Handling Data Consistency through Spatial Data Integrity Rules in Constraint Decision Tables

Fei Wang

Vollständiger Abdruck von der Fakultät für Bauingenieur- und Vermessungswesen der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines Doktor-Ingenieurs (Dr.-Ing.) genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Wilhelm Caspary

1.Berichterstatter: Univ.-Prof. Dr.-Ing. Wolfgang Reinhardt

2.Berichterstatter: Univ.-Prof. Dr.-Ing. Anders Östman

Diese Dissertation wurde am 31. Jan. 2008 bei der Universität der Bundeswehr münchen eingereicht.

Tag der Mündlichen Prüfung: 13. Mai 2008

Abstract

With the rapid development of the GIS world, spatial data are being increasingly shared, transformed, used and re-used. The quality of spatial data is put in a high priority because spatial data of inadequate quality is of little value to the GIS community. Several main components of spatial data quality were indentified by international standardization bodies such as ISO/TC 211, OGC and FGDC, which consists of seven usual quality elements: lineage, positional accuracy, attribute accuracy, semantic accuracy, temporal accuracy, logical consistency and completeness (two different names for similar aspects of quality are grouped in the same category).

In this dissertation our work focuses on the data consistency issue of the spatial data quality components, which involves the logical consistency as well as semantic and temporal aspects. Due to complex geographic data characteristics, various data capture workflows and different data sources, the final large datasets often result in inconsistency, incompleteness and inaccuracy. To reduce spatial data inconsistency and provide users the data of adequate quality, the specification of spatial data consistency requirements should be explicitly described. Data producers can follow the defined data consistency requirements to validate how well a geospatial dataset meets the criteria set, and then produce desirable geospatial data.

Current approaches for specifying data consistency requirements are through the definition of spatial data integrity constraints or rules. Nevertheless, those existing approaches are not well structured or not sufficient to deliver all user needed contents. Consequently the complex contents make it difficult for humans and computers understand the defined requirements. To distribute these spatial data integrity rules in the GIS community and to apply these into quality-aware GIS applications become a challenge.

Spatial data consistency as one component of data quality is considered as an indispensable part in an ISO metadata model. Thus, the specification of spatial data consistency requirements should be treated the same as spatial data and be considered during geographic data modeling work. However, existing geographic data modeling methods lack the specification of data consistency requirements or leave it to a separate work. This causes the difficulty for users to discovery the specification, to access the entire metadata information from a single geospatial portal, and to perform corresponding data consistency checking tasks.

This thesis proposes a straightforward approach for considering spatial data consistency requirements at the conceptual data model level of geographic data modeling. Spatial data integrity rules are adopted to describe spatial data consistency requirements. In order to represent spatial data integrity rules in a standardized and structured way, we propose a new concept of “constraint decision tables” that are based on the Event Condition Action rule. These rules are managed together with the normal conceptual schema in a consistent and cooperative way. In this way, normal geographic data modeling is extended through the defined spatial data integrity rules.

To cover comprehensive spatial data integrity rules in the real world phenomena, the literature and various GI organizations that define spatial data quality elements are reviewed. Extensive well-defined quality elements are introduced and the logical consistency with semantic and temporal information is emphasized.

To make the defined spatial integrity rules and normal conceptual data schema easily understandable and accessible by humans and computers through standardized geospatial web services, this thesis first investigates the method of data model transformation in web compatible formats, and then explains how to deploy the transformed formats using OGC Transactional Web Feature Service.

Spatial data integrity rules can be used to assure the data consistency in different GIS applications. Especially those GIS applications creating large mounts of geospatial data need criteria sets to improve and control the quality of data such as in the field survey tasks. To show the use of data integrity rules in GIS applications, a new trend of the field survey using Mobile GIS techniques is introduced. A generic field-based Mobile GIS architecture using modern hardware and software is provided.

To apply the defined spatial data integrity rules to GIS applications for handling data consistency, the concept of the quality-aware geocomputational tool is proposed. This tool is able to interpret the spatial data integrity rules and the normal conceptual schema, also more importantly to execute the rules in GIS applications in order to avoid inconsistencies.

For the sake of the proof of concept, a Mobile GIS application for monitoring a landslide area is adopted. Field tests of the landslide GIS application near Balingen City in Germany are described. The feasibility of the overall concept in this theist is testified. The results show that the field-based Mobile GIS supported by the quality-aware geocomputational tool is able to effectively assure the data consistency of collected data.

Altogether, the proposed spatial data integrity rules in constraint decision tables together with the normal conceptual data schema can describe the real world phenomena and spatial data consistency requirements in an explicit and structured way. The implemented geocomputational tool brings many benefits to the quality-aware GIS. Field tests of the Mobile GIS data capture are adopted to demonstrate the use of the proposed methodology. Finally, a proposal for future research is identified.

Table of Contents

List of Figures

List of Tables

List of Acronyms

CAD	Computer Aided Design
ECA rule	Event Condition Action rule
ER	Entity Relationship
FGDC	US Federal Geographic Data Committee
GML	Geographic Markup Language
GIS	Geographic Information System
GIScience	Geographic Information Science
ICA	International Cartographic Association
ISO	International Organization for Standardization
ISO/TC	ISO Technical Committee
JTS	Java Topology Suite
MOF	Meta-Object Facility Metamodel
OCL	Object Constraint Language
OGC	Open Geospatial Consortium
OMG	Object Management Group
Prolog	Programming in Logic
PVL	Plug-in for Visual Languages
RuleML	Rule Markup Language
SWRL	Semantic Web Rule Language
UML	Unified Modeling Language
GUI	Graphical User Interface
WFS-T	Transactional Web Feature Service
WMS	Web Map Service
XMI	XML Metadata Interchange
XML	Extensible Markup Language
XSLT	XML Stylesheet Language Transformations

Chapter 1. Introduction

Nowadays, as the web technologies are booming, the use of geospatial data increases. The rapidly growing use of geospatial data is already beyond the geographic information system (GIS) community and comes to the public. People use web mapping services of Google Maps and Microsoft Live Maps to search for addresses, plan travel routes and find places of interests on the road. People rely on the personal navigation systems to guide them to reach their destinations. As a consequence, many efforts address data collection, data process and data exchange to provide users reliable data.

However, geospatial data of inadequate quality such as incompleteness, inaccuracy, inconsistency still happen due to various factors involved in data creation such as different surveyors for handling the data and different surveying technologies. In this thesis, data consistency issue is emphasized in our work. In general, the creation of geospatial data should follow the data consistency requirements stated in product specifications in order to achieve data with expected results. Ambiguous and imperfect definitions of data consistency requirements are of little value to support those data producers creating geospatial data.

In order to represent data consistency requirements not only in an explicit and unambiguous way, but also in a standardized and structured way, the definition of data consistency requirements should be considered and implemented during geographic data modeling. However, the current geographic data modeling methods lack data consistency information or leave it to a separate work. Existing formulations of data consistency requirements through constraints or rules are not sufficient to deliver all needed contents.

In this thesis, a new methodology to define spatial data integrity rules during geographic data modeling in a standardized way is investigated. The created geographic data model aims to combine essential geographic data characteristics (like feature types and attributes information) and detailed spatial data logical consistency with semantic and temporal information as an integrated modeling framework. In this way, data users can access, maintain and update the integrated geographic data peculiarities using a common model. Thus there would be one single portal to get model information and no extra work to retrieve data integrity rules from other sources. This will improve the data and information interoperability and reduce the redundancies.

1.1 Background

The need for the specification of spatial data consistency requirements as an indispensable part of geographical data modeling work is one of the major reasons why we propose spatial data integrity rules to extend the normal geographic data modeling method. When discussing the geographic data modeling approach including the proposed spatial data integrity rules within a common model, different research areas need to be investigated.

This thesis consists of three areas (Figure 1): "Geographic Data Modeling" extended by "Spatial Data Integrity Rules", the "Geographic Data Model Transformation", and "Field Data Capture in Mobile GIS" for demonstrating the uses of the proposed methodology. All three areas deal with "Spatial Data Consistency" as the common denominator.

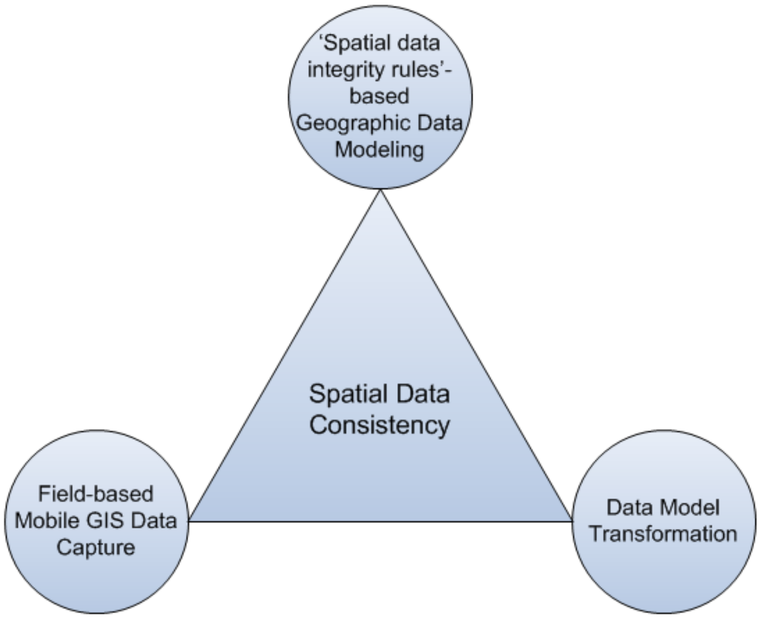


Figure 1 Thesis research areas

Spatial data consistency can be regarded as one of the aspects of spatial data quality as documented by the standardization bodies such as ISO/TC 211 (Section 2.3.2). Spatial data quality deals with all quality aspects relating to geospatial data. In this thesis, spatial data consistency refers to the logical consistency containing semantic and temporal information. Logical consistency deals with logical rules of structure and attributes for spatial data and describes the compatibility between dataset items, for example the topological relations. Semantic information indicates the pertinence of the meaning of the geographical object rather than the geometrical representation. Temporal information includes temporal attributes and temporal relationships of features.

Spatial data consistency problems occur in geospatial data involved activities, such as data collection, data processing and data usage. It can be applied at different data levels such as within a single dataset and between diverse combined datasets. We will mainly deal with consistency problems within a single dataset.

Geographic data modeling (Chapter 3) is the process for mapping geographical phenomena of the real world into digital representations in the information system. In order to handle spatial data consistency problems, spatial data integrity rules are defined extend the normal geographic data modeling for describing data consistency requirements in detail.

Spatial data integrity rules (Section 3.4) are very useful in expressing geospatial data characteristics, such as spatial relationships between feature types, in an unambiguous and standardized way. Although many existing data modeling languages like the well-known Unified Modeling Language (UML) provide a good foundation for describing geospatial data, they are still not adequate to ensure that the feature's meaning is correctly interpreted and to express complex relationships between features. Therefore, spatial data integrity rules can be adopted by current data modeling languages for supplementing insufficient description of comprehensive spatial data consistency requirements.

The possibilities of **geographic data model transformation** (Chapter 4) are discussed in this thesis too. There can be many different formats for depicting and storing a geographic data model. The desirable one is in the machine readable format, so it can be integrated into GIS applications and its meaning can be interpreted.

Field-based Mobile GIS data capture (Chapter 5) is a new trend for field survey because of advanced hardware and software architecture. Original geospatial data are produced at this stage. In this thesis, we show how the "spatial data integrity rule" based data modeling methodology and data model format transformation approach apply to a real application scenario of the field-based Mobile GIS that improves and supports spatial data quality control.

1.2 Motivation for Research

The specification of spatial data consistency requirements increasingly becomes a high priority with the rapid growing use of geospatial data. Every day, large amounts of up-to-date geospatial data are created by different persons and different agencies in different countries. Explicit specification of spatial data consistency requirements is crucial for those GIS applications to produce suitable data.

In existing GIS applications, either the spatial data consistency requirements are absent in geographic data models or they are expressed in an ambiguous way, or they do not cover all needed information. These weaknesses need to be resolved. This brings challenges for proposing a new method to handle spatial data consistency.

1.2.1 Data Consistency and Spatial Data Integrity Rules

Logical consistency is the core of the spatial data consistency issues. ISO/TC 211 standard 19113 defines the logical consistency as "*degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical)*". It was firstly used to check the integrity of non-spatial data. When the concept of topology was introduced in geospatial domain, the logical consistency including topological consistency and other components was applied to spatial data.

To examine the spatial data integrity, data consistency requirements have to be clearly described. Constraints or rules were early introduced in the database domain to express the data consistency requirements. An integrity constraint or rule can be understood as a condition or restriction that you must follow to find a solution to a specific problem. A spatial integrity constraint or rule is especially relevant to the problem involving geospatial data.

Based on the semantics of the integrity constraints and the integrity rules, the difference between them is that an integrity constraint could be reduced to "denials" or special rules whose only possible kind of action is to signal inconsistency when certain conditions are fulfilled (perhaps after recognizing a trigger event). For example, an integrity constraint "*a road is not allowed to cross a building*" can also be expressed as a detailed rule "*When any event happens, if a road crosses a building, then inconsistency happens*".

Therefore, in this thesis we adopt the term "spatial data integrity rule" for describing data consistency requirements in its widest sense. The definition of the spatial data integrity rule is proposed as:

A formal and accepted statement, definition or qualification for describing data consistency requirements in order to constrain the spatial data to correctly represent the reality in the context of the GIS applications

For non-spatial data, integrity rules are always used to express the consistency requirements of a dataset application. For example, "*the salary of an employee must be bigger than zero*" is a simple case. Complex case may include the relation of database records like "*if the working period of an employee is over three years, then the salary of this employee is increased by 10 percent*".

For spatial data, integrity rules also contain geographic data peculiarities such as geometric and topologic information. For example, "*a simple lineString feature can not be self intersected*" denotes a simple geometric case. "*A lineString feature can not intersect with a certain polygon feature*" shows a simple topologic case.

In practical GIS applications, logical consistency also contains semantic and temporal information. For example, a complex case may consist of topologic and semantic information such as "*the boundary of a city must be within the boundary of the state which it belongs to*". The complex case may also contain topologic and temporal information like "*between year 1993 and 1995, the industry area disjoins with residential area in this city*".

In order to show the defined spatial integrity rules to users and to apply the rules to GI systems for finding inconsistencies, the method of structuring, organizing and managing spatial integrity rules needs to be explored. In the GIScience literature, many people have contributed to the studies of spatial integrity rules.

Cockcroft and Servigne [Cockcroft, 1997; Servigne et al., 2000] discussed the taxonomy of spatial integrity rules such as structural, geometric and topo-semantic rules (which address errors that are predicated by the meaning behind the topological objects). Ubeda [Ubeda and Egenhofer, 1997] proposed a structured way for specifying topological integrity rule between two feature types. Cockcroft [Cockcroft, 2004] then extended Ubeda's approach to allow the consideration of the topological relationship on the basis of attribute values. Casanova et al. [Casanova et al., 2000; Casanova et al., 2002; Louwsma et al., 2006; Pinet et al., 2005] investigated using the Object Constraint Language (OCL) express spatial data integrity rules. More details of the comparison of these researches are described in section 3.4.1.

Although those approaches provide various methods for defining spatial integrity constraints or rules, they still could not satisfy all the necessary user requirements as summarized in the following:

- Spatial data integrity rules have to be defined, organized and managed in an explicit and standardized way. Thus, the international standards which relate to the data consistency and integrity rules have to be concerned. The free texts of expressions in above examples do not have the syntactic and ordered format. Therewith it may bring difficulties for users' readability, result in the interoperability problems of the data integrity rules when data fusion or data integration is performed, and lead to more efforts to apply data integrity rules to GI systems.

  The implementation of the graphic user interface (GUI) to assist users for defining and viewing spatial integrity rules is out of the scope of this thesis, but it is useful in spatial data consistency research. Therefore, we put it to the discussion part of this thesis.

- Spatial data integrity rules should consider not only the logical consistency, but also semantic and temporal information. Practical GIS applications often encounter these aspects. Moreover, different semantic notes of geospatial objects can change the meanings of spatial data integrity rules, for example, in a two dimensional map, two lineString feature types with the different semantic meaning denotes the different integrity rules: "*a road is not allowed to intersect with a lake, but a bridge can be authorized to intersect with a lake*". Therewith, semantic and temporal aspects are necessary to be taken into account when investigating spatial data integrity rules.

- Existing methods for defining spatial data integrity rules only care about how to find data inconsistencies, but ignore the information relating to the real world situations. In practical GIS applications, when data inconsistencies happen or namely when an integrity rule is violated, a consequent action to avoid this inconsistency or to correct the errors that may be caused by the operator also needs to be expressed. Explicit and detailed definition of spatial data integrity rules is always of great importance for the data operators like data producers.

  For example, if an integrity rule of "*a road is not allowed to intersect with a lake*" is trigged when a surveyor updates the geometry of a lake feature. The surveyor may just ignore this violation and continue other data capture tasks, which leads to the data inconsistency between the road and lake features. Or he may perform some appropriate actions like updating the geometry of the road, which helps to avoid the inconsistency immediately. The actions which the surveyor takes rely on the experience of the surveyor himself if no specific guidance information is provided.

  In order to help the data users such as the surveyor in the above example to solve data inconsistencies problems with the good confidence under different practical situations, it is necessary to concern such actions or instructions information when defining spatial data integrity rules. Those instructions can be achieved through experienced data users or experts. Therewith, the method of expressing the useful information within spatial data integrity rules in a structured way needs to be investigated.

  Depending on different real world situations, the policy of executing the actions of spatial data integrity rules may be different too. For example, if some rules like "*a road is not allowed to cross a building*" with strict policy are violated, data users must follow the defined actions to correct the errors. But when other rules with loose policy are violated, data users may ignore the actions and just record current situations. In this thesis, a brief study about the regulation dealing with the violation of spatial data integrity rules is performed.

- Spatial data integrity rules need to be considered together with the geographic conceptual data schema during geographic data modeling. The normal conceptual data schema is not able to deliver complex data consistency information, therewith data integrity rules should be used to supplement the content of the normal conceptual data schema during geographic data modeling. Detailed explanations on this point are given in the section 1.2.2.

- To improve the data consistency checking processes and errors awareness of existing GI systems, it is necessary to investigate the storage and format of spatial data quality rules. With rapid usages of Internet/Intranet in geospatial domain, geospatial web services are developed to provide a standardized and flexible way for publishing and sharing geospatial data. As the description of spatial data consistency, spatial data integrity rules should be treated the same as geospatial data. This point is described in detail in the section 1.2.3.

Existing approaches about spatial data integrity rules expose the shortcomings based on the above necessary conditions. Therewith, this thesis focuses on those important prerequisites and explores a new methodology for defining spatial data integrity rules to handle data consistency.

1.2.2 Spatial Data Integrity Rules and Geographic Data Modeling

Geographic data modeling is the process of selecting phenomena of the real world (source domain) and organizing them in a spatial information system (target domain) [Worboys, 1995]. The general steps of geographic data modeling can be described as: first the modeling expert uses the requirements of data users to define the content and the details of the model. For the description of the model in a conceptual schema, conceptual modeling languages are used. Afterwards the logical data model based on the conceptual schema is implemented into internal structures of a spatial database and finally used in the context of a geographic information system. In GIScience, if not explicitly pointed out, the term "geographic data modeling" always means "conceptual data modeling" which is the most important data modeling stage.

Data consistency as one of the important spatial data quality components should be considered during geographic data modeling. On the one hand, the GIScience research work has proved that data quality information is one necessary requirement to the geographic data model [Bédard et al., 2004; Borges et al., 2001; Friis-Christensen et al., 2001; Joos, 2000]. On the other hand, in the ISO/TC 211 reference model standard [ISO, 2001] and metadata standard [ISO, 2002b], data quality information is regarded as an essential part of geographic information and a necessary component of metadata information as shown in Figure 2.

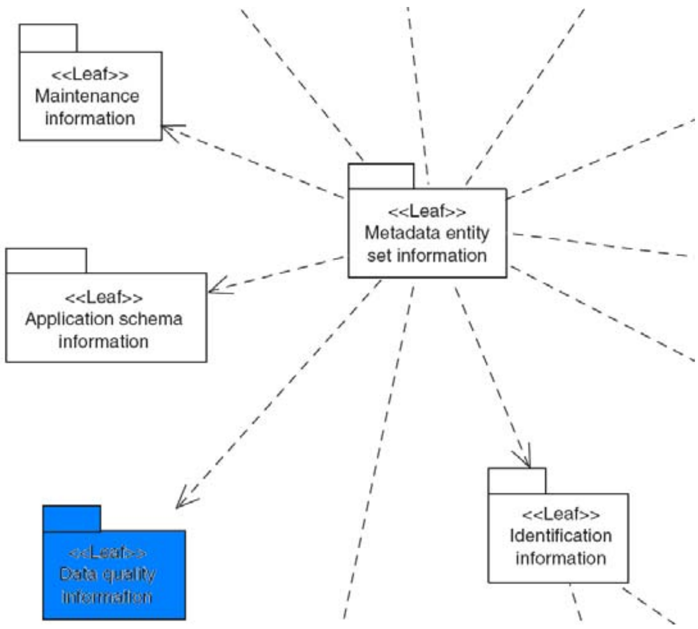


Figure 2: ISO metadata packages (partly from [ISO, 2002b])

When data consistency requirements through spatial integrity rules are specified independently with conceptual data models, data users have to access conceptual data models and spatial data integrity rules from different sources. This is a waste of time and data users are confronted with the difficulties for the data and semantics interoperability. Two examples explain this problem:

1) In a landslide GIS application, the data user gets a feature type named “Ditch” in the normal conceptual data schema and a feature type named “Trench” in the spatial data integrity rules from two separate places. “Ditch” and “Trench” are different names defined by different persons, but actually they have the same semantic meaning in the context of this application and should denote same real world objects (*a long, narrow, or deep furrow on the ground. Here it means the natural phenomena caused by land movements especially in landslide areas*). These kinds of mismatches happen to semantic meanings of feature types, and also can happen to attribute names, geometry types and spatial relationships.

2) In some cases, the data user may even get multilingual contents from different sources because persons that are responsible for defining the conceptual schema and spatial data integrity rules do not have a common work plan. Terms like “Ditch” in the normal conceptual data model, and “Spalten” (German word for Ditch) in the spatial data integrity rules cause further confusions for the data user.

Ambiguous definitions and mismatches in different sources are caused by the lack of a cooperative framework for considering normal conceptual data schemas and spatial data integrity rules. Therefore, it is necessary to consider spatial data integrity rules during geographic data modeling and build a cooperatively consistent framework in order to avoid problems like term mismatches and semantic confusions.

However, it is difficult to represent spatial data integrity rules using existing data modeling techniques. Entity Relationship (ER) model and Object Modeling Technique (OMT) are the two major data modeling techniques introduced to the GIS community.

OMT deals with an object-oriented data model and is often used now for conceptual data modeling. For example, UML is a widely used modeling language that has been accepted by ISO/TC 211 (e.g. e.g. ISO 19103 Conceptual schema language, ISO 19109 Rules for application schema and ISO 19118 Encoding) and OGC for the specification of standardized geographic information. Although UML is adopted in GI standards, it is still not sufficient to provide all needed geographic information, such as geometrical primitives and spatial relationships.

A normal UML class diagram (for details, see section 2.2.2) for describing the feature type “Ditch” is given in Figure 3, which contains only the basic information of this feature type, like class name, attribute names and operations. More detailed descriptions of this feature type are hardly represented by this normal UML class diagram, such as the geometry type of “Ditch”, temporal information of “Ditch”, domain of attribute values, allowed spatial relationships of “Ditch” with other feature types, and other detailed data consistency requirements of “Ditch”.

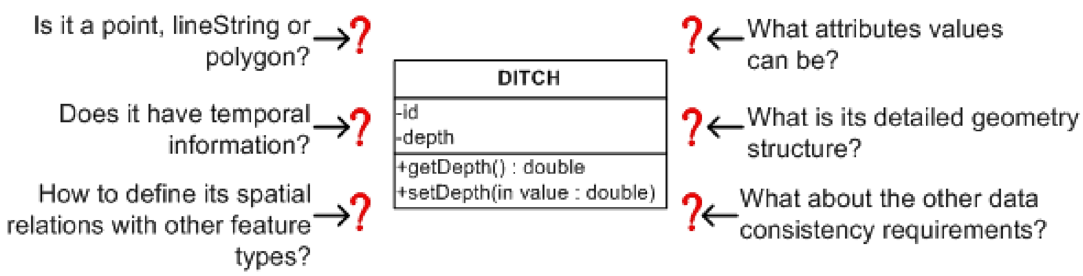


Figure 3: A normal UML class diagram example

As shown by this example, the normal UML modeling technique does not satisfy all geospatial requirements. That is why many GIScience researches have developed especially for extending UML conceptual data modeling [Bédard, 1999; Bédard et al., 2004; Belussi et al., 2004a; Borges et al., 2001; Brodeur et al., 2000; Devillers et al., 2007; Friis-Christensen et al., 2001] in the geospatial domain. It is necessary to compare those conceptual data modeling methods to discover the advantages and disadvantages. To answer the questions in Figure 3, the way to combine normal conceptual data schemas and spatial data integrity rules needs investigation.

1.2.3 Spatial Data Integrity Rules and Quality-aware GIS

Spatial data quality is increasingly becoming a high priority and the importance of enhancing GIS applications with quality-aware abilities becomes evident. Correspondingly, some developments with the terms error-sensitive GIS [Unwin, 1995], error-aware GIS [Duckham, 2002] or quality-aware GIS [Devillers et al., 2007] have been proposed.

This thesis puts the emphasis on handling data consistency problems of quality-aware GIS. As explained, spatial data integrity rules can be used to specify principles or regulations to which the created geospatial data should conform, thus to improve GIS with the quality awareness. Spatial data integrity rules contain both simple and complex descriptions of a dataset. Simple description generally means the essential geographic information, such as feature types, geometry types and attributes information. Complex description of a dataset depicts the cases which include multiple feature types, various attributes information, as well as spatial relationships.

Both simple and complex definitions of spatial data integrity rules provide the information for data producers to validate how well a geospatial dataset meets the criteria set. For example, in data capture GIS applications, surveyors should collect the data by following the defined rules in order to avoid data inconsistencies.

However to deal with data consistency problems in GIS applications, spatial data integrity rules and normal UML conceptual schema (e.g. as shown in Figure 3) should be understood by both human and computer. Therefore the format transformation and storage of the spatial data integrity rules and UML conceptual schema needs to be examined.

In the mean time, data consistency information as a part of spatial data quality belongs to the metadata as shown in Figure 2. Metadata has been proven as an integral part of any geospatial dataset and should be treated with the same tools as geospatial data [Najar, 2006]. When data users access the geospatial data, they also should be able to retrieve the content of a conceptual data model via the same way. As the growing use of the Internet/Intranet for information exchange, geospatial web services are popularly used to provide data services. For example, OGC Transactional Web Feature Service (WFS-T) allows users to download and transact large amounts of geospatial datasets in "real time".

Nevertheless, existing standard geospatial web services like WFS-T do not have the capabilities to deliver detailed spatial data integrity rules, because there is a lack of presenting data consistency information in existing web services standards. Extended method for current geospatial web services needs investigations.

Finally, when data users get defined spatial data integrity rules and normal conceptual schema data within their geographic information systems, the questions about how to apply them for improving data consistency in the applications appear. However not every GIS has the quality awareness ability. Therefore a special tool or a GIS module is needed to be developed to fulfill the practical problems where normal GIS software can not do.

1.3 Hypothesis and Objectives of Research

The definition of spatial data integrity rules together with normal geographic conceptual data models will provide data producers and vendors with the criteria to validate and assess a geospatial dataset. This affords potential data users information whether a geospatial dataset is of sufficient quality or not in the context of their particular GIS applications.

Therefore, the hypothesis of this thesis is that the constraint decision table is a suitable method to handle spatial data integrity rules as an extension to standardized data modeling, as well as to support the data consistency checking in GIS when the data model is applied.

The objectives of the research are therewith to investigate:

1. A new method to define and specify spatial data integrity rules for describing data consistency requirements in a structured and standardized way, as well as the approach to express extensive instruction information for data users when data inconsistencies are found.
2. The capability of the format transformation of spatial data integrity rules and normal conceptual data schema, for the purpose of better publishing and distributing them through a single geospatial portal.
3. The practicability of spatial data integrity rules evaluated by a quality-aware field-based Mobile GIS data capture application.

1.4 Research Overview

In order to achieve the aims of the research, the following steps will be carried out (see Figure 4):

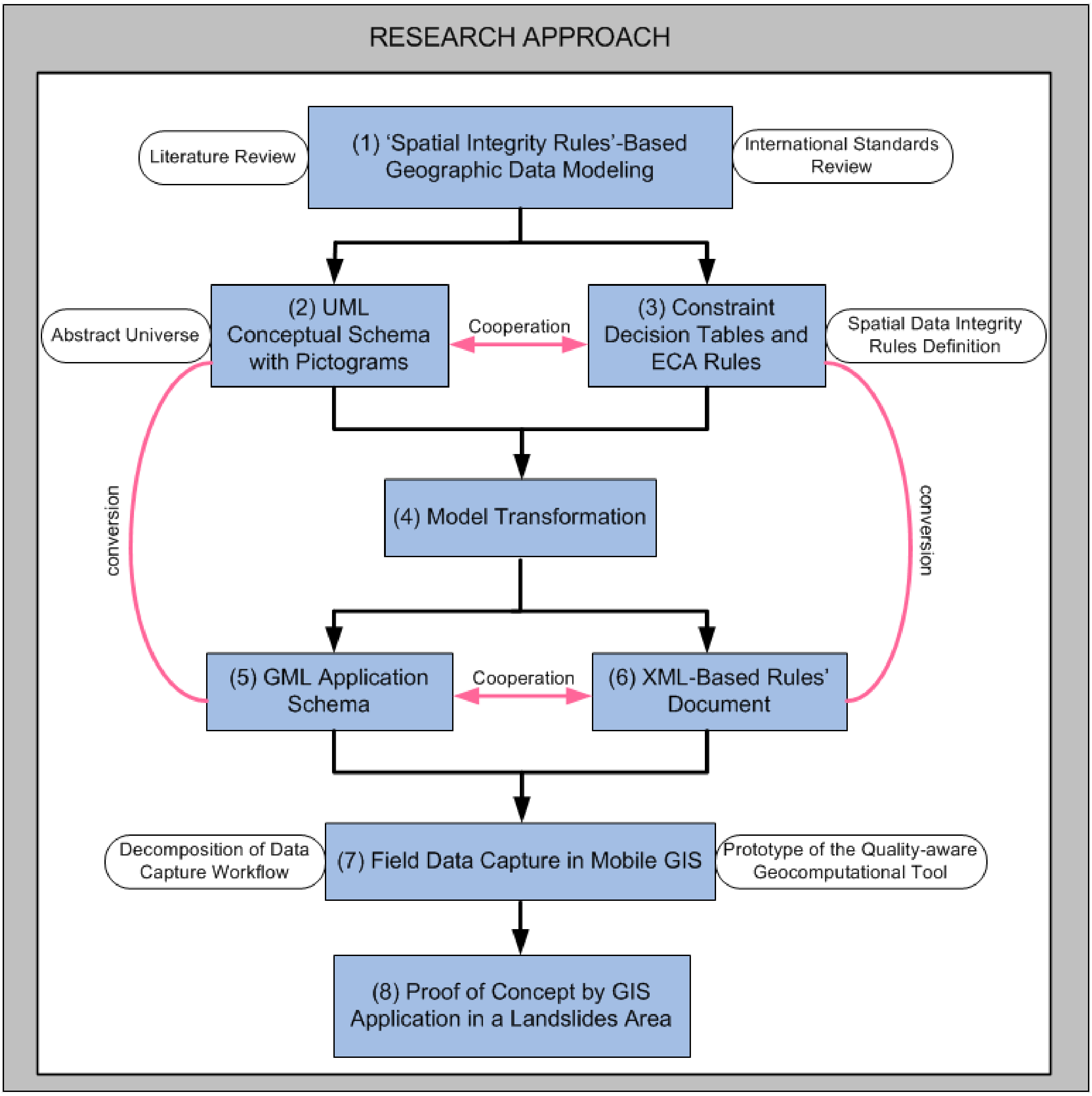


Figure 4: Thesis overview

From the extensive literature and international standards review, the method of 'spatial data integrity rules'-based geographic data modeling is proposed to meet the requirements of data modeling in geospatial world and handle data consistency (**Step 1**).

In more detail, the extended UML conceptual schema is adopted to contain essential geographic data entities, their attributes and simple relationships (**Step 2**), constraint decision tables are developed to express spatial data integrity rules (**Step 3**). In **Step 3,** specifications of spatial data quality elements from different standardization organizations are compared. Based on the most comprehensive one, the availability of considering different data quality elements when defining spatial data integrity rules is examined.

The possibility of data model transformation is studied in **Step 4**. The GML application schema is used for the conversion of UML conceptual schema (**Step 5**). The XML-based rule language is employed to convert constraint decision tables (**Step 6**).

After the data model transformation, to show the use of the converted data model for quality-aware GIS applications, field data capture in Mobile GIS is adopted here (**Step 7**). In this step, the prototype of a quality-aware geocomputational tool is implemented to communicate between the conceptual data model and the Mobile GIS data collection client software, so as to support and improve the data consistency checking within data capture workflow.

Last but not least, the overall concept is tested through the GIS application in a landslide area. The processes and the results of the case study are given to demonstrate the feasibility of the proposed approaches (**Step 8**).

1.5 Organization of the Dissertation

This thesis has five parts which refer to seven chapters. The introduction and thesis motivation are given in **Chapter 1**. This contains the background, hypothesis and objectives of the research. Finally, it explains the detailed research steps and the organization of the thesis. In **Chapter 2** basic knowledge and related issues about geographic data modeling and spatial data quality are provided. The international standards and organizations who work on these topics are introduced.

The method and implementation of this thesis are linked separately to different chapters, according to the research steps mentioned in section 1.4, described in Chapters 3, 4, and 5. In **Chapter 3**, existing data modeling techniques are examined and the 'spatial data integrity rules'-based geographic data modeling method is proposed. The availability of definition for spatial data integrity rules for a set of well known data quality elements is also introduced in Chapter 3.

In **Chapter 4**, the data model transformation method and its deployment with geospatial web services are introduced. In **Chapter 5**, the method of integrating the contents of the data model into the data capture workflow of Mobile GIS is illustrated. The prototype of a quality-aware geocomputational tool is developed. A case study is offered to test and demonstrate the above proposed concepts.

The synthesis part summarizes the results and gives a conclusion as well as the future work in **Chapter 6**.

