A Model-Driven Approach to Management of Integrated Metadata - Spatial Data in the Context of Spatial Data Infrastructures

For the degree for

Doctor of Sciences

Presented by

Christine Ruth Najar

Dipl.-Ing. für Landeskultur und Umweltschutz
University of Rostock

Born 15th December 1976
Citizen of USA, Poland

Examiner:
Christine Giger

Co-Examiners:
Alain Buogo
Francois Golay
Abbas Rajabifard

-2006-

THE RIVER IS A PIECE OF SKY
(JOHN CIARDI)

FROM THE TOP OF THE BRIDGE

THE RIVER BELOW

IS A PIECE OF SKY-

UNTIL YOU THROW

A PENNY IN

OR A COCKLESHELL

OR A PEBBLE OR TWO

OR A BICYCLE BELL

OR A COBBLESTONE

OR A FAT MAN'S CANE-

AND THEN YOU CAN SEE

IT'S A RIVER AGAIN

THE DIFFERENCE YOU'LL SEE

WHEN YOU DROP YOUR PENNY:

THE RIVER HAS SPLASHES,

THE SKY HASN'T ANY

Table of Contents

List of Figures and Tables

FIGURES

TABLES

Abstract

The demand for spatial metadata to describe spatial data is growing in the networked environment. Yet, currently metadata acquisition and management often play a subordinate role in many organizations and are considered overhead. If at all, metadata are acquired much after the spatial data and are stored in separate repositories. Consequently, there are two independent data sets to manage and update: spatial data and metadata. These are often redundant and inconsistent, as it is not always clear which information is metadata and which is spatial data.

Looking at the interoperability in Spatial Data Infrastructures (SDIs) from a technical point of view, the specialized spatial search engine needs spatial data which are labeled and indexed by metadata. The more reliable and consistent such metadata are, the better they support an SDI as an enabling platform to search, exchange and process spatial data. This leaves a gap between the status-quo of metadata and the demand for metadata which needs to be accounted for with new metadata management concepts.

Consequently, the main focus of this thesis concerns the optimization of metadata management by integrating metadata and spatial data in a common file or database. This common metadata-spatial data set can be considered to be 'comprehensive spatial data'.

The concept of metadata-spatial data integration enables the spatial data to carry their own metadata description with them. The approach distinguishes between already existing spatial data models, which have to be extended and newly planned data models and sets, which can managed commonly from the beginning. The different groups of metadata which can be integrated are discussed (implicitly derivable, explicitly derivable and new metadata attributes) and the principles how these might be placed in a model (top-down and bottom-up). The three steps of integration include firstly a semantic analysis and translation, secondly a structural analysis and thirdly hierarchical integration as well as a semantic transformation.

Provided that common metadata-spatial data sets exist, the concept of views offers the possibility to extract metadata and spatial data according to various standards and other excerpts from the comprehensive data set. This gains flexibility and interoperability for using common metadata-spatial data sets in an SDI environment in which different services and users need different extracts and structures of a certain data or metadata set.

In order to review the feasibility of the concept of metadata-spatial data integration three test data models and their data sets are integrated with their corresponding metadata. The Swiss federal cadastral model, a water supply model of the City of Zurich and an environmental data model vary in complexity, size modeling structure, modeling language as well as in the question whether they are standardized. The results of this case study show that the integration of metadata in existing models and data sets is feasible. For each of the chosen models certain top-down and bottom-up metadata attributes are defined. Furthermore, the implicitly and explicitly derivable attributes are ascertained. In a workshop the common models were verified with experts who know the original models well.

In order to be able to use the concept of integration on any data set in a similar way, rules for the integration are necessary. Therefore, general principles are derived for object-oriented and relational modeling languages by comparing the results from the case study and abstracting them to a general case of any spatial data set. A group of general, automatic principles to insert certain metadata at a specific place in the model has been defined. These automatic principles consist of general top-down metadata that are valid for the whole model and bottom-up metadata that mirror the changes and heterogeneity of data within the model. Notwithstanding, it is also necessary for the modeling expert who knows the spatial data well to choose which metadata can be derived implicitly and explicitly.

In order to support the common management of metadata and spatial data with tools and to support new metadata-spatial data sets in their common handling, two prototype implementations are realized. The first prototype is an existing open source modeling software called INTERLIS/UML Editor, which is extended by implementing the principles for metadata-spatial data integration. For example, a new functionality is that certain metadata are added automatically when a new model is generated. Consequently common modeling for spatial data and metadata is supported in a harmonized way. The second prototype implementation explores the possibilities of creating views and functionalities of views in the relational database management system Oracle 9i. Views according to different profiles of ISO 19115 are extracted from integrated data sets.

Altogether, integrating, managing and modeling existing metadata and spatial data commonly is possible. Furthermore, it is realistic to model and produce new metadata and spatial data in the same working process. This common management is supported in terms of flexibility by extracting views supporting different metadata schemas without changing the common metadata-spatial data set. In order to realize this new management concept in a harmonized way for any data set, general principles are designed which define which metadata to put where. The best automatic integration according to the general principles is discussed. Furthermore, the effects of common management on organizational aspects of metadata acquisition as well as interoperability in SDIs and quality issues are elaborated. Finally, an outlook identifies points of future development and research.

