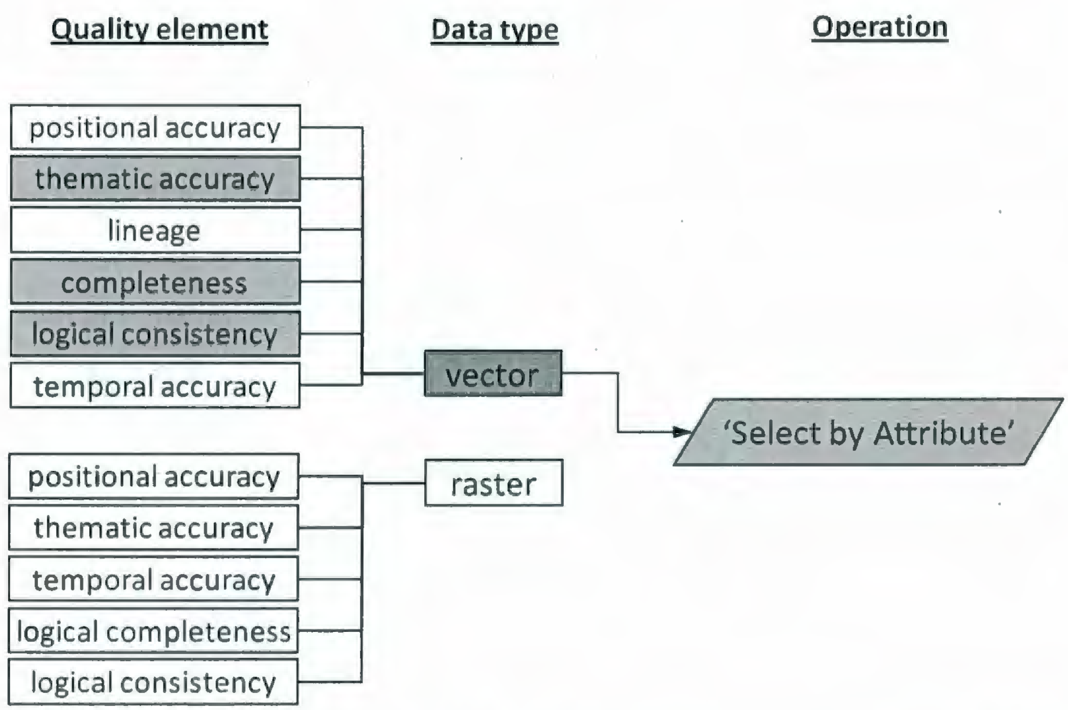


Figure 3-8: Association of spatial data and data quality elements relevant to the operation "Select by Attribute". The quality elements are taken from ISO 19115 and ISO 19115-2 (ISO/TC, 2003, ISO/TS, 2007)

2. Output stage: At the output stage (after the use of the operation), input data have been transformed and new data are produced. For example, an *Intersect* operation produces a new *feature class* and a *Global Polynomial Interpolation* outputs a continuous surface for the study area. At this stage, users may want to know about the effects of the deficiencies in the input data on the output of the operation. While models to compute such knowledge are currently largely absent (see Section 3.2), users can be provided with information describing the magnitude of the effects of the operation on quality (see Giordano *et al.*, 1994). In this regard, literature can help in assessing the quality of the outcome of some operations. There are studies that provide general information about the effects of GIS operations on data (e.g., Giordano *et al.* 1994), while others, such as Heuvelink (1998) and Crosetto *et al.* (2001), provide more detailed methods for measuring the magnitude of uncertainty in the outcome of some GIS operations. Along with the information extracted from this literature (which help improve the awareness of error in the outcome of an operation), providing real-life examples of the usage of quality elements is also helpful for the users (Gould, 2005) in addition to helping them become familiar with the terminology used in spatial data quality (Hunter *et al.*, 2005).

3.6 Operation Quality Metadata (OQM)

The OQM can describe GIS operations in two sections for linking GIS operations to quality information that are: a) with regard to the applicable quality elements on GIS operations, and b) description of the sensitivity of GIS operations to error and effects of GIS operations on the quality of the outcome. The following sections describe the information that can be stored in the OQM.

1. Applicable quality elements: This section of the OQM specifies the quality elements applicable to individual GIS operations. Table 3-3 shows the applicable quality elements for three of the Albrecht (1996) operations.

Table 3-3: ISO 19115 quality elements applicable to three universal GIS operations

Quality element \ Operation	Distance Measurement	Select by Attribute	Select by Location
Positional accuracy	X		X
Thematic accuracy		X	
Lineage			
Completeness	X	X	X
Logical consistency	X	X	X
Temporal accuracy			

This information will indicate which quality elements should be provided to the user.

2. Effects of quality issues on GIS operations: The information stored in this section describes the operations from the perspective of the sensitivity to, and effects of, each applicable quality element. It will be represented similarly to the *help* section typically found in desktop software, and will help users to view the obtained values together with a description of the effects of applicable quality elements. For each operation the user is provided with information, summarizing the effects of error on the operation with possible real-life examples of the consequences of using such data. By including the Giordano *et al.* (1994) classification, users can be provided with a ranking of the operation with regard to its effects on the accuracy of spatial data in thematic, temporal and spatial dimensions.

3.7 A model for communicating spatial data quality within GIS operations

Following the conceptualization of the link between GIS operations and quality information in the previous section, this link may be incorporated in the setup of GIS software for communicating quality information with GIS operations. GIS software can thus be modified in the necessary sections to include mechanisms for retrieving and communicating quality information during the execution of a GIS operation.

These modifications are presented in Figure 3-9: steps ①, ② and ③ represent the normal execution of a GIS operation where users request an operation. The system retrieves the relevant data and processes the program related to the operation, and finally provides users with the results. The next steps concern the extension proposed in this thesis. At step ④ the operation executed is extended to take the data quality into consideration. At steps ⑤ and ⑥ the extended operation retrieves the name of the applicable quality elements from the OQM database. At step ⑦ based on the OQM data, the extended operation retrieves the information for the applicable quality elements from the metadata. Once the data quality elements are obtained, this information, along with the information describing the quality elements effects on the operation (i.e., the second part of the OQM; see Section 3.6), can be sent back to the extension for display on an interface. At step 8, the user is provided with the applicable quality information and their descriptions. Various communication techniques may be applied in order to warn users of the data quality issues once the relevant quality information for the operation is retrieved (see Section 3.4.1). Researchers have commonly considered approaches such as the display of interfaces with quality elements and warning signs (see Section 3.4.1; Agumya and Hunter, 2002; Devillers *et al.*, 2005; Huth *et al.*, 2009).

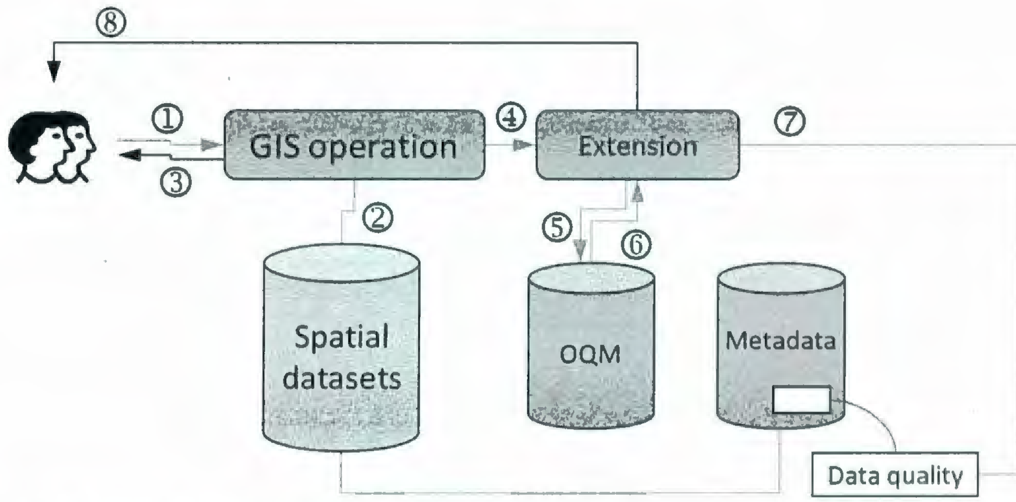


Figure 3-9: Conceptual architecture for the communication of spatial data quality with GIS operations. Black arrows represent the steps at which the users are provided with information. All other arrows present the flow of the process.

3.7 Summary

This chapter presented a conceptual framework for communicating quality information with GIS operations. First, a conceptual link between GIS operations and quality information was established, involving, a) studying the characteristics of both GIS operations, and quality information and b) finding suitable approaches in the literature to conceptually relate the two components. Second, the link was used to conceptualize an extension to GIS operations to provide relevant quality information to the user once an operation has been executed.

4. Prototype

4.1 Introduction

This chapter describes the development of a software prototype that implements the concepts presented in Chapter 3. The first part of this chapter discusses a generic implementation framework, applicable for different types of operations used within GIS software. Subsequently, this framework is applied to two specific operations that are commonly used in commercial GIS software to showcase its effectiveness.

4.2 Architecture

This section describes the architecture of the prototype. First, it provides an overview of the establishment of the various components in a GIS software and discusses their interaction with each other. Second, it describes the components' interaction in a new setting where GIS operations are extended to perform additional tasks for communicating data quality.

4.2.1 Prototype components

In Chapter 3, three components were described that are associated with the development of the prototype (GIS operations, spatial data and metadata). These components are present in most commercial GIS software and are individually studied below for the implementation of the conceptual framework.

Among different GIS software, these components have been developed in varying configurations and formats. For example, spatial data formats for one software may not be usable by another software. Additionally, GIS operation codes are often inaccessible, and are consequently, unusable within other software environments. In order to be generic, the common characteristics of these components are discussed and the development of the prototype is based on those common characteristics.

1. GIS operations: GIS operations perform different user tasks in GIS software. They are termed differently, for example *tools* or *geoprocesses* in ESRI ArcGIS, *modules* in IDRISI and *modules/command* in Intergraph Geomedia. GIS software provide a number of preset operations that can be expanded with the programming capabilities provided in the software. GIS operations are often developed exclusively to be executable within the framework of their own software. Programming frameworks such as Python provide the capability to share operation codes among GIS software.

2. Spatial data: Spatial data are transformed by various operations to produce the outputs sought by the users. Generic and legacy data formats can be used for modeling spatial data (for example the ASC file format created by the American Standard Code for Information Interchange, ASCII for the former, and the Shapefile format created by the Environmental Systems Research Institute, ESRI for the latter). Currently, most GIS software include tools for making spatial data files usable across different platforms.

3. Metadata: The standard and file format by which metadata are stored vary (see Section 2.3). A number of guidelines or "standards" have been published for storing metadata (see Section 2.3). Among the available standards, the ISO 19115 standard (ISO/TC-211, 2003) in particular has been widely used (Kazakos *et al.*, 2005). The XML (Extensible Markup Language) file format is a popular format for storing metadata (Green and Bossomaier, 2002). Given the structural representation of data in XML, this language has proven flexible and versatile. The XML representation of ISO-19115 provided by the proposed ISO-19139 standard (ISO/TS, 2007, presented by earlier drafts) has become the *"de-facto* standard" for the storage and exchange of spatial metadata (Kazakos *et al.*, 2005).

4.2.2 Prototype programming

As explained in Section 3.5 (page 50), the development of an extension to GIS operations for communicating data quality requires a number of modifications to be implemented into GIS operations (cf. Figure 3-9, page 59). The resulting *extension* to GIS operations is designed to activate following the execution of an operation. Once this extension is activated, a *quality report* is produced in a number of steps.

The programming of the prototype comprised of two tasks: a) programming an extension to GIS operations, which is an executable section that activates once an operation is performed, and b) creating Operations Quality Metadata (OQM), primarily storing information for applicable quality elements to various GIS operations (see Section 3.6, page 56).

Each task is described in the following section:

4.2.2.1 Programming the executable module

The functioning of the executable part of the prototype can be described in four steps (Figure 4-1).

> ① In order for the prototype to operate, it requires receiving two parameters (refer to Section 3.6 and Figure 3-9). These two parameters are 'operation type' and 'input data'.
>
> The type of operation is used to identify which spatial data quality elements are immediately required to be communicated to the user. For example, a distance measurement operation's quality elements would be: a) *positional accuracy* (inaccuracy in the position of the features can lead to error in the calculated distance), b) *completeness* (missing or superfluous features can lead to errors in the measurement),

and c) *logical consistency* (problems such as line overshoots/undershoots and the unintentional crossing of lines can lead users to inaccurately measure distances). These elements are required to be collected from the metadata and communicated to the user.

Information relating to input data includes input data's storage path and the type of input data (i.e., raster vs. vector). Using this information, the prototype locates the accompanying metadata files for each of the datasets.

Once these two parameters are received, the prototype's first task is completed (i.e., step ①) and the prototype will then proceed to perform the next steps.

② Based on the operation type, the prototype uses the OQM to determine which quality elements apply to the operation performed. For each quality element, the prototype retrieves additional quality-descriptive help information that is stored in another section in the OQM (explained in Section 3.6 and in detail in Section 4.2.2.2).

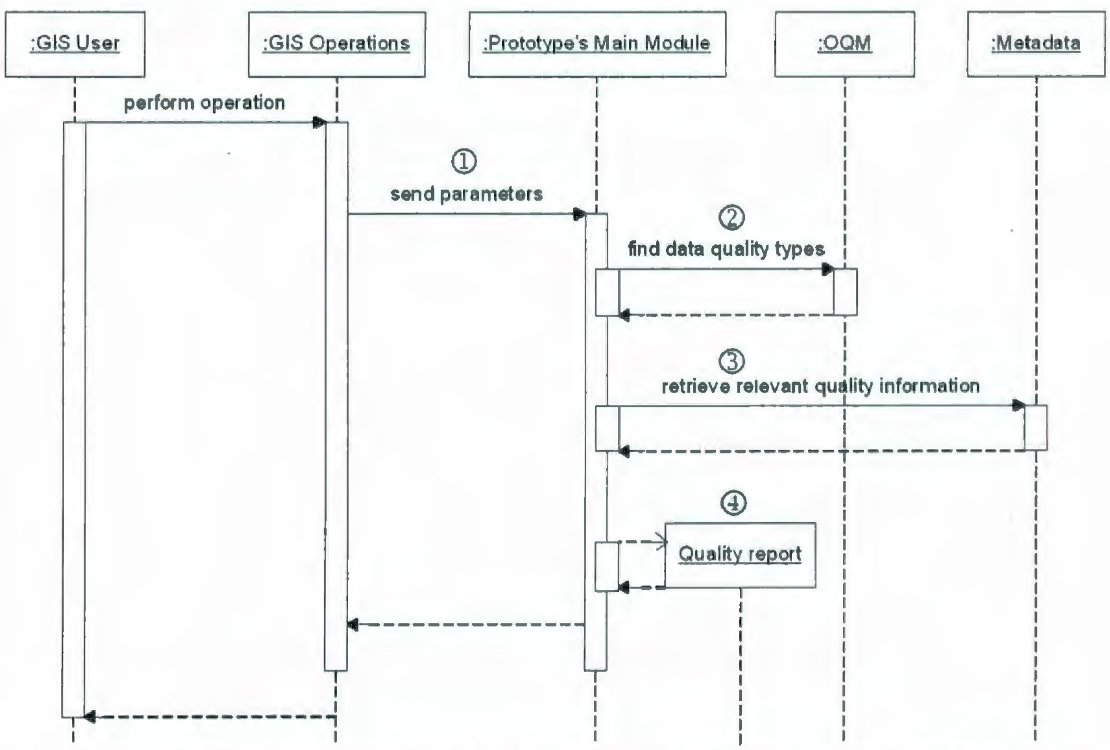


Figure 4-1: Steps involved in the collection and display of quality information presented in a UML interaction sequence diagram

③ The prototype searches for the relevant quality elements and their classes in each layer's metadata (see Section 2.3, page 17). Any information found for each of the quality elements is saved in a string variable.

④ The retrieved quality information along with quality-descriptive information are used to produce a report and display it to the user.

4.2.2.2 Creating the OQM

The OQM supports the prototype by specifying the applicable quality elements. The OQM database is shared among the prototypes that are developed using the approach described in the previous section. OQM also contains information for describing the effects of error on the operation. For each of the operations included in the OQM, a record is provided with a number of columns (fields/attributes) for describing the operation in the above mentioned context. The following attributes are included in the OQM:

1. The primary role of the OQM is to present the prototype with information describing applicable quality elements. This attribute is termed *Quality_Elements*. Depending on the type of the operation, this field specifies elements from the six spatial data quality elements which are immediately applicable to the operation (see Section 3.5 for more details related to the selection of the quality elements immediately applicable for a given operation). For example, this field will contain the following information for the *Measure* operation:

```
PA, LC, Co
```

which means the three quality elements of *positional accuracy (PA), logical consistency (LC),* and *completeness (Co)* are applicable for this operation and need to be communicated to the user.

2. A field, specifying the group that the operation belongs to according to Giordano *et al.*'s (1994) classification (termed *Operation_Group*). For the *Measure* operation the following data describes the operation's classification:

```
Analysis--Geometric Property Operations,2--4/6
```

indicating that the operation belongs to the *Geometric Property Operations* group of *Analysis* operations (the second stage in GIS functioning, see Section 3.4.2). The Geometric Property Operations group ranks fourth among the six groups of Analysis operations (see Section 3.4.2). This indicates that the *Measure* operation is among the more sensitive operations to uncertainty and that the data produced by the operation is more susceptible to effects of uncertainty.

3. A field describing the effects of error on the operation (*Error_Information*). This field contains technical information describing the effects of each quality element and the rank of the operation in Giordano *et al.* (1994) classification. The *Error_Information* field for the *Measure* operation reads:

```
The assessment of the final error on the measurement can be
complex and depends on the errors on the start and end points
and other factors related to the operation used.
```

In a case when the measurement's start or end points are based on the features on the screen, omission and commission error both undermine the accuracy of the application. The absence or presence of missing or superfluous features undermines the accuracy of the intended measurement.

Users may also consult a separate help section that describes the quality elements (e.g., positional accuracy, omission and commission). Together, this help information can aid users to better interpret the effects each applicable quality element has on the data used.

For more advanced users the same field contains more technical information:

The MEASURE operation ranks 4th (Geometric Property Operations) among the 6 groups of operations with regard to the graveness of its effect on the magnitude of uncertainty in the outcome [1].

The assessment of the final error on the measurement can be complex and depends on the errors on the start and end points and other factors related to the operation used.

In a case when the measurement's start or end points are based on the interface features, Omission and Commission error both undermine the accuracy of the application. The absence or presence of missing or superfluous features undermines the accuracy of the intended measurement.