Chapter 2. Basic Concepts and Related Research

In this chapter, an introduction of the basic concepts and relevant research areas, which are important for understanding the background, is provided. In Section 2.1 the definition of basic terms for achieving a common point of view is described. In Section 2.2 the fundamental geographic data modeling concept and literature are introduced, as well as international standards relating to the data modeling work. Data consistency belongs to scope of spatial data quality, so the historical contentions and current issues in spatial data quality are reviewed in section 2.3. This shows the growing importance of spatial data quality in GIScience.

The geographic data modeling technology and data consistency issues will be investigated in this thesis for building 'spatial data integrity rules' based geographic data models. The summary of this chapter is given in Section 2.4.

2.1 Basic Terms

Here we introduce the basic definitions and concepts used in this thesis:

First of all, the term **Geographic Information System** (**GIS**) is the most often used word in the geographic or spatial world. Sometimes people regards *geographic* as a subset of *spatial*, since *spatial* is commonly defined as a generalization of *geographic* to any space, including outer space.

GIS evolved from centuries of map-making to become the current geographic location related information system. There are lots of definitions for the term GIS, each focused on different perspectives or disciplinary origins. Definitions are Toolbox-based definitions, Database-based definitions, Organization-based definitions and Information System (IS)-based definitions [Burrough and McDonnell, 2000; Chrisman, 1996; Reinhardt, 2004].

The **Toolbox-based** definition emphasizes that GIS is a tool, which has the following definition: "a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world" [Burrough, 1986].

The **Database-based** definition emphasizes the geographic data organization in database context, for instance: "any manual or computer based set of procedures used to store and manipulate geographically referenced data" [Aronoff, 1989].

The **Organization-based** definition focuses on the institutes and people handling spatial information rather than the tools they need, for example: "a decision support system involving the integration of spatially referenced data in a problem solving environment" [Cowen, 1988].

The **IS-based** definition points out that GIS is one kind of information systems dealing with different geospatial data. Thus, it can acquire, manage, analyze and present geospatial data through the integration of hardware, software and geospatial data [Reinhardt, 2004].

With diverse capabilities and new computer technologies, GIS has become overwhelmingly a worldwide phenomenon. Systems are employed in Agriculture, Archaeology, Environment Sciences, Emergency Services, Navigation, Marketing, Real Estate, City Planning, Transportation Networks, Tourism and Utilities [Burrough and McDonnell, 2000]. Books, journals and conferences are devoted entirely to the design, technology, use and management of GIS.

Geographic Information Science (abbreviated as GI Science or **GIScience**) is the research field which addresses the general principles underlying the acquisition, representation, analysis, processing, management and storage of geographic data [Goodchild, 2003]. Geographic information system is always used to carry out the GIScience-involved activities. Study on the GIS and GIScience concepts and the differences in the definitions has been clarified in the literature [Goodchild, 1992; Goodchild, 2003; Goodchild, 2006; Mark, 2003; Wright et al., 1997].

In many ways, GIS represents a simplification of the real world. GIS can not directly be applied to the real world. Therefore **geographic data modeling** (also called **spatial data modeling**) technology is applied for abstraction of real world phenomena into a computer accessible form. That is the process of defining and organizing geographic data about the real world into consistent digital datasets that are useful and reveal information [Bonham-Carter, 1994].

Several terms often appear in geographic data modeling: **conceptual model** means the model for defining the concepts of an abstract of the real world. **Conceptual schema** is a formal description of a conceptual model. A**pplication schema** means the conceptual schema for data required by one or more applications.

Data consistency belongs to a subset of spatial data quality research, and it is treated as logical consistency with semantic and temporal information in this thesis. **Data quality** of spatial data, also of any data, plays a significant role in GIScience. People are interested in accurate and precise data, including spatial data. However, the research of the quality of spatial data needs deep study with numerous aspects. In GIScience, scientists investigate to identify the elements that describe the quality of the spatial data, that is, to show how good and reliable the data is [Caspary, 1992; Caspary, 1993; Devillers and Jeansoulin, 2006a; Guptill and Morrison, 1995; Joos, 1994; Joos, 2000]. From the definition of ICA to the standard of ISO/TC 211 (especially in ISO/TC 211 standard 19113 and 19114), spatial data quality elements have emerged and are widely accepted.

In the last few years, the rapid development of ICT and mobile technologies enable the effective and efficient access to and collection of geo-data in the field. A new term, **Mobile GIS** means "an integrated software and hardware framework for the access of geospatial data and services through mobile devices via wireline or wireless networks" [Tsou, 2004]. Mobile GIS is more and more adopted in the GIS field and as a tool in GIScience.

End users or **domain experts** are specialists in their application domains, but do not have the adequate background in software engineering or database design, and thus are not easily able to take full advantages of available GIS tools. In this sense, GIScience research should be aware of all these meanings.

2.2 Data Modeling and GIS

"All models are wrong, some are useful" [Box, 1976]. The well known industrial statistician points out that no model is possible that can fully describe the reality, but characteristics of good models can have many benefits.

A model is usually a representation of the real world phenomena. People use models to represent things from one domain (source domain) to another domain (target domain). The purpose of a model is to abstract and simplify from the source domain. The contents of the source domain are translated by the model in the target domain through the hypothetic rules, and then viewed and analyzed in the new context

[Worboys, 1995]. In the GIS domain, a geographic model is used to simplify the geographic phenomena of the real world.

When someone asks you how to reach your apartment without missing the nice dinner, it is likely that you will tell him the street number, building number, floor level, where he should “go straight”, and at which corner he should “turn left or turn right”. If he sees a supermarket “Lidl”, he is close to finding your place. Actually, you are modeling the real world into a conceptual model. Your interpretation of the real world scenes depends on your own experiences, your geographic knowledge, your understanding of your neighborhood, and the person to whom you are describing the route. In this sense, maps are considered as models of reality. The map consists of points, lineStrings, polygons and legends, which shows people streets, buildings, places of interests, rivers, mountains, directions and scale.

“The heart of any GIS is the data model, which is a set of constructs for representing objects and processes in the digital environment of the computer……Because the types of analyses that can be undertaken are strongly influenced by the way the real world is modeled, decisions about the type of model to be adopted are vital to the success of a GIS project. A data model is a set of constructs for describing and representing selected aspects of the real world in a computer” [Longley et al., 2005].

Data modeling is one of the main parts of any information system development. Some systems need a relatively simple data model, for example, in a company salary system, data about the employees, salaries, raising ratio and social insurances are clearly structured. A geographic information system requires a more complex data model, and has some differences from other information systems because of the complex characteristics of geographic data such as geometries and topological relationships for the two or even three-dimensional world.

General geographic data modeling normally consists of five levels: real world (reality), abstract universe (part of the reality), conceptual data model, logical data model and physical data model. These are illustrated in Figure 5.

In the following sub-chapters, the five geographic data modeling levels are explained separately, and in detail. Published international standards relating to the geographic data models are introduced.

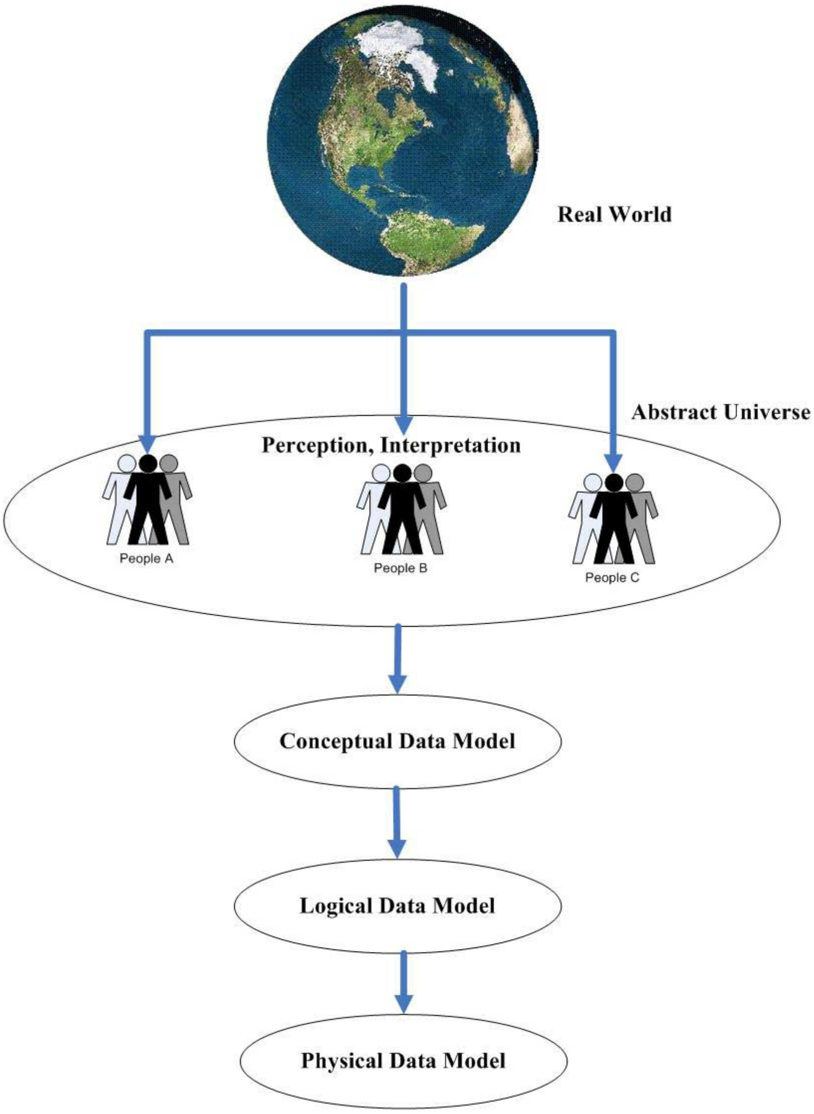


Figure 5 Geographic data modeling levels

2.2.1 Real World and Abstract Universe

The real world (reality) is the place where we live, communicate and are engaged in other social activities. The real world not only contains spatial dimensions as described via three dimensional axes, but also temporal information: past, present and future. Reality is infinite and is not deterministic. In order to model the real world, for GIS geographic data, we have to think about finite and countable things, for example, a specific area and a number of human events.

We accept that "any process which models the real world can be thought of as 'how a definite and countable model can represent the infinite and uncountable reality'" [Salge, 1995].

The model for the real world is often called Abstract Universe, or Perceived Reality, or Nominal Ground, or Terrain Nominal. ISO 19101 Reference model document [ISO, 2001] uses "Universe of discourse" for the model of the world as "view of the real or hypothetical world that includes everything of interest".

It is obvious that the abstract universe is not the real world. The abstract universe includes the definition of hypotheses on the nature of real objects, and it is a simplification of real world phenomena. It is typically described with well-defined specifications. For example, in order to show the earth in a two dimensional map, the shape and coordinates of the earth is projected on paper by longitude-latitude lines and mathematical projection approximation methods. The abstract universe is the longitude-latitude map and the specification is the projection method used to deform the shape of the earth to the map.

2.2.2 Conceptual Data Model

In ISO Reference model, conceptual data model is defined as a "model that defines concepts of a universe of discourse", and the formal description of the conceptual model is called conceptual schema. When the conceptual schema is created for the data required by one or more applications, it is called application schema.

The conceptual schema can be implemented through the conceptual data modeling techniques, and generally there are two ways: Relationship Modeling Technique and Object Modeling Technique (OMT).

Relationship modeling technique is able to define Entity Relationship (ER) model and it is considered as a straightforward way of implementing topological structures in commercial GIS [Worboys, 1995], for example, EXPRESS of ISO 10303-11 is one of the modeling languages.

OMT method deals with object-oriented data model, and one well-known modeling language is Unified Modeling Language (UML), which has been adopted by ISO/TC and OGC for specification of standardized geographic information. The extensibility mechanisms of UML make it appropriate to describe the complex peculiarities of geographic data. In the GIS community, UML modeling researches have been widely accepted [Bédard, 1999; Bédard et al., 2004; Belussi et al., 2004a; Borges et al., 2001; Brodeur et al., 2000; Devillers et al., 2007; Friis-Christensen et al., 2001].

Conceptual data modeling is the main component of the five geographic data modeling levels and also the emphasis of this thesis. Several well-known modeling languages in Table 1 are introduced and compared in the following section.

Table 1: Introduction of conceptual data modeling languages

Relationship Modeling Technique	Object Modeling Technique
Express (1)	UML (2)
	OMT-G (3), GeoUML (4), Perceptory (5), UML and OCL (6), and others (7)

1) EXPRESS:

EXPRESS is the data modeling language of "Standard for the Exchange of Product model data" (STEP) and standardized as ISO 10303-11.

With EXPRESS, a data model can be defined textually and graphically. For formal verification and as input for tools such as Standard data access interface (SDAI), the textual representation within an American Standard Code for Information Interchange (ASCII) file is the most important one. The graphical representation, called EXPRESS-G, is not able to represent all details that can be formulated in the textual format, but the graphical representation on the other hand is often more suitable for human use with prompt explanations and tutorials.

2) UML:

UML is a standardized specification language for object modeling in software engineering to specify, visualize, construct and document software-intensive systems. It is a general-purpose modeling language that includes a graphical notation to formulate an abstract model of a system.

UML is officially defined at the Object Management Group (OMG) by the UML Metamodel, a Meta-Object Facility Metamodel (MOF). UML is not restricted to systems engineering modeling. It is also used for business process modeling and representing organizational structures. UML is an implementation neutral model language that can be converted to different implementations, like XML schemas, Java codes as well as other text formats.

UML represents three prominent models:

- **Functional Model**: Showcases the functionality of the system from the user's point of view. It includes Use case diagrams.
- **Object Model**: Showcases the structure and substructure of the system using objects, attributes, operations and relationships. It includes Class Diagrams. This model is always adopted for geographic conceptual data modeling to specify the geographic objects, their attributes, geometries, topological relationships and spatial operations.
- **Dynamic Model**: Showcases the internal behavior of the system. It includes sequence diagrams, activity diagrams and state machine diagrams.

Additionally, UML is extensible, offering profiles and stereotypes for customization. UML is very suitable for describing a geographic data model.

In order to better understand the UML class diagrams used in this thesis, a basic knowledge of UML class diagrams and several simple examples are introduced here.

- **Classes**: A class represents an entity of a given system and it is regarded as a template from which other entities are created. Classes form the main building blocks of an object-oriented application. For example, there are thousands of ditches in the landslide area, but you need only model one class, called "*Ditch*", which represents the entire collection of ditches. The properties of a class are called *attributes*. The functionality that the entities support is called *operations*.

For example, a *Ditch* class containing *attributes* "id" and "depth", operations "getDepth" and "setDepth" is illustrated below:

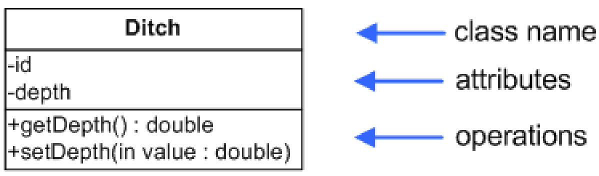


Figure 6: UML class diagram example

- **Associations**: Objects are often associated with, or related to, other objects. When you model associations in UML class diagrams, you show them as a thin line connecting two classes, as you see in Figure 7. At either end of the line, you place a role name and a multiplicity value. Role name describes the relationship from one Class to another, and multiplicity value means how many times the relationship possibly happens. Table 2 summarizes the potential multiplicity indicators you can use. Optionally, you also can add a label for the association to depict it when the association is complex or vague. When two classes are connected to each other in any way, an association relationship is established.

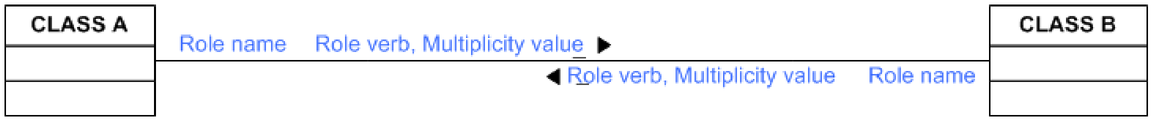


Figure 7: UML association structure

Table 2: Potential multiplicity values of UML associations

Potential multiplicity values	
Indicator	Meaning
0..1	Zero or one
1 (or 1,1)	One only
0..*	Zero or more
1..*	One or more
n	Only *n* (where *n* > 1)
0..n	Zero to *n* (where *n* > 1)
1..n	One to *n* (where *n* > 1)

- **Aggregation** and **Composition**: When a class is formed as a collection of other classes, it is called an aggregation relationship between these classes. It is also called a "*has a*" relationship. In basic aggregation relationships, the lifecycle of a part class is independent from the whole class's lifecycle. Composition is a variation of the aggregation relationship and means the life cycle is associated between the classes.

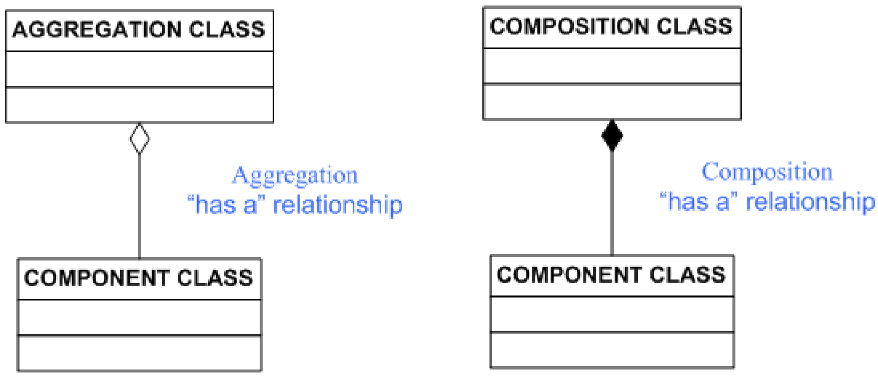


Figure 8: UML aggregation and composition relationship

- **Inheritance/Generalization**: Also called an "*is a*" relationship, because the child class is a type of the parent class. Literally, the child classes "inherit" the common functionality defined in the parent class, plus the ones defined in the child classes themselves.

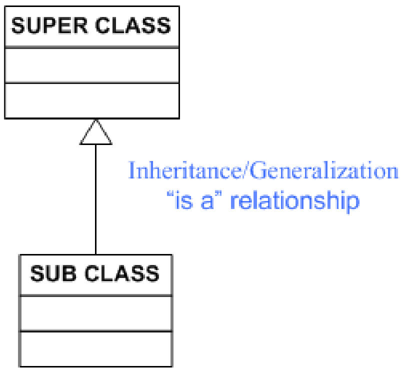


Figure 9: UML Inheritance/Generalization relationship

As mentioned in Chapter 1, a normal UML class diagram (for example, the "Ditch" in Figure 3 or in Figure 6) only contains basic geographical data information, like the class name, attribute names and operations. It is not possible to represent geographical characteristics such as geometry type of "Ditch", temporal information of "Ditch", domain of attribute values, allowed spatial relationships of "Ditch" with other feature types, and other detailed data consistency requirements.

Therefore, there have been many researches devoted to investigate the possibilities of extending normal UML modeling techniques. The following gives a brief introduction of several approaches.

3) OMT-G:

OMT-G is Object Modeling Technique for Geographic Applications [Borges et al., 2002; Borges et al., 2001]. It provides primitives for modeling the geometry and the topology of spatial data, supporting different topological structures, multiple views of objects and spatial relationships. It is a spatial extension to OMT supporting both entity and field based data. One advantage is its graphical expressiveness and representation capabilities. Textual annotations are replaced by the drawing of explicit relationships, representing the dynamics of the interaction between the various spatial or non-spatial objects.

The OMT-G model is based on three primary concepts: classes, relationships and spatial integrity constraints. Three different diagrams are adopted for designing a geographic application: class diagram, in which all classes are specified, along with their representations and relationships; transformation diagram is used to depict the multiple representations of the class; presentation diagram is able to provide guidelines for visual aspect of objects in the implementation.

4) GeoUML:

GeoUML is a conceptual data modeling technique which allows the formal specification of the content and the spatial integrity constraints of a geographic database. The design of GeoUML has two main goals: first, to be ISO compliant, and secondly, to satisfy the requirements of real life projects [Belussi et al., 2004a].

In order to assure the first goal, GeoUML is formally defined to pertain to the classes defined in ISO/TC 211 19107 Spatial Schema, and following the rules defined in ISO/TC 211 19109 Rules for Application schema. For the second goal, GeoUML has been designed with the definition of the national core schema for Italian government agencies (IntesaGIS [Belussi et al., 2004b]), and its features have been modified routinely in order to take care of the problems which were encountered by the schema design team. The main feature of GeoUML with respect to other works on spatial conceptual modeling is that it is a specialization of the ISO/TC 211 standards. It contributes to understanding how to apply the ISO standards to develop ISO profiles to be shared by projects having similar requirements.

5) UML and Plug-in for Visual Languages (PVL), Perceptory:

PVLs are sets of pictograms and grammar forming a graphical language to depict any possible geometry. In database modeling, these are used for spatio-temporal properties of object classes, of attributes, of operations and of associations via association classes [Bédard, 1999; Brodeur et al., 2000]. They offer a high level of abstraction by hiding the complexity inherent to the modeling of geometric and temporal primitives as well as implementation issues specific to each commercial GIS product or universal server. Spatial and spatio-temporal PVLs are not only applied to 2D applications, but also can be used in 3D applications through extensions [Bédard et al., 2004].

PVLs are used to extend UML class diagrams, which are implemented in a freeware Computer-aided Software Engineering (CASE) tool called Perceptory. Perceptory is an extension to UML based on stereotypes and adds support for spatiotemporal properties that are aligned with the ISO-Standards for geographic information [Bédard, 2005].

Perceptory should be used together with Microsoft Visio software (which is professional commercial software for creating a wide variety of business and technical drawings, including the UML diagrams) as a template tool. People open it in the Microsoft Visio and get extensible functionality, such as defining spatial properties and temporal properties that comply with ISO standards. One of the interesting features is that the tool provides UML stereotypes identified by the pictogram notations that are easily understood by humans.

6) UML and OCL:

Normally, most UML elements inherently imply constraints, for example, the multiplicity value of associations denotes how many times the relationship can occur, the type of attributes means which data value the attribute should contain. However, that is not enough to express constraints like the topological relationship between two associated geographic objects. Other approaches have to be applied to fulfill the shortcomings.

Object Constraint Language (OCL) as a natural supplement was developed to help UML define constraints. OCL is able to express constraints on UML schemas by attaching constraints to classes, attributes, operations on classes and associations between classes. This means that OCL has built-in constructs for navigating UML schemas, more specifically class diagrams.

The GIS community has already imposed OCL accompanying UML for geographic data modeling. Casanova et al. [Casanova et al., 2000; Casanova et al., 2002] created a knowledge model based on UML conceptual data model and OCL constraints to perform data quality control tasks. Louwsma et al. [Louwsma et al., 2006] built the UML/OCL models and put them into a repository for management. Models form the foundation for the user interface application environments, the storage data models and the exchange data models.

7) Other extensible UML approaches:

Besides the mentioned OMT-G, GeoUML, Perceptory and UML/OCL, there are some other ways for extending UML techniques, for example:

Price et al. focused on the Extended Spatiotemporal UML [Price et al., 2000], which showed how the UML is extended to capture spatially referenced, time-varying information. The Extended Spatiotemporal UML maintains language clarity and simplicity through using a small base set of fundamental modeling constructs: spatial, temporal, and thematic.

Gordillo et al. [Gordillo et al., 1999] introduced the usage of design patterns in geographic data modeling using OMT. Design patterns are an efficient strategy to record design experience. They analyzed some recurrent design problems in the GIS domain and presented some new patterns addressing those problems.

Filho and Iochpe [Filho and Iochpe, 1999] proposed GeoFrame to represent the extension of UML for defining analysis patterns for geographic applications, because analysis patterns and conceptual frameworks can both facilitate and improve geographic database design in many organizations.

Bédard et al. and Friis-Christensen et al. [Bédard et al., 2004; Friis-Christensen et al., 2001] discussed and compared the characteristics of some of the above extensible UML methods, and suggested the necessary requirements during geographic data modeling.

2.2.3 Logical Data Model

The logical data model should be based on the structures identified in the conceptual data model. At this level, the data modeler attempts to describe the data in as much detail as possible, without consideration of how this will be physically implemented in the database system. Features of logical data model include:

- All entities and relationships among them are contained.

- All attributes for each entity are specified.
- The primary key for each entity is specified.
- Foreign keys (keys identifying the relationship between different entities) are specified.
- Normalization occurs at this level.

2.2.4 Physical Data Model

At this level, the data modeler specifies how the logical data model will be realized in the database schema, like tables, indexes and records. The logical data model is mapped to the internal structures of a given system.

For example, in Oracle, a physical data model containing the implementation tables and columns is generated from the logical data model. It has indexes, constraint definitions, linking tables, partitioned tables or clusters.

In the GIS domain, people also apply Geographic Markup Language (GML) or Extensible Markup Language (XML) as the physical data models for storing, distributing, exchanging the spatial data [Najar, 2006].

2.2.5 Geographic Data Modeling and International Standards

Before introducing the international standards relating to geographic data modeling, two very famous international organizations, specifically GIS dedicated, are explained as follows.

ISO/TC 211:

ISO/TC 211 Geographic information/Geomatics is a standard technical committee formed within the International Organization for Standardization (ISO). It is responsible for preparation of a series of International Standards and Technical Specifications covering the areas of geographic information. The detailed introduction of ISO/TC 211 and its GI standards were given by Kresse [Kresse and Fadaie, 2004].

OGC:

"The Open Geospatial Consortium, Inc (OGC) is an international industry consortium of 342 companies, government agencies and universities participating in a consensus process to develop publicly available interface specifications. OpenGIS® Specifications support interoperable solutions that "geo-enable" the Web, wireless and location-based services and mainstream IT. The specifications empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications." [OGC]

These two organizations work within the same area and they have co-operative agreement to harmonize their mutual works and to develop future projects. Specifications originally defined by OGC like Web Map Service (WMS) Implementation Specification, were reviewed and examined by ISO/TC 211, and then it became ISO standard 19128: Geographic Information – Web map server interface. More detailed introduction of geographic standards can be found in [Kresse and Fadaie, 2004; Reinhardt, 2004]The complete names of ISO/TC 211 standards are listed in **Appendix I: ISO/TC 211 Geographic**

information standards. Only standards relating to geographic data modeling are explained in the next paragraphs.

ISO 19103, *Conceptual schema language,* relates to selecting a conceptual schema language that meets the needs for rigorous representation of geographic information models and ISO/TC211's needs with respect to standards development. ISO/TC211 has selected UML static structure diagram (i.e. the class diagram) with OCL. This work also provides guidance for creating standardized geographic information and service models in order to ensure interoperability between geospatial models.

ISO 19107, *Spatial schema,* specifies a set of standard spatial data types and operations for vector geometric and topological spaces. The geometry and topology are treated up to 3 dimensions. The Geometry supplies the means to describe the shapes of objects with coordinates and mathematical functions. Geometric data types are classified into three subclasses: base geometry (GM_Primitive), complex geometry (GM_Complex) and multiple geometry (GM_Aggregate). Topology describes the property of geometry that remains invariant when the space is transformed. Topological data types are subdivided into base topology (TP_Primitive) and complex topology (TP_Complex). Definition of spatial characteristics of conceptual schema should refer to this standard.

ISO 19108, *Temporal schema,* describes temporal characteristics of geographic information. It defines temporal feature attributes, feature operations, feature associations and temporal aspects of metadata about geographic information. The temporal schema provides geometric and topological data types, respectively TM_GeometricPrimitive and TM_TopologicalPrimitive. Temporal information of conceptual schema is performed according to this standard.

ISO 19109, *Rules for application schema,* defines the General Feature Model (GFM), which is the metamodel for abstracting real-world features. Application schema rules provide the principles of the abstraction process and the realization of application schemas that document one perception of the reality. This standard also explains how to integrate the standardized schemas from other ISO/TC 211 geographic information standards with the application schema. It tells how to use the conceptual schema language UML for defining application schema.

ISO 19110, *Feature cataloguing methodology,* defines a Metamodel for cataloguing feature types of the real word. It may be used as a basis for defining the universe of discourse being modeled in a particular application.

ISO 19111, *Spatial referencing by coordinates,* defines the conceptual schema for documenting one, two, three-dimensional coordinate reference systems supporting the description of a position.

ISO 19115, *Metadata,* defines the contents and structure of metadata components for describing geographic datasets. This standard indicates mandatory, conditional and minimum set of metadata required for different ranges of metadata applications.

2.3 Spatial Data Quality

Etymology of the term quality: “c.1290, from O.Fr. qualite (12c., Fr. qualité), from L. qualitatem (nom. qualitas; said to have been coined by Cicero to translate Gk. poiotes), from qualis "of what sort," from PIE pronomial base *kwo-” [OnlineEtymology, 2007].

In the ISO 9000:2005 standard "Quality management systems - Fundamentals and vocabulary" [ISO, 2005], the definition of quality is given as "degree to which a set of inherent characteristics fulfills requirements", where the requirement means "need or expectation that is stated, generally implied or obligatory".

In GIScience, quality is termed as "spatial data quality", which deals with all quality aspects of geospatial data. Geospatial data are an abstract of complex real world scenes, and the development of these involves different surveying technologies (various survey instruments), different people for handling the data (having different professional skills) and others. Thus, the quality of all geospatial data (except legally recognized as reality, such as legal cadastre of certain countries) are often not adequate for GIS applications at different levels [Devillers and Jeansoulin, 2006b].

Spatial data quality issues experienced stages from indifference to prominence in GIScience. Pioneers of this topic like Goodchild and Christman met the situation during their international conference presentations twenty and thirty years ago, that many people doubted and had little interest of data quality conceptual work. And major GIS vendors did not put emphasis on handling quality of geospatial data.

With the increasing usages of geospatial data like through the Internet GIS [Peterson, 2003; Reinhardt, 2001], people notice that data of inadequate quality may bring some social, economic, environmental and political problems. Correspondingly, data users' requirements for the quality of the geospatial data increased far beyond the geometry accuracy.

For example, at the early stage, data users only compare whether two lines representing the same entity on the earth exactly match or not when visualizing on the map. Now with the achievements of the spatial data quality research, detailed and comprehensive spatial data quality elements are introduced and defined. Data users also begin awareness on more data quality aspects, for example, they may care for when the dataset was produced, who produced the dataset, what is the bounding area of the dataset and other issues. They may not know the professional spatial data quality terms such as topological consistency and semantic accuracy, but they know what their requirements according to the specific GIS applications.

Spatial data quality also attracts eyes of the standardization bodies. During the 1990s, some international organizations have developed working groups that address spatial data quality issues. Conferences have been held for this subject, and GI-related conferences take quality as a special topic. The standardization work of quality issues in the GIS domain, specifically regarding the metadata standards, has been carried out. The most recognized are enumerated in **Appendix II: International groups, conferences, standardization organizations on spatial data quality**.

In the following paragraphs, first, sources and types of errors of geospatial data, which lead to "bad" spatial data quality, are discussed to show the importance phases in GIS which have important effects on spatial data quality. Second, the identification of the detailed spatial data quality elements from different standardization work is compared to address differences. Finally, a brief introduction of some new concepts relating to spatial data quality is provided.

2.3.1 Error and Uncertainty in Geographic Information

Error is considered as the difference between the measured data and the "real" data. The "real" data is ideal and actually does not exist. It is normally replaced by reference data or control data, e.g. the geospatial data for describing the complex earth's surface is not the real data on the earth, and thus it is obvious that a perfect replica of earth surface can not be achieved. Error all the time coexists with data and is therefore described as a "function of information" [Goodchild, 1995b] and a "fundamental dimension of data" [Chrisman, 1991].

Error of geospatial data has an important influence on spatial data quality, so the understanding of errors in GIS, like when and where the errors possibly happen, is very useful for people to better manage and eradicate.

Burrough [Burrough and McDonnell, 2000] pointed out that errors can occur at various stages in the process from observation to presentation. Hunter and Beard [Hunter and Beard, 1992] proposed to classify errors into three categories: those errors related to data collection and compilation (source error), and those related to data processing (process error), and those related to data usage (use error). More specifically, Collins and Smith [Collins and Smith, 1994] presented errors according to different stages in the form show in Table 3.

Table 3: Errors at different stages in GIS [Collins and Smith, 1994]

Stages	Sources of error
Data collection	Inaccuracies in field measurements Inaccurate equipment Incorrect recording procedures Errors in analysis of remotely sensed data
Data input	Digitizing error Nature of fuzzy natural boundaries Other forms of data entry
Data storage	Numerical precision Spatial precision (in raster systems)
Data manipulation	Wrong class intervals Boundary errors Spurious polygons and error propagation with overlay operations
Data output	Scaling Inaccurate output device
Use of results	Incorrect understating of information Incorrect use of data

The word “error” used here is in its widest sense to include all possible faults in GIS. However, spatial data quality covers a more general meaning and can not be understood to be equal to the word “error”. More accurately, the issues of spatial data quality contain the aspects of “error”. Therefore, the forms of “error” (like positional error and attribute error) caused by the above error sources correspond to the specific spatial data quality elements or parameters.

Traditionally, uncertainty and error seem to have been used almost interchangeably. Errors can be avoided or reduced to minimal levels. Errors refer conventionally to inaccuracy and imprecision, and are always considered to be computable if truth data is obtainable, or truth is recoverable if errors become known. However, ambiguity, inexactness and vagueness exist in many GIS applications. In many situations, it would be extremely difficult to believe that a truth actually exists. For example, it may be impossible to define a true soil class if the definition of soil classes is inherently uncertain or vague. Thus, the notion of uncertainty can be a more realistic view of geographic measurement than the notion of errors, which suggest attainable true values [Shi, 1994; Zhang and Goodchild, 2002].

The term error may send a misleading signal that true values are definable and retrievable, and uncertainty seems to admit the vagueness and randomness in geographic information.

However, in this thesis, we do not discuss details about their differences, and if the context does not cause confusion, the term error is used as synonymous as uncertainty. Moreover, this thesis focuses on handling data consistency errors at the field data collection stage.

2.3.2 Identification of Spatial Data Quality Elements

As stated in Appendix II, some international groups and standard organizations are devoting to the spatial data quality research. In order to describe the characteristics of quality, data quality parameters have to be identified. Many related researches have been performed in GIScience [Caspary, 1992; Caspary, 1993; Frank, 1998; Guptill and Morrison, 1995; Joos, 1994; Joos, 2000; Stanek and Frank, 1993]. Results like metadata standards have been achieved and published by data producer organization (IGN-France) or the standardization bodies like (ISO/TC 211, OGC, CEN/TC-287, FGDC).