Zusammenfassung

Der Bedarf an beschreibenden Metadaten für Geodaten wächst in der heutigen vernetzten digitalen Welt. Dennoch spielt die Metadatenerfassung und -verwaltung in den meisten Organisationen bisher eine untergeordnete Rolle. In vielen Fällen werden die Metadaten lange Zeit nach den Geodaten erfasst und in getrennten Datenbeständen gespeichert. Damit gehen Informationen verloren und es ist nicht immer klar, welche Informationen zu den Geodaten und welche zu den Metadaten gehören. Zudem ist es schwierig, bereits existierende Geodatensätze „aufzuspüren" und als Folge davon entstehen Redundanzen.

Spezialisierte Suchmaschinen einer Geodateninfrastruktur (GDI) benötigen Geodaten, welche mit Metadaten beschrieben und indiziert sind. Je verlässlicher und konsistenter solche Metadaten sind, desto besser unterstützen sie die GDI als Plattform zur Suche, zum Austausch und zur Bearbeitung von Geodaten.

Folglich gibt es eine Kluft zwischen dem momentanen Zustand der Metadaten und den Ansprüchen, welche an sie gestellt werden. Dieses Defizit kann mit neuen Verwaltungskonzepten für Metadaten behoben werden.

In dieser Dissertation wird ein neuer Ansatz zur Metadatenverwaltung durch Integration von Metadaten und Geodaten in gemeinsamen Datenbeständen oder Datenbanken untersucht. Diese gemeinsamen Metadaten-Geodaten-Datenbestände können als „umfassende Geodaten" bezeichnet werden.

Das Konzept der Integration von Meta- und Geodaten ermöglicht es den Geodaten, ihre eigene Metadatenbeschreibung zu beinhalten. Der Ansatz unterscheidet zwischen existierenden Geodatenbeständen, inklusive Modell, welche um ihre Metadaten erweitert werden müssen, und neuen, geplanten Geodatenbeständen und -modellen. Es werden verschiedene Gruppen von Metadaten unterschieden, welche integriert werden können: die implizit und explizit von Geodaten ableitbaren sowie neuen Metadatenattribute. Darüber hinaus werden verschiedene Ansätze erörtert, wie die Attribute in das Model eingefügt werden können: Top-Down und Bottom-Up.

Für die Integration werden drei Schritte definiert:

1. Semantische Analyse und Übersetzung

2. Strukturanalyse und hierarchische Integration

3. Die Semantische Transformation

Wenn das Konzept der gemeinsamen Verwaltung von Metadaten und Geodaten in Zukunft eingeführt wird, gibt das sogenannte View Konzept die Möglichkeit, Metadaten nach unterschiedlichen Standards sowie weitere Ausschnitte aus dem gemeinsamen Datensatz zu extrahieren. Dieses View Konzept unterstützt die Interoperablität und die Flexibilität im Zusammenhang mit Geodateninfrastrukturen, da hier verschiedene Diensten und Nutzern unterschiedliche Ausschnitte und Strukturen von Daten und Metadaten zur Verfügung gestellt werden können.

Um die Machbarkeit der Integration zu testen, werden drei Datenbestände mit ihren Modellen mit Metadaten ausgestattet. Als Fallstudien wurden das Modell der Amtlichen Vermessung der Schweiz, das Modell der Wasserversorgung der Stadt Zürich und ein Umweltdatenmodell des Kantons St.Gallen ausgewählt. Diese Modelle weisen Unterschiede in Komplexität, Grösse, Struktur, Standardisierungsgrad und Modellierungssprache auf und decken damit eine grosse Bandbreite ab. Die Fallstudien zeigten, dass die Integration von Metadaten machbar ist. Zudem war es möglich, für jedes Datenmodell bestimmte Top-down und Bottom-up Metadatenattribute zu definieren. Darüber hinaus werden die explizit und implizit ableitbaren Attribute festgestellt. In einem Workshop mit Experten, welche die Modelle gut kennen, wurde die Integration verifiziert.

Um das Integrationskonzept an einem beliebigen Datensatz und Modell auf eine vergleichbare Art und Weise anwenden zu können, sind allgemeine Regeln notwendig. Deshalb werden Prinzipien der Integration, basierend auf einem Vergleich der Fallstudien, für den objekt-orientierten und relationalen Fall hergeleitet. Das Ergebnis ist eine Gruppe von allgemeinen Metadatenattributen, welche automatisch an einem bestimmten Ort im Model integriert werden können. Darüber hinaus ist es jedoch auch notwendig, dass ein Experte bestimmte modellspezifische Attribute einfügt und bestimmt, welche Metadatenattribute explizit oder implizit von den Geodaten hergeleitet werden können.