[1] Giordano, A., Veregin, H., Borak, E. and Lanter, D., 1994. A Conceptual Model of GIS-Based Spatial Analysis. Cartographica, 31(1):44-51

4.2.3 Implementation into a commercial GIS

Following the development of a framework for implementing the concepts into a software prototype, this framework was implemented into a commercial GIS software. The prototype was programmed using the Visual Basic 6.0 (VB 6.0) programming language and using the ArcObjects framework. The Component Object Model (COM) architecture that is adopted in ArcObjects enables the users to program executable *add-on* tools for ArcGIS. The codes for the *Measure* and *Select* operations were modified to provide examples for the prototypes to be developed for other GIS operations. The codes for the operations were obtained from the ArcGIS Developer Online website[6]. The codes are stored in a single *class module* file. The module, along with a number of auxiliary files, is compiled into a *dynamic-link library* (dll) file. Upon adding this file as an ArcGIS extension, the GIS operation performs the regular task, however, the functioning of the operation is extended to include instructions for communicating data quality.

For the Measure operation, the process for modifying the code has been described in four steps (Figure 4-2; see Appendix A for the complete VB code).

Step 1: For measuring the distance between two points, the user clicks on two points on the ArcMap interface and, once completed, the distance between the two points is displayed at the

[6] http://edndoc.esri.com/arcobjects/9.0/. The website provides sample codes for a number of tools in ArcGIS.

left corner of the status bar. At the termination of the operation, the main module uses the *IMap* interface to access ArcMap's map objects. Using the *IMap* interface, it is possible to access the *layers* class (*ILayer*) within the ArcMap interface and its properties. The *ILayer* interface, more specifically, provides access to the individual layer-objects and their properties. In the *Measure* operation, since there is no indication as to which layers are to be included in the measurement, all visible layers are added as input data layers for communicating quality information. This step, therefore, initially involves reporting the list of visible layers[7]. Later, at the communication stage, users are provided with the capability to select the desired layers.

Step 2: The prototype searches for the metadata files for each of the involved layers. The prototype initially looks for metadata information stored as XML files with the same file name as the dataset. If the prototype is unable to locate the metadata, the user is asked to manually locate the information. After accessing the metadata the prototype separates the data quality section identified by the *DQ_DataQuality* tag. Therefore programming this part includes two tasks:

1. To locate and load metadata into the memory. A string variable is used to temporarily store all metadata for each layer.

2. Then, to locate and load the quality information, for each data layer into another string variable.

[7] For the case of most operations this step is clearer since the input data is manually selected and inputted by the user (e.g., in the case of a *Select by Attribute* operation, the user manually sets the input layers). The required metadata is then retrieved knowing the input layer.

Step 3: Once each layer's data quality information has been extracted, the prototype then searches the OQM for the applicable quality elements. Three quality elements are applicable to the *measure* operation: *positional accuracy, logical consistency,* and *completeness*. The prototype then retrieves information for these quality elements in the metadata's quality information section. The following ISO 19115 tags are used to locate the applicable quality elements:

```
<DQ_PositionalAccuracy>        </DQ_PositionalAccuracy>
<DQ_LogicalConsistency>        </DQ_LogicalConsistency>
<DQ_Completeness>              </DQ_Completeness>
```

The first tag in each line indicates the start of that quality element's text and the second points to the termination of the text. Each quality element is comprised of a number of classes (see Figure 2-3b and Appendix E for details). The next task is then to separate the information within each quality element into the comprising classes. For example, for the quality elements of *positional accuracy* (*DQ_PositionalAccuracy*) the program will search for information under the following tags:

```
DQ_AbsoluteExternalPositionalAccuracy
DQ_GriddedDataPositionalAccuracy
DQ_RelativelnternalPositionalAccuracy
```

These, respectively represent: Absolute External Positional Accuracy, Gridded Data Positional Accuracy, and Relative Internal Positional Accuracy (see Appendix E for detailed descriptions of ISO-19115 quality elements).

Step 4: The final step involves displaying the collected metadata and the relevant quality-descriptive information to the user utilizing the *quality report* (described in more detail in Section 4.2.4).

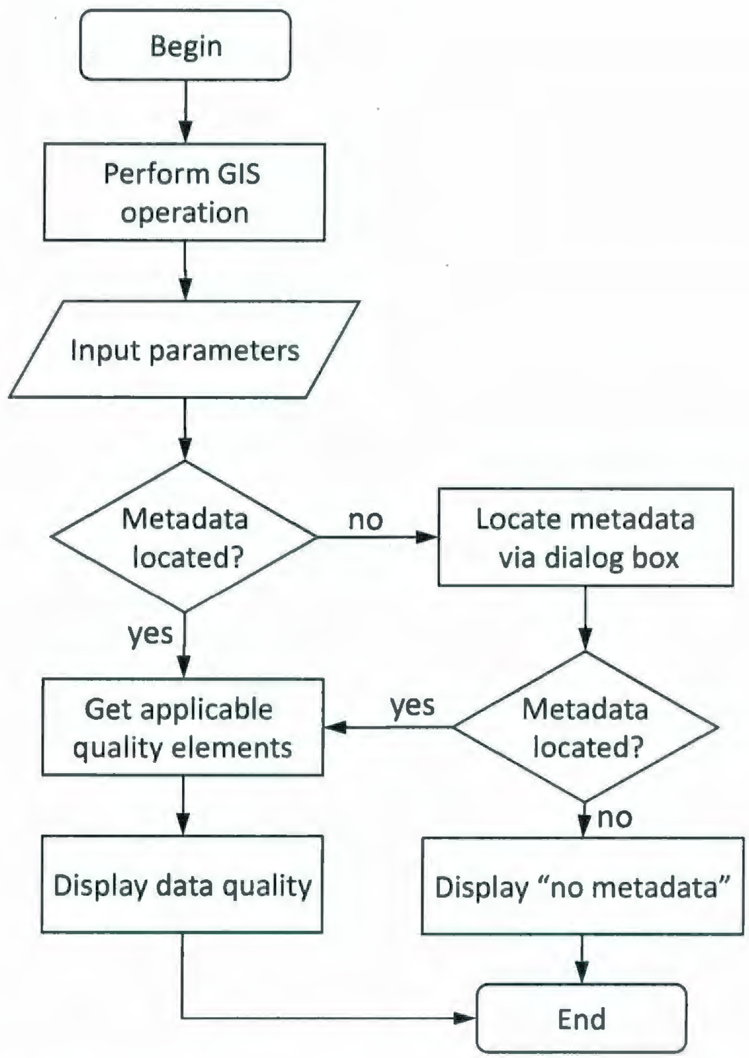


Figure 4-2: Steps involved in the production of the quality report

4.2.4 Interface for the data quality report

This section discusses the *quality report*, which communicates the spatial data quality information to the GIS user. The design and format of this report is influenced by previous approaches that have used interfaces for communicating data quality (see Section 3.4.1). This includes interfaces presented in works such as those by Devillers *et al.* (2005), Huth *et al.* (2009) and Levesque *et al.* (2007). Different colours were used for each of the quality elements. A simplified approach for warning the users of issues within quality is used: two colors in addition to a warning sign are used for this purpose. Green is used for "no problems", imitating a green traffic light (e.g., when there is 100% completeness or little spatial error). Yellow color and warning signals are displayed for warning the user in cases where data contain errors. Based on user requirements, this predetermined approach to the warning system may be enhanced to include the capability to distinguish between for instance 99% and 50% completeness. These values are indicative. Setting crisp values for thresholds between green, yellow and red is a difficult task that could be done by a more formal specification of user's tolerance to the risk.

As illustrated in Figure 4-3, the interface provides an initial warning to the user about the value obtained by the measurement (at the top section). The yellow colour has been displayed in background, warning about the issues with the positional accuracy (Figure 4-3 and Figure 4-4). In the case of 0m spatial inaccuracy or 100% completeness (i.e., 0% omission), a green light would be used (Figure 4-5 and Appendix C).

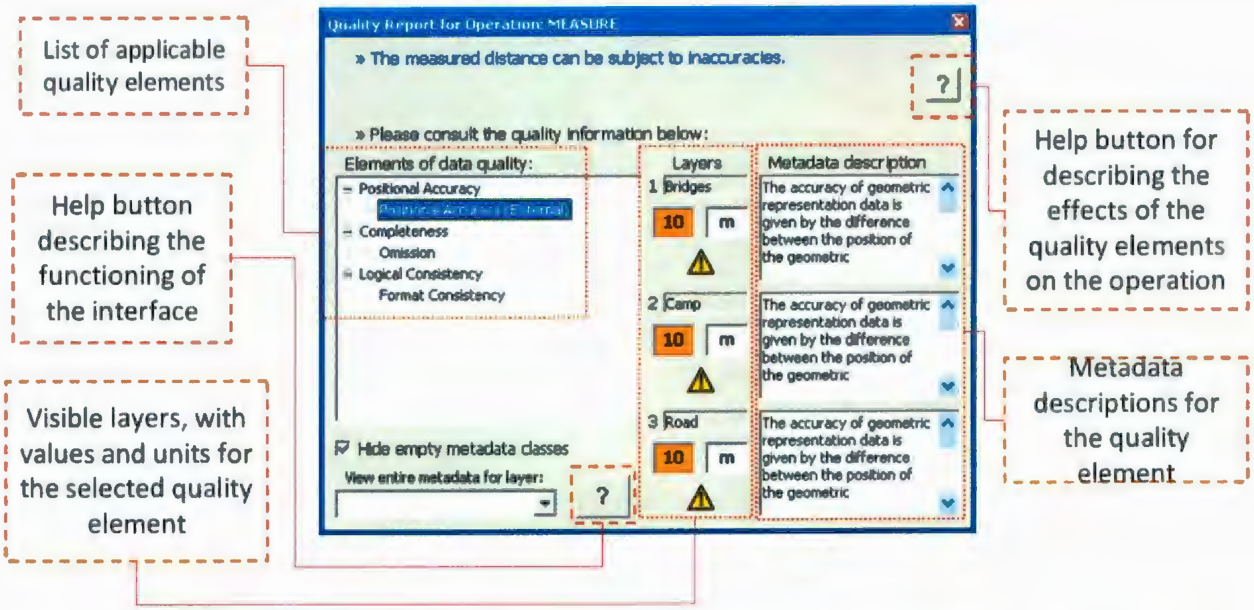


Figure 4-3: An example of the *quality report* for 'Measure' operation (Version 1.01)

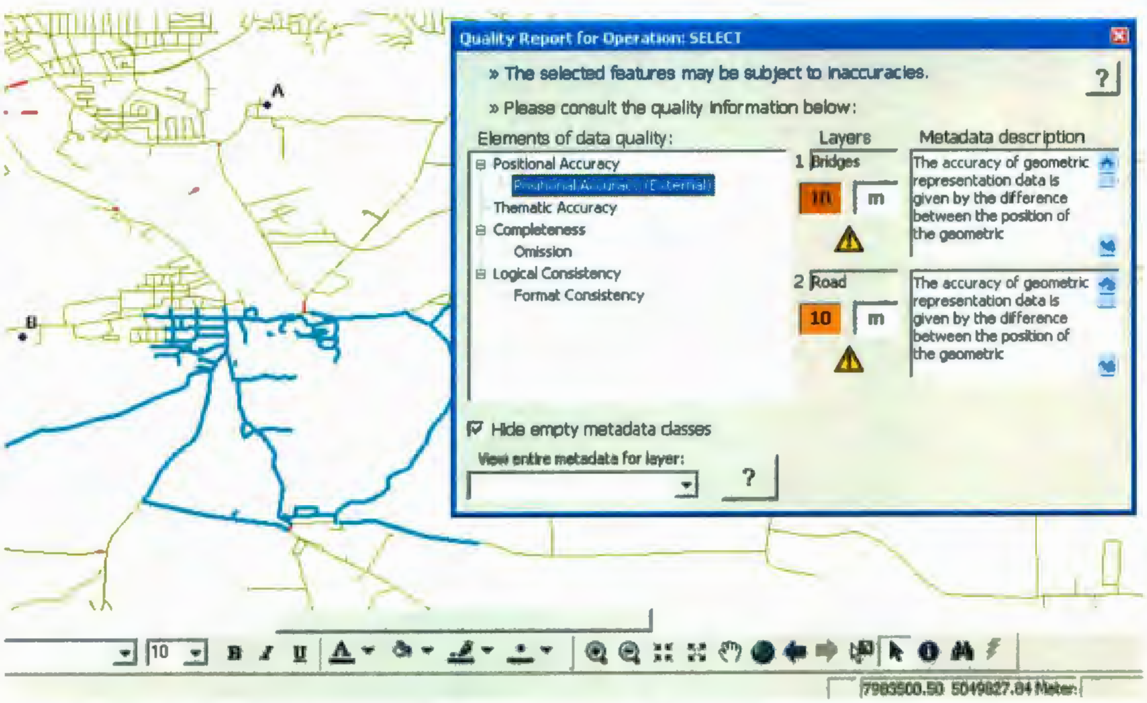


Figure 4-4: Prototype interface for operation 'Select'. Features from two layers are selected.

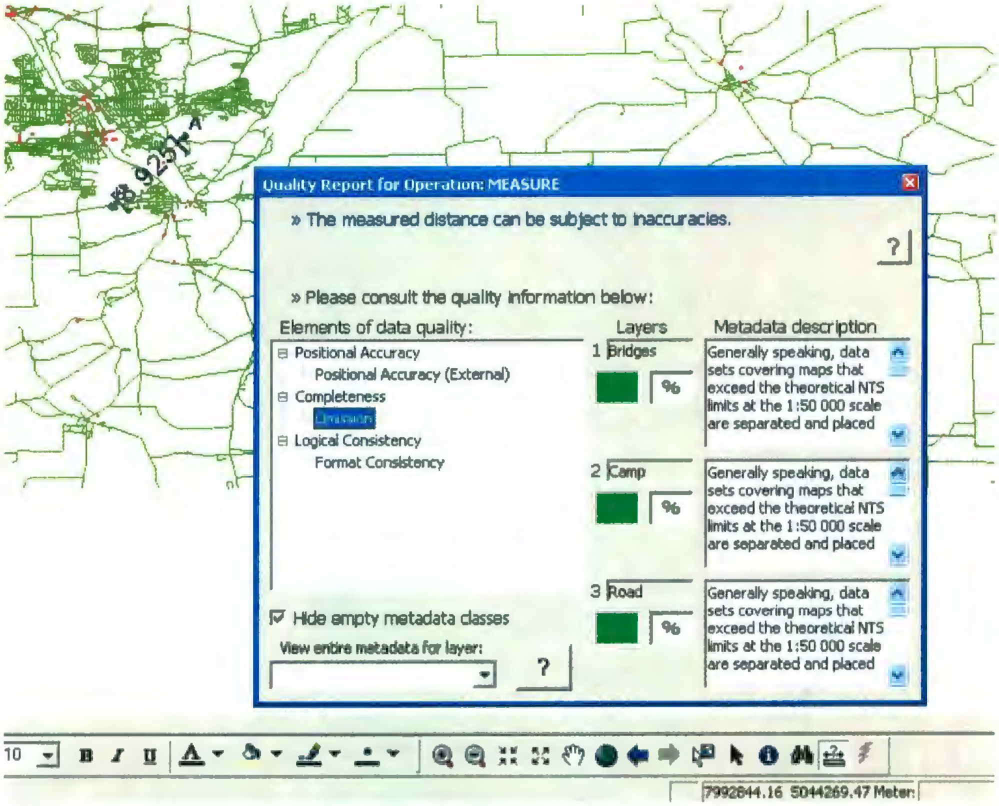


Figure 4-5: Quality report with no omission in the data ('Measure' operation)

The interface includes an educational section (or "help" section) to describe the technical terms used in the metadata documents (Figure 4-6), as clarifying metadata's technical terms was recommended by Hunter *et al.* (2005) and Boin and Hunter (2006). A Graphical User Interface (GUI) for describing the applicable quality elements is displayed. Based on the type of the operation, the "help" section also includes information for describing the effects of each of the quality elements on individual operations (Section 4.2.3). The help section for the MEASURE operation, for instance, describes the effects of positional accuracy, completeness and format consistency.

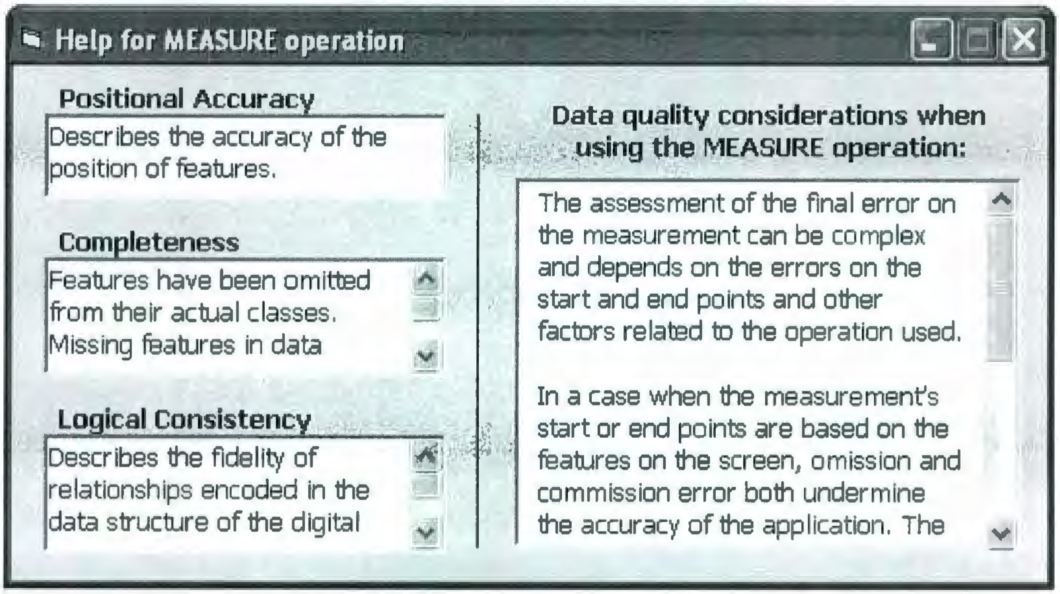


Figure 4-6: Help information for the prototype. This GUI describes the quality elements for the MEASURE operation and explains the effects of each element on the accuracy of the outcome

4.3 Testing

4.3.1 Test configurations

In order to evaluate the framework presented in this thesis, the prototype was tested among a number of GIS users. A usage test scenario was developed and 25 GIS users were asked to use the modified version of the GIS operation after having used the original operation. The configuration of the test and each of the components involved is described in the following.

1. GIS operation: The modified MEASURE operation from ESRI ArcMap was used for the test. This operation is a common operation in GIS software and mapping tools (called "measure length" in IDRISI Andes and "ruler" in Google Earth). The modified MEASURE operation was mounted onto the ArcMap interface for the test.

2. Spatial data and metadata: The data used for the prototype were obtained from the Canadian Geogratis website which provides free access to Canadian digital topographic maps[8]. The data, at 1:50,000 scale, represent an area south-east of the city of Sherbrooke located in Québec, Canada (Figure 4-7). The data include features such as watercourses, urban areas, railways, roads, vegetation, and relief. Three feature classes were included for the test: camping locations, bridges and roads (Figure 4-7).