Devillers [Devillers et al., 2005] made a comparative work of the above results, and seven usual quality elements are used for comparison: lineage, positional accuracy, attribute accuracy, semantic accuracy, temporal accuracy, logical consistency and completeness (two different names for similar aspects of quality are grouped in the same category).

	CEN[1]	ICA[2]	IGN[3]	ISO[4]	SDTS[5]
Lineage/Source	X	X		X	X
Spatial/Positional Accuracy	X	X	X	X	X
Attribute Accuracy				X	X
Semantic Accuracy	X	X	X	X	
Completeness	X	X	X	X	X
Logical Consistency	X	X	X	X	X
Temporal Information/ Accuracy	X	X		X	

[1](CEN/TC-287, 1994 and 1995), [2](Guptill and Morrison, 1995), [3](IGN, 1997), [4](ISO-TC/211, 2003), [5](FGDC, 2000)

Figure 10: Comparison of data quality elements provided by standardization or cartographic organizations [Devillers et al., 2005]

As a result in Figure 10, different standards agree on similar data quality characteristics, and standards from ISO/TC 211 have the extensive data quality elements definition. Therefore, ISO/TC 211 standards can be good choice to serve GI research for the quality elements identification.

In particular, ISO/TC 211 standard 19113 Quality principles [ISO, 2002a] and standard 19115 Metadata [ISO, 2002b] lists the defined data quality elements as shown in Table 4.

Table 4: ISO/TC 211 data quality overview elements and data quality elements

Data quality overview elements	Data quality elements
Purpose	Completeness
Usage	Logical consistency
Lineage	Positional accuracy
	Temporal accuracy
	Thematic accuracy

Although well defined data quality elements are provided in ISO/TC 211 standards, no detailed description for each element and how to apply them to GIS applications are given. This unstructured metadata information is suitable for simple uses such as browsing by a human being or free text searching, but it will bring increasing difficulties to locate accurate and relevant information and interpret its meaning correctly, as the range of sources and variety of geospatial data increase [Watson, 2007].

Therefore, it is necessary to clarify those quality elements and subelements in detail, and then formulate them in an explicit and structured way. Detailed investigation of each element and its subelements are explained in Section 3.5.

2.3.3 New Concepts relating to Spatial Data Quality

Recently in GIScience, some new concepts relating to spatial data quality were applied. In this subchapter, a brief introduction of those terms is given to explain those new concepts and to avoid confusions in spatial data quality research.

In information society, some people consider a quality product as error free, or a product which is in conformity with a specification used. Others think of it as a product that meets consumer expectations. Terms like 'data usability' [Hunter et al., 2007] or 'user-focused metadata' [Comber et al., 2007] are used to emphasize the data users' point of view to evaluate the quality of data, rather than the producer's point of view. Researches therewith suggest that more aspects relating to data consumers should be considered in future developments of existing metadata standards [Boin and Hunter, 2007; Goodchild, 2007].

For example, Bédard and Vallière [Bédard and Vallière, 1995] suggested criteria such as *legitimacy* (legal reorganization with de facto standards) and *accessibility* (costs, delays, easiness to obtain), Wang and Strong [Wang and Strong, 1996] added more characteristics based on a survey among 350 users of non-geospatial data and grouped them into four categories: *intrinsic* (e.g. believability, reputation)*, contextual* (e.g. appropriate amount of data, timeliness), *representational* (e.g. interpretability, concise representation) and *accessibility* (e.g. access security).

Devillers and Jeansouline [Devillers and Jeansoulin, 2006b] grouped spatial data quality as two category: the term "internal quality" to present the level of similarity between the perfect data ("ideal data", or called "Nominal Ground") and the produced data, and the term "external quality" to show the similarity between the characteristics of the produced data and different users' needs.

However, these definitions and classifications are still under debate and not yet widely accepted in the GIScience community. They are mentioned here to make readers know that those new terms exist but people should be aware of the usages.

2.4 Summary

In Chapter 2 the definitions of the most important terms and basic concepts of this thesis are presented and the related areas of research are introduced.

The geographic data model is the heart of a GIS, and the information of the data model will conduct the following activities in GIS, such as data collection, data processing and data usage. Data modeling in GIS normally contains five levels of abstraction: real world (reality), abstract universe (part of the reality), conceptual data model, logical data model and physical data model. These levels are explained separately, and the primary emphasis is on the conceptual data model. In GIScience, Geographic data modeling means the conceptual data modeling stage. It is the most important part of data modeling and the study of this thesis.

The ISO/TC 211 has published some standards contributing to geographic data modeling, and those standards bring many benefits for data modeling research. In ISO/TC 211 standards, spatial data quality information is suggested as one significant part of the conceptual data model. However the literature shows that there are limitations on how to represent data consistency requirements together with geographic data models. Especially when combining spatial data integrity rules and the normal conceptual data model, existing approaches expose disadvantages.

Fundamental knowledge of spatial data quality is described in this chapter. Spatial data quality elements from different standards have been compared and identified. It concludes that ISO/TC 211 Standard 19113 defines a comprehensive one. However, ISO/TC 211 Standard 19113 does not give detailed explanation of the meaning and says nothing of how to apply them into the GIS applications. Therefore, in order to discuss the data consistency problems in geographic information, deep study of the data quality elements and subelements of ISO/TC 211 Standard 19113 needs to be done.

The terms of error and uncertainty is described, as well as the sources of possible errors in geographic information. Some new terms about spatial data quality are introduced. These new concepts bring enlightened comments to spatial data quality research, but they are still at the initial stage and needs more investigation.

Chapter 3. 'Spatial Data Integrity Rules'-based Conceptual Data Modeling

As described in the previous chapter, the conceptual data model is defined as "A model that defines concepts of a universe of discourse" in the ISO reference model [ISO, 2001]. This chapter proposes a new methodology of defining spatial data integrity rules in constraints decision tables during conceptual data modeling, which is called in this thesis 'spatial data integrity rules'-based conceptual data modeling. This methodology considers UML conceptual schema and spatial data integrity rules as a cooperative framework for handling data consistency problems.

First, necessary requirements of conceptual data modeling are investigated in Section 3.1. Based on that, a 'spatial data integrity rules'-based conceptual data modeling methodology is proposed in Section 3.2. This is one of the three kernel points of the thesis research areas (see Figure 1 of the Introduction chapter). The methodology is described in two parts: **Extended UML conceptual schema** and **Spatial Data Integrity Rules**. In the part of the extended UML conceptual schema (Section 3.3), a UML modeling technique for GIS applications is introduced and some practical examples are given. In the part of the spatial data integrity rules, constraint decision table is proposed for expressing the spatial data integrity rule (Section 3.4). The theory of Event Condition Action (ECA) rule as the basis of the constraint decision table is explained, and the structure and semantics of the constraint decision table are provided. Then an explanation ensuring the consistency of the framework of the extended UML conceptual schema and spatial data integrity rule in the constraint decision table is offered in Section 3.5.

In order to show the uses of the constraint decision table for specifying comprehensive spatial data integrity rules related to the ISO/TC 211 Quality principle standard, this chapter then investigates the feasibility of the constraint decision table for each data quality element and subelement in Section 3.6. Finally, a summary of this chapter is represented.

3.1 Geographic Conceptual Data Modeling Requirements

Geospatial data have specific peculiarities that make data modeling work differently from other information systems. Before explaining the modeling method, the requirements of conceptual data modeling have to be determined. Based on previous works in the literature [Borges et al., 2001; Friis-Christensen et al., 2001; Oliveira et al., 1997; Pinet et al., 2005; Shekhar et al., 1997] and our modeling experiences on a landslide application, a set of requirements for a geographic conceptual data model are concluded. A good conceptual data model should:

- Provide a high abstraction level and consider the end users with an easily understandable and clear structure
- Support entity based and continuous field based geographic objects, which contain their thematic values e.g. the attribute "owner" of a parcel
- Represent simple geometry and multiple geometry, e.g. point (multi-points), lineStrings (multi-lineStrings), polygon (multi-polygons)
- Represent temporal extent, e.g. instantaneous existence or a period

- Support spatial relationships from simple associations to complex networks which may contain topological, metric, and directional relationships
- Contain spatial data quality requirements on geographic objects, their thematic values, their spatial relationships and other quality related aspects
- Be based on reputable standards, like ISO/TC 211 Geographic Information standards
- Be system or implementation neutral and support ready transformations into other formats, e.g. web-compatible XML format

As introduced in Section 2.2.2, there are a lot of researches that fulfill the above data modeling requirements in the geospatial domain, such as OMT-G, GeoUML, Perceptory and Extended Spatiotemporal UML. However, those approaches only partially implement the given requirements. They can not sufficiently realize every necessary point, especially the standard based way and the transformation possibility, which have been proven to be an indispensable part of geographic conceptual data modeling. Therefore, in this thesis, we will investigate a new methodology to extend the normal conceptual data modeling technique in order to resolve the highlighted points.

3.2 Proposed Conceptual Data Modeling Methodology

Consideration of the conceptual data modeling requirements introduced in the previous section and the point of view that spatial data quality information is deficient in previous geographic data modeling research, a 'spatial data integrity rules'-based conceptual data modeling methodology is proposed as illustrated in Figure 11.

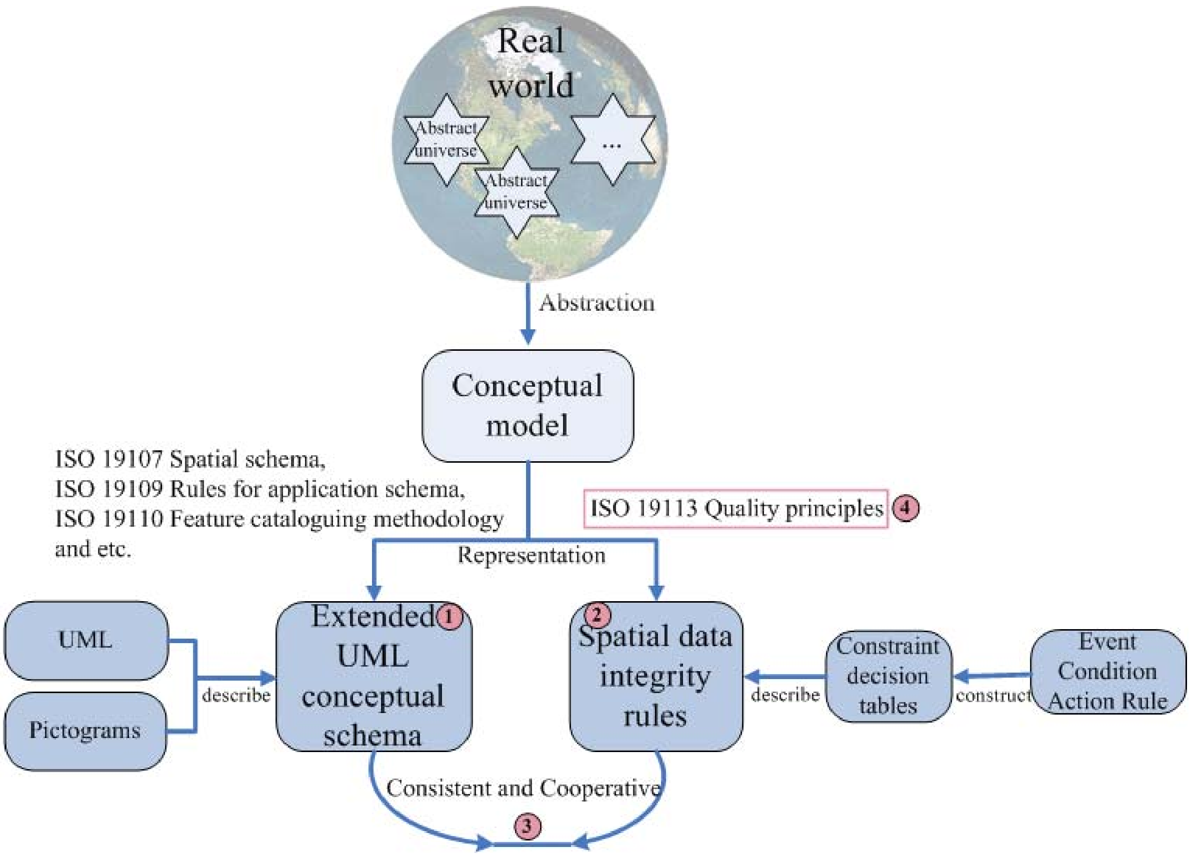


Figure 11: "Spatial data integrity rule"-based conceptual data modeling methodology

The proposed methodology for conceptual data modeling consists of two components: extended UML conceptual schema and spatial data integrity rules. Extended UML conceptual schema is employed for modeling the essential information of geographic data, like geographic entities, properties and simple relationships. Spatial data integrity rules are used to describe detailed data consistency requirements in a formal way. The two components are combined through a cooperative framework.

The following subsections in this chapter will give corresponding explanations of each component. As marked by **point 1**, we introduce the extended UML conceptual schema which is formally described by UML and Pictogram. Then **point 2** about the definition of spatial data integrity rules through constraint decision tables is explained. The theory that keeps a consistent and cooperative framework of the extended UML conceptual schema and spatial data integrity rules is demonstrated at **point 3**. In the end, in order to explore the feasibility of spatial data integrity rules for comprehensive quality elements specified in ISO/TC 211 Quality principles standard, detailed studies are provided at **point 4**.

3.3 Extended UML Conceptual Schema

In this section, the normal UML conceptual schema is introduced by an example in a landslide GIS application to show the weakness of UML modeling techniques in the geospatial domain. Then the pictogram concept is explained and the method of using pictograms for extending normal UML conceptual schema is described. In the end, we will show the data consistency requirements implicitly specified in the extended UML conceptual schema.

3.3.1 UML Conceptual Schema

In this thesis, the UML conceptual schema contains only the essentials of the geographic data like the geographic entities and their properties. In this way, the conceptual schema keeps a simple readability and its graphic notations bridge the communication between the modeler and end users.

With the normal UML conceptual schema, people can express some basic peculiarities of geospatial data such as feature types, their attributes and simple relationships. For example, two feature types "*Ditch*" and "*Extensometer*" in a landslide area are shown in Figure 12, and their class diagrams and association are represented in Figure 13.

The **Ditch** class in this example contains attributes "id" and "depth", operations "getDepth" and "setDepth", and the ditch as a landslide feature type contains zero or more than zero extensometer for measuring the distance between two points located in this ditch, in order to monitor the land change. **Extensometer** class has attributes "name" and "observedDistance", operation "getObservation", and an extensometer as a field survey instrument is installed for monitoring the land change and is located in exactly one ditch. The association between these two feature types means that one ditch feature can contain zero, one or more than one extensometer features, and one extensometer feature can be located in exact one ditch feature. The meaning of each part in the UML diagram is given by the blue comments.

Figure 12: Ditches and extensometers in a landslide area

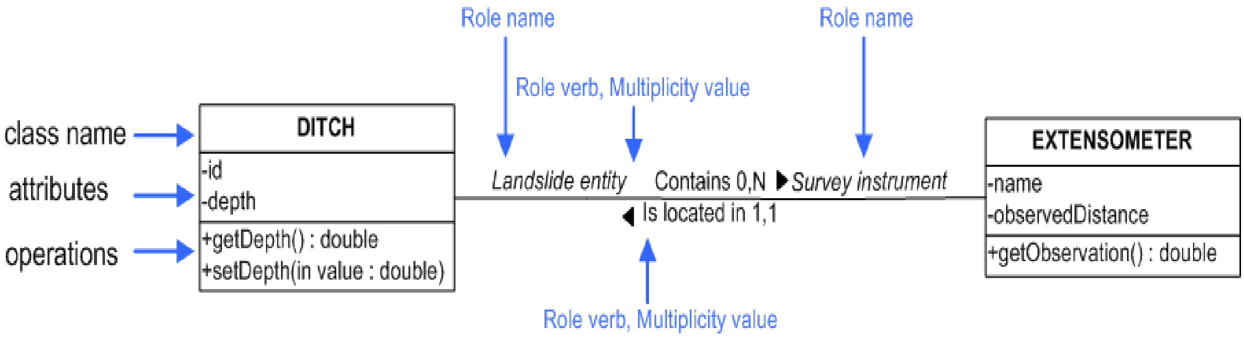


Figure 13: Ditches and extensometers represented by UML

The above UML diagram can be created by any UML compatible software tools, e.g. Microsoft Visio, Together, IMB Rational and so forth, because the attributes, operations and simple relationships of UML schema are also supported by non-geospatial data. However those UML tools can not be used to represent vividly geospatial characteristics.

In this example, you could not know the geometry type of ditch and extensometer in the graphical notations. It is possible to use the normal attributes in UML diagram to represent the geometry type of each feature type, e.g. "geometry_type" attribute is defined to contain values like "point, lineString and polygon". Or it is also possible to use textual UML stereotypes to represent geometry types. However to deliver these important geospatial characteristics in an explicit and standardized way, the UML diagram needs to be extended with a good readability. Therefore, the concept of using pictograms in UML diagram to fulfill this objective will be introduced in next subchapter.

3.3.2 Pictogram

A pictogram, also called pictograph, is a symbol representing a concept, object, activity, place or event by illustration. For example, a traffic sign is a pictogram to show or warn people something or some rules, as shown in Figure 14. Pictograms are easily understood by the public, so some international standards regarding pictograms have been developed, for instance ISO 7001: Public Information Symbols, which defines a set of pictograms and symbols for public information such as car parking and accessible vehicles. Obviously, the pictograms having straightforward and visual effects can provide users an easily understandable impression.

Figure 14: A pictogram example: traffic sign of "falling or fallen rocks"

The UML provides several extensibility mechanisms, and one of them is called stereotypes. The stereotypes are particularly useful for abstraction purposes and can extend the semantics of the metamodel. The UML also allows the use of icons for stereotypes. User-defined icons can be associated with given stereotypes for tailoring the UML to different applications. Therefore, the mentioned pictograms can be used to link the stereotypes for the abstraction and visualization of standardized modeling elements in static structure diagrams.

Tveite [Tveite, 2001] explained how to apply these intuitive and expressive icons in visual modeling, and demonstrated this approach to represent OGC simple feature specification data types. Generally, three possible ways can be used to describe geospatial stereotypes in UML as shown in Figure 15: only stereotype name, stereotype name with geo-icon, only geo-icon.

In Figure 15, the leftmost UML class does not utilize the geo-icon, so this representation does not rely on support for graphics in the modeling tools. The centre one shows the most comprehensive way with both the stereotype name and the geo-icon. The right one only uses geo-icon and this is probably the best way to apply the geo stereotype in UML classes for easy comprehension.

Figure 15: Three examples of how the geo stereotype can be applied to UML classes [Tveite, 2001]

Many other researches in GIScience also show the wide usage of pictograms to the extend normal UML modeling approach for describing geographic information, such as OMT-G and PVL-UML [Bédard, 1999; Bédard et al., 2004; Borges et al., 2002; Borges et al., 2001; Friis-Christensen et al., 2001].

Here we take the research experiences from Bédard [Bédard, 2005; Bédard et al., 2004], because one main feature is that a freeware CASE (Computer-aided Software Engineering) UML tool called Perceptory has been developed. With Perceptory, people can extend UML diagrams with more geospatial information through the predefined pictograms, such as geometry types and temporal types of geographic data. This tool also takes into account the ISO/TC 211 Geographic Information standards as listed in Figure 11. Some basic predefined geospatial constructs by Perceptory are shown in Table 5.

Table 5: Basic constructs represented by the pictograms in Perceptory

	0-dimensional geometry, e.g. a bus station represented by a point
	1 -dimensional geometry, e.g. a road segment represented by a lineString
	2-dimensional geometry, e.g. a ditch object when it is represented by a polygon
	recording a date, e.g. instantaneous existence
	recording a period, e.g. beginning date and end date, a house was built in Sep. 1930 and was demolished in Aug. 2000.

With the previously introduced example in Figure 13: Ditches and extensometers can be extended by those basic geospatial constructs shown in Figure 16. In this example, we declare the feature type ditch as a polygon geometry type and the feature type extensometer as a point geometry type (at times it is represented as the lineString or polygon geometry type, but in this thesis we simplify the geometry as its center of gravity). With pictograms and related comments, people can easily understand the intended meanings.

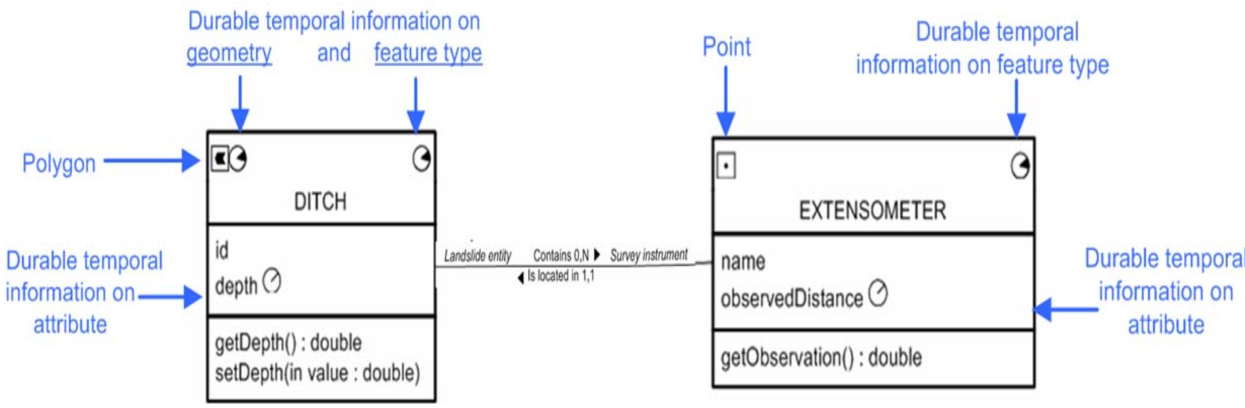


Figure 16: Extended UML example of ditches and extensometers by pictograms

3.3.3 Exposed Data Consistency Description in the UML Conceptual Schema

In last two subsections, we introduced the way to use UML and pictograms to create extended UML conceptual schema in order to meet geographic data modeling requirements. The given example in a landslide GIS application shows that the extended UML conceptual schema is able to provide basic geospatial constructs, which satisfy the aim of "essential information of the geographic data, like the geographic entities, their properties and simple relationships" in our proposed methodology.

Those essential geographic contents in the extended UML conceptual schema also implicitly show some spatial data consistency requirements that are summarized as:

1. Basic attribute accuracy, e.g. a ditch has a depth with *double* value.

2. Simple topological relationship, e.g. an extensometer is *located in* a ditch.
3. Basic temporal accuracy, e.g. a ditch should have an attribute "*start_date"* and an attribute "*end_date"* since it is a durable temporal feature.

However, data consistency requirements involve more quality elements and subelements. Therefore, the method for defining comprehensive and complex spatial data consistency requirements needs investigation and the method can be regarded as another extension of the UML conceptual schema. The next subsection will discuss the approach using "spatial data integrity rules" to express detailed data consistency requirements.

3.4 Spatial Data Integrity Rules

The definition of spatial data integrity rules for describing comprehensive data consistency requirements is investigated in this section. The existing relevant research is introduced, and based on that we propose the spatial Event Condition Action (ECA) rule based constraint decision table for specifying spatial data integrity rules. The structure of constraint decision table and its terms are explained in detail. In the end, the policy for handling violations of different data integrity rules is studied.

3.4.1 Relevant Research

This subchapter discusses the relevant research dealing with the development and management of spatial data integrity rules. To evaluate how good a method is for defining spatial data integrity rules, several main features are derived from the literature and our modeling experiences in a landslide GIS application:

- **Spatial relationships:** whether the rule gives a clear specification of topological and metric relationships (refer to Section 3.4.3 for details), which are important for representing the states between geospatial features.
- **Semantic information:** whether the rule allows the involvement of attribute values, which sometimes are used together with spatial relationships such as the semantic-topological constraint [Servigne et al., 2000].
- **Temporal information:** whether the rule can handle temporal data such as durable or instantaneous data.
- **Structured:** whether the rule is well structured and logical which keeps the rule compact and explicit.
- **Readability:** whether the documentation of the rule has a good readability that is very helpful for end users and non model experts. As announced before, we do not consider the development of the graphical user interface (GUI) for defining, visualizing and checking the rule.
- **Instruction:** whether the rule has a clear indication for data users to adjust data operations when a rule is violated, as well as detailed instructions.
- **Violation policy:** whether the policy for dealing with the violation of a rule is provided. It describes different priority levels when data users response to different rules.

In this way, the relevant research can be compared according to its fulfillment to the above identified points. They are explained separately in the following.

1. Object Constraint Language (OCL)

As previously mentioned, UML (by Object Management Group(OMG)) as a conceptual modeling language does not have the ability to express complex spatial data integrity rules such as the one containing both geometric, topologic and semantic information. Therefore, OMG developed another language called OCL to complement UML and enables one to describe expressions and constraints on UML schemas. Thus, OCL is often adopted in geographic data modeling work to express data integrity rules [Casanova et al., 2000; Casanova et al., 2002; Louwsma et al., 2006; Pinet et al., 2005].

However, the textual description of the OCL constraints severely reduces the human/computer readability of the model/schema. Vaziri [Vaziri and Jackson, 1999] confirmed the shortcomings of OCL such as at times unnecessarily verbose and often unnecessarily hard to read. Especially for non modeling experts who do not have the knowledge of modeling languages, it is difficult to understand the OCL grammar.

Moreover, OCL expressions may coexist with the UML conceptual schema, so when many data integrity rules occur and contain complex content, the UML conceptual schema may become overloaded and unreadable. For instance, the previous UML conceptual schema example in Figure 16 is extended using OCL expressions as illustrated in Figure 17.

In Figure 17, OCL expressions are defined to describe some data integrity rules, e.g. expression 1 means “if two ditch instances are not the same, their ‘id’ attribute values should not be the same either”, expression 5 depicts “the observedDistance attribute value of Extensometer is bigger than zero. More introduction of OCL and refers to [OMG]

Obviously the big problem is that for people who do not have OCL experiences, it is difficult to understand complex OCL expressions as shown in this example. Additionally, spatial relations and explanations of integrity rules are not supported by OCL.

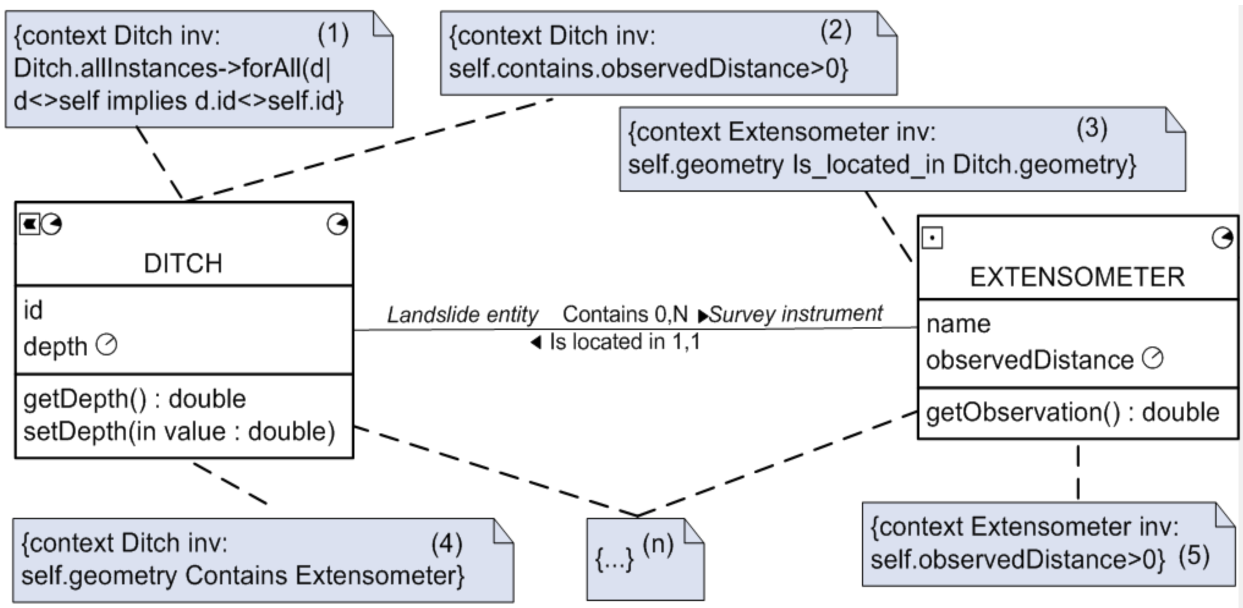


Figure 17: UML conceptual schema with OCL expressions

2. Ubeda

In order to represent topological relationships in an integrity constraint, Ubeda [Ubeda and Egenhofer, 1997] proposed a structured way for specifying topological integrity constraint as:

"*CONSTRAINT = (Entity class1, relation, Entity class2, Specification)*"

It defines the association of two geographic objects, a topological relationship between them and a specification (including "Forbidden" and "At least n times, At most n times, Exactly n times"). For example, a road not allowed to cross a building can be expressed as:

Constraint = (Road, Cross, Building, Forbidden)

This approach allows people to define topological integrity constraints in a readable way. Nevertheless, it did not study the other two spatial relationships, semantic and temporal information.

3. Cockcroft

Cockcroft [Cockcroft, 1997; Cockcroft, 1998; Cockcroft, 2004] extended Ubeda's approach to allow the consideration of the topological relationship on the basis of attribute values. For example, "*A butterfly valve may not intersect a pipe if the diameter of the pipe is greater than 40 cm*." However, her approach does not deal with temporal information and does not provide a clearly structured way to formulate data integrity rules.

4. Mostafavi

Mostafavi et al. [Mostafavi et al., 2004] proposed an ontological approach for quality assessment in a spatial database. At the ontological level, they used Prolog (Programming in Logic) as the language to manage spatial integrity constraints. Clear introduction of spatial relationships and examples were given in this study and a real dataset was automatically checked according to the defined constraints. However, a similar problem as OCL happens because understanding Prolog also applies to people who have the knowledge of that.

5. Pullar

Pullar [Pullar, 1997] suggested to use a rule-based "if-then" structure in a decision table to define the decision rules, which has advantages such as compactness, self-documentation, modifiability and completeness checking. For end users, this tabular structure has excellent readability. Moreover, the "if-then" structure provides the possibility to define a consequent instruction of the rules, e.g. "*if a tree is within 10 meters to a power line, check the height of this tree.*"

However, this "if-then" approach ignores one important element of data integrity rules called "event". For the above example, when event "we are planning a tree", the above rule makes sense for us, but when "we are getting rid of" a tree, the rule is not applicable. Additionally, this approach did not investigate the feasibility of the decision table for expressing different kinds of spatial integrity rules.

Summary:

To sum up, the comparison of above approaches can be represented in Table 6. "Yes, No, Partial" values are used to show that the approach supports, partially supports or does not support the identified feature. "Good, Middle, Bad" values are applied to show the performance in a decreasing level.

Table 6: Comparison of researches for defining spatial data integrity rules

	Spatial relationships	Semantic info	Temporal info	Structured	Readability	Instruction	Violation policy
OCL	No	Yes	Yes	Yes	Bad	No	No
Ubeda	Partial	Partial	No	Yes	Middle	No	No
Cockcroft	Partial	Yes	No	No	Middle	No	No
Mostafavi	Partial	Yes	No	Yes	Bad	No	No
Pullar	Partial	Partial	Yes	Yes	Good	Partial	No

3.4.2 Event Condition Action Rule

An ECA rule consists of events, conditions and actions and has semantics:

"*When an event occurs, check the condition and if the condition is satisfied, then execute the actions*", or shortly as "*on event if condition do actions*"

The ECA rule is also called *monitor*, or *situation-action rule* or *trigger* in active databases [Widom and Ceri, 1996]. It is an often used term since it was proposed within the active database community for monitoring state changes in database systems.

An ECA rule example in a library database is like "*when inserting a new book to the library database, if the book's keywords contain "GIS", then put it to the GIScience catalog*".

The ECA rule allows an application's reactive functionality to be defined and managed within a single rule base rather than being encoded in diverse programs, thus enhancing the modularity and maintainability of the application [Papamarkos et al., 2003]. Recently the ECA rule has been used in many settings like workflow management, network management, personalization and publish/subscribe technology, and for specifying and implementing business processes [Bailey et al., 2002; Erik Behrends et al., 2006]. The ECA rule is already in the schedule of the RuleML (Rule Markup Language) standardization initiative [RuleML], as an important contribution to the implementation of the web rule language framework.

The ECA rule was already introduced in the geospatial domain. It was used to constrain the geospatial data in the database, for example, a database ECA rule was defined as "*If the population count of a state doesn't match the sum of the populations of the associated cities, calculate a new population and set it to the state"* by Becker and Ditt [Becker et al., 1999; Ditt et al., 1997] in their geospatial systems.

However, the ECA rule that was used in the geospatial domain focused solely on the spatial database management. Any use for expressing spatial data integrity rules was not explored. Thus, the research about how to involve spatial relationships, semantic, temporal information and other data consistency characteristics in ECA rules is not found in the literature.

In this thesis we introduce the ECA rule to the geographic conceptual data modeling work for describing the spatial data consistency, therewith the proposed spatial data integrity rules can also be called **spatial ECA data integrity rules**, or spatial ECA rules.

The main advantages of the spatial ECA rule are concluded as the following:

- It has an explicit and logic structure and is applied widely in many information systems.
- It enables not only finding the inconsistency of the spatial data which the spatial integrity constraint does, but also describing various reactive functions under diverse events in the context of different GIS applications.
- Its format in XML has been put into the schedule of standardization work, which makes it more easily be distributed through the web environment.

3.4.3 Constraint Decision Table

The ECA rule provides a syntactic and semantic structure for specifying spatial data integrity rules. As a consequence, the spatial data integrity rule has to be represented in a clear way for people to view its content. In a simple way, a spatial ECA rule can be narrated as free text in the natural language such as "*When updating the geometry of a ditch, if it intersects with a hiking trail, then split the hiking trail*". The free text for expressing a spatial data integrity rule can be created quickly and is easily understood by modeling experts, but free text brings increasing difficulties for end users to comprehend its meaning, especially when integrity rules contain more complex content and the number of integrity rules increases. The important parts of an integrity rule should be explicitly organized and categorized. We propose a tabular structure named **Constraint Decision Table** to represent the spatial data integrity rule [Wang and Reinhardt, 2006a; Wang and Reinhardt, 2007].

The structure and syntax of a constraint decision table is described in Figure 18. Corresponding to the ECA rule, a constraint decision table contains three main parts "event, condition and action", with the description and specification for each part. The terms used in the constraint decision table are separately explained in the following.