Um die gemeinsame Verwaltung von Metadaten und Geodaten auch mit Software-Werkzeugen unterstützen zu können, werden zwei Prototypimplementierungen realisiert. Der erste Prototyp ist eine Erweiterung eines Open-Source Modellierungswerkzeuges mit dem Namen INTERLIS/UML Editor. Dieser wird mit neuen Funktionalitäten ausgestattet, welche zum Beispiel bestimmte Metadaten, basierend auf den definierten Prinzipien, bei der Erzeugung von neuen Geodatenklassen automatisch hinzufügen. Das Ziel ist eine gemeinsame, harmonisierte Metadaten-Geodaten Modellierung für neue Datensätze. Der zweite Prototyp ist die Implementierung und das Testen von Funktionalitäten von Views im relationalen Datenbankmanagementsystem Oracle 9i. Hier wurden Views nach verschie-

denen Metadatenprofilen erstellt und die Herleitung von Metadaten aus bestehenden Geodaten getestet.

Insgesamt ist es möglich, Metadaten und Geodaten bestehender Datensätze zu integrieren und zusammen zu verwalten. Für neue Datensätze können Metadaten und Geodaten in einem gemeinsamen Arbeitsprozess erzeugt und geplant werden. Die Flexibilität dieses neuen Verwaltungskonzept wird durch das Generieren von Views unterstützt, welches zulässt, Metadaten nach unterschiedlichen Standards zu extrahieren. Es ist darüber hinaus möglich, allgemeine Regeln für die Integration aufzustellen und die Integration automatisch durchzuführen.

In der Diskussion werden die Einflüsse einer gemeinsamen Verwaltung von Geodaten und Metadaten auf organisatorische Aspekte der Metadatenerfassung und Interoperabilität in GDI sowie Qualitätsfragen untersucht. Mögliche Fragestellungen für die zukünftige Forschung werden in einem Ausblick angesprochen.

Chapter 1   Introduction

Recently, the meaning of *spatial data*[1] has gained importance. More and more laymen use spatial data in car navigations systems, Web maps e.g. for finding the best connections in a city with public transportation, searching for addresses in digital city maps, etc. Consequently, there has been much effort in collecting digital data in various topics. Several spatial data acquisition projects were finished and now problems like updating, usage and promotion of the data occur. Spatial data users need to know about the availability and usability of the data. They need metadata (data about data) to label data. In general, spatial metadata can be compared to library metadata, which are stored in library catalogues and describe books. The spatial metadata should include complete information about suppliers, offers, prices, completeness, formats, accuracy, actuality, integration possibilities and updates of the spatial data. Undocumented data are of little or no value for possible spatial data users or Geo Web service providers in an infrastructure.

In order to make spatial data not only usable and available but also accessible, standardized and structured metadata must exist and provide a description as well as an interface. Yet, the current metadata are often inadequate and limit the interoperability e.g. in data infrastructures. This might be partly due to the effort keeping two separate data sets up to date.

This research investigates a new concept and possibilities of managing metadata and spatial data commonly, in one data set. Thus, the spatial data would be described from the beginning in a more detailed way and hence carry its own metadata. So, there would be merely one common update process. This would reduce the effort of two separate update processes and would harmonize a common update cycle: as soon as spatial data change, it is easy to transfer the necessary alterations to the corresponding metadata, too. Using a common data model, there would be no more confusion which information is stored in which data file/database and whether a certain piece of information belongs to metadata or spatial data. Consequently, there would be less redundancies.

---

[1] See also Glossary.

1.1 Background

The need of interoperability for heterogeneous spatial data, whether in IT, distributed GIS, network or *Spatial Data Infrastructure[1]* environment is one of the reasons why metadata are necessary. When discussing new metadata management concepts for supporting interoperability in a distributed network environment, different areas need to be researched. This work comprises four areas (Figure 1-1): Spatial Data Modeling, Spatial Data Infrastructures (SDI), Metadata Management and Standardization as well as Ontologies and Semantic Mapping. The common denominator are the keywords "interoperability of spatial data" which all the four areas deal with.
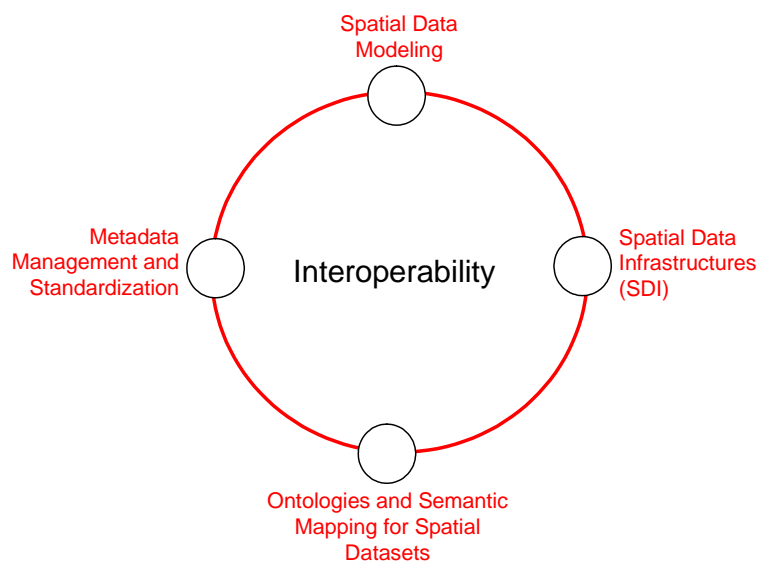
Figure 1-1: Classification of thesis