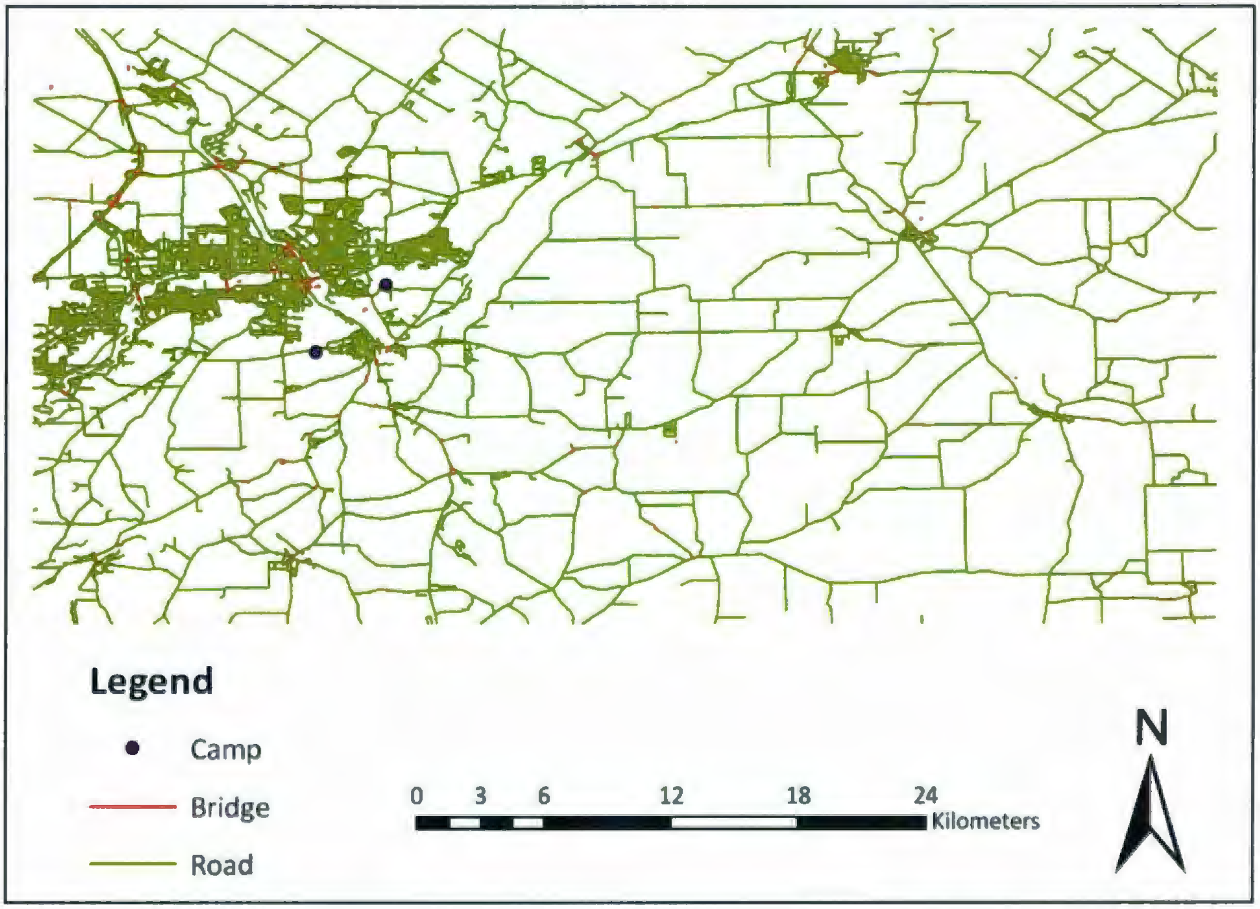


Figure 4-7: The camping locations, bridges and roads layers displayed in ArcMap

[8] http://www.geogratis.ca/

The North American profile of the ISO-19115 standard was used as schema for documenting the metadata. The metadata provide information at the datasets level (see Chapter 2) and thus a single XML file has been included for the datasets used[9].

4.3.2 Questionnaire

To evaluate the usefulness of the approach, the prototype was tested by a number of GIS users. A questionnaire was developed (presented in Appendix D) asking the users to use the GIS software under three scenarios (tests A, B and C). In each scenario, the users had to measure two Euclidean distances between two given points using the *Measure* tool and report the distances with the input accuracy associated with each measurement. The sequence of the test scenarios (A, B and then C) was designed to have the user go through different methods for accessing data quality information that are incrementally easier, requiring less GIS technical abilities. This sequence was important to follow as it allowed users to consult the metadata at the test A without needing guidance from the questionnaire.

Questions such as: "Q.2) Have you had formal education in GIS (e.g., in courses outside or within the university)?" were asked at the beginning of the questionnaire to indicate user backgrounds. A "Yes" answer would indicate that such user may have been introduced to the theory of GIS and error in spatial data. A "No" answer indicates that the user is probably a non-expert or did not receive a formal theoretical training in GIS that may have introduced them to the issues related to data quality.

[9] The data package includes 68 feature class datasets stored in ESRI's Shapefile format (*.shp).

The three data layers described in Section 4.3.1 were preloaded into the ArcMap interface and the settings were saved. The users could then simply locate the required features on the maps and perform the required operation under each scenario:

Test A. Users were asked to measure the distance (between two camp sites and then between a camp site and its nearest bridge) and report it together with its related accuracy without any indication of how to consult the metadata.

Test B. Similar to test A, except that users were asked to consult the metadata (using, for example, ArcCatalog's Metadata tool) prior to stating the error.

Test C. The "Measure" operation available from ArcMap was replaced with the modified, quality-aware one, displaying the quality report when using the operation.

Once the three tests were completed, users were asked to: a) state which approach best improved their awareness of data quality, and b) comment on their choice and explain how the preferred approach helped them.

4.3.3 Survey results

Over the course of one month (25 Sep 2008 to 28 Oct 2008), 25 people answered the questionnaire. The participants were mainly university students. The questionnaires and oral instructions were given to the participants to perform the required tasks and complete the questionnaire sheets. Users took an average of 15 minutes to complete the questionnaire. The tests were carried out in the Department of Geography at the Memorial University of Newfoundland. The results were as follows:

Q.1) How often do you use GIS software or other mapping tools (e.g., ArcGIS and Google Earth)?

Daily	Weekly	Monthly	Yearly
28%	48%	16%	8%

Q.2) Have you had formal education in GIS (e.g., in courses outside or within the university)?

Yes	No
92%	8%

Q.3) How familiar are you with any of the distance measurement tools in GIS software? (e.g. in IDRISI, in ArcGIS and in Google Earth).

I use it frequently	I have used it	I have never used it
20%	72%	8%

Q.4) Imagine a scenario where you are required to measure the distance between two objects in ArcGIS. In tests A, B and C presented in following, you will be asked to perform a simple distance measurement using the MEASURE tool . This tool measures the straight (Euclidean) distance between objects on the maps. In the following tests you will be asked to measure the distances between a) two camp sites and b) a camp site and a bridge.

(Please use <u>meters</u> for reporting all measurements).

How accurate do you think were the distances that you measured approximately?

	I don't know	Responded with a value	No answer
Test A	44%	56%	
Test B	40%	60%	
Test C	16%	76%	8%

Q.5) Information stored in the metadata, among other things, tells us about the accuracy of the spatial data we use. In the above tests, this information could be consulted by either browsing through the metadata information (if possible, Test B) or by an interface that is displayed once the operation is performed (Test C). Which test was better able to make you aware of issues regarding the quality of the data?

Test B	Test C
0%	100%

Q.6) Comparing Tests C with Test B (where you used ArcCatalog for metadata), please comment on how Test C affected your understanding and awareness of the quality of the data you used. You may also include a general explanation for your answer in Q.5.

The reasons stated for choosing Test C can be grouped as the following:

a) Providing awareness of quality issues for the operation that was otherwise missing: 60%

E.g.:

- "Test C made me much more aware of accuracy issues."
- "Test C forced me to become aware of the quality of the data by automatically displaying it."
- "It made me more aware of the accuracy & quality issue so that I don't take the distance measured for granted."
- "The display used brought the data quality issues to my attention. Without it, I might not have considered the errors associated with the data."

b) Having quality information more easily and quickly accessible: 56%

E.g.:

- "Test C is much easier to operate."
- "The data was more quickly available."
- "Test C made it much easier to access the metadata."
- "In Test C, the information appeared in on the screen, which made it very easy and there was an easy to follow explanation."

c) Providing metadata that was easier to understand by providing help information. Being more understandable and more user-friendly: 52%

E.g.:

- "Test C had the information displayed there on the screen in a user-friendly interface."
- "The descriptions of Positional Accuracy made it easier to understand the quality of the data. This is useful in report writing to demonstrate the conditionality of the data. The logical consistency descriptions were easy to understand."
- "Use of color and symbols in the tool in Test C made me more aware that there could be accuracy issues related to the measurements."

d) Bringing quality information to users' attention among all the data provided by the metadata: 20%

E.g.:

- "In Test B, I wasn't sure where to look for the exact metadata needed [...]."
- "Test C provided an easier way to view the accuracies associated with each data layers. It was much easier to understand the quality of the data being used without having to search for it."

e) Useful as it does not require experience among users prior to consulting the metadata: 20%

E.g.:

- "Test C – much better, easier to access the metadata. Particularly useful for GIS users who lack experience I think."
- "I think it would make data quality awareness much better, especially in cases where the user is not aware of the metadata in arc catalog."

4.4 Results discussion

The results showed that all users found their awareness of data quality improved by using the prototype. Improved awareness of problems in spatial data and the existence of data quality constituted the primary reason for the users to prefer using this tool over the other available method (i.e., the Metadata tool in ArcCatalog). After studying user comments, it was possible to observe that a basic warning of quality issues was a helpful reminder that made users more careful about the obtained results. Others stated that they were aware of issues in data quality but failed to consult such information as they were not appropriately provided with indications to the appropriate metadata. As such, the second reason mentioned by the users was the improved accessibility of metadata with the prototype. The efficiency in the speed and accessibility of quality information was an issue that the majority of the users found improved over the current tool. The accompanying help information and the data visualization techniques used on the interface were also cited to be helpful by the majority of the users.

A limitation of the test lies in the distribution of the participants in the test. The majority of the participants were students (both graduate and undergraduate) from the Department of Geography. This can explain the 92% response to question 2 regarding the users' technical education in GIS. To be able to accurately analyse the differences between the two groups, a more extensive survey would have to be done with a larger group showing a better balance between people having expertise in GIS and others that don't. Some recent studies also indicated that domain expertise is, at least in some context, more important than GIS expertise when it comes to assessing the risks related to the use of some geospatial data (Roth, 2009). Furthermore, more detailed analysis of user groups may be considered to adjust for a better distribution of user participants in future tests.

5. Discussion and Conclusions

5.1 Discussion

This thesis provided a partial solution to a problem that has been raised for a certain time in the scientific community and for which a number of partial solutions have been suggested. Although this work does not claim to solve all of the issues related to spatial data quality communication, it did innovate by bringing a framework that allows the communication of data quality information contextually at the time GIS operations are used. The approach was designed to support most of the typical use of a typical commercial GIS and could be easily implemented for a number of operations into a GIS. It only communicates the data quality documented in the dataset's metadata but does not takes into consideration errors from the outcome of the operation (input error and process error) nor does it support user error. Modeling the error in the outcome of operations could be a research project in itself as each operation had different ways, more or less complex, to propagate the uncertainty.

The test presented in Chapter 4 was performed among 25 users on one of two GIS operations for which a prototype was developed (the "Measure" operation). The results from the survey, combined to the feedback from the participants, indicating that the test achieved its purpose of improving user awareness of data quality issues. However, as mentioned in Section 4.4, the majority of participants in the test had formal education in GIS. The test was hence biased towards this type of users and it may be interesting in the future to extend this test to other types of profiles and hence be able to divide the users into different groups of various expertises. The technical terms used on the prototype's interface were primarily designed to be understood by users with some formal knowledge about GIS and data quality issues and could potentially confuse users without formal education in GIS (Boin and Hunter, 2006). A help section has been provided to users to offer some support to understand these technical terms. Other components of the prototype may similarly be examined including the visual aids and the estimated value of error in the outcome. As a result there may be a requirement to modify the prototype for such users or create another version of the prototype specifically for such users.

The framework presented in Chapter 3 was intended to set a general method for modifying GIS operations in order to automatically retrieve selected data quality information from the metadata and display them into the end-user interface. This framework introduced a mechanism by which the quality information that was associated with a GIS operation was simultaneously communicated to the user. Two operations were modified in this work ("Measure" and "Select"), both frequently used by GIS users and arguably straightforward. Modifying more complex operations using this approach can potentially pose challenges and requires invoking techniques such as operation-chaining discussed in Section 3.4.2 (page 43). Within the framework of their current configurations, some limitations for modifying more complex operations can already be foreseen. For example, if datasets are accessed and used by an operation at multiple points within the procedure, data quality for each of the input datasets needs to be accordingly communicated to the user. For this reason the procedures of GIS operations need to be modified at every step that involves accessing spatial data. Moreover, more complex GIS operations and computational models require a method for estimating the compound error generated from multiple operations. Subsequent operations in such models would use the outcome from preceding operations to produce results (Ganguly *et al.*, 2005).

The challenge would be to devise a mechanism to, first, estimate the quality of the outcome from such operations and, second, invoke a mechanism to communicate data quality in every step of the process where spatial data are used.

Therefore, the outcome from this work can be discussed within the scope and limitations of its results. First, the lack of implementation of larger, more complex GIS operations and processes (i.e., operation-chains), and second, addressing the requirements of users without formal education when provided with a report of the associated quality information.

5.2 Conclusions

This thesis proposed a new way to communicate contextual spatial data quality information when executing GIS operations to improve users' awareness and understanding of data quality issues related to the use of GIS. The research hypothesis was that "GIS users' awareness of spatial data quality can be improved by modifying GIS software to link GIS operations with spatial metadata". This work was able to confirm this hypothesis by providing a method that modifies the way GIS software normally operate, allowing a stronger linkage of GIS operations with spatial metadata. Within the scope of its application, the general objective of improving the "the communication of spatial data quality information by providing contextual quality information to GIS users dynamically when using selected GIS operations" was subsequently achieved (Chapter 1, page 7).

Chapter 3 presented the realization of the first objective of this research: "a conceptual framework for linking GIS operations with spatial data quality information available from metadata" (page 7). Two main aspects were considered for establishing this connection: a) the capability to link individual GIS operations with applicable quality elements, and b) accompanying the applicable quality elements with information for describing each quality element's effect on the GIS operation, and thus enabling an estimation of uncertainty in the GIS output.

The second objective was to develop "a framework for applying the conceptual framework into an existing commercial GIS software" (page 7). This objective was carried out in Chapter 4, where the developed concepts were translated into software codes and components implemented into ArcGIS 9.1. This has been presented in Sections 4.1 and 4.2 of Chapter 4, where Operation Quality Metadata (OQM) were created and the required modifications to GIS operations were implemented, expanding operation codes. The approach presented in this thesis is generic but the implementation has been adapted to work under the ArcGIS 9.1 environment.

The third and final objective was to assess "the effectiveness of the prototype with users and analyzing the results" (page 7). This was presented in the second part of the Chapter 4 (Sections 4.3 and 4.4). A test scenario was created simulating a real-world task. 25 users were asked to perform the task using two different versions of the operation; a) the standard metadata tool in ArcGIS, and, b) the modified version (prototype). The analysis of user answers and feedbacks revealed that using the modified version improved the user's awareness of spatial data quality.

In summary, this research presented an alternative, *operation-based*, approach to the communication of spatial data quality in GIS. This alternative, however, can be largely seen as complementary to previous approaches, and thus extends earlier concepts developed in the field. These include earlier works describing the effects of error on GIS operations and methods for communicating spatial data quality (for instance, Giordano *et al.*, 1994; Bédard *et al.*, 2004; Devillers *et al.*, 2005). As such, this approach was able to build on earlier advances for quality communication in GIS and realize its objective of improving awareness of data quality when using GIS operations. Later, at the implementation stage, the operation-based dimension of this work required a number of modifications to GIS operations, which enabled the communication of applicable data quality information.

The conceptual framework that was used to design the prototype was developed using earlier works such as Giordano *et al.* (1994). Giordano *et al.* (1994) mention the need for a link to connect quality information to GIS operations. Later, Duckham and McCreadie (2002) and Bédard *et al.* (2004) suggested improving the usability of quality information in various GIS functionalities. Using these works, the relation between GIS operations and data quality was studied in more detail, enabling the establishment of a conceptual framework that linked the two components together. Once conceptualized, the implementation stage used works such as Agumya and Hunter (2002), Devillers *et al.* (2005) and Huth *et al.* (2009) to bring the conceptual framework into an effective implementation framework. At the testing stage of this research, a practical implementation of the concepts with commercial GIS software, and the involved components (i.e., GIS operations, spatial data and metadata), suggested a significant improvement in users' awareness of data quality information.

Due to its conceptual nature, this method could be adapted to work under different platforms. Future research could use the same framework (for instance, extending functionality to cover more GIS operations or implementing the framework into other software packages) with minimal alterations required.

As mentioned in Chapter 2, this work is closely related to a number of other developments that try to improve the usability of spatial data quality information. This thesis mainly focused on the subject of communication of data quality in GIS, and presented an approach for configuring GIS operations to deliver contextual data quality information at the time of the execution of GIS operations. Nevertheless, in the course of this project and through the study of the answers from the questionnaire (see Section 4.4), a number of areas that could deserve more attention have been identified. The management of metadata in terms of, a) locating the related metadata resources, and b) finding the information within the metadata for the features used in the operation, remains an important problem to be addressed. In this regard, the issue of "ease of accessibility" ranked second among the concerns cited by the users in the questionnaire.

As it has been frequently argued, the separation of data and metadata resources remains a problem in GIS (Blott and Včkovski, 1995; Beard, 1997; Duckham, 2002; Qiu and Hunter, 2002; Najar, 2006). This can cause users to have difficulties in locating the metadata for resources for spatial data and, if located, difficulty in locating the relevant quality information for their usage. This problem is more evident where the varying levels of detail in the description of the metadata (or *granularity*, for example, at the dataset level vs. the primitive level; Devillers, 2004) causes inconsistencies in the mapping of metadata resources for the system. Approaches such as the QIMM (Quality Information Management Model; Devillers *et al.*, 2005) can be incorporated to aid with the management of metadata.

This work has presented an approach for implementing links between GIS operations and spatial data quality for two GIS operations: 'Measure' and 'Select'. More operations could have been tested in the context of this thesis but the two operations were considered sufficient to do a validation of the method as a proof of concept. Also, it is believed that the implementation of such an approach should not necessarily be done for all of the operations available in a GIS but for a sample of them (e.g., the most commonly used ones). A problem faced in this project was the inability to access and modify GIS operations source codes. In the case of ArcGIS operations that were used in this work (as well as other commercial software such as IDRISI and Geomedia), *entry points* within Application Programming Interface (API) for most operations remain minimal. In order to cover a larger number of GIS operations with this concept, GIS software should include methods for interaction with GIS operation interfaces. This can include an event that would declare the termination of any GIS operation that could then be followed by the activation of the extension. By gaining access to such entry points in GIS operations, the framework of this research can be expanded to cover a larger number of GIS operations.