	Description	Specification
Event	FeatureType.Operation	True/False
Condition	FeatureType1.spatialRelationship. FeatureType2	Numerical, String or character operators
	FeatureType.Operation	
Action	InstructionDialog	Text
	FeatureType.Operation	True/False

Figure 18: Constraint decision table structure

1. **Operation**:

Operational terms are defined to express specific actions for geographic data. An operational term is connected with a feature type by a "." connector, like FeatureType.operation. Operations are given as follows:

- **updateGeometry:**

Update the geometry of an existing feature, e.g. Ditch.updateGeometry means to update the geometry of a ditch feature.

- **updateAttribute:**

Update the attributes of an existing feature, e.g. Ditch.updateAttribute means to update the attribute values of a ditch feature, like to remeasure the depth of a ditch.

- **addFeature:**

Add a new feature to the datasets, e.g. Ditch.addFeature means to add a new ditch feature to the geospatial database.

- **deleteFeature:**

Delete an existing feature from the datasets, e.g. Ditch.deleteFeature means to delete an existing ditch feature from the geospatial database.

- **setAttributeValue(AttributeName):**

Set a value to a given attribute name, e.g. Ditch.setAttribute(depth) means to set a new value to the attribute “depth” of a ditch feature.

- **getAttributeValue(AttributeName):**

Get the value from a given attribute name, e.g. Ditch.getAttributeValue(depth) means to read the value of the attribute “depth” of a ditch feature.

2. **SpatialRelationship**:

Two kinds of spatial relationships are adopted herein: Topological relationship and Metric relationship. These relationships are used between two feature types by a connector “.”, for example, FeatureType1.spatialRelationship.FeatureType2 means the spatial relationship between two feature types.

In practical GIS applications, spatial relationships always contain semantic information, for example the topological relationship between a lineString and a polygon are given specific meanings in the context of its application like “the topological relationship between the ditch (polygon) and the hiking trail (lineString)”. Moreover, the attributes information of a feature type also coexists with spatial relationships, so the above example is extended to “when the attribute ‘A’ has value ‘v’, the topological relationship between the ditch (polygon) and the hiking trail (lineString)”. Therefore, semantic information should be concerned when discussing spatial relationships.

- **Topological relationship:**

Topology is a branch of geometry, and deals with a particular set of geometry properties that remain invariant under certain transformations. Topological relationships are used for describing the state between two spatial objects. Egenhofer and Franzosa [Egenhofer and Franzosa, 1991] proposed a four-intersection model comparing the boundary and the interior of two regions A and B. An extension to the four-intersection model is the one called dimensionally extended nine-intersection matrix which adds the exterior of spatial objects [Egenhofer, 1993].

People can test plenty of topological relationships by using the 9-intersection matrix, but the disadvantage is that it is theoretical and does not have a corresponding natural language equivalent. For people like IT developers as well as GIS developers, using a language like "*select all features 'topologically within' a query bounding box"* is very attractive. In order to address the needs of such users, research has been carried out by [Clementini and Felice, 1995; Clementini and Felice, 1996] to define a set of named topological relationship predicates, which were also introduced into geographic information standards: OGC simple feature specification for SQL [OGC, 1999] and ISO/TC 211 standard 19125 Simple Feature Access [ISO, 2004]). In this thesis, we adopt eight predicates from the above standards: **equals, disjoint, touches, crosses, within, overlaps, contains and intersects**.

The eight predicates of topological relationships are based on the 9-intersection model. In the 9-intersection model, there are six groups of candidate relations: A/A, A/L, A/P, L/L, L/P, P/P. Here the P is used to refer to 0 dimensional geometries like Points, L is used to refer to one-dimensional geometries like LineStrings and A is used to refer to two-dimensional geometries like Polygons. Thus, topological relationships can be applied to different groups of candidate relations. An example of the different topological relationships is illustrated in Figure 19.

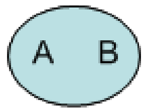
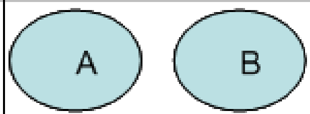
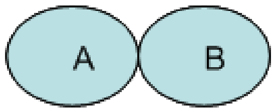
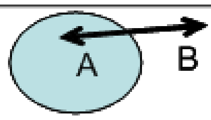
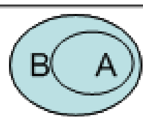
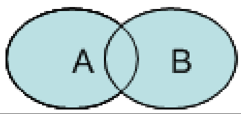
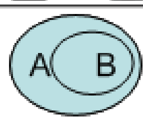
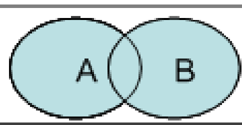
Topological relations	Examples
Equals (A/A, A/L, A/P, L/L, L/P, P/P)	
Disjoint (A/A, A/L, A/P, L/L, L/P, P/P)	
Touches (A/A, A/L, A/P, L/L, L/P)	
Crosses (A/L, A/P, L/L, L/P)	
Within (A/A, A/L, A/P, L/L, L/P)	
Overlaps (A/A, L/L, P/P)	
Contains (A/A, A/L, A/P, L/L, L/P)	
Intersects (A/A, A/L, A/P, L/L, L/P, P/P)	

Figure 19: Topological relationship predicates in OGC and ISO standards

- **Metric relationship:**

A metric relationship is defined as *distanceTo* relating to the distance. This term has the same meaning as the one called "Distance(anotherGeometry:Geometry):Double" specified in OGC simple feature specification for SQL [OGC, 1999], which returns the shortest distance between any two Points in the two geometric objects as calculated in the spatial reference system of this geometric object. It can be applied to P/P, A/A, A/L, A/P, L/L and L/P. For example, Figure 20 shows the metric relationship between a hiking trail and a ditch as *Ditch.distanceTo.HikingTrail = 5 (meters)*.

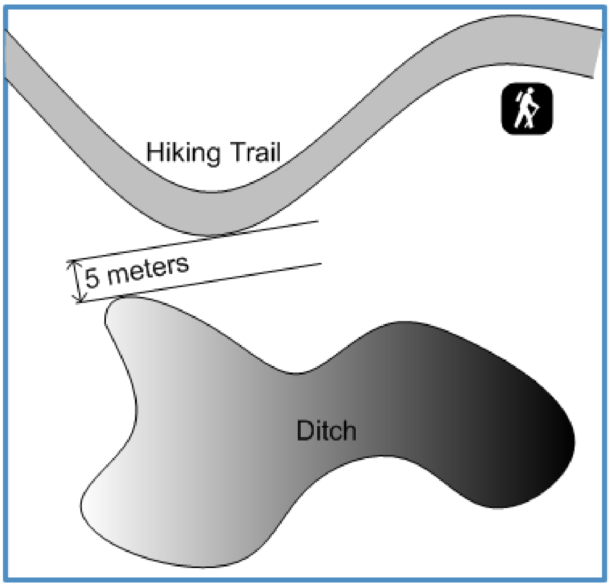


Figure 20: A distanceTo relationship example

3. InstructionDialog:

This is put under the column of "Action" for denoting the instruction message of the action part of the spatial data integrity rule. When end users get the action message of the rule, they should know unambiguously how to execute the action. For example, when a field user is collecting geographic data, a spatial data integrity rule may be violated due to data inconsistency reasons, and thus he gets an action message that tells him how to correct errors like *Road.updateGeometry.* The field user may not understand why and how he has to take this action, so confusion will lead to wrong decisions such as ignoring this action message or taking another different action.

Therefore, it is necessary to provide for the users, clear and helpful information about the spatial data integrity rule and how to execute the actions. The instructionDialog term is a specific action, which gives users a detailed description about the reasons of the rule violation, the priority of rules and steps to correct errors. The content of the instructionDialog is acquired from data modeling experts and domain experts. This improves the confidence of actual users.

4. Logical Relationships for Events, Conditions and Actions:

If more than one event, condition or action element exists in the constraint decision table, logical relationships are defined through logical operators "and" and "or". In order to keep the constraint decision table compact, we rule the expressions as:

- If two elements have "and" logical relationship, they are grouped with same number followed by a different letter, e.g. "condition1a" and "condition1b" mean both condition1a and condition1b must be satisfied.
- If two elements have "or" logical relationship, they are grouped with different numbers, e.g. "action1" and "action2" mean either action1 or action2 will be executed.

5. Description:

It includes the contents of the event, condition and action like operations and spatial relationships.

6. Specification:

It contains results of expressions in the description column, for instance, True/False, or a value with a Numerical, String or Character Operator as represented in Table 7.

Table 7: Numerical operators and String/Character operators

Numerical operators	String or Character operators
EqualTo or =	Starts
GreaterThan or >	HasSubString
LessThan or <	Ends
GreaterThanAndEqualTo or >=	
LessThanAndEqualTo or <=	
NotEqualTo or !=	

Therefore, a simple spatial data integrity rule as in the landslide application "*when updating the geometry of ditches, if the ditch feature intersects with a hiking trail feature, then one possible action is to update the geometry of the hiking trail feature*" can be expressed in a constraint decision table as shown in Table 8.

Table 8: A simple constraint decision table example

	Description	Specification	
Event	Ditch.updateGeometry	True	
Condition	Ditch.intersects.Road	True	False
Action1a	Road.updateGeoemtry	True	False
Action1b	InstructionDialog	"Remeasure the Road object, and split its geometry as two separate parts!"	False

3.4.4 Extension of Constraint Decision Table Terms

The terms defined for the constraint decision table are not limited to those mentioned here. People can define new terms according to their needs in particular GIS applications. Statements already included in the international standards are encouraged to be adopted to provide a commonly accepted knowledge base for potential data users. For example, the topological relationships above are adopted from OpenGIS Simple Features Specification for SQL. However, the standardization work does not afford enough requested matters based on different user requirements, so extra definitions need to be implemented.

In this subsection, three examples demonstrate the possibilities of the extensible constraint decision table terms, for topological relationship, direction relationship and for the term “updateGeometry”.

1) Topological relationship extension

The operators in OGC simple feature specification for SQL [OGC, 1999] and ISO/TC 211 standard 19125 Simple Feature Access [ISO, 2004] deal with the topological relationships in a general manner. In order to express special cases for certain operators, additional notations need to be offered. Vallières et al [Vallières et al., 2006] proposed three extension predicates: **tangent, borders and strict**, which are combined with the predicates **touches, within and contains**. For example, a refinement of *within* predicates for a polygon is shown in Figure 21. The left diagram represents three different cases for *within* topological relationship. The right diagram depicts the refinement of *within* predicate with three extensions. In this way, we describe topological relationships more specifically and precisely.

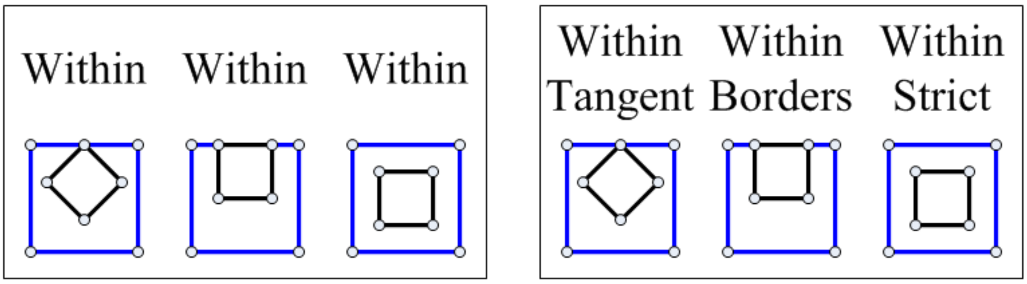


Figure 21: Topological predicates with tangent, borders and strict extensions [Vallières et al., 2006]

2) Direction relationship

Some research proposed the directional relationship in order to represent the vague relationship relating to the directions between features. For example, predicates like “left, above, beside, east, south, west, north” were discussed by Peuquet and Zhang [Peuquet and Zhang, 1987] and Frank [Frank, 1996]. For example, Figure 22 shows the directional relationship between two cities.

When data users would like to describe the relative directions between features in a rough meaning, they can adopt this concept. For example, it can be termed as “*directionTo*”, so the example in Figure 22 is expressed as *City_Nürnberg.directionTo.City_München = north*.

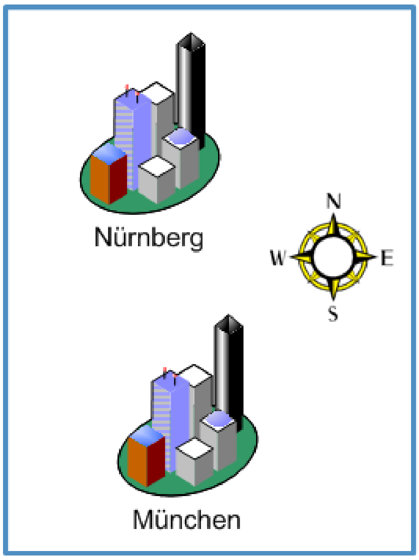


Figure 22: A directionTo relationship example

3) UpdateGeometry extension

The term "updateGeometry" contains many different interpretations, such as the detailed description about how to correct geographical data errors. For example, Ubeda and Servigne [Servigne et al., 2000; Ubeda and Egenhofer, 1997] suggested several points for modifying topological errors as:

- **Object modification**:
  - Moving the objects: The surface area of objects remains unchanged.
  - Reshaping the objects: Moving a part of an object and leaving the other part unchanged.
- **Deleting one object**: remove the object. E.g., if an object has been digitized or measured twice.
- **Object splitting**: split one object into several parts, or create a new part.

Besides these suggestions, experiences are borrowed from other sources. In particular, Computer Aided Design (CAD) systems such as AutoCAD and Microstation offer comprehensive and powerful cases for dealing with the geometries of objects. For instance, the "align" command allows people to adjust two objects to specific positions, such as to the parallel or perpendicular positions. In this way, the term "updateGeometry" can be extended with new terms such as: **Move, Reshape, Split, and Align**.

To define new terms in detail, some parameters may be used. For example, a rotation parameter may be used to describe the rotation angle when changing the geometry of a feature.

To sum up, the above three examples demonstrate the possible way of creating more terms in the constraint decision tables to meet different user requirements. However, user requirements vary a lot depending on different GIS applications. Therefore it is very hard to discuss all kinds of user requirements here, and to provide all potential terms that might be used in the constraint decision tables in this thesis. People who are interested in this topic can follow the suggested way to add their own needed terms.

3.4.5 Constraint Decision Table Category

The constraint decision table provides an explicit and structured way for not only defining the spatial integrity constraints proposed by [Cockcroft, 1997; Ubeda and Egenhofer, 1997], but also including the critical instructionDialog information which has been proved as indispensable for many GIS users. To show the coverage of the constraint decision table on geographic data, we use the classification method for spatial database constraints as introduced in [Ditt et al., 1997] to categorize the constraint decision table as:

- All instances of a feature type "A".
- A specific instance "a" of a feature type "A".
- An attribute of a feature type "A".
- All instances of different feature types "A and B", or even more than two feature types.
- An instance "a" of feature type "A" and another instance "b" of feature type "B" or even more than two instances from different feature types.
- Spatial relationships between two feature types or even more than two feature types.
- Spatial relationships involving semantic and temporal information between two feature types or even more than two feature types.

3.4.6 Policy for Spatial Data Integrity Rules' Violation

Spatial data integrity rules are defined to check the data consistency and validate the data. This means if one of the rules is violated, a certain action or actions might be performed under different situations. Depending on different real world situations, the priority of the rules might be implemented differently.

In this thesis, we simply adopt the traffic light policy for rule violations. Three types of priority values are defined as "**red**", "**yellow**" and "**green**". Three different values guide users to perform corresponding actions as shown in Figure 23.

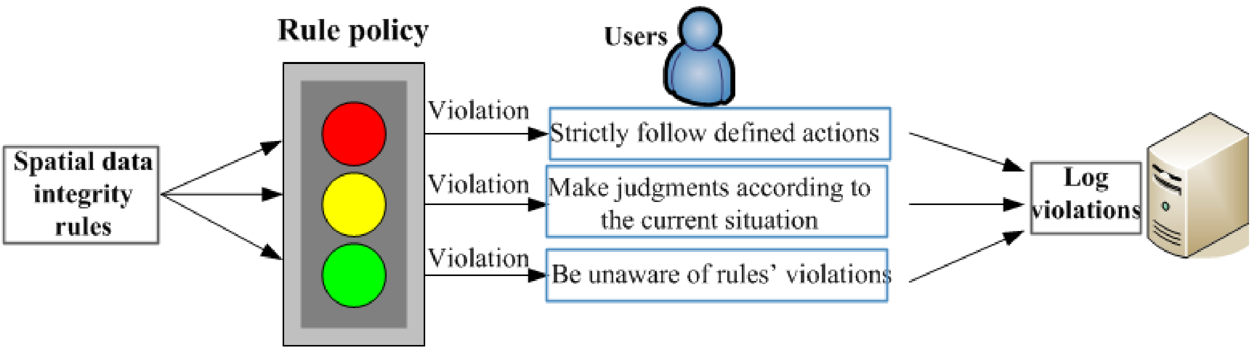


Figure 23: The policy of spatial data integrity rules' violation

It is apparent that the data integrity rules labeled with the "red" means if this rule is violated at any step of the data process, users should unconditionally correct this error or follow the suggested action defined in the rule because there is no compromise.

Spatial data integrity rules marked with “yellow” describe if the rule is violated, the user is free to adjust the result, but he should write down the current situation and may follow the suggested actions in the rule. In the mean time, the system records this event automatically in the log file.

The last one marked with “green” does not provide the user warning information when the rule is violated. But the system will record the on-site information automatically in the log file for future checking. This one is similar with “yellow”, but the difference is the “yellow” rule reminds the user of the rule’s content and expects that he makes own judgments according to the on-site situation and comments in the rule.

Thus, every spatial data integrity rule may be marked with one of three priority values. Accordingly, different priority values for the same spatial data integrity rule denote different meanings. An example of a spatial data integrity rule with “red” or “yellow” is provided to show the differences.

In the landslide area, a real world scene is like in Figure 24. There are ditches and roads objects in the landslide area. A road may be accessible for the public, e.g. a hiking trail for hikers and mountaineers. A ditch may become bigger and wider due to land movement as time elapses, then it is closer and closer to the road object. Visibly, if these two objects are too close, it is dangerous for the people who walk on the road. For the field user, it might be necessary to take certain actions to avoid the dangers.

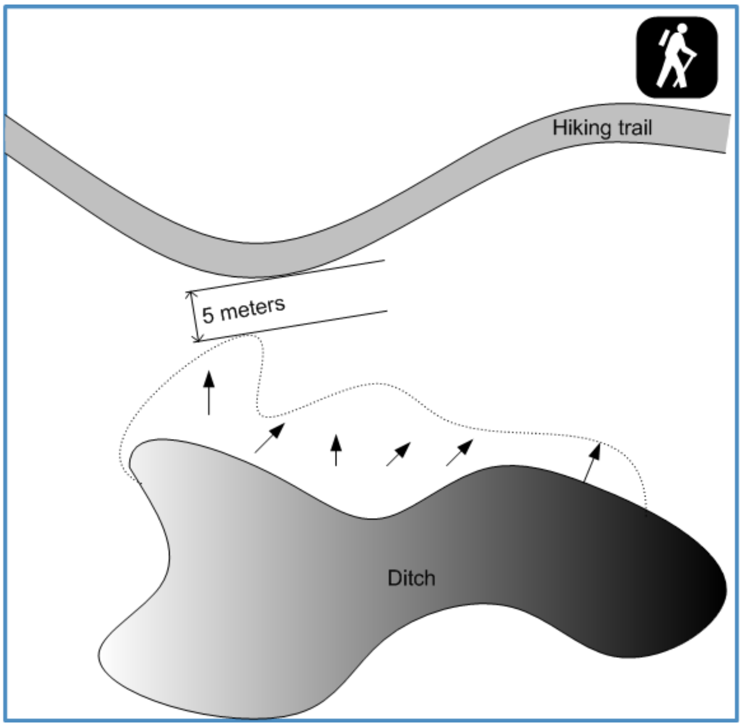


Figure 24: An example in a landslide scene: Ditch and Hiking trail

Thus, a spatial data integrity rule for the above situation is narrated as:

“*when update the geometry of the ditch feature, if a hiking trail feature has distance less than 5 meters to a ditch and the hiking trail is supposed to be publicly accessible, then update the geometry of the hiking trail or set its publicAccessibility attribute value to False*”. Table 9 shows how this example is described by a constraint decision table.

Table 9: An example in a landslide scene: Ditch and Hiking trail

	Description	Specification			
Event	Ditch.updateGeometry	True			
Condition1a	Ditch.distanceTo.HikingTrail	<=5 (meters)		>5 (meters)	
Condition1b	HikingTrail.getAttributeValue(PublicAccessibility)	True	False	True	False
Action1	HikingTrail.updateGeometry	True	False	False	False
Action2	HikingTrail.setAttributeValue(PublicAccessibility)	False	False	True	False

When this rule is violated, field user's actions depend on different policy values. If "red" policy value is applied to this rule, the field user should strictly follow one of suggested actions, e.g. either to update the geometry of the road like moving, or to change its attribute value. If "yellow" policy value is adopted, the field user gets the warning information about this rule violation. But he need not take the suggested actions in the rule. He can make own decisions according to current real world situation, e.g. he does nothing because fences or other protection devices were already installed in this area, or he wants to place a warning sign at the road entrance.

The traffic light policy provides a good foundation for dealing with rule violations. Three simple choices guide users on what to do. However, for complex rule system or decision system, this policy is too simple to handle all cases. For example, the violation of one rule may trigger the different priority of another rule, or the violation of a rule may result in loop actions to other rules. In this thesis, we do not investigate more intricate rule violation policies.

3.5 Consistency between the UML Conceptual Schema and Spatial Data Integrity Rules

The extended UML conceptual schema (see subsection 3.3) and spatial data integrity rules (represented by the constraint decision table, see subsection 3.4) are integrated to represent the conceptual data model in the proposed modeling methodology. Therefore their content, structure and semantics are required to be consistent, e.g. the feature type names, attribute types and geometry types used in the extended UML conceptual schema have to coincide with the ones defined in constraint decision tables.

To prevent conflict and inconsistency in the whole conceptual data modeling work, four principles are established to control the cooperative framework between the extended UML conceptual schema and spatial data integrity rules:

- Semantic match: shows the match between the meanings of the terms used in both sides. E.g. if a feature type "Ditch" is defined in the UML conceptual schema, then corresponding terms used in the constraint decision table should not be "Trench" or "Spalten" (German word for a ditch) but the exact term "Ditch".

  In case people want to use different terms, for example, when multilingual information needs to be provided at one part of the conceptual model, a semantic dictionary file is put forward to resolve this problem. The dictionary file offers the function to check for different terms with same semantic meanings in the conceptual model. The dictionary file is a text file containing relating terms and detailed explanations, e.g. the terms ("Ditch" ←→ "Spalten") in the dictionary file with the comments "*Spalten is the corresponding word for the Ditch, which means a long, narrow or deep furrow on the ground. Here it means the natural phenomena caused by land movements especially in a landslide area".* These are used in both UML conceptual schema and constraint decision tables, because humans or computers can find matching information from the dictionary file.

- Geometry type match: describes the match of the geometry type of a feature type. E.g. if the Ditch feature type is defined as Polygon geometry type and the Extensometer is defined as Point geometry type in the UML conceptual schema, a spatial data integrity rule including the topological relationship "Crosses" between Ditch and Extensometer will cause inconsistency, because the topological relationship "Crosses" does not apply to the Polygon/Point elementary case. So either the Extensometer is defined as a lineString feature or the spatial data integrity rule should not use the "Crosses" topological relationship between Ditch and Extensometer.
- Temporal type match: represents the match of the temporal type of a feature. E.g., the Ditch feature type is a durable temporal type in the UML conceptual schema implicitly means it has "start_date" and "end_date" attributes, which are allowable in the spatial data integrity rule.
- Attribute data type match: shows the match of the attribute data types of a feature type. E.g., the Ditch feature has an attribute "depth" with a double value defined in the UML conceptual schema, so in the spatial data integrity rule, the descriptive part of this attribute can have only Numerical operators not String operators.

The four principles should be ensured when a data modeler creates the UML conceptual schema and spatial data integrity rules as an integrated framework. In the future, a computer aided tool might be necessary to help data modelers check those principles and find inconsistencies between the rules. As mentioned, the graphical user interface for defining spatial integrity rules as well as checking the consistency of rules are beyond of the scope of this thesis.

3.6 Spatial Data Integrity Rules and ISO/TC 211 Data Quality Elements

Existing methods for considering spatial data integrity rules only focus on logical consistency, however, the real world phenomena are always complex and consist of different other data consistency requirements such as semantic and temporal information. It is very hard to classify spatial data integrity rules by a single data quality element or subelement.

Therefore, this subchapter discusses the detailed spatial data quality elements and subelements, and shows the possibility of applying them to the definition of spatial data quality rules.

The comparison of data quality elements definition from different standardization work shows that ISO/TC 211 standards provide the comprehensive definition of spatial data quality elements. In ISO/TC 211 standard 19113-Quality principles, besides three data quality overview elements **Purpose, Usage and Lineage,** data quality elements and subelements are represented in Table 10.

Table 10: ISO/TC 211 data quality elements and subelements

Data quality elements	Data quality subelements
Completeness	commission
	omission
Logical consistency	conceptual consistency
	domain consistency
	format consistency
	topological consistency
Positional accuracy	absolute or external accuracy
	relative or internal accuracy
	gridded data position accuracy
Temporal accuracy	accuracy of a time measurement
	temporal consistency
	temporal validity
Thematic accuracy	classification correctness
	non-quantitative attribute correctness
	quantitative attribute accuracy

Although this standard offers above extensive data quality elements and subelements, no detailed and explicit explanations for them are provided. Many quality elements are only specified as free texts without expanding. To correctly interpret those elements and define corresponding spatial data integrity rules, their particular meanings need investigations.

In the remaining subsections, quality elements and subelements are explained separately and diverse examples are given. Finally, a conclusion is achieved to show the propriety of defining spatial data integrity rules for each quality element and subelement.

3.6.1 Data Quality Overview Elements (Non-quantitative Quality Elements):

Three overview elements are defined to provide general, non-quantitative quality information:

- **Purpose**: "shall describe the rationale for creating a dataset and contain information about its intended use." You will always find the purpose information when you search geospatial datasets in the internet. It tells you the basic information of datasets, for example, datasets deployed in a OGC Web Feature Service for students' practices and exercises can be described as:

  Purpose: "Datasets are intended to support students to do practices on the OGC Web Feature Service."

- **Usage**: "shall describe the application(s) for which a dataset has been used. Usage describes uses of the dataset by the data producer or by other, distinct, data users." For instance, the usage of above datasets is described as:

  Usage:

"Usage 1: for providing the information to the Europe population statistics research

Usage 2: for providing the information to the Europe shoreline investigation"

- **Lineage**: "shall describe the history of a dataset and, in as much as is known, recount the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form." Two unique components are defined in ISO/TC 211 standard 19113: **source information** shall describe the parentage of a dataset; **process step or history information** shall describe a record of events or transformations in the life of a dataset, including the process used to maintain the dataset whether continuous or periodic, and the lead time.

  Clarke and Clark [Clarke and Clark, 1995] gave detailed introduction of lineage element which involves the following:

  - Source of data: origin, reference fields, spatial data characteristics, coordinate systems, map projections, and associated corrections and calibrations
  - Acquisition, complication and derivation: they correspond to four different data stages, data collection stages, parameter generations stage, data conversion stage and product stage.
  - Conversion of data: the detailed steps of conversion to the secondary format.
  - Transformation and analyses of data: all parameters during transformation have to be clearly described.
  - Date information of data processing stages.

  For example, previous WFS datasets are depicted in term of ISO/TC 211 standard 19113:

  Lineage:

  "Source: WFS datasets for 'States' and 'Cities' are from Tiger data of U.S. census bureau.

  Process step: Direct usage without processing.

  Source: WFS dataset for 'beer garden' is from the free KML dataset.

  Process step: The KML dataset is converted to the ESRI shape file by using ArcScript. Process was completed in April, 2006."

These three data quality overview elements are vital information for data users, because they concern different data processing stages and record fundamentals of data quality. Although the collection of them is an onerous and difficult task, they are still most useful and indispensable.

3.6.2 Completeness

In ISO/TC 211 standard 19113, Completeness is defined as "presence and absence of features, their attributes and relationships", and it has sub elements **commission** (excess data) and **omission** (absent data). At the beginning, Morrison [Morrison, 1988] defined it as a fundamental data quality component which describes "the relationship between the objects represented in the dataset and the abstract universe of all such objects". Later, Brassel [Brassel et al., 1995] proposed more detailed definition as "Completeness describes whether the entity objects within a dataset represent all entity instances of the abstract universe. And the degree of completeness describes to what extent the entity objects within a dataset represent all entity instances of the abstract universe."

Implicitly, above definitions have two meanings based on the variable abstract universe and therefore there are two completeness terms (see Figure 25). First one means the abstract universe is specified through data capturing rules and the measure of data completeness is application independent. It is called data completeness. The second one means the abstract universe is determined by user's requirements when fitness of use has to be assessed. In this context, the model completeness is named as the comparison between the abstract universe corresponding to the dataset and the one corresponding to the application [Brassel et al., 1995]. Model completeness is strongly application-dependent, and has to be assessed separately. Data completeness is then classified into:

- **Formal completeness**: concerns the formal data structure, adherence to data standards and formats used. For example: "If the dataset is GML 2.1 data format, all the data must contain the mandatory information which defined in GML 2.1 standard."
- **Entity object completeness**: specifies to what degree the abstract universe is present in the data. For example: "whether all ditches that meet the requirements in the specification are represented in the dataset."
- **Attribute completeness**: is subordinated to object completeness and normally expresses missing attribute values of objects. For instance: "whether the attribute value of 'id' for a ditch is null."

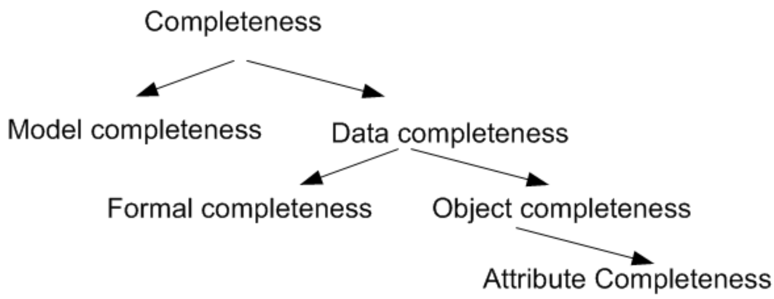


Figure 25: Different types of completeness [Brassel et al., 1995]

3.6.3 Logical Consistency

It is defined in ISO/TC 211 standard 19113 as "degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical)."

Logical consistency deals with logical rules of structure and attributes for spatial data and describes the compatibility between dataset items. This notion was used earlier in data integrity checks for non-spatial data. Its use for geographic data started when first analyses in the domain of topology [Servigne et al., 2006]. There exist several different levels of logical consistency according to its definition, thus it is subdivided into four components: **conceptual consistency, domain consistency, format consistency and topological consistency**.

First three components also apply to non-spatial data, but the last one *topological consistency* is a specific characteristic to geographic data. These are explained in the following, and the emphasis is put to topological consistency.

- **Conceptual consistency**: adherence to rules of conceptual schema. The data must follow the information stated in a conceptual schema. For example, a ditch defined in the conceptual schema as shown in Figure 16 has the polygon geometry type, so collected ditches must be polygon features.
- **Domain consistency**: adherence of values to value domains. The values of data have to be accordant with value domains, for instance, values of attribute “category” of the road feature type may be defined: hiking trail, federal road, and highway. This restricts the “category” within three values.
- **Format consistency**: degree to which data is stored in accordance with the physical structure of the dataset. For example, geographic data should be stored as the ESRI shape file format. Even some standards may consider the reliability of the medium on which the file is stored.
- **Topological consistency**: correctness of the explicitly encoded topological characteristics of a dataset. For example, every arc of the same network should be connected by a node to another arc. To check the topological consistency, topological relationships of two spatial objects are needed. The knowledge of topology and topological relationship were explained already in Section 3.4.3. However, topological relationship is only one of the spatial relationships, so metric relationships in terms of distance and direction are also introduced in this thesis.

When defining spatial data integrity rules, pure topological consistency subelement is not used alone. Semantic information of geospatial data is always combined together with topological relationships. For example, in Figure 26, the left diagram shows a lineString intersects a polygon, which may not cause the topological inconsistency. However, if in a landslide GIS application, the lineString represents a hiking trail feature and the polygon represents a ditch feature as illustrated in the right diagram, topological inconsistency happens. The left diagram has only the topological information, but the right diagram concerns the meaning of geographic features too.

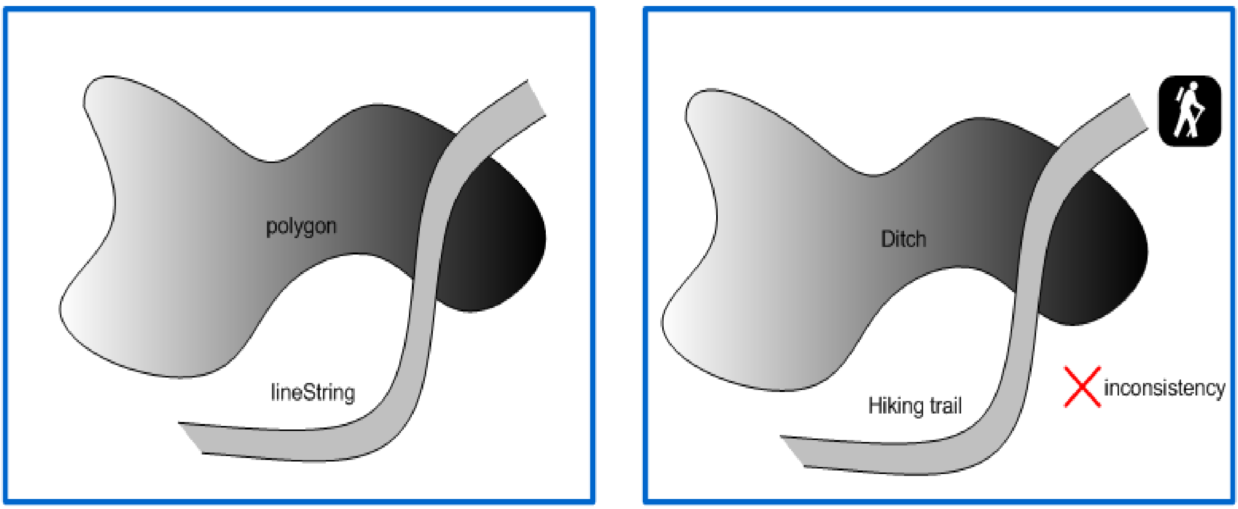


Figure 26: Topological and topo-semantic consistency

The above example "*when updating the geometry of ditches, if a ditch feature intersects with a hiking trail feature, then one possible action is to update the geometry of the hiking trail feature*", which contains the topological relationship and semantic information of two feature types, can be expressed in the constraint decision table as shown in Table 11.