Interoperability of spatial data is a long known problem in the GIS (Geographic Information System) world: it is the challenge of exchanging and combining different spatial data from various sources. Usually, GIS experts need various data sets to solve a spatial problem, e.g. a planning expert might be interested in finding the best location for a camping site. He then needs information about infrastructure, land coverage, protected areas i.e. national parks, DEMs (Digital Elevation Models) and topographical maps. All these different data sets need to have a common reference and coordinate system, compatible formats so that the GIS software can import them as well as information about their quality. Data quality is used in this context as the fitness of use of certain data for a specific task. Therefore, one of the main challenges of working with different, unknown data sets from different sources is integrating them.

Spatial data modeling (Section 2.5) tries to solve this problem by offering specialized modeling languages, which are adapted to the needs of spatial data and, therefore, describe them as precisely as possible in relationships and characteristics, as well as tools, that can cope with the transfer and processing between two different GIS applications in a system independent way. Thus, a data set is described by its model.

The field of **ontologies and semantics** advances a step further in describing spatial data. Here, not only the relationships and characteristics of the spatial data are described, but also their inherent meaning (semantics) and purpose. This can be important when data sets from different cultures and languages are integrated. Different attempts to describe ontologies with logic and other formalizations are currently researched (Section 2.6.5). Alternatively to a model, a data set is described by its ontology.

Metadata records are **standardized** and **managed** in order to inform the user about the purpose, quality, actuality and accuracy of a data set (Section 2.3.3). According to the international definition of ISO/TC 211 document ISO 19115, metadata comprises:

"... a schema required for describing geographic information and services.
Information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data…."
and metadata are

"….applicable to the cataloguing of datasets, clearinghouse activities, and the full description of datasets."
Therefore, spatial metadata have the purpose to:

- Label the spatial data sets and describe the contents
- Help finding the right spatial data sets.
- Ascertain the spatial data quality

In the context of this thesis, they are mainly specialized on describing the content and use of a spatial data sets. In order to do so in a consistent and unified way, standards define which attributes are needed and describe the structure in a metadata schema (model). Here, metadata are not only used for human interpretation of data sets, but are also considered to be computer processable information which can be used for, e.g. search engines. In an ideal case, metadata should include their own model.

A **Spatial Data Infrastructure** (SDI) is an enabling platform for the exchange, search and processing of spatial data sets and, thus supports interoperability (Section 2.2 and 2.3). The ideal SDI, therefore, is based on interoperable data. In order to achieve this ideal situation the previously mentioned areas of research are utilized.

Therefore, in the ideal situation of interoperability in an SDI, there would be data sets with model/ontology and metadata description.

1.2 Motivation

The importance of metadata and metadata management has been increasing with the development from insular GIS work stations to a distributed network environment. Metadata are inherently necessary for the interoperability between spatial data systems.

In current SDIs, the heterogeneity of data and systems is large and the demand for better interoperability between spatial data sets grows and has not yet been fully achieved on a technical level. The pressure to have adequate metadata and metadata management grows. Future development towards the use of more Web services in a network of SDI needs the foundation of new ideas for metadata management.

1.2.1 Metadata and GIS

In practice, metadata often play a subordinate role. A typical procedure for a spatial data project is to first plan, model and acquire digital spatial data by various methods (e.g. photogrammetry, digitalizing, measuring in the field etc.) and to consider and acquire the metadata only much later, sometimes even years afterwards (Najar, 2004). This means extra and lengthy work. Thus, metadata are expensive and considered an overhead by many organizations. Metadata usually do not exist for every data set.

If metadata do exist, then in a historically grown metadata management environment. As a consequence, the metadata repositories are often inconsistent. In interviews with various administrations in Switzerland, for example, ((Golay et al, 2003); (Stopper, 2004)) it was stated that about ten years ago the metadata description was created with a small group of attributes and then more metadata attributes were added and others taken out over the years, but not all data sets were updated at the same time. Therefore, metadata description can consist of different information and it is not always clear which information is stored in the metadata and which in the spatial data, e.g. projection and coordinate system.

Moreover, the metadata are managed in various stand-alone files and formats:

- Unstructured text files (e.g. Microsoft Word)
- Structured files: html, xml, etc.
- Databases

The mentioned human-readable forms of storing metadata are nowadays all managed separately from the spatial data. If the spatial data changes, the metadata must be updated independently. This may result in extra work (time consuming and costly) and in inconsistencies.

There are many different GIS domains which have created independent, competing metadata schemas. These metadata schemas are large and not flexible and even experts have difficulty staying up-to-date with them. Also, using any current metadata system, for example the ISO 19115 Metadata Standard, requires systematic study and practice (Onsrud et al, 2004).

From an organizational point of view, it is not always clear, especially for large organizations with many units, departments, various users and producers of spatial data, who is in charge of acquiring metadata.