This work, therefore, presented a partial solution to the problem of communicating spatial data quality information to GIS users, and suggests to be more efficient than existing approaches currently used in GIS software for the communication of data quality and metadata (i.e., ArcCatalog). This approach can also be used together with other approaches suggested in the literature for data quality communication. User awareness of elements of spatial data quality was improved, enabling users to make more informed decisions when using spatial data. Nevertheless, in order to further improve the usability of quality information, this work can benefit from incorporating the outcome from research such as improving data quality management and developing error propagation models (Devillers *et al.*, 2005; Heuvelink *et al.*, 2006).

5.3 Future Work

With regard to the limitations mentioned throughout this work, future research could be pursued in the following areas:

- **Customization of the display of quality elements:** The prototype developed communicated spatial data quality based on the applicability of specific quality elements to the operation. For this reason, the OQM was programmed to contain a predefined list of applicable elements for each operation. Users may require adding emphasis to certain aspects of the quality for their applications, for instance, on the temporal accuracy, or the completeness of data. Users may therefore be provided with the capability to define profiles for data quality information similar to Duckham and McCreadie (2002), where users individually define their settings. In such a profile, users may be given the ability to add weights to each quality element, and thus increase or decrease the importance for their particular applications (Devillers, 2004).

- **Incorporating a metadata management system:** The metadata used for this work are raw, as provided by the metadata provider. The prototype was programmed to make use of this specific format. As discussed in the previous section, the method used for embedding the metadata (in this case, XML metadata files) and the tools available in GIS software for making use of these resources (e.g., ArcCatalog's Metadata Tool) are currently limited. As such, the incorporation of a metadata management system (such as Devillers *et al.*, 2005) would aid in better making use of the metadata resources for accessing, querying and manipulating purposes.

- **Covering specific issues in spatial data and data quality:** This work used spatial data from a reliable spatial data provider with relatively high-quality metadata. Certain issues in spatial data and data quality require enhancements in the proposed methods of this work to support a wider range of spatial data types and metadata. These include:

    1. This work assumed a homogenous error in the dataset level where quality information was given. However, spatial datasets with spatially- and temporally-varying data quality would require approaches that would address variations in particular zones or temporal extents. Also, in certain datasets, there is a dependency between attribute errors, or there is a dependency between attribute error and geometry error (e.g., the value of a house depends on its area), leading to varying errors. Future research should similarly take into consideration these concerns.

    2. Issues of fuzziness (e.g., marsh) and vagueness (e.g., some natural objects) in spatial data are not supported by this approach as they are not yet supported by commercial GIS. In order for users to comprehend other known dimensions of uncertainty, and minimize making undesired decisions, information about such uncertainty in spatial data could be communicated.

3. Spatio-temporal databases are used by certain users and require a system that would address the complexity involved in such databases. The current approach does not deal with databases managing temporal features.

- **Expanding the number of operations covered by the approach:** The prototype developed here requires access to two parameters to operate: operation type, and data used by the operation. Two operations from two Albrecht (1996) groups were addressed. By accessing the code of more GIS operations, a global mechanism could be set up in GIS to input these parameters and send them to the prototype in order to prepare the quality report. Since this project focused on commercial software, and the source codes for most operations provided by such software are inaccessible, open-source GIS software (such as GRASS GIS) may be used instead for future developments, showcasing the implementation of the concept to similar operations in GIS software.

- **Providing the users with error estimates and expanding the mechanism to GIS applications:** A number of error models exist for certain GIS operations (see Section 3.3). This does not only take into consideration the input error, but also looks at the process error. Once GIS operations are modified by the method presented in this research, the resultant modified operations could be accompanied by models for estimating error in the outcome of the operation. The error outcome from one operation may be used as input error for the next operation in the sequence (i.e., operation-chaining). This subsequently enables error modeling for the entire application (Ganguly *et al.*, 2005). The resultant measures of error could then be used by users to determine the acceptability of their products.

- **Testing the prototype among users without formal education in GIS and testing the long-term effectiveness of the prototype:** The majority of the participants (92%) in the prototype test (Chapter 4) had formal education in GIS. The other group, lacking such education, was represented by a smaller sample (8%). The test was not designed to understand the difference of behaviour between expert and non-expert GIS users but it would be interesting to extend this study to understand the difference in behaviours between these two groups. To do so, a larger number of users with a better representation would have been necessary. The test also studied the short-term effect of the prototype. There is a possibility that in long-term, users will get used to the prototype and ignore the warnings provided (Boin and Hunter, 2009). Future tests could thus be conducted in a longer time-frame and more extensively among users.

- **Enhancing visualization:** MacEachren *et al.* (2005) provide a comprehensive overview of the visualization techniques that can be incorporated for communicating uncertainty. This study could be improved upon by providing an interactive system in which users could select a visualization tool and its parameters for their applications (e.g., warning signs and change in color on the map). This would enable the customization of the visual components (e.g., signs, arrows, lights) to emphasize user-defined aspects of the quality information. This customization may include a single composite warning measure derived from the weights given by the user to the quality elements (refer to the first point in this section). As a result, users may use data quality information in a fashion that is more accurately directed towards their needs.

References

Aalders, H.J.G.L. 1996. Quality metrics for GIS. In *Proceedings of the Seventh International Symposium on Spatial Data Handling*, edited by M.J. Kraak and M. Molenaar. London: Taylor & Francis, pp. 5B.1-5B.10.

Aalders, H.J.G.L. 2002. The registration of quality in a GIS. In *Spatial Data Quality*, edited by W. Shi, P.F. Fisher and M.F. Goodchild. London: Taylor & Francis, pp. 186-199.

Aditya, T., and R.L.G. Lemmens. 2003. Chaining distributed GIS services. In *Prosiding Pertemuan Ilmiah Tahunan XII, Masyarakat Penginderaan Jauh Indonesia*. July 29-30, Bandung, Indonesia, 8 p.

Agumya, A., and G.J. Hunter. 1999. A risk-based approach to assessing the fitness for use of spatial data. *URISA Journal* 11 (1):33-44.

Agumya, A., and G.J. Hunter. 2002. Responding to the consequences of uncertainty in geographical data. *International Journal of Geographical Information Science* 16 (5):405-417.

Aktas, M.S., G. Aydin, G.C. Fox, H. Gadgil, M. Pierce, and A. Sayar. 2005. Information Services for Grid/Web Service Oriented Architecture (SOA) Based Geospatial Applications. Technical Report [accessed 6/1/2009]. Available from http://grids.ucs.indiana.edu/ptliupages/publications /UDDI_Aktas_Final_Fix.pdf

Aktas, M.S., G.C. Fox, and M. Pierce. 2007. Fault tolerant high performance Information Services for dynamic collections of Grid and Web services. *Future Generation Computer Systems* 23 (3):317-337.

Alameh, N. 2003. Chaining geographic information web services. *IEEE Internet Computing* 7 (5):22-29.

Albrecht, J.H. 1996. Universal GIS operations for environmental modeling. In *Proceedings of the 3rd International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modeling CD-ROM*. September 26-30, UC Santa Barbara, Santa Barbara: National Center for Geographical Information and Analysis.

Albrecht, J.H. 1997. Universal analytical GIS operations: A task-oriented systematization of data-structure-independent GIS functionality. In *Geographic Information Research - Transatlantic Perspectives*, edited by M. Craglia and H. Onsrud. London: Taylor & Francis, pp. 577-591.

Anderson, J.T., and M. Stonebraker. 1994. Sequoia 2000 metadata schema for satellite images. *SIGMOD Record* 23 (4):42-48.

Beard, M.K. 1989. User Error: The Neglected Error Component. In *Proceedings of the Ninth International Symposium on Computer-Assisted Cartography, American Congress on Surveying and Mapping*. Falls Church, VA, pp. 808-817.

Beard, M.K. 1997. Representations of data quality. In *Geographic Information Research: Bridging the Atlantic*, edited by M. Craglia and H. Couclelis. London: Taylor & Francis, pp. 280-294.

Beard, M.K., and B.P. Buttenfield. 1999. Detecting and evaluating errors by graphical methods. In *Geographical Information Systems*, edited by P.A. Longley, M.F. Goodchild, D.J. Maguire and D. Rhind, New York: John Wiley & Sons, pp. 219-233.

Beard, M.K., B.P. Buttenfield, and S.B. Clapham. 1991. NCGIA research initiative 7: Visualization of spatial data quality. NCGIA Technical Paper 91-26.

Beard, M.K., and W. Mackaness. 1993. Visual access to data quality in geographic information systems. *Cartographica* 30 (2/3):37-45.

Bédard, Y., R. Devillers, M. Gervais, and R. Jeansoulin. 2004. Towards Multidimensional User Manuals for Geospatial Datasets: Legal Issues and their Considerations into the Design of a Technological Solution. In *Third International Symposium on Spatial Data Quality (ISSDQ)*. April 15-17, Bruck an der Leitha, Austria, pp. 183-195.

Bennett, B. 2001. What is a forest? On the vagueness of certain geographic concepts. *Topoi* 20 (2):189-201.

Bernard, L., U. Einspanier, M. Lutz, and C. Portele. 2003. Interoperability in GI-Service Chains - The way forward. In *7th AGILE Conference on Geographic Information Science*, edited by M. Gould, R. Laurini and S. Coulondre. April 24-26, Lyon, France.

Berry, J.K. 1987. Fundamental operations in computer-assisted map analysis. *International Journal of Geographical Information Systems* 1 (2):119-136.

Blott, S., and A. Včkovski. 1995. Accessing geographical metafiles through a database storage system. In *Advances in Spatial Databases. Fourth International Symposium SSD'95*, edited by M. Egenhofer and J. Herring. Portland, ME: Springer Verlag.

Boin, A.T., and G.J. Hunter. 2006. Do spatial data consumers really understand data quality information? In *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environment Sciences*, edited by M. Caetano and M. Painho. July 5-7, Lisbon, Portugal, pp. 215-224.

Boin, A.T., and G.J. Hunter. 2009. What communicates quality to the spatial data consumer? In *Quality Aspects in Spatial Data Mining,* edited by A. Stein, J. Shi and W. Bijker. Boca Raton, FL: Taylor & Francis, pp. 285-296.

Bose, R., and F. Reitsma. 2005. Advancing Geospatial Data Curation. In *Conference on Ensuring Long-term Preservation and Adding Value to Scientific and Technical Data (PV 2005)*. November 21-23, The Royal Society, Edinburgh.

Burrough, P.A. 1996. Natural objects with intermediate boundaries. In *Geographic Objects with Indeterminate Boundaries*, edited by P.A. Burrough and A.U. Frank. London: Taylor & Francis, pp. 3-28.

Burrough, P.A. 1999. GIS and geostatistics: Essential partners for spatial analysis. In *Proceedings of the International Symposium on Spatial Data Quality (ISSDQ '99)*. Hong Kong, China, pp. 10-20.

Burrough, P.A., and A.U. Frank. 1995. Concepts and paradigms in spatial information: are current geographical information systems truly generic? *International Journal of Geographical Information Systems* 9 (2):101-116.

Burrough, P.A., and R.A. McDonnell. 1998. *Principles of Geographical Information Systems*. Oxford: Oxford University Press.

Burrough, P.A., R. van Rijn, and M. Rikken. 1996. Spatial data quality and error analysis issues: GIS functions and environmental modeling. In *Environmental Modeling and GIS, Proceedings 2nd Conference on Integrating Geographic Information System and Environmental Modeling*, edited by M.F. Goodchild, B.O. Parks and L. Steyaert. Santa Fe, New Mexico: GIS-World Publications, pp. 29-34.

Buttenfield, B.P., and M.K. Beard. 1991. Visualizing the quality of spatial information. In *Auto-Carto 10*. April 2-7, Baltimore, MD, pp. 145-154.

Cheng, T. 2001. Quality Assessment of Model-Oriented Generalization. In *Fourth Workshop ACI on Progress in Automated Map Generalisation*. August 2-4, Beijing, China.

Chrisman, N.R. 1983. The Role of Quality Information in the Long Term Functioning of a Geographical Information System. In *Proceedings of International Symposium on Automated Cartography (Auto-Carto 6)*. October 16-21, Ottawa, Canada, pp. 303-312.

Chrisman, N.R. 1994. Metadata required to determine the fitness of spatial data for use in environmental analysis. In *Environmental Information Management and Analysis: Ecosystems to Global Scale*, edited by W. Michener, J. Brunt and S. Stafford. Philadelphia: Taylor & Francis, pp. 177-190.

Chrisman, N.R. 1999. A transformational approach to GIS operations. *International Journal of Geographical Information Science* 13 (7):617-637.

Chrisman, N.R. 2006. Development in the treatment of spatial data quality. In *Fundamentals of Spatial Data Quality*, edited by R. Devillers and R. Jeansoulin. London: ISTE, pp. 21-29.

Clark Labs. 2006. IDRISI 15.0 User's Guide [accessed 6/1/2009]. Available from http://www.idrisi.com/products/loader.cfm?url=/commonspot/security/getfile.cfm&pageid=2417.

Comber, A.J., P.F. Fisher, F. Harvey, M. Gahegan, and Wadsworth R.A. 2006. Using metadata to link uncertainty and data quality assessments. In *12th International Symposium on Spatial Data Handling (SDH06)*, edited by A. Riedl, W. Kainz and G. Elmes. July 12-14, Vienna, Austria: Springer Berlin, pp. 279-292.

Comber, A.J., P.F. Fisher, F. Harvey, and R.A. Wadsworth. 2007. Land Cover: to standardise or not to standardise? Comment on 'Evolving standards in land cover characterization' by Herold *et al. Journal of Land Use Science* 2 (4):287-291.

Couclelis, H. 1992. Two Perspectives on Data Quality for quality monitoring. NCGIA Technical Technical Report 92-12 [accessed 6/1/2009]. Available from http://www.ncgia.ucsb.edu/Publications/Tech_Reports/92/92-12.PDF.

Couclelis, H. 2003. The certainty of uncertainty: GIS and the limits of geographic knowledge. *Transactions in GIS* 7 (2):165-175.

Crosetto, M., and S. Tarantola. 2001. Uncertainty and sensitivity analysis: Tools for GIS-based model implementation. *International Journal of Geographical Information Science* 15 (5):415-437.

Curry, M.R. 1998. *Digital Places: Living with Geographic Information Technologies*. New York: Routledge.

Dangermond, J. 1983. A classification of software components commonly used in geographic information systems. In *Proceedings of the US/Australia Workshop on the Design and Implementation of Computer-based Geographic Information Systems*, edited by D. Peuquet and J. O'Callaghan. Amherst, New York: IGU Commission on Geographic Data Sensing and Processing.

De Bruin, S., A. Bregt, and M. van de Ven. 2001. Assessing fitness for use: The expected value of spatial data sets. *International Journal of Geographical Information Science* 15 (5):457-471.

Devillers, R. 2004. Conception d'un système multidimensionnel d'information sur la qualité des données géospatiales. PhD Thesis, Département des sciences géomatiques, Université Laval, Canada, & Institut Francilien des Géosciences, Université de Marne-la-Vallée, France.

Devillers, R., Y. Bédard, and R. Jeansoulin. 2005. Multidimensional management of geospatial data quality information for its dynamic use within GIS. *Photogrammetric Engineering and Remote Sensing* 71 (2):205-215.

Devillers, R., Y. Bédard, R. Jeansoulin, and B. Moulin. 2007. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *International Journal of Geographical Information Science* 21 (3):261-282.

Devillers, R., M. Gervais, Y. Bédard, and R. Jeansoulin. 2002. Spatial data quality: From metadata to quality indicators and contextual end-user manual. In *Proceedings of the OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management*. March 21-22, Istanbul, Turkey.

Devillers, R., and R. Jeansoulin. 2006. Spatial Data Quality: Concepts. In *Fundamentals of Spatial Data Quality*, edited by R. Devillers and R. Jeansoulin. London: ISTE, pp.31-42.

Drecki, I. 2002. Visualisation of uncertainty in geographic data. In *Spatial Data Quality*, edited by W. Shi, P.F. Fisher and M.F. Goodchild. London: Taylor & Francis, pp. 140–159.

Duckham, M. 1999. Implementing an object-oriented error-sensitive GIS. In *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*, edited by K. Lowell and A. Jaton. Michigan: Ann Arbor Press, pp. 209-215.

Duckham, M. 2000. Error-Sensitive GIS Development: Technology and Research Themes. In *Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.* July 12-14, Amsterdam, The Netherlands, pp. 183-190.

Duckham, M. 2002. A user-oriented perspective of error-sensitive GIS development. *Transactions in GIS* 6 (2):179-193.

Duckham, M., and J. Drummond. 1999. Implementing an object-oriented approach to data quality. In *Integrating Information Infrastructures With GI Technology*, edited by B. Gittings. London: Taylor & Francis, pp. 53-64.

Duckham, M., K. Mason, J. Stell, and M. Worboys. 2001. A formal approach to imperfection in geographic information. *Computers, Environment and Urban Systems* 25 (1):89-103.

Duckham, M., and J.E. McCreadie. 2002. Error-aware GIS Development. In *Spatial Data Quality*, edited by W. Shi, P.F. Fisher and M.F. Goodchild. London: Taylor & Francis, pp. 63-75.

Egenhofer, M.J., and A.U. Frank. 1992. Object-oriented modeling for GIS. *URISA Journal* 4 (2):3-19.

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

Einspanier, U., M. Lutz, K. Senkler, I. Simonis, and A. Sliwinski. 2003. Toward a process model for GI. service composition. In *Proceedings of Muenster GI-Days 2003. IfGI Prints Band 18, "Geodaten- und Geodienste-Infrastrukturen - Von der Forschung zur praktischen Anwendung".* June 26-27, Münster, Germany, pp. 31-46.

Elsha Thrall, S., and G.I. Thrall. 1999. Desktop GIS software. In *Geographic Information Systems*, edited by F. Longley, M.F. Goodchild, D.J. Maguire and D. Rhind. New York: John Wiley & Sons Inc, pp. 331-345.

ESRI. 2007. *Array of Geoprocessing Options in ArcGIS.* Environmental Systems Research Institute 2004 [accessed 6/1/2009]. Available from http://www.esri.com/news/arcuser/0104/geoprocess.html.

FGDC. 1994. Content standards for digital Geospatial Metadata. Washington, D.C.: Federal Geographic Data Committee.

FGDC. 1998. Content standards for digital Geospatial Metadata. Washington, D.C.: Federal Geographic Data Committee.

Fischer, M.M., and P. Nijkamp. 1992. Geographic information systems and spatial analysis. *The Annals of Regional Science* 26 (1):3-17.

Fisher, P.F. 1993. Algorithm and implementation uncertainty in viewshed analysis. *International Journal of Geographical Information Systems* 7 (4):331-347.

Fisher, P F. 1993. Visualizing uncertainty in soil maps by animation. *Cartographica* 30 (2-3):20-27.