Table 11: Constraint decision table example: topo-semantic information

	Description	Specification	
Event	Ditch.updateGeometry	True	
Condition	Ditch.intersects.HikingTrail	True	False
Action	HikingTrail.updateGeometry	True	False

To more explicitly guide the user to correct this kind of error, detailed "Action" operations and instructions need to be provided. As mentioned in section 3.4.4, the term "updateGeometry" can be extended by new terms such as: **Move, Reshape, Split and Align**. One reasonable detailed action is to split this lineString feature as two parts, and remove the parts within the ditch feature as shown in Figure 27.

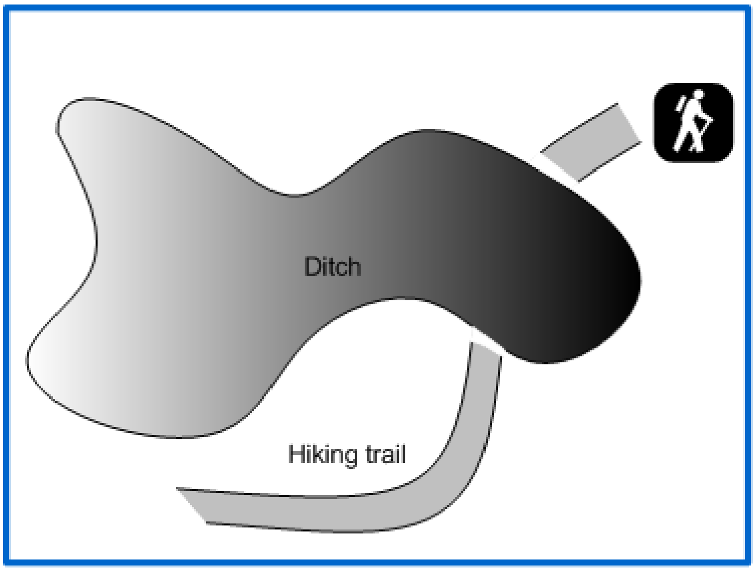


Figure 27: Action operation in constraint decision tables: split the geometry

In this case, the above constraint decision table is expanded as in Table 12. Furthermore, depending on different user requirements, the split operation can be defined with specified parameters, for example the distance between the ditch and the divided hiking trails.

Table 12: Constraint decision table example extension: topo-semantic information

	Description	Specification	
Event	Ditch.updateGeometry	True	
Condition	Ditch.intersects.HikingTrail	True	False
Action1a	HikingTrail.split()	True	False
Action1b	InstructionDialog	"Remeasure the hiking trail, and split its geometry to two separate parts."	null

To show the use of proposed metric relationships, the example representing *distanceTo* relationship in Figure 20 is described as a spatial data integrity rule: "*when updating the geometry of ditches, if a hiking trail's distance to the ditch is less than 5 meters, then move the hiking trail and tell users 'the hiking trail is dangerous for walking'*" is expressed by constraint decision table in Table 13.

Here, we do not discuss details about the parameters used in the operation terms. In this example, user can define the moving distance or direction in the context of the current situation. To understand the capability of extending the constraint decision table with more terms, please refer to section 3.4.4.

Table 13: Constraint decision table example with metric spatial relationship

	Description	Specification	
Event	Ditch.updateGeometry	True	
Condition	Ditch.distanceTo.HikingTrail	<=5 (meters)	>5 (meters)
Action1a	HikingTrail.move()	True	False
Action1b	InstructionDialog	"The hiking trail is dangerous for walking, because it is too close to a ditch!"	Null

3.6.4 Positional Accuracy

It is shortly defined in ISO/TC 211 standard 19113 as "accuracy of the position of features". The position of features in the database is a set of cardinal values that allow objects to be positioned in Cartesian or polar coordinates. If we only consider two dimensional data, for a point feature it has x and y coordinates to represent its location in the real world. Sometimes the third dimension altitude is stored as an attribute value of features, then it becomes to the issue of attribute accuracy.

To measure positional accuracy, we have to compare the dataset with the "true" dataset, where "true" dataset means "control" or "referenced" dataset with better quality. Position information is achieved through different instruments that measure distances, angles or time or some combination of these such as leveling staves, tapes, airborne, digitizers, theodolites, total stations and GPS receivers. Boot Mean Square Error for calculating accuracy and Standard Deviation for precision have a long history in the mapping sciences for representing position quality. The methods of determination of positional accuracy like using variance propagation mathematical theories were explained by Drummond [Drummond, 1995]. Positional accuracy was also discussed in many other GIS studies [Caspary and Scheuring, 1993; Veregin, 1999].

ISO/TC 211 standard 19113 divides positional accuracy into three types:

- **Absolute or external accuracy**: closeness of reported coordinate values to values accepted as or being true. For example, the accuracy of the ditch feature should be no more than plus or minus 0.5 meter.
- **Relative or internal accuracy**: closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true. It is a measure of the positional consistency of a data point in relation to other near points of detail. Relative accuracy compares the scaled distance between features measured from the map data with distances measured between the same features on the ground. It is always used in local area purposes, e.g. an example from Ordnance Survey "If the distances between two well defined points of detail 60.0 meters apart were measured in the real world, there would be an expectation that this distance would be represented in the map data by a scaled distance of between 59.2 meters and 60.8 meters."
- **Gridded data position accuracy**: closeness of gridded data position values to values accepted as or being true. This element relates to the gridded data. For example Fekete et al [Fekete et al., 2001] investigated the gridded data position accuracy based on gridded networks for macroscale hydrology.

3.6.5 Temporal Accuracy

"Accuracy of the temporal attributes and temporal relationships of features" describes temporal accuracy in ISO/TC 211 standard 19113. As geospatial data are being increasingly shared, produced and interchanged in the GIS domain. A single application may use heterogeneous data coming from different parties. To know whether the data is outdated or up-to-date, temporal information has to be provided, such as when a feature was collected or revised and when it expires. Although up-to-date data are most required by field users, history data are also indispensable for many GI researches like reconstruction of period of data in order to find the regular changes.

Based on the time types distinguished in [Servigne et al., 2006], we propose four kinds of time concepts:

- **Logical or factual time**: represents the dates on which the phenomenon in the reality took place. Sometimes, logical time is recorded when a specific event is triggered automatically, e.g. if the height level of river reaches a certain value, the time will be immediately saved.
- **Time of observation of phenomenon**: sometimes is different from logical time since the data collection preparation took time as well. Normally, it is stored when the surveyor performs data collection task.

- **Transfer time**: means a period of time that the data was transferred from observation result to database. It has a durable value which contains the start time and end time of data transferring, e.g. data was collected in the field and then uploaded to a remote database. Transfer time may take seconds or hours or even more, which mostly depends on the dataset size and network connection. It is used for some online data transmitting instruments, e.g. extensometers in the landslide area are installed to transfer the data to a remote database through wireless connection. Moreover it is often considered for the new field data capture method through spatial web services, e.g. using OGC WFS-T to insert, update and delete data of remote databases.
- **Transactional time**: indicates the time the data was entered into database. It should equal the summary of observation time and transfer time.

Temporal information can be used to describe geographic features as well as their attributes and geometries, such as the UML conceptual schema with temporal dimension in Figure 16.

Three kinds of sub elements are given in ISO/TC 211 standard 19113.

- **Accuracy of a time measurement**: correctness of the temporal references of an item (reporting of error in time measurement). It generally depends on the time recording system, e.g. computer clocks are used to display the time to computer users and can be accessed by computer programs. The temporal accuracy varies with the computer hardware and software. Most computer clocks' accuracy is plus or minus 5 to 15 second per day. A usual way to improve its accuracy is synchronizing it with the atomic clock through the Internet.
- **Temporal consistency**: correctness of ordered events or sequences, if reported. The time information for spatial component should be in a correct sequence. E.g. if the start time of a ditch feature is later than its end time, temporal inconsistency happens.
- **Temporal validity**: validity of data with respect to time. It means whether the data expires or not. For example, a road segment in München is valid until 06.11.2008, since this road segment needs to be removed because the parcel is reserved for creating a building according to the city plan.

3.6.6 Thematic Accuracy

"- accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships" is declared in ISO/TC 211 standard 19113.

As described above, thematic accuracy is used to deal with attributes information of features, so sometimes it is also named as "Attribute accuracy". Goodchild [Goodchild, 1995a] defined attribute as a fact about some location, set of locations, or feature on the surface of the earth in the context of the GIS domain. The fact can be the result of measurement with different kinds of surveying instrument, or the result of interpretation by a trained observer, such as a land use or soil class, or the outcome of historical or political consensus such as names given to roads. Attributes are used to distinguish one location or set of locations from another or one feature from another, and they are expressed by various data types such as Boolean, integer, nominal, ordinal, and real data types.

Thematic inaccuracy happens since attributes can be complex and uncertain. For example, inaccuracies of surveying instruments may lead to bad measurements, historical uncertainty may produce confusion of road name assignment or misinterpretation of observer can cause wrong feature classification.

In order to evaluate thematic accuracy, it is necessary to develop a classification of scales of measurement [Goodchild, 1995a]. This classification contains four types of scales which apply to both quantitative and non-quantitative (qualitative) attributes:

- **Nominal scales**: is used to distinguish different things, quantitative attribute like a class numbered by 4 is distinct from one numbered by 3 and non-quantitative attribute like residential zone or industrial zone of a class.
- **Ordinal scales**: is to assign order or rank, such as land suitability level: 1, 2, 3, etc (quantitative), or criminal level: high, middle, low (non-quantitative).
- **Interval scales**: the Celsius scale of temperature is interval, since the change between 10 and 20 is the same as between 20 and 30.
- **Ratio scales**: the Kelvin scale of temperature is ratio, since a temperature of 200K = 2*100K. The Celsius scale is not ratio since it is based on an arbitrary zero point, which means 200°C is not two times of 100°C.

Interval and Ratio scales are strictly quantitative attributes and Nominal and Ordinal scales can be either quantitative or non-quantitative attributes. Followed by the classification of ISO/TC 211 standard 19113, which differentiates attribute accuracy as three types.

- **Classification correctness**: comparison of the classes assigned to features or their attributes to a universe of discourse. For example, during digitizing of aerial photograph two different interpreters may assign a feature to different feature types.
- **Non-quantitative attribute correctness**: correctness of non-quantitative attributes. For example, if the name of a road is correctly identified.
- **Quantitative attribute accuracy**: accuracy of quantitative attributes, e.g. for road features, the attribute value of width is numeric and bigger than zero.

3.6.7 Summary of the ISO/TC 211 Quality Elements and Subelements

Through the detailed study of ISO/TC 211 standard 19113 Quality principles, all defined data quality elements and subelements are explained in detail. Consequently, for each quality element, whether it is reasonable to be applied to the definition of spatial data integrity rules is concluded in Table 14.

Table 14: The propriety of spatial data integrity rules for each data quality element

Data quality overview elements and data quality elements	Applicable
1. **Purpose**	No
Usage	No
Lineage	No
2. **Completeness**	No
3. **Logical consistency** conceptual consistency domain consistency format consistency topological consistency	Yes Yes Yes Yes Yes
4. **Positional accuracy**	No
5. **Temporal accuracy** accuracy of a time measurement temporal consistency temporal validity	Partial No Yes Yes
6. Thematic accuracy classification correctness non-quantitative attribute correctness quantitative attribute accuracy	Partial No No Yes

1. Data quality overview elements (Purpose, Usage and Lineage):

Evidently, quality overview elements can not be defined and ensured by spatial data integrity rules because they are normally stored as documentation. Our suggestion is to enhance GI systems by providing a digital form containing three elements and sub parts with detailed explanations. Data producers are asked to fill in the form when producing data. The texts in the form should be automatically converted into metadata information in the final report.

2. Completeness:

For formal and attribute completeness subelements, existing GIS software or its extended functions are able to avoid the incompleteness easily, such as checking the data format and missing attribute values. For entity object completeness subelement, an experienced surveyor is the key to reduce incompleteness. Moreover, well-designed software GUI and data collection procedures also greatly reduce observer's bias and improve the data capture result.

For example, when field users collect data from the real world, GIS software allows the captured data immediately to be visualized and highlighted in the computer screen. Field users can see and check the captured data comparing with the reality. In this way, missing or excess objects can be found directly during the field task.

Of course, comparison of collected data with other datasets in the office is another way to check the completeness (e.g. using satellite images or aerial photos). However this costs more labors and time.

3. Logical consistency:

As previously shown, logical consistency especially the topological consistency containing semantic information is always ensured through spatial data integrity rules. Conceptual consistency, domain consistency and format consistency also can be expressed through integrity rules, but they are usually already described in the normal UML conceptual schema. Therefore, spatial data quality rules for checking those inconsistencies are not necessarily repeatedly defined.

4. Positional accuracy:

It is often improved via updating surveying instruments, getting average value from more times measurements and employing an experienced surveyor.

5. Temporal accuracy:

The accuracy of a time measurement relies on the time instruments. Spatial data integrity rules can be defined to ensure the temporal consistency and validity, such as the start_time value of a feature must be smaller than its end_time value like the example in Table 15.

Table 15: Constraint decision table: temporal inconsistency

	Description	Specification	
Event	Ditch.updateAttribute	True	
Condition1a	Ditch.getAttributeValue(start_time)	= t (YYYY-MM-DD,hh:mm:ss)*	
Condition1b	Ditch.getAttributeValue(end_time)	<= t	> t
Action	InstructionDialog	"start_time must be smaller than end_time, please give a reasonable value"	null

*time format complies with ISO 8601(2000) standard: Data elements and interchange formats – Information interchange – Representation of dates and times. E.g. 2007-06-06, 16:30:59Z, here Z stands for time zone of "zero meridian" which goes through Greenwich in London.

6. Thematic accuracy:

Classification and non-quantitative errors can not be ensured by spatial data integrity rules, because they are produced by the operators or observers who make decisions, e.g. one digitizing operator identifies an area for commercial land use and another one takes it as residential land use. Or those errors are caused by the uncertainty because the truth value in the real world is very difficult to be achieved. Quantitative attribute errors can be controlled by spatial data integrity rules such as attribute data types and attribute values range checking.

3.6.8 Interactions of Quality Elements/Subelements in Spatial Data Integrity Rules

In last subsection, the availability of spatial data integrity rules definition for ISO quality elements/subelements is described. However, in the real world, it is very difficult to make a clear border line for each data quality element, because real world phenomena usually comprise several different quality elements and subelements. Those data quality elements and subelements always interact with each other.

For example, if a ditch object in the reality is not represented in the database, the problem can be incompleteness because the field user ignored to measure it, or classification errors of thematic accuracy because the field user put it to the wrong feature type category, or temporal inconsistency of temporal accuracy because the dataset in the database is too outdated, or positional inaccuracy that the measured one is far from the reality.

When defining spatial data integrity rules, we have to consider all possible conditions although it is an expensive and time consuming task. Servigne [Servigne et al., 2006] also commented the information concerning quality should be relevant so that producers and end-users accept the rules that quality evaluation entails and understand its utility. Therefore, all possible related quality elements or subelements have to be concerned when defining spatial data integrity rules.

To show the complexity of real world phenomena, an example is provided. It is derived from a city plan. As we know, industry areas normally should be located far from the residential regions under a certain rule. In a city, the law stated that *“the distance between the chemical factory and residential region should be more than 300 meters from year 2003 to 2006*”. Then this law was revised to “*the distance should be more than 500 meters*” and would be valid for new chemical factories and residential regions from year 2007 to 2009.

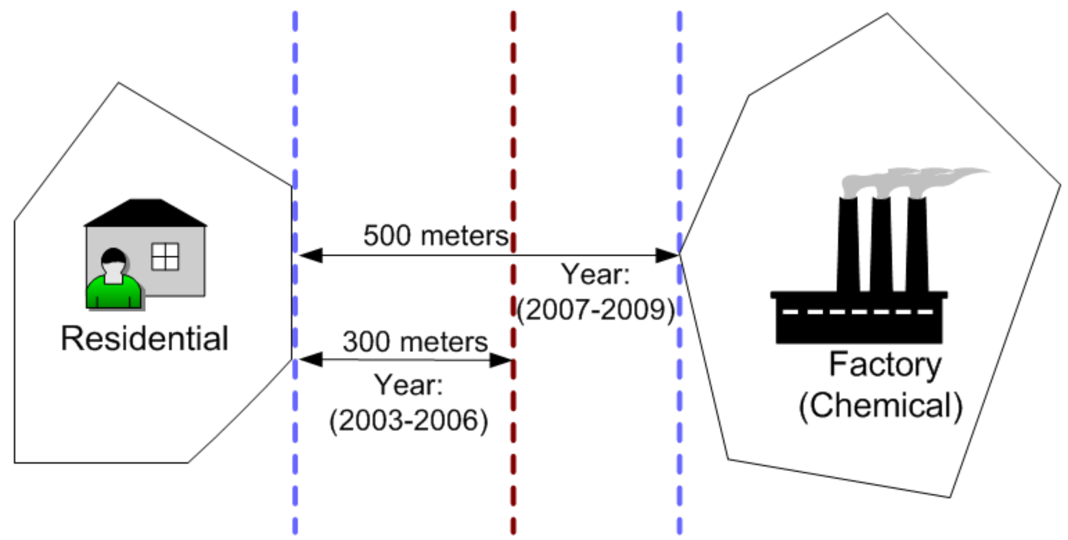


Figure 28: A city plan example

This example contains quality elements: **thematic accuracy, metric relationship, topological and temporal consistency** and is described in Table 16.

Table 16: The city plan example in the constraint decision table

	Description	Specification			
Event	Factory.updateGeometry	True			
Condition1a	CurrentTime	>2003 (Year), <2007(Year)		>2007 (Year), <2010 (Year)	
Condition1b	Factory.getAttributeValue(category)	"chemical"		"chemical"	
Condition1c	ChemicalFactory.distanceTo.ResidentialArea	<=300	>300	<=500	>500
Action	ChemicalFactory.Move()	True	False	True	False

3.6.9 Advice of Using ISO/TC 211 Standard 19113 - Quality Principles

ISO/TC 211 standard 19113 does not provide any detailed explanation for its quality elements and subelements. It is difficult for data users to understand four examples given in the annex of this standard. In this way, it brings significant problems for different users to implement this standard. Users should find explicit descriptions from other sources in GIScience to help themselves to comprehend this standard. In this thesis, above subchapters are used to provide an explanation of those ISO quality elements and subelements in order to help the study of spatial data integrity rules.

Other shortcomings of this standard were discussed by Kresse and Fadaie, and Servigne [Kresse and Fadaie, 2004; Servigne et al., 2006] such as the quality of imagery data has to be enhanced and the introduction of meta-quality is missing. Moreover ISO/TC 211 [ISO, 2006] review teams unveiled the discrepancy between ISO/TC 211 standard 19113 and ISO/TC 211 standard 19115 Metadata standard, which although some mismatches happen to ISO/TC 211 standard 19115, ISO/TC 211 standard 19113 still needs improvements.

Therefore, we would suggest that people who want to follow current ISO standards in spatial data quality research should adhere to ISO/TC 211 standard 19113 independently if there are inconsistencies existing between it and other ISO standards. In this way, people state clearly which standard they have adopted to avoid confusions and the work can be easily merged to future revised new standards.

3.7 Summary

In Chapter 3, we discuss existing conceptual data modeling methods and point out disadvantages. To provide the GIS the ability of checking the data consistency, other approach has to be implemented during conceptual data modeling. Thus, a new methodology called "spatial data integrity rule"-based conceptual data modeling is proposed, consisting of two main components: extended UML conceptual schema and spatial data integrity rules.

Detailed descriptions of both model components are given. For extended UML conceptual schema, it contains main geographic data information, and pictograms or symbols are adopted in order to represent vividly geographic data characteristics. For spatial data integrity rules, they are expressed by constraint decision tables which are based on the ECA rule. The structure, semantic meanings of defined terms and extensible terms in constraint decision tables are explained.

The real world phenomena are always complex, and thus the data consistency requirements contain many kinds of spatial data quality issues. In order to handle data consistency in a wide range, spatial data quality elements and subelements from ISO/TC 211 standard 19113 are introduced.

Then we discuss the possibility of applying those ISO quality elements and subelements to spatial data integrity rules. It shows that the proposed approach of using constraint decision tables mainly can be employed to handle logical consistency, temporal accuracy and thematic accuracy. Other quality elements either can not be described by this approach or has been specified at other steps like the UML conceptual schema.

Finally, the difficulties of interactions of different data quality elements are represented, and the shortcomings of ISO/TC 211 standard 19113 are exposed to make people in related research aware.

Chapter 4. Transformation of Conceptual Data Model

In the literature, conceptual data models are mostly used afterwards in different proprietary systems, thus any detailed transformation steps from the conceptual data model to the system executable format are not clearly introduced and results can not be adopted for other different systems. For example, Casanova [Casanova et al., 2002] created the UML and OCL knowledge model and then converted these to executable pieces of program code, however the program codes only worked for a specific geographic system, and not for any others.

Thus, the transformation of the conceptual data model into computer understandable language and open format is very important. Especially today the Internet is considered as a booming platform for GIS to share, publish, interchange and process geographic data. The transformation compatible to web usage is critical.

In this chapter, first several critical transformation principles are identified. Then corresponding to the proposed conceptual data modeling methodology by the extended UML conceptual schema and the spatial data integrity rules in constraint decision tables, a standard based transformation approach is provided. Afterwards, due to the limited ability of existing geospatial web services, the deployment of the transformed results via the standardized or extended geospatial web services is described. A summary is provided in the end.

4.1 Transformation Principles

In order to make the 'spatial data integrity rules' based conceptual data model accessible to the generic web geographic information system and deliver the specification of spatial data consistency requirements to data producers to support the data consistency checking task, the transformation of the conceptual data model should follow several principles:

- The whole content of the conceptual data model should be preserved, including essential geographic information and the detailed spatial data integrity rules.
- The converted model should be in machine readable and web compatible formats for the generic web geographic information system, in order to facilitate distribution, sharing and interchange of the model.
- The transformation method should be a standard based way, not the proprietary way. Thus, it should consider the existing standardization work on data formats and geospatial web services, such as the GI standards specified in ISO/TC 211 and OGC, in order to support the data interoperability.

Based on the above principles, a transformation approach is proposed as represented in Figure 29.

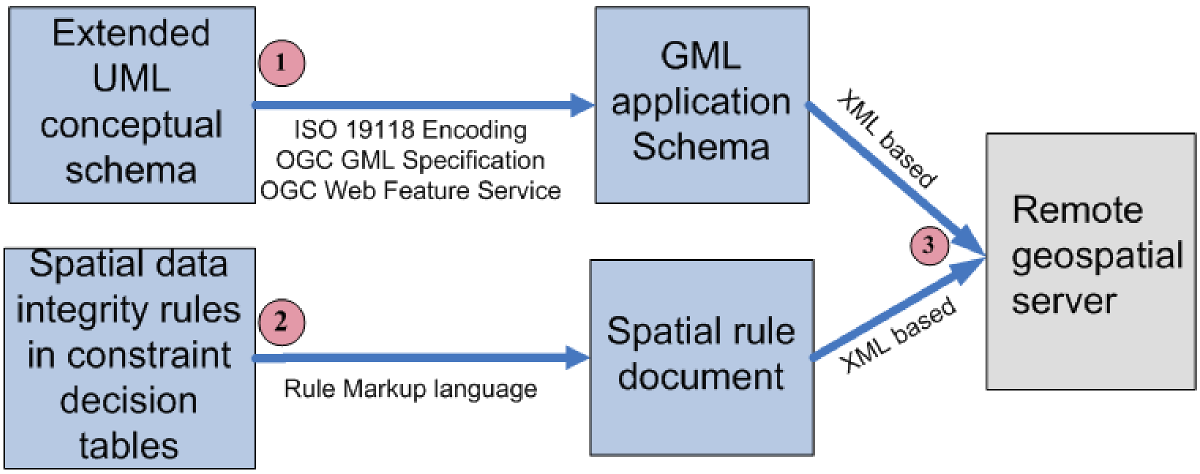


Figure 29: Transformation approach

1. The extended UML conceptual schema as an implementation neutral model is converted to an OGC GML application schema to comply with geospatial web services.

2. Spatial data integrity rules in the tabular constraint decision tables which have ECA syntaxes are embodied in a XML based spatial rules' document.

3. Both transformed formats having the XML structure and finally are stored in the remote geospatial server for consequent data access. The three points are explained separately in the following subsections.

4.2 Extended UML Conceptual Schema to GML Application Schema

The UML conceptual schema is an implementation neutral model, so there are various possibilities for transformed output formats such as XML schema, Java codes, CORBA, SQL and other textual formats. The basic approach of mapping UML conceptual schema to other data formats is shown in Figure 30.

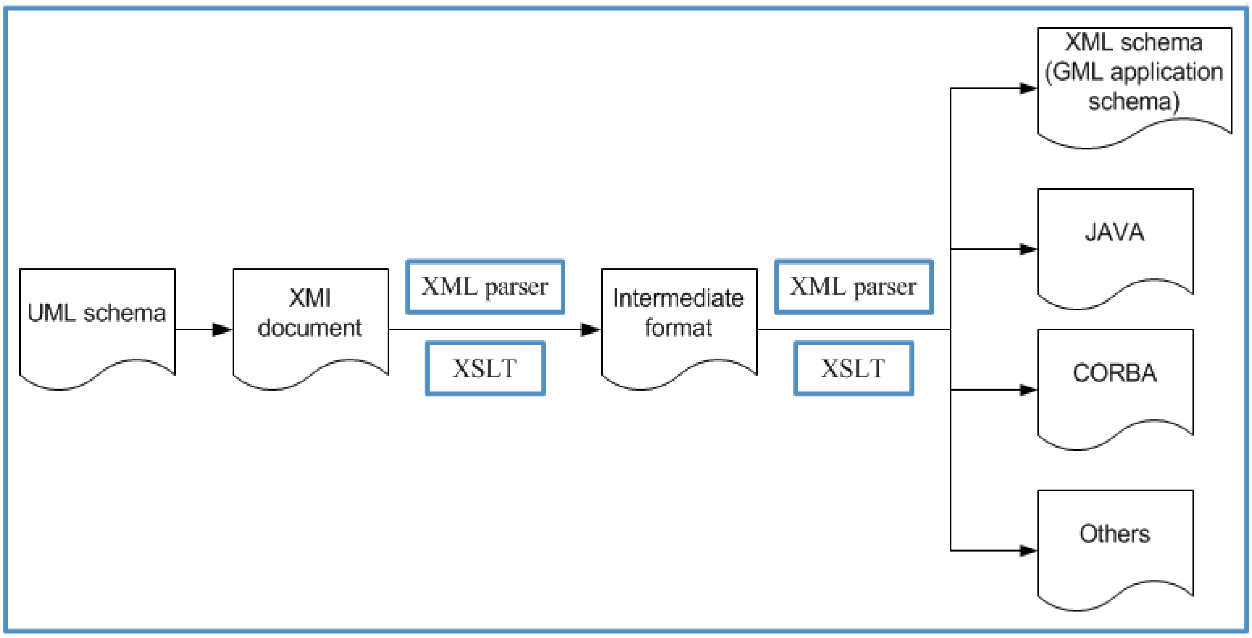


Figure 30: Transforming the UML conceptual schema to other formats

Most UML software allows exporting the UML schema into XML Metadata Interchange (XMI) [OMG, 2005], which at the UML schema level is an OMG specification to describe the UML schema in XML format, and one of its objectives is making UML schemas easy to be interchanged themselves or with different data standards. Once the XMI document is created, technologies for handling XML data can be adopted since the XMI document is in XML format, e.g. XML Stylesheet Language Transformations (XSLT) [W3C, 1999] that is a W3C specification for transforming XML documents into other XML documents can be used.

XML parsers in different programming languages can also be used to transform the XMI document. Some people use the intermediate document between XMI and the other formats to improve this transformation, because XMI is a very complex format with long XML Lag names and lots of nested XML tags [Grønmo, 2001].

The **GML application schema** is one mapping possibility adopted for this thesis. A GML application schema is based on the normal XML schema for exposing an application's geographic data with GML. It describes object types whose data the community is interested in and which community applications may expose. As a result, it is used to contain the essential geographic information specified in the extended UML conceptual schema such as the feature types like Ditch, Road and Extensometers and their attributes types [Lu et al., 2007; Shi, 2004]. The defined contents in turn refer to primitive types defined in the GML standard.

The literature has shown the mapping encoding strategy will be different if the application is not the same [Bernauer et al., 2004; Carlson, 2001; Salim et al., 2004]. Grønmo and Portele [Grønmo et al., 2002; Portele, 2005] set up some mapping rules for conversion from the normal UML conceptual schema to the GML application schema. However, the UML conceptual schema in this thesis is extended by the pictograms in order to represent geometric and temporal geospatial characters, so the approach and experiences from Grønmo and Portele need enhancement.

The ISO/TC 211 standard 19118 Encoding [ISO, 2002c] provides a basis for creating conversion rules to translate a UML conceptual schema into a corresponding XML Schema. The GML implementation specification [OGC, 2004a] gives the detailed explanation for representing geospatial data in GML. OGC Web Feature Service implementation specification [OGC, 2005b] shows the use of GML application schema in a web feature server.

Based on the above standards, the conversion rules from the extended UML conceptual schema to the GML application schema are created in the following tables: the basic data types from ISO/TC 211 standard 19103 Conceptual Schema Language to XML schema basic types are depicted in Table 17.

The geometry pictograms complying with ISO/TC 211 standard 19107 Spatial Schema and the temporal pictograms complying with ISO/TC 211 standard 19108 Temporal Schema are converted to GML types as shown in Table 18 and Table 19.

Table 17: ISO/TC 211 basic data types to XML schema basic data types

Basic data types according to ISO 19103	XML schema basic data type
CharacterString	string
Integer	integer
Date	date
Boolean	boolean
Real	decimal

Table 18: Spatial pictograms to GML2 types

Spatial pictograms according to ISO 19107	GML 2 specification Type
	PointPropertyType
	LineStringPropertyType
	PolygonPropertyType

Table 19: Temporal pictograms to XML schema types

Temporal pictograms according to ISO 19108	XML schema
 the UML class	“className”_instant:date
 the geometry of UML class	“geometry”_instant:date
 the attributes of UML class	“attributeName”_instant:date
 the UML class	“className”_start:date “className”_end:date
 the geometry of UML class	“geometry”_ start:date “geometry”_ end:date
 the attributes of UML class	“attributeName”_ start:date “attributeName”_ end:date

Names and data types of feature attributes will be encoded into XML elements and types in the GML application schema, and the enumeration data type is converted to corresponding enumeration XML simple types.

In this way, an example of conversion from an extended UML conceptual schema to a GML application schema is given in Figure 31. The converted application schema can be then stored in a remote web feature server.

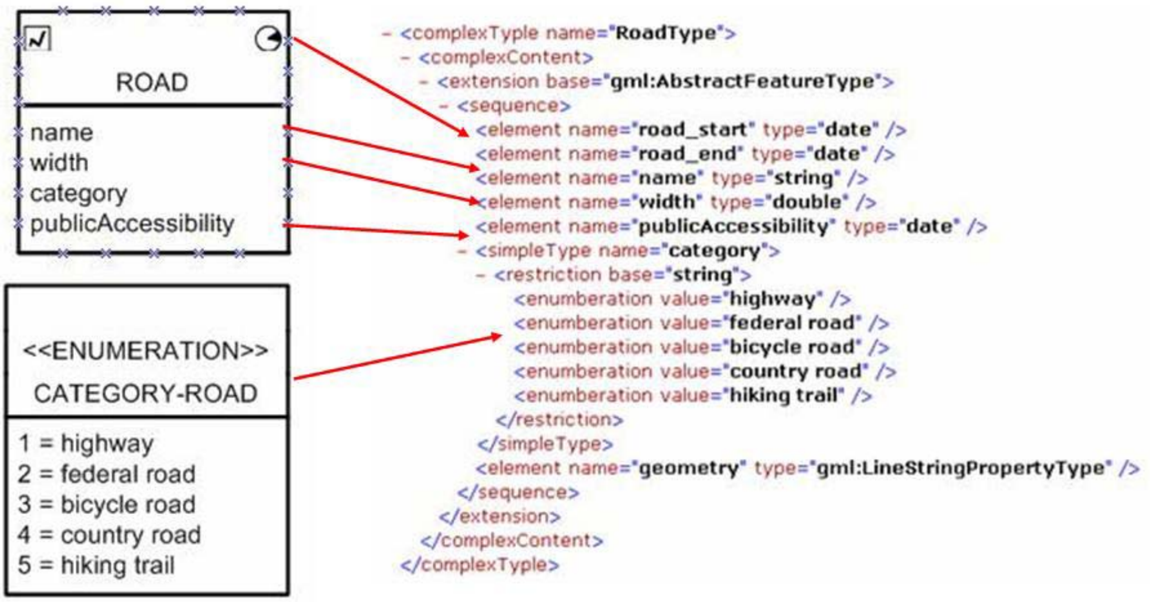


Figure 31: UML conceptual schema to GML application schema example

4.3 Spatial Data Integrity Rules to XML-based rules' Document

The spatial data integrity rules are defined by the constraint decision tables which are based on the ECA rule. Therefore, the transformation from spatial data integrity rules in the constraint decision tables to the XML-based rules' document is equivalent to from the ECA rules to an XML-based document. However, the existing work for representing the ECA rules in XML data format is still in the initial phases, so the ECA rule is not yet fully supported. Moreover, ECA rules used here for the definition of spatial data integrity rules are applied to the geospatial domain, but geospatial characters in those XML data formats are not yet widely applied. For example:

- The Rule Markup Language (RuleML) standardization initiative [RuleML], contributing to the implementation of the web rule language framework, has put the ECA rule into its development schedule. But it only provides partial support for events, conditions and actions. And it has not considered geospatial characters in its future plans as of this time.
- The Semantic Web Rule Language (SWRL) [SWRL, 2004] whose syntax is based on a combination of OWL and the sublanguage of RuleML, has been developed recently by the Joint US/EU ad hoc Agent Markup Language Committee. But it cannot deal with partial information and with closed predicates [Wagner, 2004], nor with spatial operators.
- Watson [Watson, 2007] developed a formal rule language that used a concept similar to SWRL to add additional support for spatial operators. Nevertheless, the language is used in the proprietary

system and is not expressed in any detail. Moreover, it cannot express the action part of the ECA rule which has been confirmed as one indispensable component for spatial data integrity rules.

Therefore, we have to propose this new approach to perform the conversion of spatial ECA data integrity rules. The experiences of how to build a general rule Markup language by Wagner [Wagner, 2002], who listed common problems when people set about to design a rule markup language and gave suggestions to deal with them, and how to design and implement an ECA rule markup language by Paschke and Seirio [Paschke, 2006; Seirio and Berndtsson, 2005] are adopted.