The possibilities include:

- Decentralized organization: every group or unit is in charge for its own metadata
- Centralized organization: a central unit, e.g. IT centre takes over certain GIS tasks and creates metadata for the other units.

The first possibility has the disadvantage that the metadata are potentially heterogeneous: ISO 19115 defines different hierarchy levels of metadata: series, dataset, feature, attribute, etc (ISO/TC211 International Organization for Standardization ISO, 2001). Also, the update and, thus, quality of the metadata might vary strongly.

The second solution means that people who do not know the spatial data of various topics well enough have to acquire specific metadata information. This might mean that certain metadata information that are specific for the data set, like the acquisition method, precision, accuracy, date of acquisition and units might be lost.

1.2.2 Metadata and Spatial Data Infrastructures

A GIS is mainly operated by expert users. Whereas, the development of SDI means an opening of spatial data and GIS to laymen, meaning non-expert users. Especially Web services support laymen with simple spatial problems, e.g. visualizing a map.

From the point of view of SDI, metadata must fulfill different tasks for different stakeholders:

- A data provider, possibly with commercial and economic interests, might want to have metadata for laymen users which are easily readable and understandable, e.g. with images of the data set. These could be product leaflets.

- On the other hand a data provider needs metadata for experts who are interested in technical details about quality, reference system, etc.

- Computer processable metadata, which must be in accordance with an existing metadata schema, can be read and analyzed by a computer or search engine in the Web. They are usually encoded in ASCII format such as XML.

Altogether, metadata must comply in a more and more networked environment with claims to be human-readable for marketing purposes, informative for GIS-experts and computer processable for Web services. This poses high demands on the quality of metadata and their interoperability.

Looking at the interoperability in SDIs from a technical point of view, the specialized spatial search engine, needs spatial data which are labeled and indexed by metadata. These so-called catalogues are important in order to be able to search for spatial data in the Web.

The more reliable and consistent such metadata are, the better they support an SDI as an enabling platform to search, exchange and process spatial data.

At the moment, current Clearinghouses offer metadata, e.g. listed as text in a html, as a final result of a search for spatial data. There is usually no URL link between metadata and the required spatial data set. Instead, the user must find the address of the custodian and, e.g. order the data by email or telephone.

On a national level, there are many different organizations (private and public sector) involved in producing, maintaining and offering spatial data sets (Figure 1-2). Often they work independently or are only partly interconnected. As a result tow organizations might acquire the same data: Organization B did not know about the existence of the data set in Organization A and therefore acquired them again.

There are many data providers who manage a great variety of data in different scales, quality, topic, format, acquired by different methods and for many purposes. This heterogeneity is partly due to missing technical regulations, but also caused by institutional obstacles, e.g. communication between different national administrations.

Figure 1-2 shows different ways institutions can connect to a national Clearinghouse in the Web: some have a common, centralized database for the data and others use different, decentralized servers with various data and tasks. Also, within one agency, the same

complex situation can occur with internal departments or groups that have various functions. Furthermore, the interfaces for access can differ strongly. From the technical point of view, different existing SDIs such as on communal or state level which possibly use different network protocols (see also Section 2.3.1), but also different systems, data formats and Web services can have various interface properties.

Many organizations are simultaneously data providers and users and so have two roles to fulfill and will connect to a Clearinghouse with two different purposes. Additionally, in order to comply with the different needs of users, many different Web services are needed (Najar and Giger, in process).

In conclusion, metadata management for a network environment has a technical aspect concerning the usability of metadata for search engines and an institutional aspect which involves coordinating many different producers of spatial data and metadata and canalize them. They need guidance and tools to produce useful, high quality metadata. For SDIs, both the technical and the institutional aspect of metadata management must be clarified and taken into account.
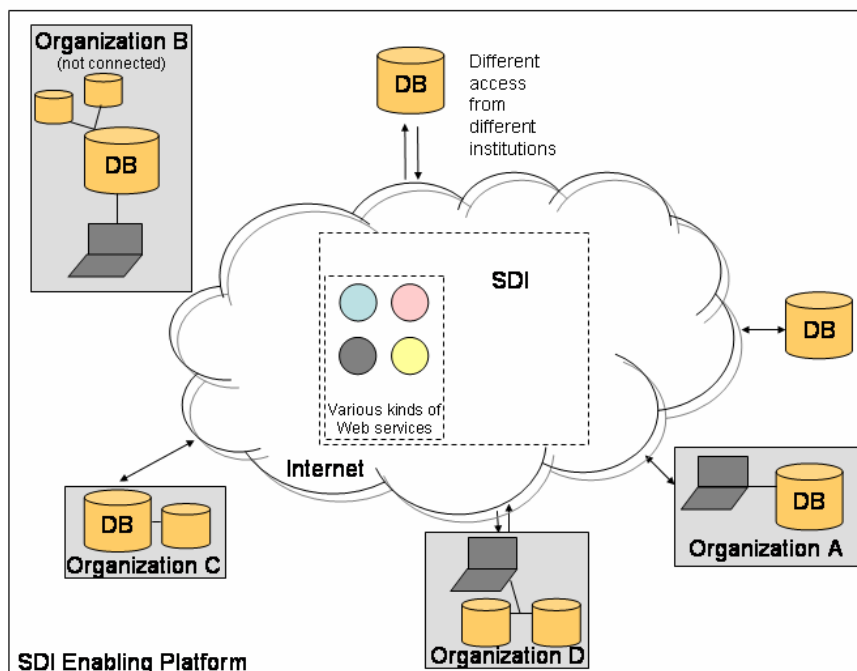
Figure 1-2: Interoperability between different organizations in a national SDI