Fisher, P.F. 1994. Hearing the reliability in classified remotely sensed images. *Cartography and Geographic Information Systems* 21 (1):31-36.

Fisher, P.F. 1998. Is GIS hidebound by the legacy of cartography? *Cartographic Journal* 35 (1):5-9.

Fisher, P.F. 1999. Models of uncertainty in spatial data. In *Geographical Information Systems: Principles, Techniques, Management and Applications*, edited by P. Longley, M.F. Goodchild, D. Maguire and D. Rhind. New York: John Wiley and Sons, pp. 191-205.

Fisher, P.F., A.J. Comber, and R.A. Wadsworth. 2006. Approaches to uncertainty in spatial data. In *Fundamentals of Spatial Data Quality*, edited by R. Devillers and R. Jeansoulin. London: ISTE, pp. 43-59.

Foresman, T.W., H.V. Wiggins, and D.L. Porter. 1996. Metadata myth: Misunderstanding the implications of federal metadata standards. In *Proceedings of the First IEEE Metadata Conference*. April 16-18, Silver Spring, MD.

Frank, A.U. 2007. Data quality ontology: An ontology for imperfect knowledge. In *Spatial Information Theory 8th International Conference COSIT 2007*, edited by S. Winter, M. Duckham, L. Kulik and B. Kuipers. Melbourne, Australia: Springer, pp. 406-420.

Ganguly, A.R., V. Protopopescu, and A. Sorokine. 2005. A bottom-up strategy for uncertainty quantification in complex geo-computational models. In *GeoComputation 2005: 8th International Conference on GeoComputation*. August 1-3, Ann Arbor, MI.

Gervais, M. 2006. On the Importance of External Data Quality in Civil Law. In *Fundamentals of Spatial Data Quality*, edited by R. Devillers and R. Jeansoulin. London: ISTE, pp. 283-300.

Giordano, A., H. Veregin, E. Borak, and D. Lanter. 1994. A conceptual model of GIS-based spatial analysis. *Cartographica* 31 (4):44-51.

Gomez Benitez, F.J. 2000. User interface for the integration of GIS component. Master's Thesis, ITC, The Netherlands.

Goodchild, M.F. 1987. Towards an Enumeration and Classification of GIS Functions. In *Proceedings of the International Geographic Information Systems Symposium: The Research Agenda II*. November 15-18, Arlington, VA: National Aeronautics and Space Administration, Washington, DC, pp. 67-78.

Goodchild, M.F. 1995. Sharing imperfect data. In *Sharing geographic information*, edited by H. Onsrud and G. Rushton. New Jersey: Rutgers, pp. 413-425.

Goodchild, M.F. 1998. Uncertainty: The Achilles heel of GIS? *Geo Info Systems* 8 (11):50-52.

Goodchild, M.F., M.J. Egenhofer, and R. Fegeas. 1997. Interoperating GISs. In *Proceedings of the Report of the Specialist Meeting Held Under the Auspices of the Varenius Project.* December 5-6, Santa Barbara, CA.

Goodchild, M.F., and S. Gopal. 1989. *Accuracy of Spatial Databases*. New York: Taylor & Francis.

Goodchild, M.F., and R.P. Haining. 2003. GIS and spatial data analysis: Converging perspectives. *Papers in Regional Science* 83 (1):363-385.

Goodchild, M.F., and K.K. Kemp. 1990. NCGIA Core Curriculum in GIS. Santa Barbara, CA: National Center for Geographic Information and Analysis.

Goodchild, M.F., M. Yuan, and T.J. Cova. 2007. Towards a general theory of geographic representation in GIS. *International Journal of Geographical Information Science* 21 (3):239-260.

Gould, M. 2005. Geospatial Metadata: Part 2. *GEO:connexion*, [accessed 6/1/2009]. Available from http://www.accessmylibrary.com /coms2/summary_0286-9395607_ITM.

Green, D., and T. Bossomaier. 2002. *Online GIS and Spatial Metadata*. London: Taylor & Francis.

Guptill, S., and J. Morrison. 1995. *Elements of Spatial Data Quality*. Oxford, UK: Elsevier Science Ltd.

Herzog, M. T. 2000. GIS technology and implementation. In *Urban Planning and Development Applications of GIS*, edited by S. Easa and Y. Chan. Reston, VA: American Society of Civil Engineers, pp. 9-31.

Heuvelink, G.B.M. 1998. *Error propagation in environmental modelling with GIS*. London: Taylor & Francis.

Heuvelink, G.B.M., P.A. Burrough, and A. Stein. 1989. Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems* 3 (4):303-322.

Heuvelink, G.B.M., P.A. Burrough, and A. Stein. 2006. Developments in analysis of spatial uncertainty since 1989. In *Classics from IJGIS: 20 Years of the International Journal of Geographical Information Science and Systems*, edited by P.F. Fisher. Boca Raton: CRC, pp 91-95.

Hunter, G.J. 1999. New tools for handling spatial data quality: Moving from academic concepts to practical reality. *URISA Journal* 11 (2):25-34.

Hunter, G.J. 2001. Spatial Data Quality Revisited. In *Proceedings of GeoInfo 2001.* October 4-5, Rio de Janeiro, Brazil, pp. 1-7.

Hunter, G.J., and M.F. Goodchild. 1996. Communicating uncertainty in spatial databases. *Transactions in GIS* 1 (1):13-24.

Hunter, G.J., S. Hope, Z. Sadiq, A.T. Boin, M. Marinelli, A.N. Kealy, M. Duckham, and R.J. Corner. 2005. Next-Generation Research Issues in Spatial Data Quality. In *Proceedings of SSC 2005 Spatial Intelligence, Innovation and Praxis*, edited by C. Bellman, W.E. Cartwright and M. Shortis. September 12-16, Melbourne, Australia: Spatial Sciences Institute, pp. 865-872.

Hunter, G.J., and K.J. Reinke. 2000. Adapting Spatial Databases to Reduce Information Misuse Through Illogical Operations. In *4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2000)*, edited by M.J.P.M. Lemmens and G.B.M. Heuvelink. August 25-26, Amsterdam, The Netherlands: Delft University Press, pp. 313-319.

Hunter, G.J., M. Wachowicz, and K. Bregt. 2003. Understanding spatial data usability. *Data Science Journal* 2 (26):79-89.

Huth, K., N. Mitchell, and G. Schaab. 2009. Judging and Visualising the Quality of Spatio-temporal Data on the Kakamega-Nandi Forest Area in West Kenya. In *Quality Aspects in Spatial Data Mining*, edited by A. Stein, J. Shi and W. Bijker. Boca Raton, FL: Taylor & Francis, pp. 297-313.

ISO/TC-211. 2003. Geographic Information - Metadata (ISO/FDIS 19115). International Organization for Standardization.

ISO/TS. 2007. ISO/TS 19139:2007 Geographic information -- Metadata -- XML schema implementation. International Organization for Standardization.

Juran, J.M., F.M. Gryna, and R.S. Bingham. 1974. *Quality Control Handbook*. New York: McGraw-Hill.

Kazakos, W., M. Briesen, R. Lehfelt, and H.C. Reimers. 2005. Using ISO 19115 metadata and web services to facilitate data access, visualization and processing. In *Proceedings of the 19th International Conference Informatics for Environmental Protection*, edited by J. Hřebíček and J. Ráček. September 7-9, Brno, Czech Republic.

Kemp, K.K. 1996. Fields as a framework for integrating GIS and environmental process models. Part 1: Representing spatial continuity. *Transactions in GIS* 1 (3):219-234.

Keukelaar, J.H.D., O. Ahlqvist, and K. Oukbir. 2000. Representing Uncertainty in Geographical Data. In *Twelfth Annual Colloquium of the Spatial Information Research Centre*, edited by P.A. Whigham. December 10-13, Dunedin, New Zealand: University of Otago, pp.187-194.

Kiehle, C., K. Greve, and C. Heier. 2006. Standardized geoprocessing - Taking spatial data infrastructures one step further. In *Proceedings AGILE 2006*. July 23-28, Visegrád, Hungary.

Kim, T.J. 1999. Metadata for geo-spatial data sharing: a comparative analysis. *The Annals of Regional Science*. 33 (2):171-181

King, J.P. 2002. Modeling Boundaries of Influence among Positional Uncertainty Fields. Master's Thesis, NCGIA, University of Maine.

Klir, G.J., and B. Yuan. 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Englewood Cliffs, NJ: Prentice-Hall.

Krivoruchko, K., and C.A.G. Crawford. 2007. *Assessing the uncertainty resulting from geoprocessing operations* 2003 [accessed 6/1/2009]. Available from http://www.esri.com /library/whitepapers /pdfs/assessing-uncertainty.pdf.

Krygier, J. 1994. Sound and Geographic Visualization. In *Visualization in Modern Cartography,* edited by A.M. MacEachren and D.R.F. Taylor. New York: Elsevier Science Ltd, pp. 149-166.

Kuhn, W. 2005. Geospatial semantics: Why, of what, and how? In *Journal on Data Semantics* edited by S. Spaccapietra and E. Zimányi. Berlin: Springer-Verlag, 3: 1-24.

Kyriakidis, P.C., and M.F. Goodchild. 2006. On the prediction error variance of three common spatial interpolation schemes. *International Journal of Geographical Information Science* 20 (8):823-856.

Lanter, D.P., and H. Veregin. 1992. A research paradigm for propagating error in layer-based GIS. *Photogrammetric Engineering & Remote Sensing* 58 (6):825-833.

Lemmens, R.L.G., and R.A. de By. 2002. Distributed GIS and metadata: methods for the description of interoperable GIS components. In *International workshop on mobile and internet GIS*. August 15-16, Wuhan, China.

Lemmens, R., A. Wytzisk, R. de By, C. Granell, M. Gould, and P. van Oosterom. 2006. Integrating semantic and syntactic descriptions to chain geographic services. *IEEE Internet Computing* 10 (5):42-52.

Lemmens, R., R. de By, M. Gould, A. Wytzisk, C. Granell, and P. van Oosterom. 2007. Enhancing geo-service chaining through deep service descriptions. *Transactions in GIS* 11 (6):849-871.

Levesque, M.A., Y. Bédard, M. Gervais, and R. Devillers. 2007. Towards Managing the Risks of Data Misuse for Spatial Datacubes. In *International Symposium on Spatial Data Quality (ISSDQ 2007)*. June 14-15, Enschede, The Netherlands.

MacEachren, A.M., A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. 2005. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science* 32 (3):139-160.

Mark, D.M. 1993. Toward a theoretical framework for geographic entity types. In *Spatial Information Theory: A Theoretical Basis for GIS,* edited by A.U. Frank and I. Campari. Berlin: Springer-Verlag, pp. 270-283.

McGranaghan, M. 1993. A cartographic view of spatial data quality. *Cartographica* 30 (2-3):8-19.

Meaden, G.J., and J.M. Kapetsky. 1991. *Geographical information systems and remote sensing in inland fisheries and aquaculture* [accessed 6/1/2009]. Available from http://www.fao.org/DOCREP/003/T0446E/T0446E00.htm.

Montello, D.R., M.F. Goodchild, J. Gottsegen, and P. Fohl. 2003. Where's downtown?: Behavioral methods for determining referents of vague spatial queries. *Spatial Cognition and Computation* 3 (2-3):185-204.

Morrison, J.L. 1995. Spatial data quality. In *Elements of Spatial Data Quality*, edited by S.C. Guptill and J.L. Morrison. New York: Elsevier Science, pp. 1-12.

Mostafavi, M.A., G. Edwards, and R. Jeansoulin. 2004. An ontology-based method for quality assessment of spatial data bases. In *ISSDQ'04 Proceedings*. April 15-17, Bruck an der Leitha, Austria.

Mowrer, H.T. 1999. Accuracy (re)assurance: Selling uncertainty assessment to the uncertain. In *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*, edited by K. Lowell and A. Jaton. Michigan: Ann Arbor Press, pp. 3-10.

Najar, C.R. 2006. A model-driven approach to management of integrated metadata-spatial data in the context of spatial data infrastructures. PhD Thesis, Institut fur Geodäsie und Photogrammetrie an der Eidgenössischen Technischen Hochschule, Zürich.

NIST. 1992. Federal Information Processing Standard, Publication 173 (Spatial Data Transfer Standard) National Institute of Standards and Technology. Washington, D.C.: U.S. Government Printing Office.

Olson, J. 2003. *Data Quality: The Accuracy Dimension*. San Francisco: Morgan Kaufmann.

Onsrud, H.J. 1995. The role of law in impeding and facilitating the sharing of geographic information. In *Sharing Geographic Information*, edited by H. J. Onsrud and G. Rushton. New Brunswick, NJ: Center for Urban Policy Research, pp. 292-306.

Openshaw, S. 1989. Learning to live with errors in spatial databases. In *The accuracy of spatial databases*, edited by M.F. Goodchild and S. Gopal. London: Taylor & Francis.

Östman, A. 1996. Quality systems for spatial data. In *Proceedings of the 1996 Joint European Conference and Exhibition on Geographical Information*. March 27-29, Barcelona, Spain Amsterdam: IOS Press, pp. 268-276.

Paradis, J., and K. Beard. 1994. Visualization of spatial data quality for the decision-maker: a data- quality filter. *URISA Journal* 6 (2):25-34.

Peng, Z.R., and M.H. Tsou. 2003. *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, NJ: John Wiley & Sons, Inc.

Qiu, J., and G.J. Hunter. 2002. A GIS with the Capacity for Managing Data Quality Information. In *Spatial Data Quality*, edited by W. Shi, P.F. Fisher and M.F. Goodchild. London: Taylor & Francis, pp. 230-250.

Radwan, M., A. Alvarez, R. Onchaga, and J. Morales. 2003. Designing an Integrated Enterprise Model to support Partnerships in the Geo-Information Industry. In *Map Asia, 2003*. October 13-15, Kuala Lumpur, Malaysia.

Reinke, K.J., S. Jones, and G.J. Hunter. 2006. Implementation of a prototype toolbox for communicating spatial data quality and uncertainty using a wildfire risk example. In *12th International Symposium on Spatial Data Handling*, edited by A. Riedl, W. Kainz and G.A. Elmes. July 10-12, Vienna, Austria: Springer-Verla, pp. 321-337.

Rhind, D.W., and N.P.A. Green. 1988. Design of a geographical information system for a heterogeneous scientific community. *International Journal of Geographical Information Systems* 2 (2):171-189.

Roth, R.E. 2009. The impact of user expertise on geographic risk assessment under uncertain conditions. *Cartography and Geographic Information Science*, 36 (1):29-43.

Salgé, F. 1995. Semantic accuracy. In *Elements of Spatial Data Quality*, edited by S.C. Guptill and J.L. Morrison. New York: Elsevier Science, pp. 139-151.

Schneider, M. 1997. Spatial data types for database systems - Finite resolution geometry for geographic information systems. *Lecture Notes in Computer Science* 1288.

Schuurman, N. 2006. Formalization matters: Critical GIS and ontology research. *Annals of the Association of American Geographers* 96 (4):726-739.

Shi, W. 2002. Introduction to Part II: Methods for Handling Spatial Data Quality. In *Spatial Data Quality*, edited by W. Shi, P.F. Fisher and M.F. Goodchild. London: Taylor & Francis, pp. 102-105.

Shi, W., C.K. Cheung, and C. Zhu. 2003. Modelling error propagation in vector-based buffer analysis. *International Journal of Geographical Information Science* 17 (3):251-271.

Shi, W., P.F. Fisher, and M.F. Goodchild. 2002. *Spatial Data Quality*. London: Taylor & Francis.

Shortridge, A.M., and M.F. Goodchild. 1999. Communicating uncertainty for global datasets. In *Proceedings of International Symposium on Spatial Data Quality*, edited by W. Shi, M.F. Goodchild and P.F. Fisher. August 18-21, Hong Kong, pp. 59-65.

Smith, B., and D.M. Mark. 2003. Do mountains exist? Towards an ontology of landforms. *Environment and Planning B: Planning and Design* 30 (3):411-427.

Timpf, S. 2003. Geographic activity models. In *Foundations of Geographic Information Science*, edited by M. Duckham, M.F. Goodchild and M.F. Worboys. London: Taylor and Francis, pp. 241-254.

Timpf, S., M. Raubal, and W. Kuhn. 1996. Experiences with Metadata. In *7th International Symposium on Spatial Data Handling, SDH'96*. August 12-16, Delft, The Netherlands:

International Geographical Union.

Tomlin, C.D. 1983. Digital Cartographic Modelling Techniques in Environmental Planning. PhD Thesis, School of Forestry and Environmental Studies, Yale University, New Haven, CT.

Tomlin, C.D. 1990. *Geographic Information Systems and Cartographic Modeling*. Englewood Cliffs, NJ: Prentice-Hall.

Tsou, M.H. 2001. A Dynamic Architecture for Distributing Geographic Information Services on the Internet. PhD thesis, Department of Geography, University of Colorado, Boulder, Colorado.

Tsou, M.H. 2002. An operational metadata framework for searching, indexing, and retrieving distributed geographic information services on the Internet. *Lecture Notes in Computer Science* 2478:313-333.

Unwin, D.J. 1995. Geographical information systems and the problem of error and uncertainty. *Progress in Human Geography* 19 (4):549-558.

Van Oort, P. 2006. Spatial Data Quality: From Description to Applications. PhD Thesis, Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, Wageningen, The Netherlands.

Van Oort, P., and A.K. Bregt. 2005. Do users ignore spatial data quality? A decision-theoretic perspective. *Risk Analysis* 25 (6):1599-1610.

Vasseur, B., R. Devillers, and R. Jeansoulin. 2003. Ontological approach of the fitness of use of geospatial dataset. In *Proceedings of 6th AGILE Conference on Geographic Information Science*. April 24-26, Lyon, France.

Veregin, H. 1989. Error modeling for the map overlay operation. In *The accuracy of spatial databases*, edited by M.F. Goodchild and S. Gopal. London: Taylor & Francis, pp. 3-18.

Veregin, H. 1995. Developing and testing of an error propagation model for GIS overlay operations. *International Journal of Geographical Information Systems* 9 (6):595-619.

Veregin, H. 1998. Data quality measurement and assessment. *NCGIA Core Curriculum in Geographic Information Science*, [accessed 9/1/2009]. Available from http://www.ncgia.ucsb.edu/giscc/units/u100/u100.html.

Veregin, H., and P. Hargitai. 1995. An evaluation matrix for geographical data quality. In *Elements of Spatial Data Quality*, edited by S.C. Guptill and J.L. Morrison. New York: Elsevier Science, pp. 167-188.

Visvalingam, M. 1991. Areal units and the linking of data: some conceptual issues. In *Spatial analysis and spatial policy using geographic information systems*, edited by L. Worral. London: Belhaven Press, pp. 12-37.

Wang, R.Y. 1996. Beyond accuracy: What data quality means to data consumers. *Journal of*

Management Information Systems 12 (4):5-33.

Warren, S.D., M.G. Hohmann, K. Auerswald, and H. Mitasova. 2004. An evaluation of methods to determine slope using digital elevation data. *Catena* 58 (3):215-233.