Before carrying out the transformation, here are several restrictions:

- In this research, we do not intend to develop a whole rule markup language, but only to use the rule-based XML structure to describe the spatial data integrity rule.
- The XML structure for our spatial ECA rules might be considered as a subset or profile of the future rule Markup language.
- The way we are encoding the ECA rule could be converted easily to future emerging standardized Rule Markup language, such as RuleML.

In order to convert the spatial ECA data integrity rules into a rules' document, the following conversion rules have been developed:

1. The XML-based rules' document has a root element called "SpatialDQRuleDoc", which means an XML document containing spatial data integrity rules. It then has a list of data integrity rules named "qualityRule" with a unique identification attribute "id" for representing various rules. Each "qualityRule" consists of "Events", "Conditions" and "Actions" XML elements, which correspond to the Event, Condition and Action components in the constraint decision table. The logical operators "And" and "Or" are included to connect the subelements of the "Events", "Conditions", and "Actions" XML elements. Thus, "condition1a" and "condition1b" in a constraint decision table are converted to "And" logical relationship, and "“action1" and "action2" are converted to "Or" logical relationship. The overview is described in Figure 32.

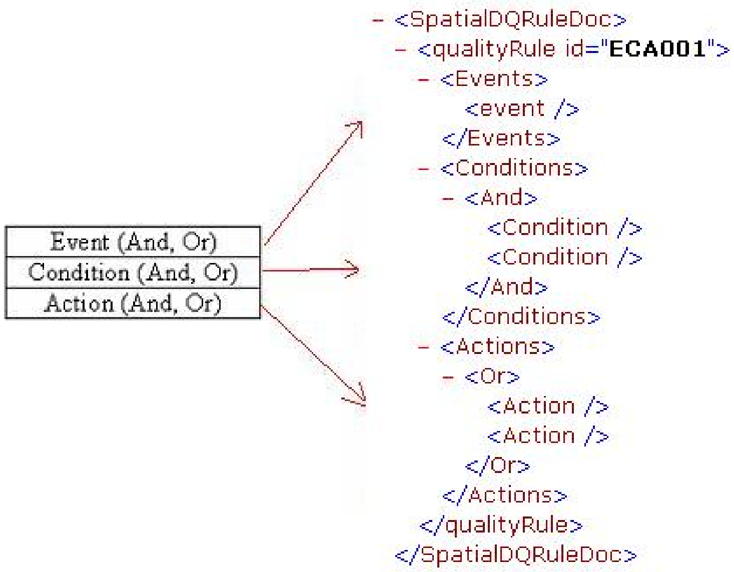


Figure 32: Event, Condition and Action mapping to the XML elements

2. A FeatureType in the constraint decision table is defined as a variable in the rules' document. The variables statement should be put in the element "arguments", and the feature type is defined in its subelement "variable" with attributes "name" and "classid" as shown in Figure 33. Then the variables can be used throughout the rules' document.

```
- <qualityRule id="ECA001">
  - <arguments>
      <variable name="ft1" classid="FeatureType1" />
      <variable name="ft2" classid="FeatureType2" />
    </arguments>
  </qualityRule>
```

Figure 33: The feature type statement in the rules' document

3. If the FeatureType has operations, the feature type will be defined by the "subject" element using the claimed variable such as "ft1", and its operations such as updateGeometry, updateAttribute and so on (see 3.4.3) are encoded in the "operation" element. And if the operation has some parameters, the parameters will be encoded by the "parameter" XML attributes of the "operation" XML element (as depicted in Figure 35).

```
<Events>
- <event>
    <subject>ft1</subject>
    <operation parameter1="para1">operation_ft1</operation>
  </event>
</Events>
```

Figure 34: The operations statement in the rules' document

4. When two feature types have a spatial relationship, one is set to the "subject" element and the other to the "object" element. The spatial relationship such as topological and metric are expressed by "spatialRelationship" XML element (see Figure 35).

```
<Conditions>
- <Condition>
    <subject>ft1</subject>
    <spatialRelation>spatialRelationship1</spatialRelation>
    <object>ft2</object>
    <dataPredicateAtom>atom_x</dataPredicateAtom>
    <dataValueAtom unit="unit_x">value_x</dataValueAtom>
  </Condition>
</Conditions>
```

Figure 35: The spatial relationships in the rules' document

5. The rule policy value is encoded as the element "rulePolicy" in the rules' document to each "qualityRule" element.

In this way, a spatial data integrity rule which contains metric relationship and attribute values defined by the constraint decision table in Table 9 is converted to a rules' document as described in Figure 36.

```
<SpatialDQRuleDoc>
- <qualityRule id="ECA001">
    <description>This is a topo-semantic quality rule</description>
    <rulePolicy>Red</rulePolicy>
  - <arguments>
      <variable name="ditch" classid="Ditch" />
      <variable name="road" classid="Road" />
    </arguments>
  - <Events>
    - <event>
        <subject>ditch</subject>
        <operation>updateGeometry</operation>
      </event>
    </Events>
  - <Conditions>
    - <And>
      - <Condition>
          <subject>ditch</subject>
          <spatialRelation>distanceTo</spatialRelation>
          <object>road</object>
          <dataPredicateAtom>lessThanOrEqualTo</dataPredicateAtom>
          <dataValueAtom unit="meter">5</dataValueAtom>
        </Condition>
      - <Condition>
          <subject>road</subject>
          <operation parameter1="publicAccessibility">getAttributeValue</operation>
          <dataPredicateAtom>euqalTo</dataPredicateAtom>
          <dataValueAtom>true</dataValueAtom>
        </Condition>
      </And>
    </Conditions>
  - <Actions>
    - <Or>
      - <Action>
          <subject>road</subject>
          <operation>updateGeometry</operation>
        </Action>
      - <Action>
          <subject>ditch</subject>
          <operation parameter1="publicAccesibility">setAttributeValue</operation>
          <dataPredicateAtom>euqalTo</dataPredicateAtom>
          <dataValueAtom>true</dataValueAtom>
        </Action>
      </Or>
    </Actions>
  </qualityRule>
</SpatialDQRuleDoc>
```

Figure 36: An example of the constraint decision table represented in the rules' document

4.4 Transformed Conceptual Schema and Geospatial Web Services

The last two sections explain the approach of the transformation from 'spatial data integrity rules' based conceptual schema to an XML based rules' document. The converted rules' document is then stored in a remote geospatial server. In this subsection, we describe how to store a rules' document in the remote server and deploy it to geospatial web services.

Nowadays, the dominant means of map distribution has changed with ever increasing Internet usage, because the Internet provides a simple way to share, interchange and distribute the geospatial data among huge multiple data users. Therefore, geospatial web services are increasingly adopted as the platform for handling geospatial data, especially standardized web services. The leading ones are from OGC, who provides distinct web services for serving different predefined web requests. In particular, the Transactional Web Feature Service (WFS-T) [OGC, 2005b] is used as a core service, which provides the standard interface that allows users to access geospatial data from heterogeneous remote databases, and uses Geography Markup Language (GML) [OGC, 2004a] for data exchange.

Moreover, the WFS-T server allows users to access data from and to transact data (insert, update and delete) to the remote server in real time. This brings many benefits to field data acquisition as employed in this thesis.

In Figure 37, we offer general techniques for integrating the transformed conceptual schema with the transactional web feature service of the remote geospatial server. Client and Remote geospatial server communicate with each other through HTTP requests, where Remote geospatial server contains an OGC WFS-T and an extended WFS-T component.

Component1 is the GML application schema, which belongs to the WFS-T specification, so it can be deployed easily by WFS-T and accessed directly through standard WFS-T HTTP requests by the client, e.g. GetDescribeFeatureType WFS-T request to the sever can return a main description of each feature type.

Component2 is the spatial data integrity rules' document, which does not belong to the WFS-T specification, so the normal WFS-T sever has to be enhanced with extended functions to support this rules' document. The rules' document has an XML-based structure, thus some pieces of server side programs need to be implemented to deal with this content. With extended server side functions, the client is able to send the corresponding extended HTTP requests to retrieve spatial data integrity rules from the geospatial server, e.g. an extended WFS-T request called **GetDQRules** can return all the data integrity rules of each feature type for the client.

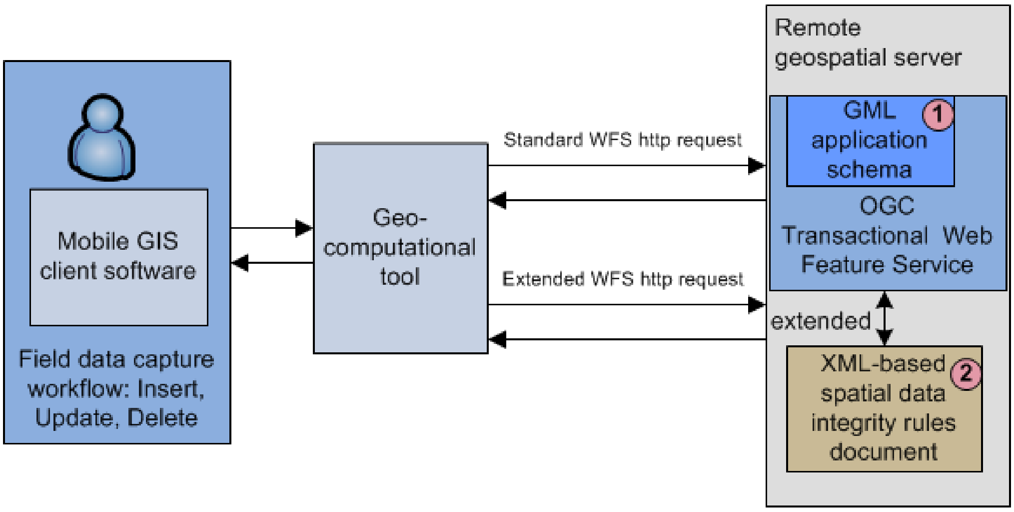


Figure 37: Integration of transformed data model with WFS-T

More sophisticated extended web feature service functions might be needed to perform complex operations on spatial data integrity rules, such as to list specific rules relating to a certain feature type, or to make spatial computations for checking rules on the server side. However, this topic is beyond this thesis. An interesting reference for this topic can be found in the ongoing work of OGC Quality Assurance Web Service.

4.5 Summary

In Chapter 4, we explain the methods for transforming of the 'spatial data integrity rules' based conceptual schema to an XML-based rules' document. Two separate means are developed to achieve the purpose.

The first explains that extended UML conceptual schema is converted to the GML application schema through a set of conversion rules, where extended pictograms for the UML conceptual schema and the specific features of GML are considered. The second describes how the constraint decision table structure is mapped to the rule markup language document. For each part, examples are given to demonstrate results.

Finally, in order to deploy the transformed conceptual schema by the standard geospatial web services, OGC WFS-T is employed. But OGC WFS-T only supports the GML application schema. Therewith, a new proposal is given to extend the current OGC WFS-T to handle the transformed spatial data integrity rules' document.

Chapter 5. Integration of the 'Spatial Data Integrity Rules'-Based Data Model into the Field-based Mobile GIS Workflow

In the above chapters, we declared that spatial data integrity rules can be used to specify data consistency requirements, and the defined rules can be applied to GI systems to find data inconsistencies and avoid potential errors. Then we proposed the constraint decision tables and UML conceptual schema to describe the geographic conceptual data model. In this chapter, in order to show the usage of the defined conceptual data model in GIS data activities, the field survey approach of data capture is adopted.

In section 5.1, the field survey using Mobile GIS techniques is introduced and a generic Mobile GIS architecture is given. In section 5.2, the three general steps for field data capture are explained. Based on that, in section 5.3, the definition of the quality-aware geocomputational tool is proposed. A prototype of the geocomputational tool is developed to control the data consistency of the collected data within the data capture workflow. In section 5.4, the field tests of the real GIS application in a landslide area are explained to demonstrate the overall concept of this thesis.

5.1 Field Survey by Mobile GIS Techniques

For GIS, data capture always starts after the data model is implemented, which means to store the data in the geospatial information system, for instance, in the database. There are three main approaches to capture the data: **field survey, photogrammetry** and **reproducing data from external data sources**. In this thesis, we emphasize the field survey in collecting the geospatial data.

The **field survey** approach with Mobile GIS techniques is rapidly growing and becoming a major trend for geospatial data capture, because it provides the field user more time and cost effective tools than traditional field data capture methods. The existing geospatial data in the remote server can be accessed directly in the field, as well as collected data like coordinates information is saved directly into the mobile field editing system so that any degradation of post-capture digitizing of detail can be avoided.

Thus, many GIScience researches on field-based Mobile GIS can be found in the literature [Breunig et al., 2005; Mäs et al., 2005a; Maula, 2001; Nusser et al., 2003; Peng and Tsou, 2003; Pundt and Brinkkötter-Runde, 2000; Tsou, 2004]. A generic field based Mobile GIS can be designed as shown in Figure 38, and it has been proven in the applications such as [Plan et al., 2004; Reinhardt et al., 2005; Wang et al., 2005].

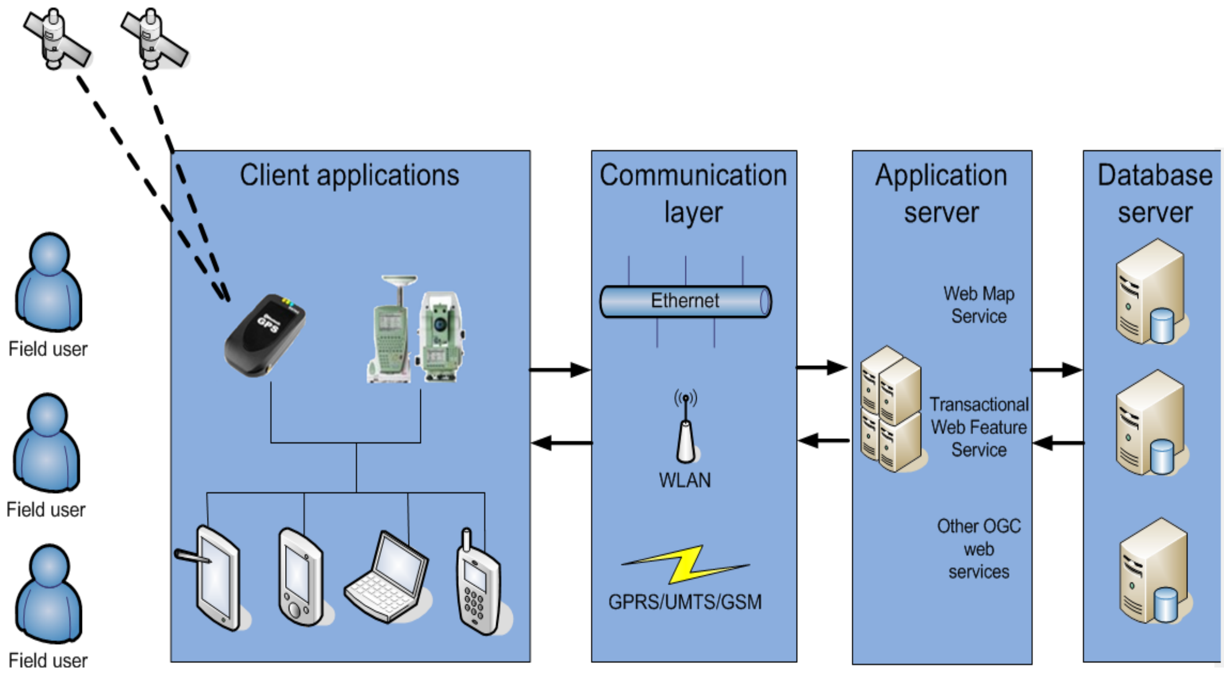


Figure 38: A field-based Mobile GIS generic architecture

By and large, field-based Mobile GIS is considered as several layers: client applications contain survey instruments and computers or handheld devices which field users employ. Communication layer provides the communication means through Ethernet, WirelessLAN or GPRS/UMTS/GSM. Application server consists of different geospatial web services like WFS-T and WMS. Database server involves heterogeneous geospatial databases. All the layers communicate with each other through specified protocols or interfaces.

Mobile GIS can be considered as an integrated software and hardware framework, which provides the access of geospatial data and services through mobile devices via wireline or wireless networks [Mäs et al., 2005b; Tsou, 2004]. In detail, its hardware and software components are explained in the following.

- **Hardware components**:

Emerging innovations like **wearable computers and rugged Tablet PCs** allow users to work in all weather, which is a remarkable advantage for field survey tasks under unexpected weather conditions. Tablet PCs have a bigger touch-sensitive screen than other handheld devices like the PDA, and have a pen-shaped controller with the same functionality as the mouse of a normal computer. The processing speed and performance of the Tablet PCs are now as good as normal PCs, and thus they are becoming better equipments to be used by field workers for survey work in Mobile GIS.

Survey instruments or **multi Geo-sensors** [Kandawasvika and Reinhardt, 2005] like GPS receiver, total stations, digital cameras and so forth are employed in the field to observe different real world phenomena.

Network hardware such as Wireless-Lan connectors provide high speed accessibility of huge amounts of geographic data, UMTS and GPRS receivers which are also suitable for transferring data in the field, and their interworking can be used to improve network performance [Schmidt, 2004].

Remote GIS server includes geospatial web servers and database servers.

- **Software components**:

WFS-T [OGC, 2005b] provides standard interfaces that allows users to access the geospatial data from heterogeneous remote databases and uses Geography Markup Language (GML) [OGC, 2004a] for data exchange.

Web Map Service (WMS) [OGC, 2004b] can be used to produce maps of spatially referenced data dynamically from geographical information.

Geospatial database provides highly efficient management of geospatial data, such as Oracle spatial and PostGIS (open source spatial database).

Web server is able to handle the HTTP protocols and requests, e.g. the Tomcat web server.

Mobile GIS client software offers various GIS functions for the field user (e.g. visualization of personalized map, communication with different kinds of survey instruments, accessing the remote database server or web services and other location based services).

5.2 Decomposition of Mobile GIS Data Acquisition Workflow

Before integrating the quality information of the data model into the field data capture workflow, we give the detailed data collection steps in this section. Based on the generic field based Mobile GIS architecture presented in Figure 38 and our field work experiences, we propose three main data capture scenarios: **insert a new feature**, **update an existing feature** and **delete an existing feature**. The breakdown of each scenario is given in the following subsections.

5.2.1 Insert (or add) a New Feature

- **Select the measured feature type**:

The field user first selects the available feature type in the databases which he wants to measure before he starts to capture the data. A feature type selection window listing possible feature types is displayed as shown in Figure 39, such as Road, Ditch and Extensometer.

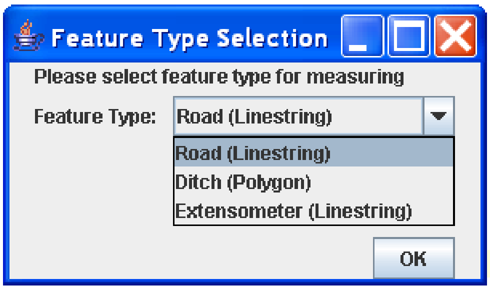


Figure 39: Feature type selection

- **Measure the geometry of the feature**:

The field user measures the real world object through surveying instruments such as GPS receiver or total station to get the coordinates of points, and successive measured points create different geometries such as LineString and Polygon. The results are visualized on the user interface of mobile GIS client software as shown in Figure 40.

Figure 40: Geometry visualization on the Mobile GIS client

- **Input the attribute values**:

After the geometry is captured, a window for giving attribute values pops up for the user. Then the user is asked to fill in the attribute values according to the real world situation, for example the name of the road and the category of the road as illustrated in Figure 41. Some attributes values are predefined according to the conceptual schema, like the road category, so the user can simply choose the item and avoid input mistakes.

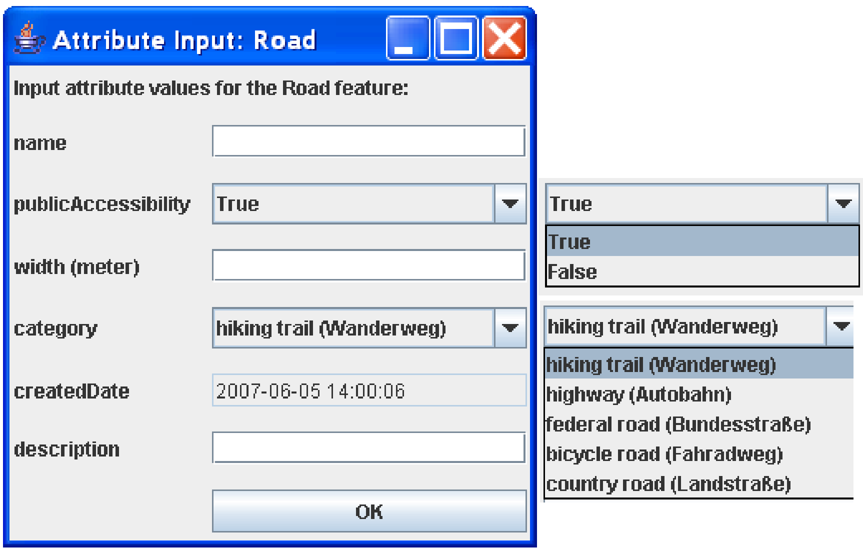


Figure 41: Attribute input window

- **Transact the feature**:

In the end, geometry and attributes information are combined together to generate a feature, which is afterwards sent back to the remote database through WFS-T. The user will get a confirmation dialog for this action as displayed in Figure 42.

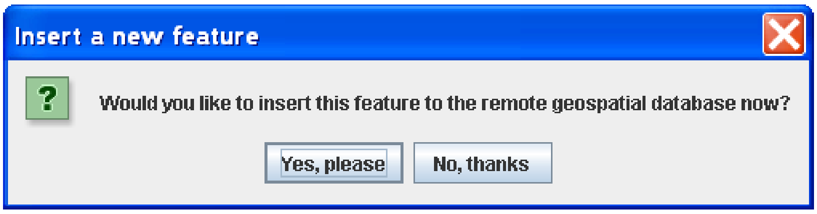


Figure 42: Insert a new feature confirmation

- **Display the response**:

The remote geospatial database will return a response through WFS-T to the user when the feature is successfully added as demonstrated in Figure 43, otherwise the user gets a failure message.

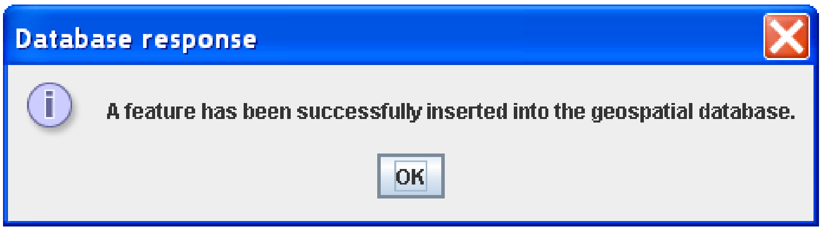


Figure 43: Database response message for insert

5.2.2 Update an Existing Feature

- **Select the feature type to be updated:**

It is as same as "**Insert a new feature**", see Figure 39.

- **Choose update type**:

Three types of update are provided for selection: **update whole feature, update geometry or update attributes** as represented in Figure 44.

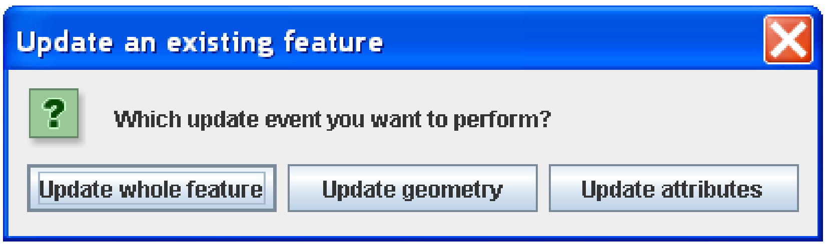


Figure 44: Select update event

If the user selects update whole feature, it works as same as the combination of "**delete an existing feature"** (refer to the next subsection) and then "**insert a new feature"** (refer to the previous subsection).

If the user selects update geometry, he only needs to measure the geometry. The attribute values are kept as original values. It means the step of "input attribute values" will be ignored. In the complex real world case, some of original attribute values of the feature may be changed if the geometry changes. Those complex cases are not discussed in this thesis.

If the user selects update attributes, he will not be asked to measure the geometry again. The attribute window will pop up with the existing old attribute values. Users then can modify the intended attribute values. In this case, one more attribute called "updateDate" is automatically added to record the current update time (Figure 45).

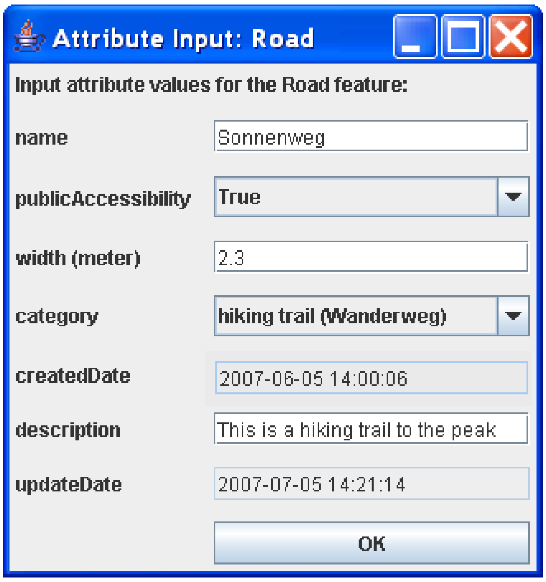


Figure 45: Update attribute window

- **Transact the feature**:

In the end, the updated geometry or attributes information of a feature is sent back to the remote database through WFS-T, and the user will get a confirmation dialog for this (Figure 46).

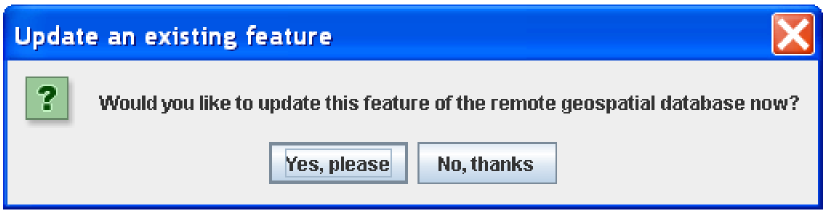


Figure 46: Update an existing feature confirmation

- **Display the response**:

The remote geospatial database will return a response through WFS-T to the user when the feature is successfully updated; otherwise the user gets a failure message (Figure 47).

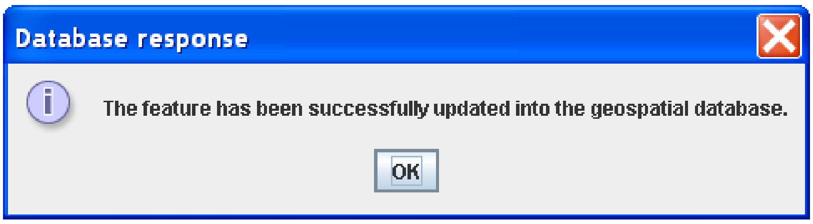


Figure 47: Database response message for update

5.2.3 Delete an Existing Feature

The user can easily delete an existing feature through the Mobile GIS client software. First, the user selects one feature or more than one feature with a mouse, and then uses a button or context menu "delete" to delete the feature or features. Deleted features are removed immediately from the client side. To remove these from the remote database, the user has to send a delete command through WFS-T (Figure 48) to the database.

Finally, the user will get the response from the remote database, either success or failure (Figure 49).

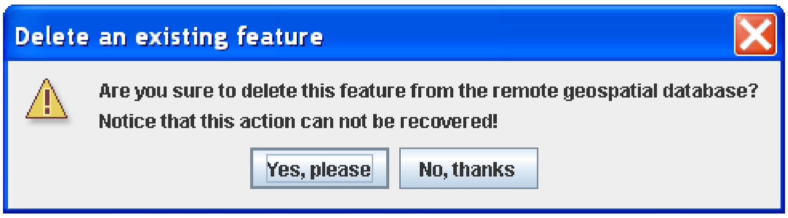


Figure 48: Delete an existing feature confirmation

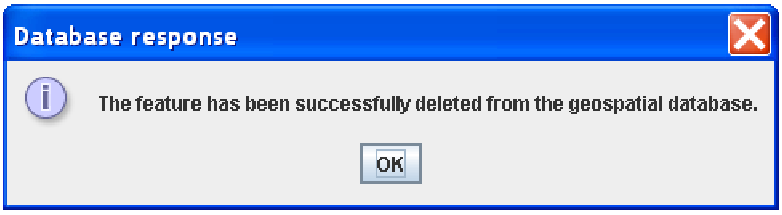


Figure 49: Database response message for delete

5.3 Quality-aware Geocomputational Tool

In this subsection, a prototype of the quality-aware or error-aware geocomputational tool is implemented, which can monitor each data capture step, discover data errors and provide a friendly and reliable user interface, support and improve the data consistency checking within the whole workflow. The basic concepts for quality-aware or error-aware geocomputational tools are introduced. The implementation of such a geocomputational tool is offered in remaining subsections.

5.3.1 Basic concepts

Longley [Longley, 1998] summarized that a geocomputational tool is a special tool for handling geospatial data in many ways for different purposes. Many people do not use proprietary GIS technology, and choose instead to write their own programs in order to fulfill their practical research problems where proprietary GIS software cannot do. Others prefer to couple the GIS module to their own specialized software in a variety of ways.

Unwin [Unwin, 1995] and Duckham [Duckham, 2002; Duckham and McCreadie, 2002] showed the current lack of error-sensitive functionality found in commercial GIS and suggested that **error-aware** or **error-sensitive** GIS is crucial to data quality related research. Devillers [Devillers et al., 2007] used the term **quality-aware** GIS which extends the term error-aware for a GIS with the added capability to manage, update, explore, assess, and communicate quality information.

We also see the same deficit of quality-aware ability in field based Mobile GIS. Without the data quality control functions, data is collected and stored in a temporary database in the field, and the quality of the data is often checked later in the office. In this way, the user cannot correct the error directly in the field, and may forget the real world scene later on [Pundt, 2002]. Moreover the time period for later quality checking delays any use of up to date data. Redo work in the field is very expensive.

Therefore, in the context of handling spatial data consistency checking, we put forward the concept of a **geocomputational tool** for Mobile GIS, which is defined as:

An independent software module with the abilities of geospatial web services communication, spatial computation, considering user interactions and others to enhance the field based Mobile GIS, in order to handle the data consistency problems within the data capture workflow

This tool helps the field user to avoid misunderstandings, mistakes and errors, and achieve qualified data as constrained by the defined spatial data integrity rules of the data model. Consequently, the implementation of such a geocomputational tool is explained in the next subsection.

5.3.2 Implementation of the Quality-aware Geocomputational Tool

Based on the explained Mobile GIS data capture steps and the concept of the quality-aware geocomputational tool, we implement the tool by using Java program language.

One core part of the tool bases on the Java Topology Suite (**JTS**) [Aquino, 2003], which is an open source Java API that provides the functions of spatial data model and spatial predicates of the OGC simple features specification [OGC, 1999]. JTS provides a complete, consistent, robust implementation of fundamental 2D spatial algorithms, and thus is adopted in this prototype for coping with spatial relationships computation.

Another core part of the tool for handling the requests and responses of the geospatial web services takes a foundation from an open source project named **deegree** [deegree, 2007] which builds a Java framework that offers the main building blocks for spatial data infrastructure and implements the existing OGC geospatial web services as well as clients and other components.

The general working principles are represented in Figure 50. The Mobile GIS client software and remote geospatial server communicate with each other through the geocomputational tool. For different Mobile GIS user's actions, the geocomputational will perform corresponding operations. The detailed data capture workflow of the "insert a new feature" is described in the following parts, and the workflow of "update or delete and existing feature" has the similar steps.

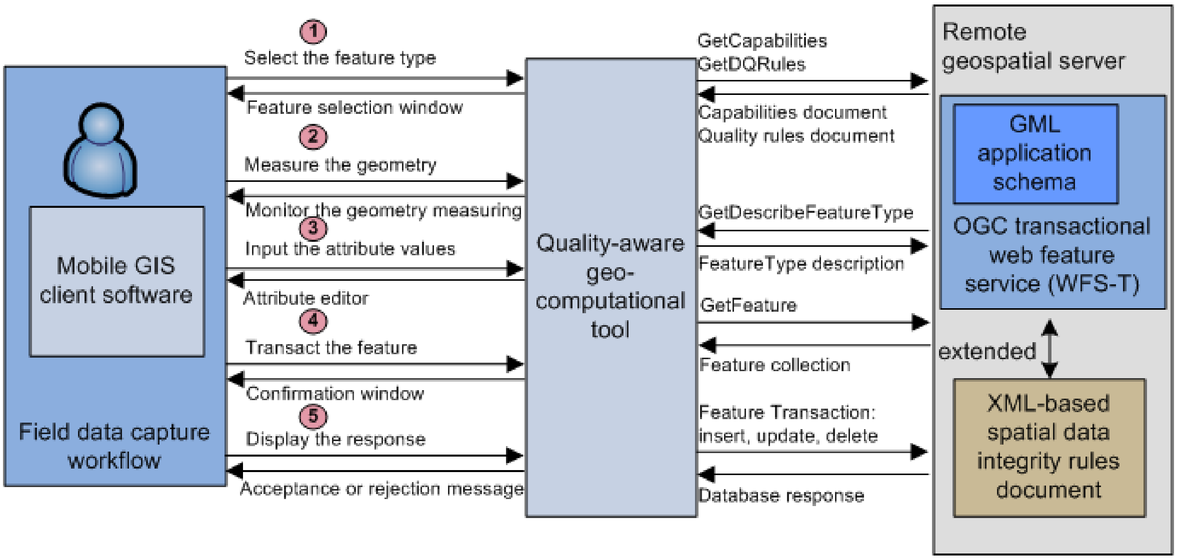


Figure 50: Working steps of the quality-aware geocomputational tool

1) **Select the feature type**

This tool first sends the **GetCapabilities** request to the WFS-T server to get the capabilities document, which contains the general description of the WFS-T server, such as a list of available feature types, supported WFS-T requests, service provider data and supported filter operators [OGC, 2005a]. Then a feature type selection window with available feature types is offered to the client.

Afterwards, this tool sends the **getDQRules** extended WFS-T request to the WFS-T server to load the integrity rules' document. The rules' document is then interpreted by this tool, and all the rules are mapped to respective Java objects for subsequent data quality control functions at different data collection steps.

2) **Measure the geometry of the feature**

When the user selects one of the feature types, this tool sends the **GetDescribeFeatureType** request to the WFS-T server to get the detailed description of this feature type, including its geometry type, attribute names and types.