1.3 Problem Statement

Currently, metadata are missing or incomplete, and are acquired in a heterogeneous way. At the same time, metadata are gaining importance in an internet environment. The need of excellent metadata description in quality, as in terms of fitness for use, and completeness to fulfill the needs of different users is growing rapidly. The objective of SDIs to facilitate interoperability is limited by inadequate metadata.

This leaves a gap between the status-quo of metadata and the demand for metadata which needs to be accounted for with new metadata management concepts.

1.4 Hypothesis and Objectives

The integration of metadata and spatial data in a common file or database, will support and improve the fitness for use of (meta-)data and as consequence the interoperability of GIS and SDI.

The hypothesis of this thesis is therefore, that metadata must be an integral part of any spatial data set and treated with the same tools as spatial data.

The consequence is a flexible interpretation of what is data and what is metadata. The common metadata-spatial data set is therefore considered to be comprehensive spatial data.

The objectives of this thesis are to investigate:
(1) The feasibility of integrating, managing and updating existing metadata and spatial data commonly.
(2) The possibility to model and produce new metadata and spatial data in the same working process
(3) The flexible creation of views, e.g. the possibility of supporting any metadata schema (e.g. ISO 19115 profiles) at any time without changing the original data.

Detailed Questions are analyzed:
- Is it possible to integrate metadata in any spatial data set?
- Is it necessary to use all metadata from ISO 19115? Or how can one avoid enlarging the spatial data set?
- Is it possible to create general rules for the integration (where to put the metadata and which metadata) ?
- Is it possible to do the integration automatically for new and existing data sets?

1.5 Approach

In order to achieve the integration of spatial data and metadata six steps were carried out (Figure 1-3).
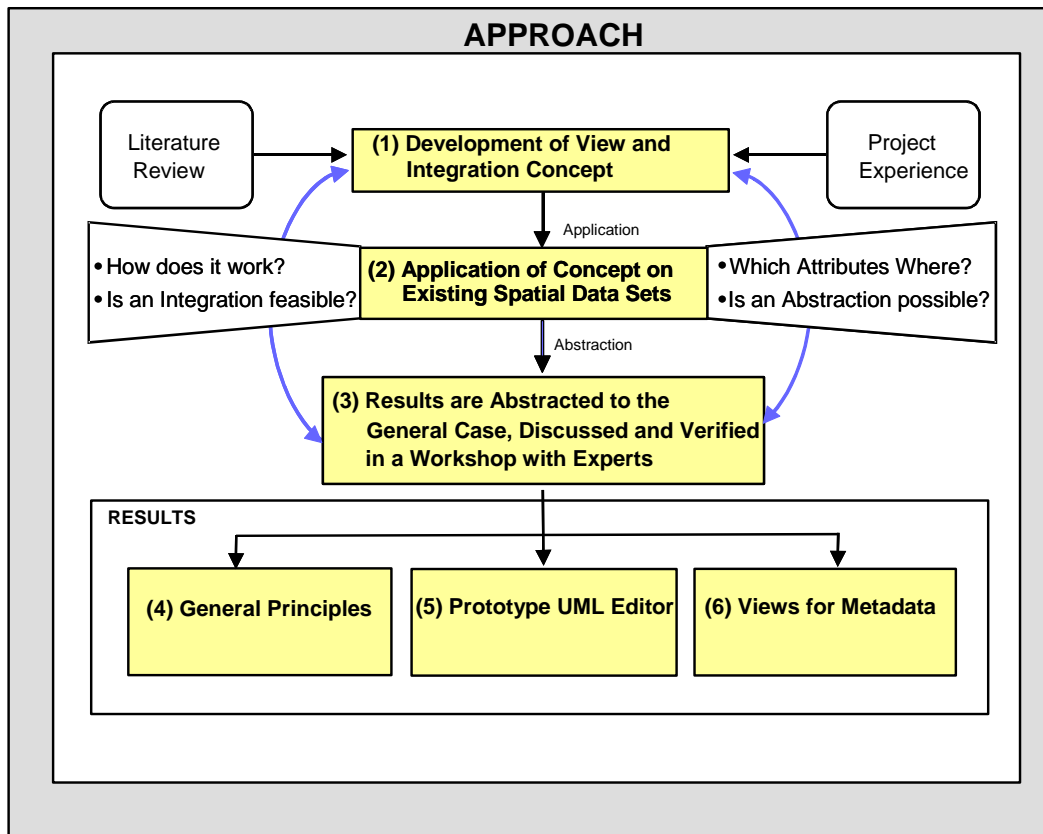
Figure 1-3: Working steps in realizing the integration of spatial data and metadata

Using the experience from an extensive literature review and projects dealing with Web applications, SDI and metadata standardization, the concept of ´Common Management of Spatial Data and Metadata` was developed (**Step 1**).