Worboys, M. 1998. Imprecision in finite resolution spatial data. *GeoInformatica* 2 (3):257-279.

Wytzisk, A., and A. Sliwinski. 2004. Quo vadis SDI? In *Proceedings of the 7th AGILE Conference on Geographic Information Science*. April 29 - May 1, Heraklion, Greece.

Yang, Y., D. Gillingwater, and C. Hinde. 2006. Soft computing in geographical Information Systems. In *Frontiers of Geographical Information technology*, edited by R. Sanjay and S. Jayant. New York: Springer, pp. 47-62 .

Yuan, M., and J.H. Albrecht. 1995. Structural analysis of geographic information and GIS operations from a user's perspective. In *Spatial Information Theory: A Theoretical Basis for GIS. International Conference COSIT'95*, edited by A.U. Frank and W. Kuhn. Ellicottville, NY: Springer-Verlag, pp. 107-122.

Zhang, J., and M.F. Goodchild. 2002. *Uncertainty in Geographical Information*. London: Taylor & Francis.

Appendix A: Prototype Code for the Measure Operation

1- Form file (frmMeasure.frm), the interface for the *quality report*.

```
' Comments are marked by (')
' Flag that signals a Drag Drop operation.
Dim indrag As Boolean

' Item that is being dragged.
Public nodX As Node

' Declare string variables for metadata, data quality and quality
elements
Dim MDFilePath, MDText, DQText, Quality_Elements As String

' Declare integer variables for quality report dimensions.
Dim FormHeight, FormWidth As Integer

' Declare integer variables for class positions in the list
Dim DQEPos, ClassListPos, DQCLength, EachLayer As Integer

' Declare temporary variables
Dim DQCPos(15), DQPos(15) As Double
Dim DQCText(15), TextTemp(15) As String
Dim dqctemp(15) As DQC
Dim DQCLassU, DQCLassV, DQCLassD As String

' Declare instances of the DQE class module for each quality element.
Declare string variables for quality elements' texts.
Dim PA As New DQE: Dim DQPAText as String
Dim AA As New DQE: Dim DQAAText as String
Dim LC As New DQE: Dim DQLCText as String
Dim TA As New DQE: Dim DQTAText as String
Dim Co As New DQE: Dim DQCoText as String
Dim Li As New DQE: Dim DQLiText as String
```

```
Private Sub Form_Load()

' Position the quality report to the lower right side of the screen
FrmMeasure.Left = Screen.Width - FrmMeasure.Width
FrmMeasure.Top = Screen.Height - FrmMeasure.Height - 500

' Load pictures for warning signs
For i = 1 To 3
    Picture1(i).Picture = LoadPicture(App.Path & "\warning 3.bmp")
Next

FrmMeasure.Label10.Visible = False
```

```
FrmMeasure.Label12.Visible = False

End Sub
```

```
Private Sub Form_Initialize()

' Store initial form height and width
FormHeight = FrmMeasure.Height
FormWidth = FrmMeasure.Width

' Retrieve the list of applicable quality elements for the Measure
operation from the OQM
Data1.DatabaseName = App.Path + "\OQM.mdb"
Data1.RecordSource = "OQTDDb"
Set OQM = Data1.Recordset
OQM.MoveFirst
OQM.FindFirst ("[Operation_Name] = 'Measure'")
Quality_Elements = OQM![Quality_Elements]

' Locate the metadata file and load the entire metadata into a single
string variable (MDText)
MDFile = FreeFile
If Dir$(App.Path & "\021E05.xml") <> "" Then
    Open App.Path & "\021E05.xml" For Input As MDFile: MDFilePath =
App.Path & "\021E05.xml"
    Else:
        LocateMetadataFile.DialogTitle = "Please locate the metadata
file.."
        LocateMetadataFile.Filter = "Metadata File (*.*)|*.*|AllFiles
(*.*)|*.*"
        LocateMetadataFile.CancelError = True
        On Error Resume Next
        LocateMetadataFile.ShowOpen
        Open LocateMetadataFile.FileName For Input As MDFile
        MDFilePath = LocateMetadataFile.FileName
        If Err Then Exit Sub
End If
MDText = Input$(LOF(MDFile), MDFile)
Close #MDFile

' From the metadata, load the spatial data quality section into
another string variable (DQtext)
DQEPos1 = InStr(1, MDText, "<gmd:DQ_DataQuality>", vbTextCompare)
DQEPos2 = InStr(1, MDText, "</gmd:DQ_DataQuality>", vbTextCompare)
DQLength = DQEPos2 - DQEPos1
DQText00 = Left(MDText, DQEPos2 + Len("<gmd:DQ_DataQuality>"))
DQText = Right(DQText00, DQLength + Len("</gmd:DQ_DataQuality>"))

' Load the preset structure for spatial data quality
```

```
    PA.AddDQC "DQ_AbsoluteExternalPositionalAccuracy", "Positional
Accuracy (External)", "", "", "", "0"
    PA.AddDQC "DQ_GriddedDataPositionalAccuracy", "Gridded Data
Positional Accuracy", "", "", "", "0"
    PA.AddDQC "DQ_RelativeInternalPositionalAccuracy", "Relative
Internal Positional Accuracy", "", "", "", "0"
    AA.AddDQC "DQ_ThematicClassificationCorrecness", "Thematic
Classification Correctness", "", "", "", "0"
    AA.AddDQC "DQ_NonQuantitativeAttributeAccuracy", "Non Quantitative
Attribute Accuracy", "", "", "", "0"
    AA.AddDQC "DQ_QuantitativeAttributeAccuracy", "Quantitative
Attribute Accuracy", "", "", "", "0"
    LC.AddDQC "DQ_ConceptualConsistency", "Conceptual Consistency",
"", "", "", "0"
    LC.AddDQC "DQ_FormatConsistency", "Format Consistency", "", "",
"", "0"
    LC.AddDQC "DQ_TopologicalConsistency", "Topological Consistency",
"", "", "", "0"
    LC.AddDQC "DQ_DomainConsistency", "Domain Consistency", "", "",
"", "0"
    TA.AddDQC "AccuracyOfATimeMeasurement", "Accuracy Of A Time
Measurement", "", "", "", "0"
    TA.AddDQC "DQ_TemporalConsistency", "Temporal Consistency", "",
"", "", "0"
    TA.AddDQC "DQ_TemporalValidity", "Temporal Validity", "", "", "",
"0"
    Co.AddDQC "DQ_CompletenessCommission", "Commission", "", "", "",
"0"
    Co.AddDQC "DQ_CompletenessOmission", "Omission", "", "", "", "0"
    Li.AddDQC "LI_Lineage", "Lineage", "", "", "", "0"

' Depending on the applicability of the quality element, retrieve the
metadata text for each quality element. Also, create the list of
quality elements
QualityElementsList.Nodes.Clear: DQEPos = 0: ClassListPos = 0
If InStr(1, Quality_Elements, "PA", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Positional Accuracy"
    For Each DQCTemp1 In PA.DQCCollection
        If InStr(1, DQText, DQCTemp1.DQClassIdentifier, vbTextCompare)
<> 0 Then
            Set DQCTemp2 =
PA.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCTemp2.DQClassIsAvailable = True
            QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
            ClassListPos = ClassListPos + 1
        End If
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "Co", vbTextCompare) <> 0 Then
```

```
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Completeness"
    For Each DQCTemp1 In Co.DQCCollection
        If InStr(1, DQText, DQCTemp1.DQClassIdentifier, vbTextCompare)
<> 0 Then
            Set DQCTemp2 =
Co.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCTemp2.DQClassIsAvailable = True
            QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
            ClassListPos = ClassListPos + 1
        End If
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "LC", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Logical Consistency"
    For Each DQCTemp1 In LC.DQCCollection
        If InStr(1, DQText, DQCTemp1.DQClassIdentifier, vbTextCompare)
<> 0 Then
            Set DQCTemp2 =
LC.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCTemp2.DQClassIsAvailable = True
            QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
            ClassListPos = ClassListPos + 1
        End If
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "TA", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Temporal Accuracy"
    For Each DQCTemp1 In TA.DQCCollection
        If InStr(1, DQText, DQCTemp1.DQClassIdentifier, vbTextCompare)
<> 0 Then
            Set DQCTemp2 =
TA.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCTemp2.DQClassIsAvailable = True
            QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
            ClassListPos = ClassListPos + 1
        End If
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "AA", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Thematic Accuracy"
    For Each DQCTemp1 In AA.DQCCollection
```

```
        If InStr(1, DQText, DQCTemp1.DQClassIdentifier, vbTextCompare)
<> 0 Then
            Set DQCTemp2 =
AA.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCTemp2.DQClassIsAvailable = True
            QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
            ClassListPos = ClassListPos + 1
        End If
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "Li", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Lineage"
    For Each DQCTemp1 In Li.DQCCollection
        If InStr(1, DQText, DQCTemp1.DQClassIdentifier, vbTextCompare)
<> 0 Then
            Set DQCTemp2 =
Li.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCTemp2.DQClassIsAvailable = True
            QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
            ClassListPos = ClassListPos + 1
        End If
    Next DQCTemp1
    DQEPos = ClassListPos
End If
QualityElementsList.Refresh

' Extract data quality class description, unit and value for
applicable quality elements
If InStr(1, Quality_Elements, "PA", vbTextCompare) <> 0 Then
    ' For each Positional Accuracy quality class, get the text
    For Each DQCTemp1 In PA.DQCCollection
        If DQCTemp1.DQClassIsAvailable = True Then
            Set DQCTemp10 =
PA.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCPos1 = InStr(1, DQText, DQCTemp1.DQClassIdentifier,
vbTextCompare)
            DQCPos2 = InStr(DQCPos1 + 1, DQText,
DQCTemp1.DQClassIdentifier, vbTextCompare)
            DQPAText = Mid(DQText, DQCPos1 + Len(DQCPos1), DQCPos2 -
DQCPos1)

             ' Get the value
            DQCPos4 = InStr(1, DQPAText, "<gco:Integer>",
vbTextCompare)
            DQCPos5 = InStr(DQCPos4 + 1, DQPAText, "</gco:Integer>",
vbTextCompare)
            DQCLength = Len("<gco:Integer>")
            DQCPos6 = DQCPos5 - DQCPos4 - DQCLength
```

```
            DQCTemp10.DQClassValue = Mid(DQPAText, DQCPos4 +
DQCLength, DQCPos6)

             ' Get the unit
            'DQCPos7 = InStr(1, DQPAText, "<gml:identifier
codeSpace="">", vbTextCompare)
            DQCPos7 = InStr(1, DQPAText, "<gml:identifier codeSpace=",
vbTextCompare)
            DQCPos8 = InStr(DQCPos7 + 1, DQPAText,
"</gml:identifier>", vbTextCompare)
            DQCLength = Len("<gml:identifier codeSpace=") + 3
            DQCPos9 = DQCPos8 - DQCPos7 - DQCLength
            DQCTemp10.DQClassUnit = Mid(DQPAText, DQCPos7 + DQCLength,
DQCPos9)
            If InStr(1, DQCLassU, "Meters") <> 0 Then
DQCTemp2.DQClassUnit = "m" ' This vocabulary can be extended

            'Get the description
            DQCPos10 = InStr(1, DQPAText, "<gco:CharacterString>",
vbTextCompare)
            DQCPos11 = InStr(DQCPos10 + 1, DQPAText,
"</gco:CharacterString>", vbTextCompare)
            DQCLength = Len("<gco:CharacterString>")
            DQCPos12 = DQCPos11 - DQCPos10 - DQCLength
            DQCTemp10.DQClassDescription = Mid(DQPAText, DQCPos10 +
DQCLength, DQCPos12)
        End If
    Next DQCTemp1
End If

If InStr(1, Quality_Elements, "Co", vbTextCompare) <> 0 Then
        ' For each Completeness data quality class, get the entire
text
    For Each DQCTemp1 In Co.DQCCollection
        If DQCTemp1.DQClassIsAvailable = True Then
            Set DQCTemp11 =
Co.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCPos1 = InStr(1, DQText, DQCTemp1.DQClassIdentifier,
vbTextCompare)
            DQCPos2 = InStr(DQCPos1 + 1, DQText,
DQCTemp1.DQClassIdentifier, vbTextCompare)
            DQCoText = Mid(DQText, DQCPos1 + Len(DQCPos1), DQCPos2 -
DQCPos1)

             ' Get the value
            DQCPos4 = InStr(1, DQCoText, "<gco:Decimal>",
vbTextCompare)
            DQCPos5 = InStr(DQCPos4 + 1, DQCoText, "</gco:Decimal>",
vbTextCompare)
            DQCLength = Len("<gco:Decimal>")
            DQCPos6 = DQCPos5 - DQCPos4 - DQCLength
```

```
            DQCTemp11.DQClassValue = Mid(DQCoText, DQCPos4 +
DQCLength, DQCPos6)

             ' Get the unit
            'DQCPos7 = InStr(1, DQCoText, "<gml:identifier
codeSpace="">", vbTextCompare)
            DQCPos7 = InStr(1, DQCoText, "<gml:identifier codeSpace=",
vbTextCompare)
            DQCPos8 = InStr(DQCPos7 + 1, DQCoText,
"</gml:identifier>", vbTextCompare)
            DQCLength = Len("<gml:identifier codeSpace=") + 3
            DQCPos9 = DQCPos8 - DQCPos7 - DQCLength
            DQCTemp11.DQClassUnit = Mid(DQCoText, DQCPos7 + DQCLength,
DQCPos9)
            If InStr(1, DQCLassU, "Percentage") <> 0 Then
DQCTemp11.DQClassUnit = "%" ' This vocabulary can be extended

            'Get the description
            DQCPos10 = InStr(1, DQCoText, "<gco:CharacterString>",
vbTextCompare)
            DQCPos11 = InStr(DQCPos10 + 1, DQCoText,
"</gco:CharacterString>", vbTextCompare)
            DQCLength = Len("<gco:CharacterString>")
            DQCPos12 = DQCPos11 - DQCPos10 - DQCLength
            DQCTemp11.DQClassDescription = Mid(DQCoText, DQCPos10 +
DQCLength, DQCPos12)
        End If
    Next DQCTemp1
End If

If InStr(1, Quality_Elements, "LC", vbTextLCmpare) <> 0 Then
        ' For each Logical Consistency data quality class, get the
entire text
    For Each DQCTemp1 In LC.DQCCollection
        If DQCTemp1.DQClassIsAvailable = True Then
            Set DQCTemp12 =
LC.DQCCollection(DQCTemp1.DQClassIdentifier)
            DQCPos1 = InStr(1, DQText, DQCTemp1.DQClassIdentifier,
vbTextLCmpare)
            DQCPos2 = InStr(DQCPos1 + 1, DQText,
DQCTemp1.DQClassIdentifier, vbTextLCmpare)
            DQLCText = Mid(DQText, DQCPos1 + Len(DQCPos1), DQCPos2 -
DQCPos1)

             ' Get the value
            DQCPos4 = InStr(1, DQLCText, "<gco:Decimal>",
vbTextLCmpare)
            DQCPos5 = InStr(DQCPos4 + 1, DQLCText, "</gco:Decimal>",
vbTextLCmpare)
            DQCLength = Len("<gco:Decimal>")
            DQCPos6 = DQCPos5 - DQCPos4 - DQCLength
```

```
            DQCTemp12.DQClassValue = Mid(DQLCText, DQCPos4 +
DQCLength, DQCPos6)

             ' Get the unit
            'DQCPos7 = InStr(1, DQLCText, "<gml:identifier
codeSpace=""">", vbTextLCmpare)
            DQCPos7 = InStr(1, DQLCText, "<gml:identifier codeSpace=",
vbTextLCmpare)
            DQCPos8 = InStr(DQCPos7 + 1, DQLCText,
"</gml:identifier>", vbTextLCmpare)
            DQCLength = Len("<gml:identifier codeSpace=") + 3
            DQCPos9 = DQCPos8 - DQCPos7 - DQCLength
            DQCTemp12.DQClassUnit = Mid(DQLCText, DQCPos7 + DQCLength,
DQCPos9)
            If InStr(1, DQCLassU, "Percentage") <> 0 Then
DQCTemp12.DQClassUnit = "%" ' This vocabulary can be extended

            'Get the description
            DQCPos10 = InStr(1, DQLCText, "<gco:CharacterString>",
vbTextLCmpare)
            DQCPos11 = InStr(DQCPos10 + 1, DQLCText,
"</gco:CharacterString>", vbTextLCmpare)
            DQCLength = Len("<gco:CharacterString>")
            DQCPos12 = DQCPos11 - DQCPos10 - DQCLength
            DQCTemp12.DQClassDescription = Mid(DQLCText, DQCPos10 +
DQCLength, DQCPos12)

        End If
    Next DQCTemp1
End If

' Code for Attribute Accuracy
If InStr(1, Quality_Elements, "AA", vbTextLCmpare) <> 0 Then
End If

' Code for Temporal Accuracy
If InStr(1, Quality_Elements, "TA", vbTextLCmpare) <> 0 Then
End If

' Code for Lineage
If InStr(1, Quality_Elements, "Li", vbTextLCmpare) <> 0 Then
End If

End Sub
```

```
Private Sub AllMdCheck_Click()

' If unchecked, display applicable quality elements and all their
classes. If checked, only display classes for which information is
available
```

```
QualityElementsList.Nodes.Clear: DQEPos = 0: ClassListPos = 0
If InStr(1, Quality_Elements, "PA", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Positional Accuracy"
    For Each DQCTemp1 In PA.DQCCollection
        If AllMdCheck.Value = 1 Then If DQCTemp1.DQClassIsAvailable =
False Then GoTo 100
        QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
        ClassListPos = ClassListPos + 1
100
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "Co", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Completeness"
    For Each DQCTemp1 In Co.DQCCollection
        If AllMdCheck.Value = 1 Then If DQCTemp1.DQClassIsAvailable =
False Then GoTo 200
        QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
        ClassListPos = ClassListPos + 1
200
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "LC", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Logical Consistency"
    For Each DQCTemp1 In LC.DQCCollection
        If AllMdCheck.Value = 1 Then If DQCTemp1.DQClassIsAvailable =
False Then GoTo 300
        QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
        ClassListPos = ClassListPos + 1
300
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "TA", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Temporal Accuracy"
    For Each DQCTemp1 In TA.DQCCollection
        If AllMdCheck.Value = 1 Then If DQCTemp1.DQClassIsAvailable =
False Then GoTo 400
        QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
        ClassListPos = ClassListPos + 1
400
    Next DQCTemp1
    DQEPos = ClassListPos
```

```
End If
If InStr(1, Quality_Elements, "AA", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Thematic Accuracy"
    For Each DQCTemp1 In AA.DQCCollection
        If AllMdCheck.Value = 1 Then If DQCTemp1.DQClassIsAvailable =
False Then GoTo 500
        QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
        ClassListPos = ClassListPos + 1
500
    Next DQCTemp1
    DQEPos = ClassListPos
End If
If InStr(1, Quality_Elements, "Li", vbTextCompare) <> 0 Then
    DQEPos = DQEPos + 1: ClassListPos = ClassListPos + 1
    QualityElementsList.Nodes.Add , , , "Lineage"
    For Each DQCTemp1 In Li.DQCCollection
        If AllMdCheck.Value = 1 Then If DQCTemp1.DQClassIsAvailable =
False Then GoTo 600
        QualityElementsList.Nodes.Add DQEPos, tvwChild, ,
DQCTemp1.DQClassName, 1
        ClassListPos = ClassListPos + 1
600
    Next DQCTemp1
    DQEPos = ClassListPos
End If
QualityElementsList.Refresh

End Sub
```

```
Private Sub MDCombo_Change()

Call DisplayAllMetadata

End Sub
```

```
Private Sub MDCombo_Click()

Call DisplayAllMetadata

End Sub
```

```
Private Sub DisplayAllMetadata()

' Open entire metadata for the layers
Dim XMLopen As String
XMLopen = "explorer " & MDFilePath
```

```
Shell XMLopen

End Sub
```

```
Private Sub Form_Resize()

' Make form resizable only in width
If FrmMeasure.Height <> FormHeight Then FrmMeasure.Height = FormHeight
If FrmMeasure.Width < FormWidth Then FrmMeasure.Width = FormWidth

' Resize metadata descriptions textboxes
For i = 1 To 3
    MDTBText(i).Width = FrmMeasure.Width - 5145
Next

End Sub
```

```
Private Sub HelpBtn1_Click()

' Display the help for the prototype's interface
Load frmMeasureHelp
frmMeasureHelp.Show vbModal
End Sub
```

```
Private Sub HelpBtn2_Click()

' Display the help for the Measure operation
Load InterfaceHelp
InterfaceHelp.Show vbModal
End Sub
```

```
Private Sub QualityElementsList_MouseUp(Button As Integer, Shift As
Integer, x As Single, y As Single)

On Error GoTo 100

Label10.Visible = False
Label12.Visible = False

' Clear the textboxes and load information for the selected class of
quality element
For i = 1 To 3 'NoOfLayers
    FrmMeasure.MDTBValue(i).Text = ""
    FrmMeasure.MDTBUnit(i).Text = ""
    FrmMeasure.MDTBText(i).Text = ""
    FrmMeasure.MDTBValue(i).BackColor = &HFFFFFF
    FrmMeasure.Picture1(i).Visible = False
```

```
Next
If QualityElementsList.SelectedItem.Parent.Text = "Positional
Accuracy" Then
    For Each DQCTemp1 In PA.DQCCollection
        Set DQCTemp5 = PA.DQCCollection(DQCTemp1.DQClassIdentifier)
        If QualityElementsList.SelectedItem.Text =
DQCTemp5.DQClassName And DQCTemp5.DQClassIsAvailable = True Then
            For EachLayer = 1 To 3 'NoOfLayers
                FrmMeasure.MDTBValue(EachLayer).Text =
DQCTemp5.DQClassValue
                FrmMeasure.MDTBText(EachLayer).Text =
DQCTemp5.DQClassDescription
                If InStr(1, DQCTemp5.DQClassUnit, "Meters",
vbTextCompare) > 0 Then DQCTemp5.DQClassUnit = "m"
                FrmMeasure.MDTBUnit(EachLayer).Text =
DQCTemp5.DQClassUnit
                If Val(DQCTemp5.DQClassValue) > 0 Then
FrmMeasure.MDTBValue(EachLayer).BackColor = &H80FF&
                If Val(DQCTemp5.DQClassValue) = 0 Then
FrmMeasure.MDTBValue(EachLayer).BackColor = &HFF00&
            Next EachLayer
            Label10.Visible = True
            Label12.Visible = True
        Exit For
        End If
    Next DQCTemp1

If QualityElementsList.SelectedItem.Parent.Text = "Completeness" Then
    For Each DQCTemp1 In Co.DQCCollection
        Set DQCTemp5 = Co.DQCCollection(DQCTemp1.DQClassIdentifier)
        If QualityElementsList.SelectedItem.Text =
DQCTemp5.DQClassName And DQCTemp5.DQClassIsAvailable = True Then
            For EachLayer = 1 To 3 'NoOfLayers
                FrmMeasure.MDTBValue(EachLayer).Text =
DQCTemp5.DQClassValue
                FrmMeasure.MDTBText(EachLayer).Text =
DQCTemp5.DQClassDescription
                If InStr(1, DQCTemp5.DQClassUnit, "Percentage",
vbTextCompare) > 0 Then DQCTemp5.DQClassUnit = "%"
                FrmMeasure.MDTBUnit(EachLayer).Text =
DQCTemp5.DQClassUnit
                If Val(DQCTemp5.DQClassValue) = 0 Then
FrmMeasure.MDTBValue(EachLayer).BackColor = &HFF00&
                If Val(DQCTemp5.DQClassValue) > 0 Then
FrmMeasure.MDTBValue(EachLayer).BackColor = 33023
            Next EachLayer
            Exit For
            End If
    Next DQCTemp1
End If
If QualityElementsList.SelectedItem.Parent.Text = "Logical
Consistency" Then
```

```
    For Each DQCTemp1 In LC.DQCCollection
        Set DQCTemp5 = LC.DQCCollection(DQCTemp1.DQClassIdentifier)
        If QualityElementsList.SelectedItem.Text =
DQCTemp5.DQClassName And DQCTemp5.DQClassIsAvailable = True Then
            For EachLayer = 1 To 3 'NoOfLayers
                FrmMeasure.MDTBValue(EachLayer).Text =
DQCTemp5.DQClassValue
                FrmMeasure.MDTBText(EachLayer).Text =
DQCTemp5.DQClassDescription
                If InStr(1, DQCTemp5.DQClassUnit, "Percentage",
vbTextCompare) > 0 Then DQCTemp5.DQClassUnit = "%"
                FrmMeasure.MDTBUnit(EachLayer).Text =
DQCTemp5.DQClassUnit
                If Val(DQCTemp5.DQClassValue) < 100 Then
FrmMeasure.MDTBValue(EachLayer).BackColor = &H80FF&
                If Val(DQCTemp5.DQClassValue) = 100 Then
FrmMeasure.MDTBValue(EachLayer).BackColor = &HFF00&
            Next EachLayer
        Exit For
        End If
        Next DQCTemp1
End If

' Code for Attribute Accuracy
If QualityElementsList.SelectedItem.Parent.Text = "Attribute Accuracy"
Then
End If

' Code for Temporal Accuracy
If QualityElementsList.SelectedItem.Parent.Text = "Temporal Accuracy"
Then
End If

' Code for Lineage
If QualityElementsList.SelectedItem.Parent.Text = "Lineage" Then
End If

100

End Sub
```