For example, when the user selects the Road feature type, this tool knows the Road feature is LineString geometry and must consist of at least two points. This tool then monitors every move of the geometry collection according to the defined integrity rule, e.g. if a rule "*The lineString feature is not allowed to be self-intersected*" is defined, this tool checks the geometry of this lineString feature every time the field user adds a new point. If the self intersection is found, a warning message box pops up with the rule's content and asks to correct the last lineString segment.

When the collection of the whole geometry is finished, this tool checks the integrity rules that contain the spatial relationships, e.g. "*a road is not allowed to intersect with a building*". First, this tool sends a **GetFeature** request to the WFS-T to download all nearby building features in the database, and then it calculates the topological relationship "intersects" between the measured road feature with all building features. If one or more than one building is found "intersects" with the road, this tool immediately informs the error information to the user.

3) **Input the attribute values**

This tool reads the attributes information from the **GetDescribeFeatureType** response document at previous step, so it generates a user-friendly attributes editor to collect attribute values (Figure 41). This editor is error-sensitive and can deal with the integrity rules that include the attributes expressions, e.g. the "width" value of the Road feature is numerical, so input values with characters will be rejected. Additionally, the editor is able to fill in system values itself like the observation time to reduce the input mistakes.

When all attribute values are done, this tool checks the complex integrity rules that include not only attributes information, but also spatial relationships and temporal information, such as the examples containing interactions of data quality elements represented in Section 3.6.8. This tool checks those rules and finds the corresponding errors before the feature is created. The user has the chance to modify the feature to avoid errors.

4) **Transact the feature**

At this step, the whole feature is created and ready for uploading to the remote database. The field user is asked to confirm this action. If the user agrees to upload the feature, this tool sends this feature to the WFS-T sever through the transactional "insert" WFS-T http post request, otherwise, this tool cancels this action and ignores this feature.

5) **Display the response**

Finally, this tool provides a dialog with the acceptance or rejection message from the WFS-T server about whether the feature is successfully transacted or not. If it fails due to the network traffic or connection, this tool will resend the transactional request again in some minutes. In the mean time, the user can continue the field work.

Through the plausibility checks by the quality-aware geocomputational tool within the whole data capture workflow, the field user can immediately receive the error messages and the tutorial of possible reactions for correcting or avoiding errors. In such a way, the quality of collected data can be improved and the field user is confident about the performed actions (the field tutorial for correcting errors bases on many experts' knowledge). The errors can be greatly reduced in the field before the geospatial data is transferred to the remote database.

The implemented geocomputational tool is a portable Java software component with open and simple architecture, and therewith it can be easily integrated into other different GIS applications with minor changes.

5.4 Proof of Concept

In this subchapter, a real Mobile GIS application is demonstrated, and the proposed methodology to handle data consistency is evaluated by this application. First, the introduction of the application and the detailed architecture of the field-based Mobile GIS are given. Then examples are provided about how the proposed approach works for the data capture workflow.

5.4.1 Background

Within the project "Advancement of Geoservices" (http://www.geoservices.uni-osnabrueck.de/), a prototype for a mobile data acquisition system has been developed which is designed particularly for an application in landslide monitoring and decision making. In this specific scenario the application supports monitoring of the geological phenomena and acquisition of geological objects like ditches and edges.

In cooperation with local authorities, two test areas have been selected. One of the test areas is located near Balingen in Germany (http://www.balingen.de/). This area has serious environmental problems with landslides as demonstrated in Figure 51. Due to slope instability, rock masses, soil and other materials may fall down the mountain slopes and damage the nearby infrastructure.

Figure 51: The landslide area near Balingen in Germany

To get immediate warning information about slope movements, permanent survey instruments have been installed by local authorities For example, the on-site extensometers and automatic total stations shown in Figure 52 are normally installed in cracks, gaps or ditches in sensitive areas.

If significant surface movement measured by these instruments is sent out to the control station, an alarm is triggered. In this case, the geologists or surveyors have to go out to the field to validate the alarm, because extensometers are very sensitive to disturbances from animals or falling objects. New measurements are then collected in the field to check whether the alarm is justified, if so, to estimate the severity of the landslide and then to inform the public.

Figure 52: The on-site extensometer

5.4.2 Field Work Introduction

Field tests were carried out during the summer at the landslide area near the Balingen in Germany. The test area can be described as:

- It covers an area around 150*250 square meters
- It is situated on a steep slope with about 50 meters elevation difference
- It is a forested area with a high density of tall trees

Based on these characteristics and the generic field-based Mobile GIS introduced in Section 5.1, the system configuration of the field tests is shown in Figure 53. Its components are:

- The geospatial server containing the Apache Tomcat (the servlet container and the web server), deegree web feature server (compatible to the OGC WFS-T), PostGIS installed in a laptop.
- A local WLAN network with two access points and one antenna in order to cover the entire area.
- The Mobile GIS client software including the data quality-aware geocomputational module running on a Tablet PC.
- A GPS receiver and a total station as the surveying instruments.

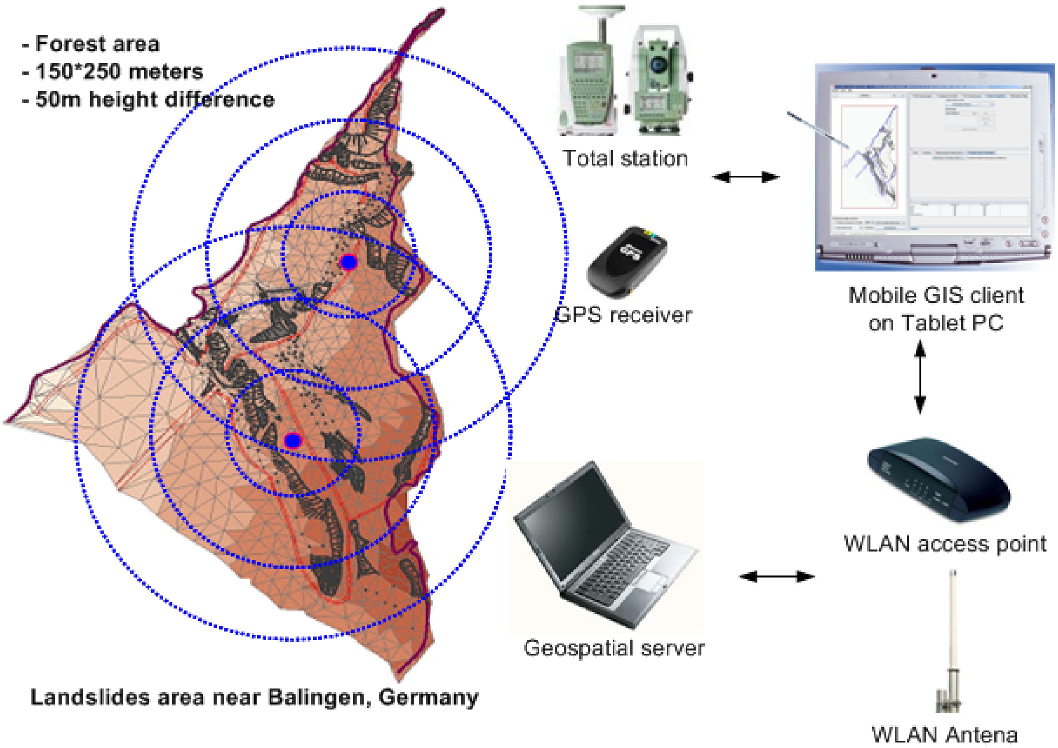


Figure 53 Balingen field tests system configuration

Two pictures illustrate the field test work in Figure 54 and Figure 55. The first picture shows the field user of the Mobile GIS data capture software on the tablet PC to measure real world objects through a total station. The second picture shows the geospatial sever and the WLAN network, including a battery power supply.

Figure 54: Field users were measuring ditch objects

Figure 55: The geospatial sever and the WLAN network

Through field tests, the following points are concluded:

- The high density of tall trees with plenty of branches and leaves in the summer made GPS availability marginal, so measurements were captured primarily by the total station.
- Two WLAN access points and an antenna were good enough to cover the whole area, and field users had a stable network connection.
- The Mobile GIS graphical user interface provided visualization of existing surrounding environmental data and also the newly collected data for the field users, which helped them know the real world and complete the data capture task.
- During the data capture workflow, the quality-aware geocomputational module successfully enabled field users to find immediately data errors relating to the defined spatial data integrity rules, and to correct errors with the well defined instructions.

In the future, mobile communications with GSM, GPRS or UMTS can be tested, the professional GPS receiver can be adopted to enhance the receiving signal, a network of multiple sensors can be set to collect the data [Kandawasvika and Reinhardt, 2005], and different geosciences applications can be considered.

5.4.3 Demonstration

As explained in Chapter 1, a quality-aware geocomputational tool was developed and integrated in the Mobile GIS client software for data quality assurance during field data collection. In this section, several screenshots are provided to demonstrate the GUI of the Mobile GIS client, and to show how the quality-aware functions work.

Figure 56 gives an overview of the Mobile GIS client GUI. The left panel is used to visualize geospatial datasets that include existing datasets and the immediate collected data. The left panel contains several tabs for different GIS functions (also called PlugIns). The developed quality-aware geocomputational tool is integrated with the PlugIn "Feature Acquisition". This PlugIn controls the field data collection workflow, such as surveying instruments connections and data recording. This Mobile GIS architecture can easily be extended to accommodate additional Functional PlugIns.

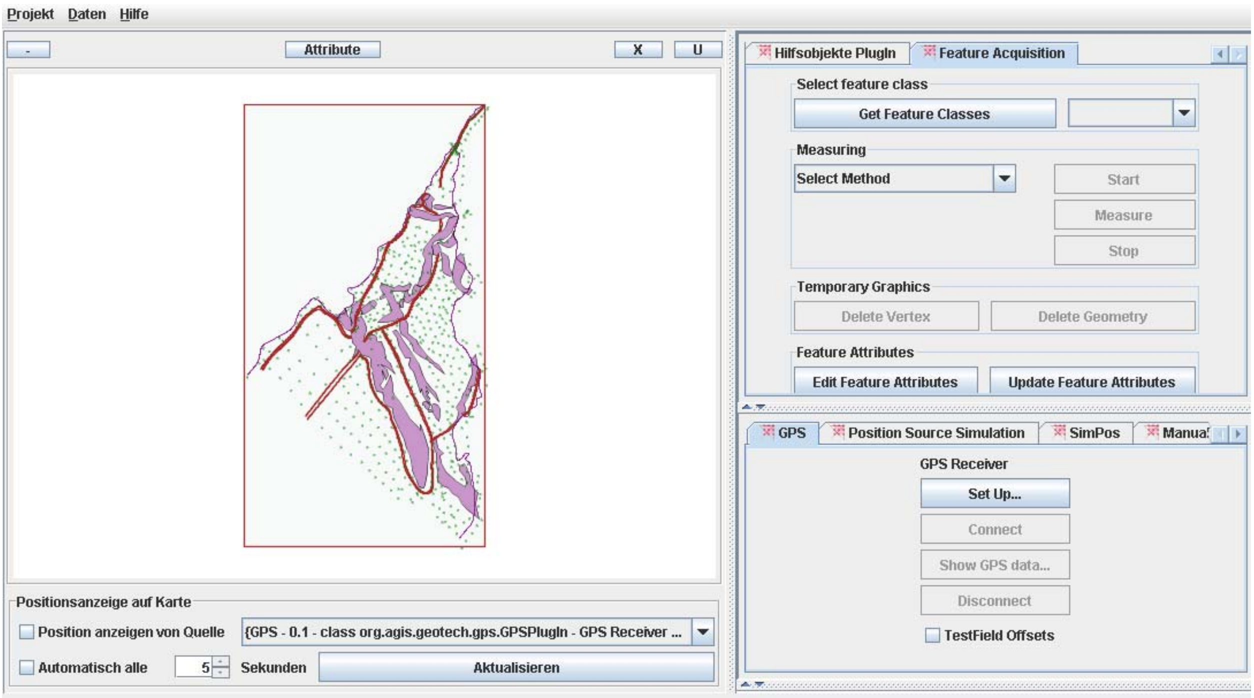


Figure 56: Overview of Mobile GIS client GUI

Figure 57 represents attribute values input for a Road feature. The user can input values or select predefined values from list boxes.

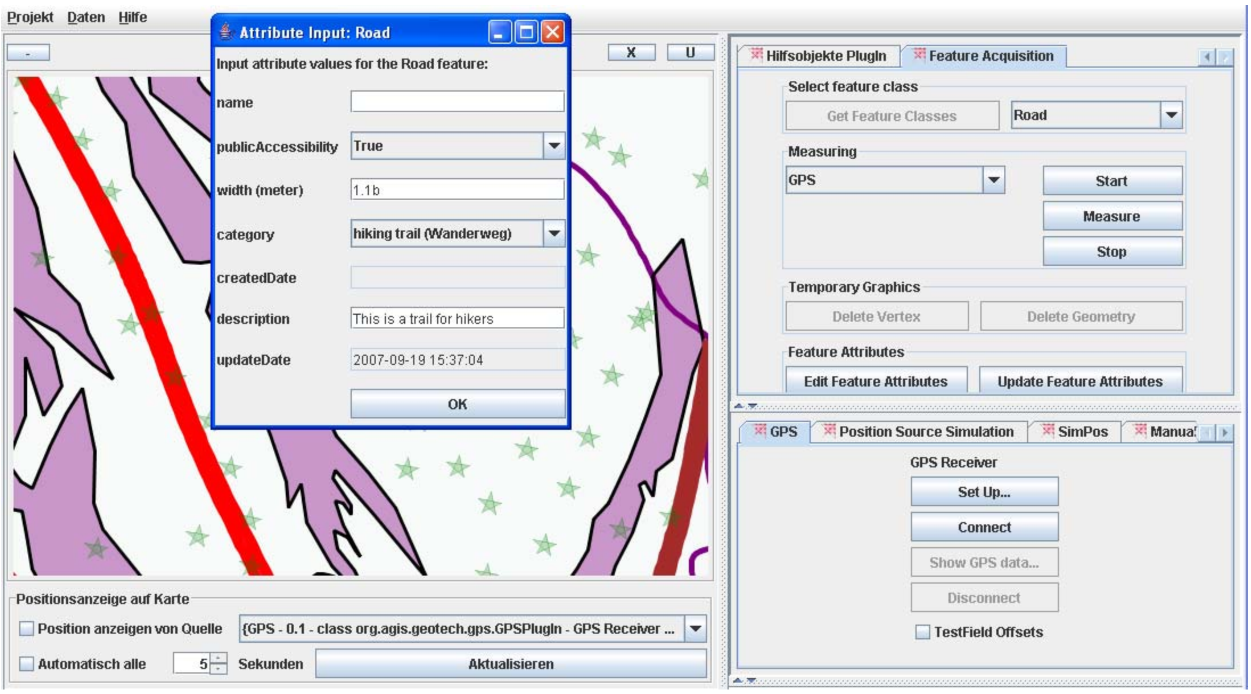


Figure 57: Attributes input window

As described in Section 5.3, the quality-aware geocomputational tool enables the “Feature Acquisition” PlugIn to monitor each data collection step during the whole workflow. If one of the defined integrity rules is violated, a dialog is generated to warn the user. Figure 58 illustrates the result of the violating integrity rules about attribute values. The user forgets to give the attribute value, and inputs a character value for a defined numerical data type. Thus, he gets a warning dialog with the comment for correcting the error.

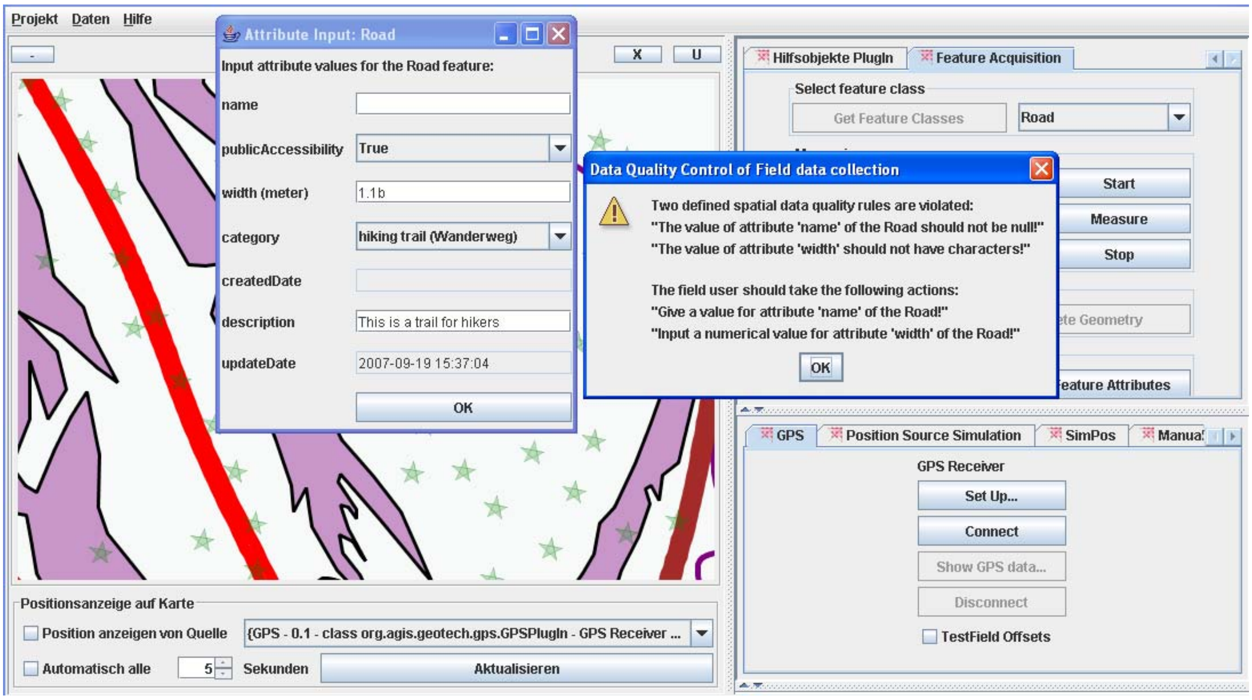


Figure 58: Quality-aware module demo 1

Figure 59 depicts that the topological integrity rule represented in Table 8 is violated. The user gets the error information with the detailed explanation. The popup dialog also contains the clear instructions for modifying this error.

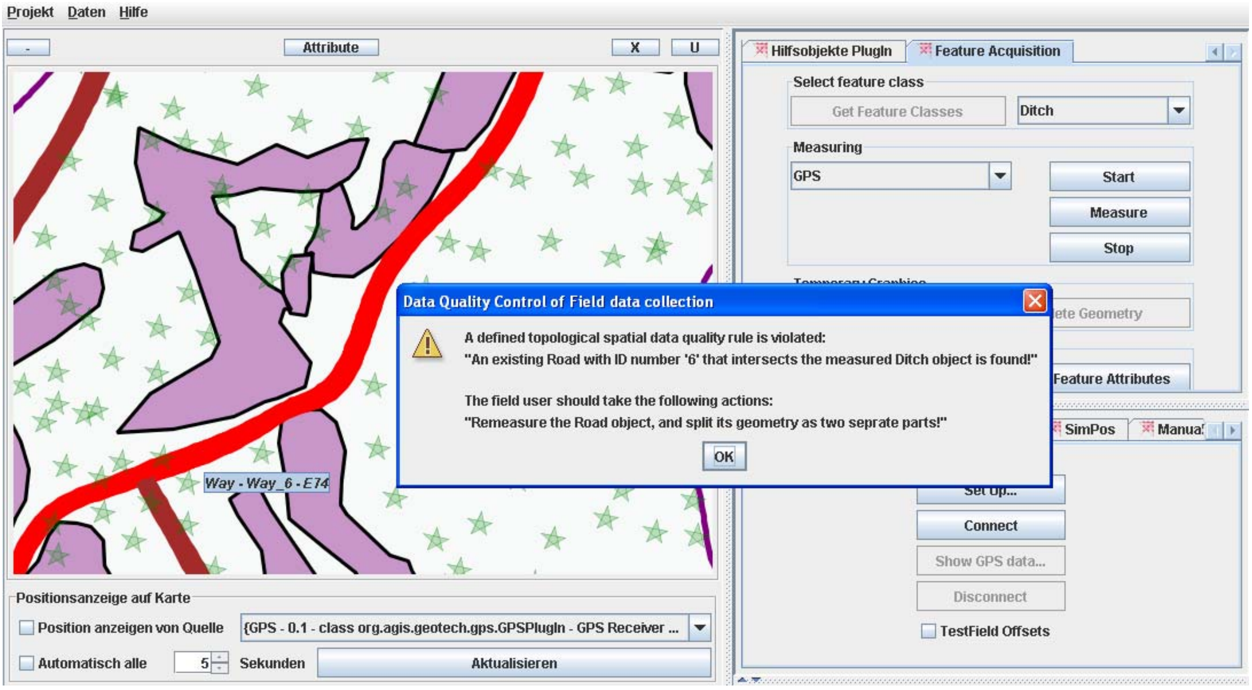


Figure 59: Quality-aware module demo 2

If none of the defined spatial data integrity rules is violated or the errors are corrected, the collected data can be successfully transacted to the remote database through the WFS-T.3

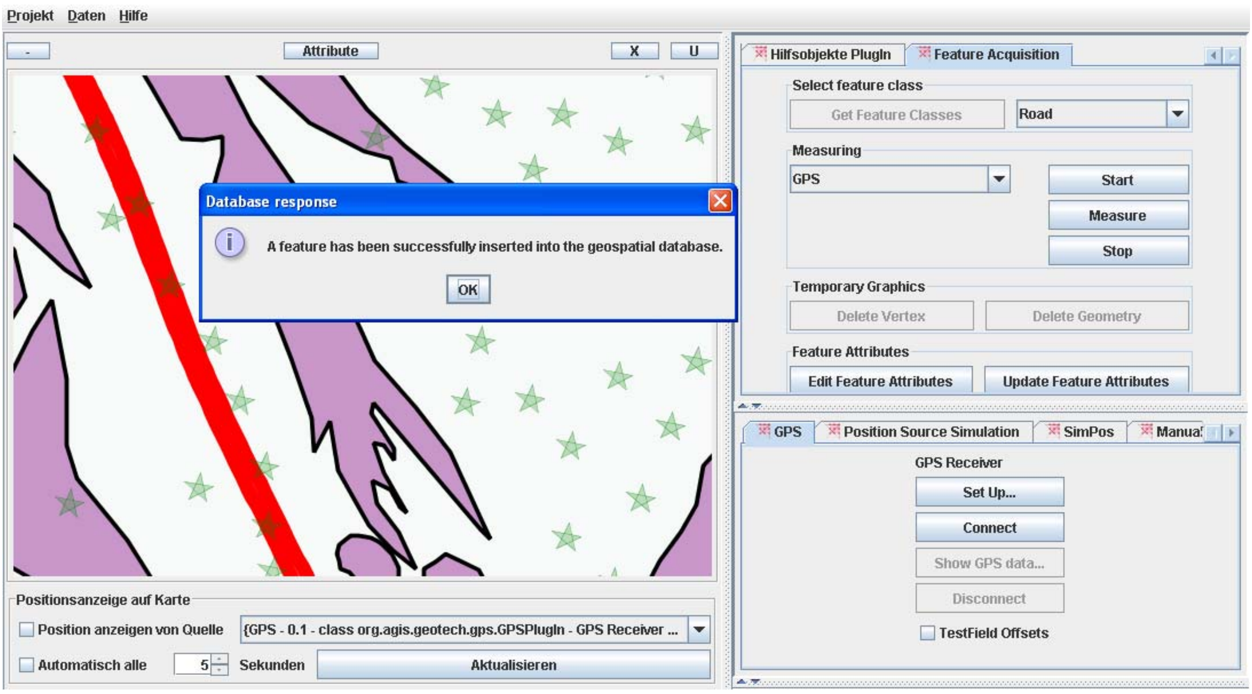


Figure 60: Successful data transaction

More introductions of the field tests and the software demonstration for this landslide monitoring system can be found in [Wang et al., 2005; Wang and Reinhardt, 2006b].

5.5 Summary

In Chapter 5, the field based Mobile GIS is adopted for testing the use of the proposed 'spatial data integrity rules'-based data model. First, a generic Mobile GIS architecture for field survey data capture is provided, and its main hardware and software components are explained.

Then the detailed steps of the data capture in the field are given as three separate scenarios "insert a new feature", "update an existing feature" and "delete an existing feature". Examples about how these steps are executed by Mobile GIS client software are offered.

Afterwards, the concept of the quality-aware geocomputational tool for Mobile GIS is defined. This tool is able to apply the defined spatial data integrity rules to the detailed steps of the data capture workflow, in order to support the data consistency checking of the data collected directly in the field. A prototype of the geocomputational tool is implemented and how its quality-aware functionality works for the data capture workflow is depicted.

Finally, the field tests of a landslide Mobile GIS application for evaluating the proposed concepts are described. The background and practical problems of the field tests are explained, and the pictures and examples in the field survey are provided. For handling the data consistency of the Mobile GIS using the methodology of this thesis, the field tests achieved very desirable results.

Chapter 6. Discussion of Results

In this chapter, first a summary of the achievements of this thesis is provided. According to the established objectives, the conclusions and explanations of them are described. The limitations of the proposed methodology are also discussed. Then the future research and outlook relating to this research are given.

6.1 Summary of Results

Altogether, the objectives of this thesis have been fulfilled, and the hypothesis of using constraint decision tables to specify spatial data integrity rules has been testified and proven to be possible. Constraint decision tables are treated as the extension to the standardized data modeling, and then are applied to the field-based Mobile GIS application after the format transformation.

In detail, the achieved results and limitations can be summarized according to the established three objectives in this thesis.

1) **A new method to define and specify spatial data integrity rules for describing data consistency requirements in a structured and standardized way, as well as the approach to express extensive instruction information for data users when data inconsistencies are found.**

Spatial data integrity rules are often used to specify data consistency requirements and to help checking the data integrity in GIS. With the review of the literature, different methods of defining and organizing spatial data integrity rules are compared, and their shortcomings such as non-standardized structure and inadequate expression are exposed. To better handle spatial data consistency, the term "spatial data integrity rule" is entitled to have the following meaning:

A formal and accepted statement, definition or qualification for describing data consistency requirements in order to constrain the spatial data to correctly represent the reality in the context of the GIS applications

To realize the aim, the proposed methodology of defining spatial data integrity rules concerns several important issues:

- The rules should give a clear specification of spatial relationships including semantic and temporal information
- The rules should be well structured and logical which keeps the rules compact and explicit, and should have a good readability
- The rules should have the clear instructions information for data users
- The policy for dealing with the rules violation should be provided

Then, the concept of the ECA rule is introduced, and based on that we propose constraint decision tables to contain spatial data integrity rules. Different terms with spatial abilities are defined in constraint decision tables. With such a logical tabular structure, end users or non-modelers can easily define and understand spatial data integrity rules.

As a part of geographic data modeling components, the cooperative usage of spatial data integrity rules and UML conceptual schema is described. To extensively represent the complex data consistency requirements in the real world, spatial data quality elements and subelements of the ISO/TC211 standard 19113 Quality principles are investigated. Diverse examples are provided to show the feasibility of the proposed method.

However, users have to define constraint decision tables through carefully manual input, since no computer assisted software tool is provided in this research. This causes the difficulty for users when large numbers of constraint decision tables are created or a constraint decision table contains complex contents.

In this thesis, we only focus on the cases that a spatial data integrity rule is expressed through one constraint decision table, and the provided examples only contain one or two feature types. The complex spatial data integrity rules that include more than two feature types or need to be expressed through more than one constraint decision table are not discussed.

The relationships among different spatial data integrity rules are not investigated, such as finding mistakes or inconsistencies of rules.

2) **The capability of the format transformation of spatial data integrity rules and normal conceptual data schema, for the purpose of better publishing and distributing them through a single geospatial portal.**

In this thesis, a method of transforming both UML conceptual schema and constraint decision tables into web compatible formats is offered. The UML conceptual schema is converted into the GML application schema. Constraint decision tables are encoded into the XML based rules' document.

To distribute spatial data integrity rules through the existing OGC WFS-T server, extended functions for WFS-T are implemented. Thus, data users can access all contents of the data model from the same geospatial portal through standard HTTP requests and responses.

The transformation approach only aims to solve basic rule language syntaxes. The development of a comprehensive spatial data integrity rules Markup language and the introduction of it to the standardization process needs more investigations and efforts.

3) **The practicability of spatial data integrity rules evaluated by a quality-aware field-based Mobile GIS data capture application.**

To apply the proposed methodology to GIS for supporting data consistency checking, a prototype of the quality-aware geocomputational tool is developed. This tool is an independent software module that can be easily integrated into existing GIS.

In order to testify the overall concept, a practical field based Mobile GIS application for the data capture in a landslide area (Balingen in Germany) is introduced. The results of real world field tests are then provided. It shows the proposed methodology is able to efficiently support and improve the spatial data consistency checking ability of Mobile GIS.

The field tests only used a small amount of dataset and emphasized on the simple geometry types. The practicability of the concept for large amounts of dataset and complex geometries in different GIS applications was not investigated.

6.2 Outlook

Although the proposed methodology leads to the desirable results, it still has the limitations as shown in last subsection. Thus, there are several open issues for future researches.

- **The software tool, which helps end users to define, view and organize the spatial data integrity rules, needs to be developed.**

In this thesis, spatial data integrity rules are created in constraint decision tables without developing any computer aided tool. Users have to create constraint decision tables according to the provided terms in a very careful manner, in order to avoid mistakes. In future, a well designed tool with a user-friendly GUI for creating spatial data integrity rules may be necessary to improve the efficiency and correctness of the rules' definition. In this way, end users or non experts can define data integrity rules in a simple and flexible way.

In the literature, Nottrott et al. [Nottrott et al., 1999] implemented a metadata editor which can generate XML-encoded ecological metadata information of the data. With the metadata editor, users also can define some simple constraints, like the range value of a variable. Later on, Servigne et al. [Servigne et al., 2000] developed a tool to allow defining topological constraints as shown in Figure 61. Users can define the topological constraints without knowing the logic way of specifying the constraints, for example selecting the spatial objects and their relationships from drop down lists and viewing the result from the picture. Similar approaches were also done by Cockcroft [Cockcroft, 2001] and [Mäs, 2007], who considered constraints with semantic information.

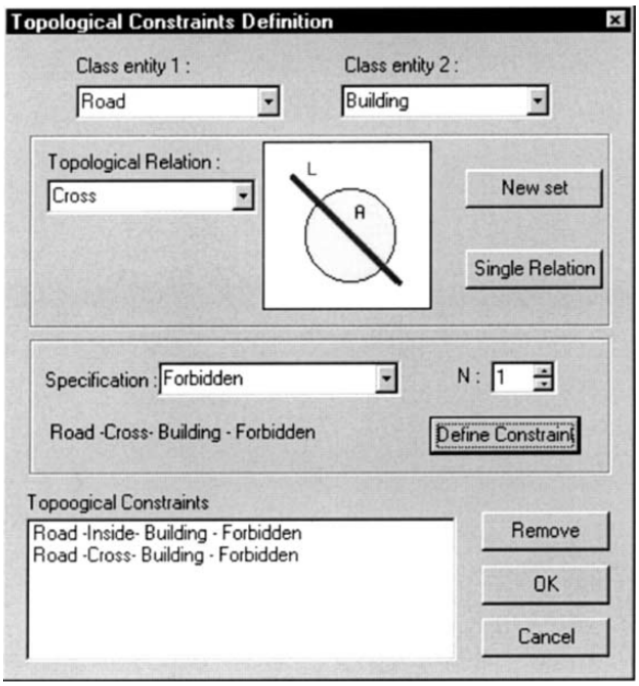


Figure 61: Topological constraints definition tool [Servigne et al., 2000]

With a similar approach, a tool may be developed to provide users a friendly GUI to define spatial data integrity rules and automatically generate the rules in constraint decision tables or in XML-based rules' document.

- **Management of complex spatial data integrity rules in constraint decision tables needs to be studied.**

When we study the method to specify spatial data integrity rules, we expect the rules should be well structured with explicit contents in order to have a good readability. The proposed 'constraint decision tables' method that bases on Event Condition Action Rule is aimed to provide a clear and logical structure. As shown by the examples of this thesis, people can easily comprehend the meanings of defined constraint decision tables.

However, the given examples of constraint decision tables only contain one or two feature types, which make them compact and intelligible. End users are able to understand those examples without the difficulty. In the reality, spatial data integrity rules may involve more than two feature types, their attributes and other semantic and temporal information. Thus, a complex spatial data integrity rule may be expressed through more than one constraint decision table, and those tables may consist of many elements. It will become hard for people especially end users or non-experts to find out the meanings of those constraint decision tables. Moreover, different constraint decision tables may have connections or relations with each other, for example, one constraint decision table can not be activated unless another special constraint decision table is activated.

- **The effort to introduce spatial data integrity rules into the standardization process needs to be carried out.**

For this point, there are two different standardization sessions that need to be considered. The first one is about the XML-based rule language to express spatial data integrity rules. As described in this thesis, some terms with spatial abilities are defined in the constraint decision tables. Those terms are then encoded into the rules' document based on the given conversion conditions. However, this approach only aims to solve basic rule language syntaxes. More efforts of considering existing or upcoming rule language standards have to be carried out. For example, to extend the standardized Rule Markup Language (RuleML) with built-in spatial terms such as spatial relationships and geometry operations.

The second session refers to distribution the data integrity rules' document through standardized geospatial web services. As explained in section 4.4, current widely used OGC WFS-T does not have the ability to handle the rules' document through its own standard HTTP requests. Extended functions of WFS-T are suggested in this thesis. However, to make standardization bodies like OGC aware of this issue and update the current specification, a great deal of work needs to be done, such as the spatial data integrity rules' discovery, the rules' search, sever-side data consistency checking, as well as OGC standardization procedures.

Recently, OGC paid attention to this flaw and set up a Data Quality Working Group:

"The mission of the DQ WG to establish a forum for describing an interoperable framework or model for OGC Quality Assurance measures and Web Services to enable access and sharing of high quality geospatial information, improve data analysis and ultimately influence policy decisions" [OGC, 2007]

One of its aims is to propose a “Web Based Data Quality Assurance Service” for handling spatial data quality information through web services. This is in the initial stage and the process is still under construction.

- **The method for reasoning the defined spatial data integrity rules needs to be studied, and the reasoning ability may be supported by a software tool.**

Spatial data integrity rules are defined in this thesis without checking the relationships among the rules through the rule reasoning method. In this way, it may bring the shortcomings such as: duplicate rules with same meanings, conflicts between different rules and so forth.

For example, one rule is given as "The building with ID 37 is topologically *Within* the campus of Universität der Bundeswehr München". Another rule is described as "The campus of Universität der Bundeswehr München topologically *Contains* the building with ID 37". Obviously, the two rules have same meanings and can be expressed only one time. Without considering the reasoning of those topological relations, duplicate cases can not be easily avoided.