In order to see whether the concept is feasible and to verify how it works in technical detail, three existing data models and metadata according to the Swiss Profile of ISO 19115 (ISO/TC211 International Organization for Standardization ISO, 2001) were used in a case study (**Step 2**).

The result of the case study are three extended metadata-spatial data models, which have specialized metadata description. In order to make the integration of metadata possible with any data set, these results are generalized. The case study models were discussed in a workshop with modeling and GIS experts from industry and administration (**Step 3**).

As a result, general principles (**Step 4**) for integrating metadata and spatial data were designed.

These general principles are implemented and verified in a prototype: an existing UML modeling tool is extended in order to support metadata-spatial data modeling according to the generated principles (**Step 5**).

The view concept is tested and applied on the integrated metadata-spatial data model as well as set. The implementation is realized in the Database Management System (DBMS) Oracle and different profiles of ISO 19115 are extracted (**Step 6**).

1.6 Structure of the Thesis

This thesis consists of five parts which are linked to seven chapters. The Introduction and Problem Statement relates to Chapter 1 and includes the research the motivation, problem statement, objectives and hypothesis of the research. Finally, it explains the approach and relevant steps of the thesis. Chapter 2 and Chapter 3 provide background information and basic notions about current issues in research related to SDIs, spatial data modeling and semantic processing of spatial data and especially in the area of metadata management.

Chapter 4 introduces the two concepts of research: the concept of metadata-spatial data integration and the Concept of Creating Views. At the same time it defines relevant methods such as semantic mapping which is used for both concepts.

Part four explains method, design and implementation of the work and consists of Chapter 5 and Chapter 6. The feasibility of concepts is verified in a case study and abstracted so that general principles for integration can be designed. This is followed by prototype implementations to support common metadata-spatial data management and verify the general principles.

The synthesis part of the thesis summarizes the results and provides a conclusion as well as outlook (Chapter 7).

PART	CHAPTER
I Introduction and Problem Statement	1 Introduction
II Background Information and Current Research	2 Basic Concepts and Related Literature 3 State of the Art in Metadata
III Concept of Research	4 Concept of Integration and Views
IV Method, Design and Implementation	5 Creating Principles for Spatial Data-Metadata Integration 6 Prototype Implementation
V Synthesis	7 Discussion of Results

Table 1-1: Schematic outline of the chapters

1.7 Summary

Chapter 1 first provides background information why metadata management in general and in the context of SDIs is of relevance and offers an overview of the four areas which are of importance for this thesis. The research motivation surveys the current issues when dealing with metadata for spatial data with regards to GIS and SDIs.

The problem of this thesis is stated: there is a gap between current, inadequate metadata and the demand for metadata which needs to be accounted for with new metadata management concepts.

Subsequently, the hypothesis asserts that metadata must be an integral part of any spatial data set.

The integration of metadata and spatial data in a common file or database, will support and improve the fitness for use of (meta-)data and as a consequence the interoperability of GIS and SDI.

Therefore, the objectives are to investigate:

(1) The feasibility of integrating, managing and updating existing metadata and spatial data commonly.

(2) The possibility to model and produce new metadata and spatial data in the same working process

(3) The flexible creation of views, e.g. the possibility of supporting any metadata schema (e.g. ISO 19115 profiles) at any time without changing the original data.

This is followed by an overview of the approach of the research and the necessary steps taken. The chapter concludes with a brief description of the structure of the thesis.

Chapter 2   Basic Concepts and Related Literature

This chapter introduces the basic concepts for understanding the situation in relevant areas of research which are important for this thesis. In Section 2.1 the definitions of important terms are provided in order to have a common point of view. Section 2.2 discusses SDIs and their purpose as means of enabling interoperability. The components of SDIs and their relevance to metadata are illustrated in Section 2.3. Current SDI initiatives (Section 2.4) are important to understand the situation which is provided currently in real world SDI projects. Section 2.5 gives an overview of current issues in spatial data modeling which is later applied to metadata modeling. In order to exchange, analyze and restructure the modeled spatial data, semantic processing is necessary. This is especially interesting for the transferring data from different sources and integrating them (Section 2.6).

2.1 Basic Definitions

Here, we introduce basic definitions and concepts as used in this thesis.

The term "**Information Community**" was promoted by the Open Geospatial Consortium to mark the differences of creating different models for different topics and also within the same topic. It is defined as "a collection of people (e.g. government agency, profession, group of scientists in the same discipline, etc.) who…share common digital geographic information language and share common spatial feature definitions. This implies a common world view as well as common abstractions, feature representations, and metadata (Open (GIS) Geospatial Consortium (OGC), 1998). Each information community has its own "world view"-called conceptualization of the world.