2- Class Module: clsMeasure.cls

```
' Copyright 2006 ESRI
'
' All rights reserved under the copyright laws of the United States
' and applicable international laws, treaties, and conventions.
'
' You may freely redistribute and use this sample code, with or
```

```
' without modification, provided you include the original copyright
' notice and use restrictions.
'
' See use restrictions at /arcgis/developerkit/userestrictions.

' Use the ICommand interface to query the properties of a COM command
or to create COM commands. Use ITool to access to members that define
a tool in ArcGIS.
Implements ICommand
Implements ITool

Private m_pApp As IApplication
Private m_bInUse As Boolean
Private m_pLineSymbol As ILineSymbol
Private m_pLinePolyline As IPolyline
Private m_pTextSymbol As ITextSymbol
Private m_pStartPoint As IPoint
Private m_pTextPoint As IPoint

' Declare variables to contain the outcome from the operation.
Public DistanceKM, DistanceM As Double
Public NoOfLayers As Integer
```

```
Private Property Get ICommand_Bitmap() As esriSystem.OLE_HANDLE
  ICommand_Bitmap = frmResources.imlBitmaps.ListImages(1).Picture
End Property
```

```
Private Property Get ICommand_Caption() As String
  ICommand_Caption = "Modified Measure Tool Version 1.00
End Property
```

```
Private Property Get ICommand_Category() As String
  ICommand_Category = "Developer Samples"
End Property
```

```
Private Property Get ICommand_Enabled() As Boolean
  ICommand_Enabled = True
End Property
```

```
Private Property Get ICommand_Message() As String
  ICommand_Message = " Modified Measure Tool"
End Property
```

```
Private Property Get ICommand_Name() As String
  ICommand_Name = "Developer Samples_ Modified Measure Tool"
```

```
End Property
```

```
Private Sub ICommand_OnCreate(ByVal hook As Object)
  Set m_pApp = hook
End Sub
```

```
Private Property Get ICommand_Tooltip() As String
  ICommand_Tooltip = " Modified Measure Tool"
End Property
```

```
Private Property Get ITool_Cursor() As esriSystem.OLE_HANDLE
  ITool_Cursor = frmResources.imlBitmaps.ListImages(2).Picture
End Property
```

```
Private Function ITool_Deactivate() As Boolean
  ' stop doing operation
  Set m_pTextSymbol = Nothing
  Set m_pTextPoint = Nothing
  Set m_pLinePolyline = Nothing
  Set m_pLineSymbol = Nothing
  m_bInUse = False

  ITool_Deactivate = True
End Function
```

```
Private Sub ITool_OnMouseDown(ByVal Button As Long, ByVal Shift As
Long, ByVal x As Long, ByVal y As Long)
  m_bInUse = True
  Dim pMxDoc As IMxDocument
  Dim pActiveView As IActiveView
  Set pMxDoc = m_pApp.Document
  Set pActiveView = pMxDoc.FocusMap

  'Get point to measure distance from
  Set m_pStartPoint =
pActiveView.ScreenDisplay.DisplayTransformation.ToMapPoint(x, y)

End Sub
```

```
Private Sub ITool_OnMouseMove(ByVal Button As Long, ByVal Shift As
Long, ByVal x As Long, ByVal y As Long)
  If (Not m_bInUse) Then Exit Sub

  Dim pMxDoc As IMxDocument
  Dim pActiveView As IActiveView
```

```
  Set pMxDoc = m_pApp.Document
  Set pActiveView = pMxDoc.FocusMap

  Dim bfirstTime As Boolean
  If (m_pLineSymbol Is Nothing) Then bfirstTime = True

  'Get current point
  Dim pPoint As IPoint
  Set pPoint =
pActiveView.ScreenDisplay.DisplayTransformation.ToMapPoint(x, y)

  pActiveView.ScreenDisplay.StartDrawing
pActiveView.ScreenDisplay.hDC, -1

  If bfirstTime Then
    Dim pRGBColor As IRgbColor
    Dim pSymbol As ISymbol
    Dim pFont As IFontDisp

    'Line Symbol
    Set m_pLineSymbol = New SimpleLineSymbol
    m_pLineSymbol.Width = 2
    Set pRGBColor = New RgbColor
    With pRGBColor
      .Red = 223
      .Green = 223
      .Blue = 223
    End With
    m_pLineSymbol.Color = pRGBColor
    Set pSymbol = m_pLineSymbol
    pSymbol.ROP2 = esriROPXOrPen

    'Text Symbol
    Set m_pTextSymbol = New TextSymbol
    m_pTextSymbol.HorizontalAlignment = esriTHACenter
    m_pTextSymbol.VerticalAlignment = esriTVACenter
    m_pTextSymbol.Size = 16
    Set pSymbol = m_pTextSymbol
    Set pFont = m_pTextSymbol.Font
    pFont.Name = "Arial"
    pSymbol.ROP2 = esriROPXOrPen

    'Create point to draw text in
    Set m_pTextPoint = New Point

  Else
    'Use existing symbols and draw existing text and polyline
    pActiveView.ScreenDisplay.SetSymbol m_pTextSymbol
    pActiveView.ScreenDisplay.DrawText m_pTextPoint,
m_pTextSymbol.Text
    pActiveView.ScreenDisplay.SetSymbol m_pLineSymbol
    If (m_pLinePolyline.Length > 0) Then _
```

```
    pActiveView.ScreenDisplay.DrawPolyline m_pLinePolyline
  End If

  'Get line between from and to points, and angle for text
  Dim pLine As ILine
  Set pLine = New esriGeometry.Line
  pLine.PutCoords m_pStartPoint, pPoint
  Dim angle As Double
  angle = pLine.angle
  angle = angle * (180# / 3.14159)
  If ((angle > 90#) And (angle < 180#)) Then
    angle = angle + 180#
  ElseIf ((angle < 0#) And (angle < -90#)) Then
    angle = angle - 180#
  ElseIf ((angle < -90#) And (angle > -180)) Then
    angle = angle - 180#
  ElseIf (angle > 180) Then
    angle = angle - 180#
  End If

  'For drawing text, get text(distance), angle, and point
  Dim deltaX As Double
  Dim deltaY As Double

  deltaX = pPoint.x - m_pStartPoint.x
  deltaY = pPoint.y - m_pStartPoint.y
  m_pTextPoint.x = m_pStartPoint.x + deltaX / 2#
  m_pTextPoint.y = m_pStartPoint.y + deltaY / 2#
  m_pTextSymbol.angle = angle

  ' The measured distance is converted to meters (DistanceM) and
kilometers (DistanceKM)
  DistanceKM = Round(95.33167 * Sqr((deltaX * deltaX) + (deltaY *
deltaY)), 3)
  DistanceM = Round(95331.67 * Sqr((deltaX * deltaX) + (deltaY *
deltaY)), 2)
  m_pTextSymbol.Text = "[" & DistanceKM & "]"
  m_pApp.StatusBar.Message(0) = Str(DistanceM) & " Meters"

  'Draw text
  pActiveView.ScreenDisplay.SetSymbol m_pTextSymbol
  pActiveView.ScreenDisplay.DrawText m_pTextPoint, m_pTextSymbol.Text


  'Get polyline with blank space for text
  Dim pPolyLine As IPolyline
  Set pPolyLine = New Polyline
  Dim pSegColl As ISegmentCollection
  Set pSegColl = pPolyLine
  pSegColl.AddSegment pLine
  Set m_pLinePolyline = GetSmashedLine(pActiveView.ScreenDisplay,
m_pTextSymbol, m_pTextPoint, pPolyLine)
```

```
  'Draw polyline
  pActiveView.ScreenDisplay.SetSymbol m_pLineSymbol
  If (m_pLinePolyline.Length > 0) Then _
    pActiveView.ScreenDisplay.DrawPolyline m_pLinePolyline

  pActiveView.ScreenDisplay.FinishDrawing

End Sub
```

```
Private Sub ITool_OnMouseUp(ByVal Button As Long, ByVal Shift As Long,
ByVal x As Long, ByVal y As Long)

If (Not m_bInUse) Then Exit Sub
m_bInUse = False

If (m_pLineSymbol Is Nothing) Then Exit Sub

Dim pMxDoc As IMxDocument
Dim pActiveView As IActiveView
Set pMxDoc = m_pApp.Document
Set pActiveView = pMxDoc.FocusMap

'Draw measure line and text
pActiveView.ScreenDisplay.StartDrawing pActiveView.ScreenDisplay.hDC,
-1
pActiveView.ScreenDisplay.SetSymbol m_pTextSymbol
pActiveView.ScreenDisplay.DrawText m_pTextPoint, m_pTextSymbol.Text
pActiveView.ScreenDisplay.SetSymbol m_pLineSymbol
If (m_pLinePolyline.Length > 0) Then
pActiveView.ScreenDisplay.DrawPolyline m_pLinePolyline
pActiveView.ScreenDisplay.FinishDrawing

Set m_pTextSymbol = Nothing
Set m_pTextPoint = Nothing
Set m_pLinePolyline = Nothing
Set m_pLineSymbol = Nothing

Dim pMap As IMap
Set pMap = pMxDoc.FocusMap

' Load the quality report
Load FrmMeasure

' Set the quality reports layers to ArcMap layers
FrmMeasure.MDCombo.Clear

Dim i As Long
For i = 0 To pMap.LayerCount - 1
     Dim pLayer As ILayer
```

```
      Set pLayer = pMap.Layer(i)
        Dim pFeatLayer As IFeatureLayer
        Set pFeatLayer = pLayer
     If pLayer.Visible = True Then
         FrmMeasure.MDCombo.AddItem pLayer.Name
         FrmMeasure.Label1(i + 1).Caption = pLayer.Name
     End If
Next

'Available in Version 1.01
FrmMeasure.Label10.Caption = "(0) m"
If DistanceM > 0 Then
    FrmMeasure.Label12.Caption = "The measured distance with the
consideration of spatial accuracy is: "
    FrmMeasure.Text1.Text = DistanceM
End If

FrmMeasure.Label10.Visible = False
FrmMeasure.Label12.Visible = False

NoOfLayers = pMap.LayerCount

' Show the quality report
  FrmMeasure.Show

End Sub
```

```
Private Function GetSmashedLine(pDisplay As IScreenDisplay,
pTextSymbol As ISymbol, pPoint As IPoint, pPolyLine As IPolyline) As
IPolyline
  'Returns a Polyline with a blank space for the text to go in
  Dim pSmashed As IPolyline
  Dim pBoundary As IPolygon
  Set pBoundary = New Polygon
  pTextSymbol.QueryBoundary pDisplay.hDC,
pDisplay.DisplayTransformation, pPoint, pBoundary
  Dim pTopo As ITopologicalOperator
  Set pTopo = pBoundary

  Dim pIntersect As IPolyline
  Set pIntersect = pTopo.Intersect(pPolyLine, esriGeometry1Dimension)
  Set pTopo = pPolyLine
  Set GetSmashedLine = pTopo.Difference(pIntersect)
End Function
```

```
Private Property Get ICommand_Checked() As Boolean

End Property
```

```
Private Property Get ICommand_HelpContextID() As Long

End Property
```

```
Private Property Get ICommand_HelpFile() As String

End Property
```

```
Private Sub ICommand_OnClick()

End Sub
```

```
Private Function ITool_OnContextMenu(ByVal x As Long, ByVal y As Long)
As Boolean

End Function
```

```
Private Sub ITool_OnDblClick()

End Sub
```

```
Private Sub ITool_OnKeyDown(ByVal keyCode As Long, ByVal Shift As
Long)

End Sub
```

```
Private Sub ITool_OnKeyUp(ByVal keyCode As Long, ByVal Shift As Long)

End Sub
```

```
Private Sub ITool_Refresh(ByVal hDC As esriSystem.OLE_HANDLE)

End Sub
```

3- Class Module: DQE.cls

```
Public DQCCollection As New Collection
```

```
Public Function AddDQC(DQCI As String, DQCN As String, DQCDesc As
String, DQCU As String, DQCV As String, DQCAv As String)

' Add class identifier, name, description, unit, value and
availability for each quality element class
Dim dqctemp As New DQC
    dqctemp.DQClassIdentifier = DQCI
    dqctemp.DQClassName = DQCN
    dqctemp.DQClassDescription = DQCDesc
    dqctemp.DQClassUnit = DQCU
    dqctemp.DQClassValue = DQCV
    dqctemp.DQClassIsAvailable = DQCAv
DQCCollection.Add dqctemp, DQCI

End Function
```