Another example: the first rule defines "Object A *contains* object B, and object A *disjoints* with object C", and the second rule gives "Object B *intersects* with object C". This case is totally wrong and these two rules can not coexist, as shown in Figure 62. But the system can not automatically recognize this error without having such rules reasoning method.

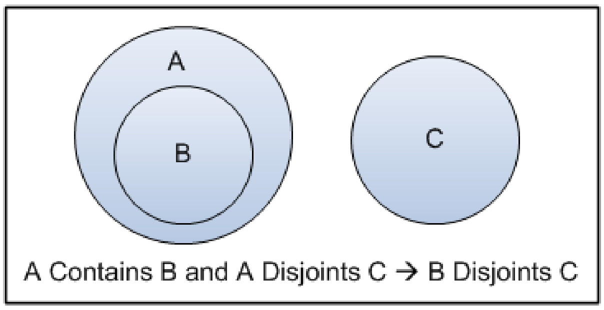


Figure 62: Topological relations inference

This is an interesting field of research in future. Some researchers in this area such as Grigni et al. [Grigni et al., 1995] and Mäs [Mäs, 2007] have already carried out.

A software tool may be necessary to perform the rules’ reasoning and find their inconsistencies. The tool may be regarded as a subpart of the tool mentioned in last point.

- **Applying the overall concept to different GIS applications needs to be evaluated.**

In this thesis, a field-based Mobile GIS application is adopted for evaluating the practicability of the proposed methodology. The achieved results show that our method can work very well for Mobile GIS field data capture applications. Thus, to find out whether the proposed method can also bring the same success in other types of GIS applications, more practical tests need to be performed. For example, applications of the quality assessment and control for data in large geospatial databases [Caspary and Joos, 1998; Joos, 1996; Joos, 1999; Mostafavi et al., 2004].

References

Aquino, J., 2003. JTS Topology Suite Developer's Guide. VIVID SOLUTIONS.

Aronoff, S., 1989. Geographic Information Systems: A Management Perspective. WDL Publications, Ottawa, Canada.

Bailey, J., Poulovassilis, A. and Wood, P.T., 2002. An event-condition-action language for XML, Proceedings of the 11th international conference on World Wide Web. ACM Press, Honolulu, Hawaii, USA, pp. 486-495.

Becker, L., Ditt, H., Hinrichs, K. and Voigtmann, A., 1999. Constraints and triggers: a method for ensuring data quality in an object-oriented geo database kernel, 7th ACM international symposium on Advances in geographic information systems. ACM Press, Kansas City, USA, pp. 160 - 161.

Bédard, Y., 1999. Visual Modelling of Spatial Database Towards Spatial PVL and UML. Geomatica, 53(2): 169-185.

Bédard, Y., 2005. Perceptory Web Site, http://sirs.scg.ulaval.ca/Perceptory/.

Bédard, Y., Larrivée, S., Proulx, M.-J. and Nadeau, M., 2004. Modeling Geospatial Databases with Plug-Ins for Visual Languages: A Pragmatic Approach and the Impacts of 16 Years of Research and Experimentations on Perceptory. In: S. Wang et al. (Editors), CoMoGIS'04. Lecture Notes in Computer Science. Springer, Shanghai, China, pp. 5-16.

Bédard, Y. and Vallière, D., 1995. Qualité des données à référence spatiale dans un contexte gouvernemental, Laval University, Quebec.

Belussi, A., Negri, M. and Pelagatti, G., 2004a. GEOUML: A Geogaphic Conceptual Model Defined through Specialization of ISO TC211 Standards, 10th EC GI & GIS Workshop,ESDI State of the Art, Warsaw, Poland.

Belussi, A., Negri, M. and Pelagatti, G., 2004b. an ISO TC 211 compatible data model for the conceptual design of geographic databases. Internal report N. 2004.21, Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy.

Bernauer, M., Kappel, G. and Kramler, G., 2004. Representing XML Schema in UML - A Comparison of Approaches. LECTURE NOTES IN COMPUTER SCIENCE, 3140: 440-444.

Boin, A.T. and Hunter, G.J., 2007. What communicates quality to the spatial data consumer?, ISSDQ 2007, Enschede, Netherland.

Bonham-Carter, G.F., 1994. Geographic Information Systems for Geoscientists: Modelling with GIS. COMPUTER METHODS IN THE GEOSCIENCES, 13. Elsevier, Kidlington UK.

Borges, K.A.V., Clodoveu A. Davis, J. and Laender, A.H.F., 2002. Integrity constraints in spatial databases, Database integrity: challenges and solutions. Idea Group Publishing, pp. 144-171.

Borges, K.A.V., Davis, C.A. and Laender, A.H.F., 2001. OMT-G: An Object-Oriented Data Model for Geographic Applications. GeoInformatica, 5(Volume 5, Number 3): 221 - 260.

Box, G.E.P., 1976. Science and Statistics. Journal of the American Statistical Association, 71: 791-799.

Brassel, K., Bucher, F., Stephan, E.-M. and Vckovski, A., 1995. Completeness. In: I.C. Association (Editor), Elements of Spatial Data Quality. Elsevier.

Breunig, M. et al., 2005. Further developing mobile spatial services for the Geosciences, 19th International CODATA Conference. Data Science Journal, Berlin.

Brodeur, J., Bédard, Y. and Proulx, M.-J., 2000. Modelling geospatial application databases using UML-based repositories aligned with international standards in geomatics, Proceedings of the 8th ACM international symposium on Advances in geographic information systems. ACM Press, Washington, D.C., United States, pp. 39-46.

Burrough, P.A., 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford Universtity Press, Oxford, 194 pp.

Burrough, P.A. and McDonnell, R.A., 2000. Principles of Geographical Information Systems. Oxford Universtity Press, Oxford, 333 pp.

Carlson, D., 2001. Modeling XML Applications with UML: Practical e-Business Applications. Addison Wesley, 368 pp.

Casanova, M., Wallet, T. and D'Hondt, M., 2000. Ensuring Quality of Geographic Data with OCL and UML. In: A. Evans, S. Kent and B. Selic (Editors), UML 2000 - The Unified Modeling Language, Advancing the Standard: Third International Conference, York, UK, pp. 225-239.

Casanova, M., Wallet, T. and D'Hondt, M., 2002. Integrating an Explicit Knowledge Model into Geographic Information Systems, The 14th international conference on Software engineering and knowledge engineering, Ischia, ITALY.

Caspary, W., 1992. Qualitätsmerkmale von Geodaten. Zeitschrift für Vermessungswesen (ZfV), 7: 360-367.

Caspary, W., 1993. Qualitätsaspekte bei Geo-Informationssystemen. Zeitschrift für Vermessungswesen (ZfV), 8/9: 444-449.

Caspary, W. and Joos, G., 1998. Quality Criteria and Control for GIS Databases. In: H. Kahmen, E. Brückl and T. Wunderlich (Editors), IAG SC4 Symposium Eisenstadt/Austria.

Caspary, W. and Scheuring, R., 1993. Positional accuracy in spatial databases. Computers, Environment and Urban Systems, 17: 103-110.

Chrisman, N., 1991. The error component in spatial data. In: D.J. Maguire, M.F. Goodchild and D.W. Rhind (Editors), Geographical information systems: principles and applications. Longman Scientific and Technical, Harlow, UK, pp. 165-174.

Chrisman, N., 1996. Exploring Geographic Information Systems. John Wiley & Sons, Inc, New York, US, 298 pp.

Clarke, D.G. and Clark, D.M., 1995. Lineage. In: I.C. Association (Editor), Elements of Spatial Data Quality. Elsevier.

Clementini, E. and Felice, P.D., 1995. A comparison of methods for representing topological relationships. Inf. Sci. Appl., 3(3): 149-178.

Clementini, E. and Felice, P.D., 1996. A model for representing topological relationships between complex geometric features in spatial databases. Inf. Sci., 90(1-4): 121-136.

Cockcroft, S., 1997. A Taxonomy of Spatial Data Integrity Constraints. Geoinformatica, 1(4): 327-343.

Cockcroft, S., 1998. Spatial Data Quality Control through the Use of an Active Repository, Geocomputation 98.

Cockcroft, S., 2001. Modelling Spatial Data Integrity Rules at the Metadata Level, Geocomputation 01, University of Queensland, Brisbane, Australia.

Cockcroft, S., 2004. The Design and Implementation of a Repository for the Management of Spatial Data Integrity Constraints. GeoInformatica, 8(1): 49-69.

Collins, F.C. and Smith, J.L., 1994. Taxonomy for error in GIS. In: R.G. Congalton (Editor), International symposium on spatial accuracy in Natural Resource Data Bases "Unlocking the puzzle,". American Society for Photogrammetry and Remote Sensing, Williamsburg, US, pp. 1-7.

Comber, A.J., Fisher, P.F. and Wadsworth, R.A., 2007. User-focused metadata for spatial data, geographical information and data quality assessments. In: S. Fabrikant and M. Wachowicz (Editors), AGILE 2007. Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Aalborg, Danmark.

Cowen, D.J., 1988. GIS versus CAD versus DBMS: what are the differences? Photogrammetric Engineering and Remote Sensing, 54(1551-4).

deegree, 2007. www.deegree.org.

Devillers, R., Bédard, Y. and Jeansoulin, R., 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., Bédard, Y., Jeansoulin, R. and Moulin, B., 2007. Towards Spatial Data Quality Information Analysis Tools for Experts Assessing the Fitness for Use of Spatial Data. International Journal of Geographical Information Sciences, 21(3): 261-282.

Devillers, R. and Jeansoulin, R., 2006a. Fundamentals of Spatial Data Quality. ISTE Ltd, 309 pp.

Devillers, R. and Jeansoulin, R., 2006b. Spatial Data Quality: Concepts. In: R. Devillers and R. Jeansoulin (Editors), Fundamentals of Spatial Data Quality. ISTE Ltd, pp. 309.

Ditt, H., Voigtmann, A., Becker, L., Voigtmann, A. and Hinrichs, K.H., 1997. Constraints and Triggers in an Object-Oriented Geo Database Kernel, 8th International Workshop on Database and Expert Systems Applications (DEXA '97), pp. 508-513.

Drummond, J., 1995. Positional accuracy. In: I.C. Association (Editor), Elements of Spatial Data Quality. Elsevier.

Duckham, M., 2002. A user-oriented perspective of error-sensitive GIS development. Transactions in GIS, 6(2): 179-194.

Duckham, M. and McCreadie, J., 2002. Error-aware GIS development. In: W. Shi, M.F. Goodchild and P.F. Fisher (Editors), Spatial Data Quality. Taylor and Francis, pp. 63-75.

Egenhofer, M., 1993. A model for detialed binary topological relationships. Geomatica, 47: 261-273.

Egenhofer, M.J. and Franzosa, R.D., 1991. Point-Set Topological Spatial Relations. International Journal of Geographical Information Systems, 5(2): 161-174.

Erik Behrends, Oliver Fritzen, Wolfgang May and Schenk, F., 2006. Combining ECA Rules with Process Algebras for the Semantic Web, RuleML 2006: Rules and Rule Markup Languages for the Semantic Web, Athens, U.S.A.

Fekete, B.M., Vorosmarty, C.J. and Lammers, R.B., 2001. Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error. Water Resources Research, 37(7): 1955-1968.

Filho, J.L. and Iochpe, C., 1999. Specifying analysis patterns for geographic databases on the basis of a conceptual framework, Proceedings of the 7th ACM international symposium on Advances in geographic information systems. ACM Press, Kansas City, Missouri, United States.

Frank, A.U., 1996. Qualitative spatial reasoning: cardinal directions as an example. International Journal of Geographical Information Systems, 10(3): 269-290.

Frank, A.U., 1998. Metamodels for data quality description. In: R. Jeansoulin and M. Goodchild (Editors), Data Quality in Geographic Information - from Error to Uncertainty. Hermes, Paris.

Friis-Christensen, A., Tryfona, N. and Jensen, C.S., 2001. Requirements and Research Issues in Geographic Data Modeling, The Ninth ACM International Symposium on Advances in Geographic Information Systems, Atlanta, GA, USA.

Goodchild, M.F., 1992. Geographical information science. International Journal of Geographical Information Systems, 6(1): 31–45.

Goodchild, M.F., 1995a. Attribute accuracy. In: I.C. Association (Editor), Elements of Spatial Data Quality. Elsevier.

Goodchild, M.F., 1995b. Sharing imperfect data. In: H.J. Onsrud and G. Rushton (Editors), Sharing Geographic Information. Rutgers University Press, New Brunswick, NJ, pp. 413-425.

Goodchild, M.F., 2003. GEOGRAPHIC INFORMATION SCIENCE AND SYSTEMS FOR ENVIRONMENTAL MANAGEMENT. Annual Review of Environment and Resources, 28: 493-519.

Goodchild, M.F., 2006. Geographical information science: fifteen years later. In: P.F. Fisher (Editor), Classics from IJGIS: Twenty years of the International Journal of Geographical Information Science and Systems. CRC Press, Boca Raton, pp. 199-204.

Goodchild, M.F., 2007. Beyond Metadata: Towards User-Centric Description of Data Quality, ISSDQ 2007, Enschede, Netherland.

Gordillo, S., Balaguer, F., Mostaccio, C. and Neves, F.D., 1999. Developing GIS Applications with Objects. A design Patterns Approach. Geolnformtica, 3(1): 7-32.

Grigni, M., Papadias, D. and Papadimitriou, C., 1995. Topological inference, International Joint Conference of Artificial Intelligence (IJCAI), pp. 901-906.

Grønmo, R., 2001. Supporting GI standards with a model-driven architecture, Proceedings of the 9th ACM international symposium on Advances in geographic information systems. ACM Press, Atlanta, Georgia, USA, pp. 100-105.

Grønmo, R., Solheim, I. and Skogan, D., 2002. Experiences of UML-to-GML Encoding, 5th AGILE Conference on Geographic Information Science, Palma (Balearic Islands, Spain).

Guptill, S.C. and Morrison, J.L. (Editors), 1995. Elements of Spatial Data Quality. Elsevier.

Hunter, G.J. and Beard, M.K., 1992. Understanding error in spatial databases. Australian Surveyor, 37(2): 108-119.

Hunter, G.J., De Bruin, S. and Bregt, A., 2007. Improving the Usability of Spatial Information Products and Services. In: S. Fabrikant and M. Wachowicz (Editors), The European Information Society: Leading the Way with Geo-information (AGILE 2007). Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Aalborg, Danmark, pp. 405-418.

ISO, 2001. ISO 19101 Geographic information -- Reference model.

ISO, 2002a. Geographic information -- Quality principles, ISO 19113:2002, pp. TC 211; ISO Standards.

ISO, 2002b. ISO 19115 Geographic information -- Metadata.

ISO, 2002c. ISO 19118 Geographic information -- Encoding.

ISO, 2004. ISO 19125-1 Geographic information -- Simple feature access -Part1: Common architecture.

ISO, 2005. ISO 9000: Quality management systems - Fundamentals and vocabulary.

ISO, 2006. Review summary of preliminary work item Amendment to ISO 19113:2002 Geographic information - Quality principles and ISO 19115:2003 Geographic information - Metadata.

Joos, G., 1994. Quality Aspects of Geo-Informations, the fifth European conference on geographical information systems, EGIS, Paris, France.

Joos, G., 1996. Konsistenzprüfung für das ATKIS-Datenmodell. Internet Bericht Nr. 5, Institut für Geodäsie und Geoinformatik.

Joos, G., 1999. Assessing the quality of geosdata by testing consisteny with respect to the conceptual data schema. In: M. Craglia and H. Onsrud (Editors), Geographic Information Research: Trans-Atlantic Perspectives. Taylor & Francis, pp. 509-519.

Joos, G., 2000. Zur Qualität von objektstrukturierten Geodaten, Universität der Bundeswehr München, München, 141 pp.

Kandawasvika, A. and Reinhardt, W., 2005. Concept for interoperable usage of multi-sensors within a landslide monitoring application scenario, 8th AGILE Conference on Geographic Information Science.

Kresse, W. and Fadaie, K., 2004. ISO Standards for Geographic Information. Springer, 322 pp.

Longley, P.A., 1998. Foundations. In: P.A. Longley, S.M. Brooks, R.A. McDonnell and B. Macmilian (Editors), Geocomputation: A Primer. John Wiley & Sons Ltd, pp. 278.

Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W., 2005. Geographical Information Systems and Science. John Wiley & Sons, 536 pp.

Louwsma, J., Zlatanova, S., Lammeren, R.v. and Oosterom, P.v., 2006. Specifying and Implementing Constraints in GIS - with Examples from a Geo-virtual Reality System. GeoInformatica, 10: 531-550.

Lu, C.-T., Dos Santos, R.F., Sripada, L.N. and Kou, Y., 2007. Advances in GML for Geospatial Applications. Geoinformatica, 11: 131-157.

Mark, D.M., 2003. Geographic information science: defining the field. In: M. Duckham, M.F. Goodchild and M.F. Worboys (Editors), Foundations of Geographic Information Science. Taylor and Francis, New York, pp. 3-18.

Mäs, S., 2007. Reasoning on Spatial Semantic Integrity Constraints, Conference on Spatial Information Theory (COSIT 2007). Lecture Notes in Computer Science 4736, Melbourne, Australia, pp. 285–302.

Mäs, S., Reinhardt, W., Kandawasvika, A. and Wang, F., 2005a. Concepts for quality assurance during mobile online data acquisition, Proceeding 8th AGILE Conference on GIScience, Estoril, Portugal.

Mäs, S., Wang, F. and Reinhardt, W., 2005b. Using Ontologies for Integrity Constraint Definition, 4th International Symposium on Spatial Data Quality, Beijing, China.

Maula, H., 2001. Mobile GIS and Multimedia as Fieldwork Support in Helsinki Water, 4th AGILE conference, Brno, Czech Republic, pp. 205-208.

Morrison, J.L., 1988. the proposed standard for digital cartographic data. The American Cartographer, 15(1): 129-135.

Mostafavi, M.A., Edwards, G. and Jeansoulin, R., 2004. An ontology-based method for quality assesment of spatial data bases, ISSDQ'04 Proceedings. Frank,A. and Grum,E.

Najar, C.R., 2006. A model-driven approach to management of integrated metadata-spatial data in the context of spatial data infrastructures.

Nottrott, R.W., Jones, M.B. and Schildhauer, M., 1999. Using XML-Structured Metadata to Automate Quality Assurance Processing for Ecological Data, Advance Program for META-DATA '99, 3rd IEEE META-DATA CONFERENCE, Bethesda, Maryland.

Nusser, S., Miller, L., Clarke, K. and Goodchild, M., 2003. Geospatial IT for mobile field data collection. Commun. ACM, 46(1): 45-46.

OGC, http://www.opengeospatial.org/.

OGC, 1999. OpenGIS Simple Features Specification For SQL, Revision 1.1, OpenGIS Project Document 99-049.

OGC, 2004a. Geography Markup Language (GML) Implementation Specification 3.1.1.

OGC, 2004b. Web Map Service Version 1.3. Reference number of this OGC. project document: OGC 04-024.

OGC, 2005a. Filter Encoding Implementation Specification, Version 1.1. OpenGIS Project Document: 02-058.

OGC, 2005b. Web Feature Service Implementation Specification 1.1. Reference number of this OpenGIS project document: OGC 04-094.

OGC, 2007. http://www.opengeospatial.org/projects/groups/dqwg.

Oliveira, J.L., Pires, F. and Medeiros, C.M.B., 1997. An environment for modeling and design of geographic applications. GeoInformatica, 1(1): 29-58.

OMG, Object Mangement Group: Unified Modeling Language (UML), Object Constraint Language (OCL).

OMG, 2005. MOF 2.0/XMI Mapping Specification.

OnlineEtymology, 2007. http://www.etymonline.com/index.php.

Papamarkos, G., Poulovassilis, A. and Wood, P., 2003. Event-Condition-Action Rule Languages for the Semantic Web, Workshop on Semantic Web and Databases, at VLDB'03, Berlin.

Paschke, A., 2006. ECA-LP / ECA-RuleML: A Homogeneous Event-Condition-Action Logic Programming Language, RuleML 2006: Rules and Rule Markup Languages for the Semantic Web, Athens, U.S.A.

Peng, Z.-R. and Tsou, M.-h., 2003. Internet GIS : distributed geographic information services for the internet and wireless networks. Wiley, Hoboken, NJ, xxxvi, 679 pp.

Peterson, M., 2003. Maps and the Internet: An Introduction. In: M. Peterson (Editor), Maps and the Internet. Elsevier Science.

Peuquet, D.J. and Zhang, C.-X., 1987. An algorithm to determine the directional relationship between arbitrarily-shaped polygons in the plane. Pattern Recogn., 20(1): 65-74.

Pinet, F., Kang, M.-A. and Vigier, F., 2005. Spatial Constraint Modelling with a GIS Extension of UML and OCL: Application to Agricultural Information Systems, Metainformatics. Lecture Notes in Computer Science. Springer.

Plan, O., Mäs, S., Reinhardt, W., Kandawasvika, A. and Wang, F., 2004. Konzepte für die mobile Erfassung von Geodaten, Geoinformation und Mobilität von der Forschung zur praktischen Anwendung, Tagungsband Münsteraner GI-Tage 2004. IFGI prints, pp. 97-105.

Portele, C., 2005. Mapping UML to GML Application Schemas: Guidelines and Encoding Rules.

Price, R., Tryfona, N. and Jensen, C., 2000. Extended Spatiotemporal UML: Motivations, Requirements, and Constructs. Journal of Database Management. Special Issue on UML, 11(4): 14-27.

Pullar, D., 1997. Rule-based modelling in GIS, 2nd annual conference of GeoComputation, University of Otago, New Zealand, pp. 15-23.

Pundt, H., 2002. Field Data Acquisition with Mobile GIS: Dependencies Between Data Quality and Semantics. GeoInformatica, 6(4): 363-380.

Pundt, H. and Brinkkötter-Runde, K., 2000. Visualization of spatial data for field based GIS. Comput. Geosci., 26(1): 51-56.

Reinhardt, W., 2001. Zur Bedeutung der Internettechnologie in Geoinformationssystemen. In: Internationale Geodätische Woche - Tagungsband.

Reinhardt, W., 2004. Raumbezogene Informationssysteme. Handbuch Ingenieurgeodäsie Herbert Wichmann Verlag, Heidelberg, 226 pp.

Reinhardt, W., Kandawasvika, A., Mäs, S. and Wang, F., 2005. Mobile Geospatial Data Collection - concepts and first experiences, International Cartographic Conference, A Coruna, Spain.

RuleML, http://www.ruleml.org/.

Salge, F., 1995. Semantic accuracy. In: I.C. Association (Editor), Elements of Spatial Data Quality. Elsevier.

Salim, F.D., Price, R., Indrawan, M. and Krishnaswamy, S., 2004. Graphical Representation of XML Schema. Lecture Notes in Computer Science, 3007: 234 - 245.

Schmidt, A.L., 2004. UMTS and WLAN interoperability, Technical University of Denmark.

Seirio, M. and Berndtsson, M., 2005. Design and Implementation of an ECA Rule Markup Language, RuleML-2005, international Conference on Rules and Rule Markup Languages for the Semantic Web and OWL Workshop, Galway, Ireland.

Servigne, S., Lesage, N. and Libourel, T., 2006. Quality Components, Standards, and Metadata. In: R. Devillers and R. Jeansoulin (Editors), Fundamentals of Spatial Data Quality. ISTE Ltd, pp. 309.

Servigne, S., Ubeda, T., Puricelli, A. and Laurini, R., 2000. A Methodology for Spatial Consistency Improvement of Geographic Databases. GeoInformatica, 4(1): 7-34.

Shekhar, S., Coyle, M., Goyal, B., Liu, D. and Sarkar, S., 1997. Data models in geographic information systems. Communications of the ACM, 40(4): 103-111.

Shi, W., 2004. XML -Basierter Austausch von Geodaten, Universität der Bundeswehr München, München, 111 pp.

Shi, W.Z., 1994. Modelling Positional and Thematic Uncertainties in Integration of Remote Sensing and Geographic Information Systems. ITC Publication No.22, Enschede, The Netherlands.

Stanek, H. and Frank, A.U., 1993. Data Quality Requirements for GIS defined by Law: A Case Study, UDMS'93, Wien, Austria.

SWRL, 2004. Semantic Web Rule Language (SWRL).

Tsou, M.-H., 2004. Integrated Mobile GIS and Wireless Internet Map Servers for Environmental Monitoring and Management. Cartography and Geographic Information Science, 31(3): 153-165.

Tveite, H., 2001. Structural Modelling for Geographical Information Systems using Icons, 8th Scandinavian Research Conference on Geographical Information Science, Norway.

Ubeda, T. and Egenhofer, M.J., 1997. Topological Error Correcting in GIS. Springer Verlag , Advances in Spatial Databases, 1262(SSD?7): 283-297.

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

Vallières, S., Brodeur, J. and Pilon, D., 2006. Spatial Integrity Constraints: A Tool for Improving the Internal Quality of Spatial Data. In: R. Devillers and R. Jeansoulin (Editors), Fundamentals of Spatial Data Quality. ISTE Ltd, pp. 309.

Vaziri, M. and Jackson, D., 1999. Some Shortcomings of OCL, the Object Constraint Language of UML, Proceedings of the Technology of Object-Oriented Languages and Systems (TOOLS 34'00). IEEE Computer Society, pp. 555.

Veregin, H., 1999. Data quality parameters. In: P.A. Longley, M. Goodchild, D.J. Maguire and D.W. Rhind (Editors), Geographical information systems. Wiley, New York, pp. 177–189.

W3C, 1999. Extensible Stylesheet Language Transformations (XSLT).

Wagner, G., 2002. How to Design a General Rule Markup Language?, in Proceedings of the Workshop XML Technologies for the Semantic Web (XSW 2002). Lecture Notes in Informatics, Gesellschaft f. Informatik, HU Berlin, Institut für Informatik.

Wagner, G., 2004. RuleML, SWRL and REWERSE: Towards a General Web Rule Language Framework. Semantic Web Technologies Column of the SIGSEMIS Bulletin, 1(1).

Wang, F., Mäs, S., Reinhardt, W. and Kandawasvika, A., 2005. An Ontology-based Quality Assurance Method for Mobile Data Acquisition. In: J. HŘEBIČEK and J. RÁČEK (Editors), 19th international conference on Informatics for Environmental Protection: Networking Environmental Information, EnviroInfo 2005. Masaryk University, Brno, Czech Republic.

Wang, F. and Reinhardt, W., 2006a. The integration of spatial integrity constraints in geographic data modeling. In: M. Raubal, H.J. Miller, A.U. Frank and M.F. Goodchild (Editors), GIScience 2006. IfGI prints, Muenster, Germany, pp. 395-398.

Wang, F. and Reinhardt, W., 2006b. Spatial data quality concerns for field data collection in Mobile GIS. In: J. Gong and J. Zhang (Editors), Geoinformatics 2006. Geospatial Information Science. SPIE, Wuhan, China, pp. 64201C.

Wang, F. and Reinhardt, W., 2007. Extending Geographic Data Modeling by Adopting Constraint Decision Table to Specify Spatial Integrity Constraints. In: S. Fabrikant and M. Wachowicz (Editors), The European Information Society: Leading the Way with Geo-information (AGILE 2007). Lecture Notes in Geoinformation and Cartography. Springer, Berlin, Aalborg, Danmark, pp. 435-454.

Wang, R.Y. and Strong, D.M., 1996. Beyond accuracy: what data quality means to data consumers. Journal of Management Information Systems, 12: 5-34.

Watson, P., 2007. Formal Languages for Expressing Data Consistency Rules and Implications for Reporting of Quality Metadata, ISSDQ 2007, Enschede, Netherland.

Widom, J. and Ceri, S., 1996. Active database systems: triggers and rules for advanced database processing. Morgan Kaufmann series in data management systems, xvii. Morgan Kaufmann, San Francisco, Calif., 332 pp.

Worboys, M.F., 1995. GIS: A Computing Perspective. Taylor & Francis Ltd, London, 376 pp.

Wright, D.J., Goodchild, M.F. and Proctor, J.D., 1997. Demystifying the persistent ambiguity of GIS as 'tool' versus 'science'. Annals of the Association of American Geographers 87(2): 346-362.

Zhang, J. and Goodchild, M., 2002. Uncertainty in Geograhic Information. CRC Press, 266 pp.

Appendix I: ISO/TC 211 Geographic information standards

ISO 6709:1983	Standard representation of latitude, longitude and altitude for geographic point locations
ISO 19101:2002	Geographic information -- Reference model
ISO/TS 19103:2005	Geographic information -- Conceptual schema language
ISO 19105:2000	Geographic information -- Conformance and testing
ISO 19106:2004	Geographic information -- Profiles
ISO 19107:2003	Geographic information -- Spatial schema
ISO 19108:2002	Geographic information -- Temporal schema
ISO 19108:2002/Cor 1:2006	
ISO 19109:2005	Geographic information -- Rules for application schema
ISO 19110:2005	Geographic information -- Methodology for feature cataloguing
ISO 19111:2003	Geographic information -- Spatial referencing by coordinates
ISO 19112:2003	Geographic information -- Spatial referencing by geographic identifiers
ISO 19113:2002	Geographic information -- Quality principles
ISO 19114:2003	Geographic information -- Quality evaluation procedures
ISO 19114:2003/Cor 1:2005	
ISO 19115:2003	Geographic information -- Metadata
ISO 19115:2003/Cor 1:2006	
ISO 19116:2004	Geographic information -- Positioning services
ISO 19117:2005	Geographic information -- Portrayal
ISO 19118:2005	Geographic information -- Encoding
ISO 19119:2005	Geographic information -- Services
ISO/TR 19120:2001	Geographic information -- Functional standards
ISO/TR 19121:2000	Geographic information -- Imagery and gridded data
ISO/TR 19122:2004	Geographic information / Geomatics -- Qualification and certification of personnel
ISO 19123:2005	Geographic information -- Schema for coverage geometry and functions
ISO 19125-1:2004	Geographic information -- Simple feature access -- Part 1: Common

	architecture
ISO 19125-2:2004	Geographic information -- Simple feature access -- Part 2: SQL option
ISO/TS 19127:2005	Geographic information -- Geodetic codes and parameters
ISO 19128:2005	Geographic information -- Web map server interface
ISO 19133:2005	Geographic information -- Location-based services -- Tracking and navigation
ISO 19134:2007	Geographic information -- Location-based services -- Multimodal routing and navigation
ISO 19135:2005	Geographic information -- Procedures for item registration
ISO/TS 19138:2006	Geographic information -- Data quality measures

Appendix II: International groups, conferences, standardization organizations on spatial data quality

Working groups	Conference	Standardization of metadata
International society for Photogrammetry and Remote Sensing (ISPRS): WG II/7: Quality of Spatio-Temporal Data and Models) http://igm.univ-mlv.fr/~jeansoul/ISPRS_WGII_7/	Accuracy (International Symposium on Spatial Accuracy Assessment in Natural Resources & Environmental Sciences): Organized by a research group that focuses on how to measure, model and manage uncertainty in spatial data, specifically the one that comes from natural resources and the environment. It has been held every two years since 1994.	ISO/TC 211
Associations of Geographic Information Laboratories in Europe (AGILE): WG on Spatial Data Usability http://www.modulobus.org/agile-sdu/default.html	ISSDQ (International Symposium on Spatial Data Quality): Addresses all aspects of spatial data quality in general. It has been held every two years since 1999.	OGC
International Cartographic Association (ICA): WG on Spatial Data Uncertainty and Map Quality http://www.icaci.org/en/commissions.html	Others: GIScience (every even year), AGILE (every year), ISPRS Congress (every year)	European Committee for Standardization (CEN)
EuroGeographics (European National Mapping and Cadastral Agencies Association): Expert Group on Quality http://www.eurogeographics.org/eng/05_quality.asp	Many other GI-related conferences have a specific topic on data quality	Federal Geographic Data Committee (FGDC)

Curriculum Vitae

Name: Fei Wang
Birthday: 1979, April 06
Birthplace: XiangCheng, HeNan, P.R. China
Contact: fei.wang@hotmail.de

Education:

2002 – 2008 Doctor of Engineer (Dr. –Ing) on Geoinformatics
Universität der Bundeswehr München, Deutschland

2000 – 2002 Master Degree of Science on Photogrammetry and Geoinformatics
Stuttgart University of Applied Sciences, Stuttgart, Deutschland

1996 – 2000 Bachelor Degree of Science on Geographic Information System (GIS)
Wuhan Technical University of Surveying and Mapping, Wuhan, P.R. China

Professional Experience:

2002 – 2008 Wissenschaft Mitarbeiter (Scientific research assistant) in Arbeitsgemeinschaft Geoinformationssysteme (AGIS), Universität der Bundeswehr München

1. 2002 - 2004
   PARAMOUNT (Public Safety & Commercial Info-Mobility Applications & Services in the Mountains) – IST program of the EU project
2. 2004 – 2006
   Advancement of Geoservices (A joint project as part of Geotechnologie) – BMBF (German Federal Ministry of Education and Research) and DFG (German Research Council) project
3. 2007 – 2008
   ASYS (Automatisches System zur Aufnahme und Dokumentation von Gebäudeentwässerungsanschlüssen)
4. 2006 – 2008
   3D Indoor Navigation – BMBF

Acknowledgements

I am full of gratitude to my supervisor, professor and mentor Wolfgang Reinhardt. His open-mindedness, motivation, insight and kindness are the foundation of the great success of my dissertation work. He always offered many useful comments through our discussions, and guided me to the right objectives.

My gratitude also goes to Prof. Anders Östman and Prof. Wilhelm Caspary for their valuable advice and discussions.

My former Master degree supervisor Prof. Rainer Kettemann from Stuttgart University of Applied Science also deserves my gratitude. His kindness and encouragement was also the key of my success.

Dr. James and Hua Gao provided me the opportunity of checking and improving the wording of my dissertation. The patience and carefulness of Dr. James made the dissertation more graceful.

During the working time in AGIS, I had many times of informative symposiums with my colleagues, such as Admire Kandawasvika, Simone Stürmer, Stephan Mäs, Stefan Strobel, and many others. Their advice from different aspects brought me the wide scope of my dissertation work.

Many thanks also go to my Chinese friends in this university, who helped me a lot during my work and private life.

My parents and my little brother in China always give me a big support in my whole life. I give my great appreciation to them.

My highest gratitude is to my wife Quan Zhou. First, she brought the best gift to me, that is, my daughter, named Zizhou. My lovely daughter made a great deal of happiness to my spare time. Second, my wife’s patience, carefulness, femininity and gentilesse make the family very harmonious. In the mean time, she also contributed many good suggestions and ideas to my dissertation.