4- Class Module: DQC.cls

```
Public DQClassName, DQClassIdentifier, DQClassUnit,
DQClassDescription, DQClassValue As String
Public DQClassIsAvailable As Boolean
```

Appendix B: Prototype Code for the Select Operation

Class Module: clsSelect.cls

```
' Copyright 1995-2004 ESRI

' All rights reserved under the copyright laws of the United States.

' You may freely redistribute and use this sample code, with or
without modification.

' Disclaimer: THE SAMPLE CODE IS PROVIDED "AS IS" AND ANY EXPRESS OR
IMPLIED
' WARRANTIES, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY AND
FITNESS
' FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL ESRI OR
' CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL,
SPECIAL, EXEMPLARY,
' OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT
OF
' SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR
BUSINESS
' INTERRUPTION) SUSTAINED BY YOU OR A THIRD PARTY, HOWEVER CAUSED AND
ON ANY
' THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT
ARISING IN ANY
' WAY OUT OF THE USE OF THIS SAMPLE CODE, EVEN IF ADVISED OF THE
POSSIBILITY OF
' SUCH DAMAGE.

' For additional information contact: Environmental Systems Research
Institute, Inc.

' Attn: Contracts Dept.

' 380 New York Street

' Redlands, California, U.S.A. 92373

' Email: contracts@esri.com

Option Explicit

Implements ICommand
Implements ITool

Private m_pApp As IApplication
Private m_pFeedbackEnv As INewEnvelopeFeedback
```

```
Private m_pPoint As IPoint
Private m_bIsMouseDown As Boolean

Dim NumLayersWithSelFeatures As Integer
```

```
Private Property Get ICommand_Bitmap() As esriSystem.OLE_HANDLE
  ICommand_Bitmap = frmResources.imlBitmaps.ListImages(4).Picture
End Property
```

```
Private Property Get ICommand_Caption() As String
  ICommand_Caption = "Modified Select Tool Version 1.00"
End Property
```

```
Private Property Get ICommand_Category() As String
  ICommand_Category = "Developer Samples"
End Property
```

```
Private Property Get ICommand_Checked() As Boolean
  ICommand_Checked = False
End Property
```

```
Private Property Get ICommand_Enabled() As Boolean
  ICommand_Enabled = True
End Property
```

```
Private Property Get ICommand_HelpContextID() As Long
  ' No help implemented for this tool
End Property
```

```
Private Property Get ICommand_HelpFile() As String
  ' No help implemented for this tool
End Property
```

```
Private Property Get ICommand_Message() As String
  ICommand_Message = "Selects features in the map by rectangle"
End Property
```

```
Private Property Get ICommand_Name() As String
  ICommand_Name = "Developer Samples_ Modified Select"
End Property
```

```
Private Sub ICommand_OnClick()

End Sub
```

```
Private Sub ICommand_OnCreate(ByVal hook As Object)
  Set m_pApp = hook
End Sub
```

```
Private Property Get ICommand_Tooltip() As String
  ICommand_Tooltip = "Select Features"
End Property
```

```
Private Property Get ITool_Cursor() As esriSystem.OLE_HANDLE
  If Not m_bIsMouseDown Then ' not in the middle of rubber banding
    ITool_Cursor = frmResources.imlIcons.ListImages(7).Picture
  Else
    ITool_Cursor = frmResources.imlIcons.ListImages(8).Picture
  End If
End Property
```

```
Private Function ITool_Deactivate() As Boolean
  ITool_Deactivate = True
End Function
```

```
Private Function ITool_OnContextMenu(ByVal x As Long, ByVal y As Long)
As Boolean

End Function
```

```
Private Sub ITool_OnDblClick()

End Sub
```

```
Private Sub ITool_OnKeyDown(ByVal keyCode As Long, ByVal Shift As
Long)

End Sub
```

```
Private Sub ITool_OnKeyUp(ByVal keyCode As Long, ByVal Shift As Long)

End Sub
```

```
Private Sub ITool_OnMouseDown(ByVal Button As Long, ByVal Shift As
Long, ByVal x As Long, ByVal y As Long)
  ' Get the ActiveView for the map
  Dim pMxDoc As IMxDocument
  Dim pActiveView As IActiveView
  Set pMxDoc = m_pApp.Document
  Set pActiveView = pMxDoc.FocusMap

  'Store current point, set mousedown flag
  Set m_pPoint =
pActiveView.ScreenDisplay.DisplayTransformation.ToMapPoint(x, y)
  m_bIsMouseDown = True

End Sub
```

```
Private Sub ITool_OnMouseMove(ByVal Button As Long, ByVal Shift As
Long, ByVal x As Long, ByVal y As Long)
  On Error GoTo ErrorHandler

  If Not m_bIsMouseDown Then Exit Sub

  ' Get the ActiveView for the map
  Dim pMxDoc As IMxDocument
  Dim pActiveView As IActiveView
  Set pMxDoc = m_pApp.Document
  Set pActiveView = pMxDoc.FocusMap

  ' Create a rubber banding box, if it hasn't been created already
  If (m_pFeedbackEnv Is Nothing) Then
    Set m_pFeedbackEnv = New NewEnvelopeFeedback
    Set m_pFeedbackEnv.Display = pActiveView.ScreenDisplay
    m_pFeedbackEnv.Start m_pPoint
  End If

  'Store current point, and use to move rubberband
  Set m_pPoint =
pActiveView.ScreenDisplay.DisplayTransformation.ToMapPoint(x, y)
  m_pFeedbackEnv.MoveTo m_pPoint

  Exit Sub
ErrorHandler:
  MsgBox "An error has occurred within the Zoom Out Tool." & vbCr &
vbCr & _
        "Error Details : " & Err.Description, vbExclamation +
vbOKOnly, "Error"
End Sub
```

```
Private Sub ITool_OnMouseUp(ByVal Button As Long, ByVal Shift As Long,
ByVal x As Long, ByVal y As Long)
  Dim pEnv As IEnvelope
  Dim pActiveView As IActiveView
  Dim pMxDoc As IMxDocument

  On Error GoTo ErrorHandler

  ' Get the ActiveView for the map
  Set pMxDoc = m_pApp.Document
  Set pActiveView = pMxDoc.FocusMap

 ' Refresh the selections
  pActiveView.PartialRefresh esriViewGeoSelection, Nothing, Nothing

  ' If user dragged an envelope...
  If (Not m_pFeedbackEnv Is Nothing) Then
    'Use it to calculate new extent
    Set pEnv = m_pFeedbackEnv.Stop
    ' Select all features that intersect with that shape
    pMxDoc.FocusMap.SelectByShape pEnv, Nothing, False
  ' Else...
  Else
    ' Select by point
    pMxDoc.FocusMap.SelectByShape m_pPoint, Nothing, False
  End If

  ' Refresh the selections
  pActiveView.PartialRefresh esriViewGeoSelection, Nothing, Nothing

'    MsgBox pMxDoc.FocusMap.SelectionCount 'OK

  'Reset rubberband and mousedown state
  Set m_pFeedbackEnv = Nothing
  m_bIsMouseDown = False

' Check if any feature is selected
If pMxDoc.FocusMap.SelectionCount < 1 Then Exit Sub

' Load the prototype's interface into memory
Load FrmSelect

' Set the layer names on the prototype's interface to those with
features selected
NumLayersWithSelFeatures = 0
Dim i As Integer
For i = 1 To pMxDoc.FocusMap.LayerCount
    FrmSelect.Label1(i).Visible = False
    FrmSelect.Label10(i).Visible = False
    FrmSelect.MDTBText(i).Visible = False
    FrmSelect.MDTBUnit(i).Visible = False
    FrmSelect.MDTBValue(i).Visible = False
```

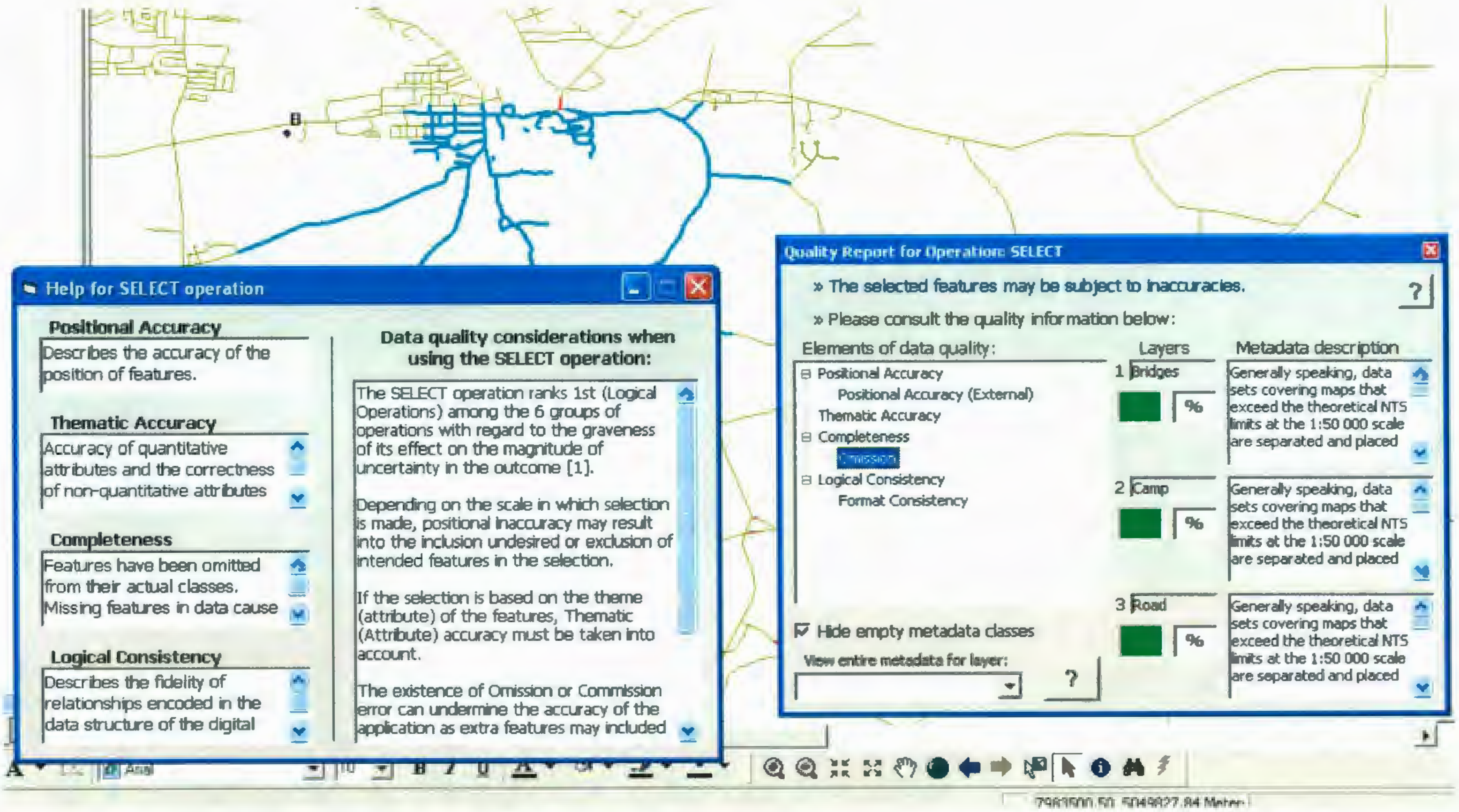
```
Next i
For i = 0 To pMxDoc.FocusMap.LayerCount - 1
    Dim pLayer As IFeatureLayer: Set pLayer = pMxDoc.FocusMap.Layer(i)
    Dim pFCursor As IFeatureCursor
    Dim pFeature As IFeature
    Dim pFeatureS1 As IFeatureSelection
    Dim pSelectionSet As ISelectionSet
    Set pFeatureS1 = pLayer
    Set pSelectionSet = pFeatureS1.SelectionSet
    If pSelectionSet.Count > 0 Then
        NumLayersWithSelFeatures = NumLayersWithSelFeatures + 1
        FrmSelect.Label1(NumLayersWithSelFeatures).Caption =
pLayer.Name
        FrmSelect.Label1(NumLayersWithSelFeatures).Visible = True
        FrmSelect.Label10(NumLayersWithSelFeatures).Visible = True
        FrmSelect.MDTBText(NumLayersWithSelFeatures).Visible = True
        FrmSelect.MDTBUnit(NumLayersWithSelFeatures).Visible = True
        FrmSelect.MDTBValue(NumLayersWithSelFeatures).Visible = True
        FrmSelect.MDCombo.AddItem (pLayer.Name)
    End If
Next
FrmSelect.Text1.Text = NumLayersWithSelFeatures
FrmSelect.Show

Exit Sub


ErrorHandler:
  MsgBox "An error has occured within the Select Tool." & vbCr & vbCr
& _
         "Error Details : " & Err.Description, vbExclamation +
vbOKOnly, "Error"
End Sub
```

```
Private Sub ITool_Refresh(ByVal hDC As esriSystem.OLE_HANDLE)

End Sub
```



Prototype interface for operation 'Select'. Features from three layers selected

Appendix D: Questionnaire for Prototype Validation

Questionnaire for the test of a software prototype
MSc Student: Amin Zargar

Date: ___/___/2008

Personal information:

First Name: _______________ **Last Name:** _______________ **Email:** _______________

(Please hand in each page once you have completed it.)

1- How often do you use GIS software or other mapping tools (e.g. ArcGIS and Google Earth)?

☐ Daily ☐ Weekly ☐ Monthly ☐ Yearly

2- Have you had formal education in GIS (e.g. in courses outside or within the university)?

☐ Yes ☐ No

3- How familiar are you with any of the distance measurement tools in GIS software? (e.g. in IDRISI, in ArcGIS and in Google Earth).

☐ I use it frequently ☐ I have used it ☐ I have never used it

4- Imagine a scenario where you are required to measure the distance between two objects in ArcGIS. In tests A, B and C presented in following, you will be asked to perform a simple distance measurement using the MEASURE tool . This tool measures the straight (Euclidean) distance between objects on the maps. In the following tests you will be asked to measure the distances between a) two camp sites and b) a camp site and a bridge. (Please use <u>meters</u> for reporting all measurements).

Test A

a) Open the "Test A" file from the "Tests" folder on the desktop.
b) Use the MEASURE tool (typically located at the lower right corner of ArcMap) to measure the distance between camps **A** and **B**.
The measured distance is: ....m
c) Use the MEASURE tool to measure the distance between camp **A** and the right side (or most easterly point) on the **nearest bridge** to camp A.
The measured distance is: ....m
d) How accurate do you think were the distances that you measured approximately?
☐±......................m ☐ I don't know
e) Exit ArcMap.

Page 1 of 3

(Q.4 continued) Test B

a) **Open the "Test B" file from the "Tests" folder on the desktop.**

b) Use the MEASURE tool to measure the distance between camps **A** and **B**. If familiar with metadata information, you may consult this information (using the **ArcCatalog** program) to find out about the issues that may accompany the features you are measuring.
The measured distance is: ....m

c) Use the MEASURE tool to measure the distance between camp **A** and the right side (or most easterly point) on the **nearest bridge** to camp A. Again, you may consult the metadata for issues regarding the data.
The measured distance is: ....m

d) **How accurate do you think were the distances that you measured approximately?**
☐±........................m ☐ I don't know

e) Exit ArcMap.

(Q.4 continued) Test C

a) Open the "Test C" file from the "Tests" folder on the desktop.
b) Use the MEASURE tool to measure the distance between camps **A** and **B**.
   The measured distance is: ....m
c) Use the MEASURE tool to measure the distance between camp A the right side (or most easterly point) on the **nearest bridge** to camp A.
   The measured distance is: ....m
d) How accurate do you think were the distances that you measured approximately?

   ☐±........................m ☐ I don't know
e) Exit ArcMap.

5- Information stored in the metadata, among other things, tells us about the accuracy of the spatial data we use. In the above tests, this information could be consulted by either browsing through the metadata information (if possible, Test B) or by an interface that is displayed once the operation is performed (Test C). Which test was better able to make you aware of issues regarding the quality of the data?

☐ Test B ☐ Test C

6- Comparing Tests C with Test B (where you used ArcCatalog for metadata), please comment on how Test C affected your understanding and awareness of the quality of the data you used. You may also include a general explanation for your answer in Q.5.

____

____

____

____

____

____

____

Privacy statement: The personal information that you provided (your first and last names and your email address) will remain confidential with the distributors of the questionnaire (Amin Zargar and Rodolphe Devillers). The other answers that you provided for this questionnaire will be compiled with those of others and published as part of an MSc thesis.

Please print your name and sign below if you agree to the privacy statement:

Name: ______________________

Date: ___ /___ / 2008

Signature: ____________________

Page 3 of 3

Appendix E: ISO 19115 Quality Elements and Classes

Name of Measure	Metadata Identifier	Description
Lineage	LI_Lineage	information about the events or source data used in constructing the data specified by the scope or lack of knowledge about lineage
Completeness	DQ_Completeness	presence and absence of features, their attributes and their relationships
Completeness Commission	DQ_CompletenessCommission	excess data present in the dataset, as described by the scope
Completeness Omission	DQ_CompletenessOmission	data absent from the dataset, as described by the scope
Logical Consistency	DQ_LogicalConsistency	degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical)
Conceptual Consistency	DQ_ConceptualConsistency	adherence to rules of the conceptual schema
Domain Consistency	DQ_DomainConsistency	adherence of values to the value domains
Format Consistency	DQ_FormatConsistency	degree to which data is stored in accordance with the physical structure of the dataset, as described by the scope
Topological Consistency	DQ_TopologicalConsistency	correctness of the explicitly encoded topological characteristics of the dataset as described by the scope
Positional Accuracy	DQ_PositionalAccuracy	accuracy of the position of features
Absolute External Positional Accuracy	DQ_AbsoluteExternalPositionalAccuracy	closeness of reported coordinate values to values accepted as or being true
Gridded Data Positional Accuracy	DQ_GriddedDataPositionalAccuracy	closeness of gridded data position values to values accepted as or being true
Relative Internal Positional Accuracy	DQ_RelativeInternalPositionalAccuracy	closeness of the relative positions of features in the scope to their respective relative positions accepted as or being true

(Continued in the next page.)

Name of Measure	Metadata Identifier	Description
Temporal Accuracy	DQ_TemporalAccuracy	accuracy of the temporal attributes and temporal relationships of features
Accuracy of a Time Measurement	DQ_AccuracyOfATimeMeasurement	correctness of the temporal references of an item (reporting of error in time measurement)
Temporal Consistency	DQ_TemporalConsistency	correctness of ordered events or sequences, if reported
Temporal Validity	DQ_TemporalValidity	validity of data specified by the scope with respect to time
Thematic Accuracy	DQ_ThematicAccuracy	accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships
Thematic Classification Correctness	DQ_ThematicClassificationCorrectness	comparison of the classes assigned to features or their attributes to a universe of discourse
Non Quantitative Attribute Accuracy	DQ_NonQuantitativeAttributeAccuracy	accuracy of non-quantitative attributes
Quantitative Attribute Accuracy	DQ_QuantitativeAttributeAccuracy	accuracy of quantitative attributes