One of the most important characteristics of the **GI community** (Geographical Information Community) is its diversity as many areas use GIS for different purposes, and, thus, data with different characteristics, e.g. archeology, hydrology, social sciences, statistics, real estate, environmental issues, geology etc. Often this spatial information may have a temporal dimension which is difficult to deal with using current GIS functions and tools. Looking at the GI information from the semantic point of view, many geographic features have multiple names, depending on social and language conventions. These are only a few examples for particularities of the GI community and the resulting necessity of modeling and handling syntactical and structural differences of heterogeneous spatial data.

A **Geographic Information System (GIS)** is commonly known as a tool for storing, managing, processing and visualizing digital spatial data.

As mentioned in Section 1.1, **metadata** are data about data. In this thesis, the considered metadata describe spatial data, therefore exclusively spatial metadata are discussed. They answer questions like: "Does a dataset on a specific topic exist ('what')?", "For a specific place ('where')?", "For a specific date or period ('when')?" and "A point of contact to learn more about or order the dataset ('who')?" (ISO/TC211 International Organization for Standardization ISO, 2001).

The value of spatial data is directly related to their **quality**. The views on quality differ among the various user communities. For example, cadastral applications require the positional accuracy within a few centimeters whereas for a nautical chart a few meters suffices. The standards ISO 19113 and ISO 19114 define the principles for describing the quality (Kresse and Fadaie, 2004). In this thesis, the selection which are the best spatial data set suited for an application depends on the fitness for use of the data.

As mentioned in Section 1.1, the author considers **Spatial Data Infrastructures (SDIs)** to be enabling platforms which facilitate the access to spatial data and comprise organizational, political, cultural and technical aspects.

Clearinghouses and **geoportals** are needed for SDIs. From a technical point of view, a Clearinghouse is considered to be a specialized, complex and important Web service which assists in offering the exchange and sharing of spatial data between different users and suppliers in an SDI. A geoportal is the access point of an SDI on the Internet (see also Section 2.3.1).

Spatial data **catalogues** are discovery and access systems for example in an SDI, that use metadata as the target for query on raster, vector, and tabular geospatial information. Indexed and searchable metadata provide a disciplined vocabulary against which intelligent geospatial search can be performed within or among SDI communities (Nebert, 2001).

Interoperability has many aspects. In this thesis the focus is on the technical exchange between spatial data and metadata as well as **integration** of data from heterogeneous sources. This integration combines data from different systems. Furthermore, a user should need no or little knowledge about the characteristics of the necessary functional units. The need of interoperability for heterogeneous spatial data, whether in IT, distributed GIS or in network or SDI environment is one of the reasons why metadata are necessary (Chapter 3).

"**Models** for GIS operate in a wide range of different situations, from models of particular application domains (e.g. transportation models) to specific computer-based models of the physical information in the system (Worboys and Duckham, 2004)". In this work, the terms model and **schema** are used as synonyms. In the context of spatial data, they describe selected aspects of reality in a computer; thereby, utilizing relationships and hierarchical structures as well as groupings. Usually, models are provided with **data dictionaries**. This is a collection of descriptions of the data objects or items in a data model. Hence, models are an important contribution to spatial data interoperability.

Other lines of research in interoperability consider different solutions such as the use of **ontologies** as the common point among diverse user communities (Wiederhold, 1994) cited in (Fonseca, 2001). An ontology is an explicit specification of a conceptualization to which the vocabulary of one information community refers (Gruber, 1993). Hence, ontologies are considered a more generic application and task-independent description of spatial data than models (Spyns et al, 2002) as well as describe functional behaviour (Section 2.6.5).

In current database systems **views** are used as virtual tables, that are linked to the real, physical data table by queries or mapping. In this thesis views are extended to the conceptual level and for metadata.

2.2 Spatial Data Infrastructures and GIS

As introduced in Section 1.1 and 2.1, Spatial Data Infrastructures (SDIs) are enabling platforms which facilitate the access to spatial data and comprise organizational, political, cultural and technical aspects. It also shows how the traditional GIS is growing together with Web applications.

Since the use of spatial data is growing rapidly it is important to understand the current and future development of the technologies and policies that process, enhance, exchange and manage spatial data and metadata as well as policies that provide a framework.

According to Williamson et al (2003), there are two major forces driving the development of spatial data: the growing need of improving decision making and increase of efficiency by government and businesses with the help of spatial data, on the one hand. On the other hand, spatial data has become ever more important and needs to be shared among different organizations.

The fact that spatial data is used nowadays for many everyday tools which can be utilized by everyone (e.g. car navigations systems, GPS for mountaineering and various maps for devices like PDAs and other hand-held devices) shows, that spatial data are needed and used more broadly, also by laymen. Studies have shown that the availability of spatial data, meaning the knowledge of where to find which data has been a problem between and within organizations (Golay et al(2003), Giger et al (2005)). It was recognized that the benefits of sharing data in an enabling platform (e.g. saving time, effort and money) would support the creation of SDI initiatives.

Rajabifard (2002) gives a good summary of the SDI as …"an initiative intended to create an environment that can cooperate with each other and interact with technology, to better achieve their objectives at different political/administrative levels. SDI initiatives around the world have evolved in response to the need for cooperation between users and producers of spatial data to nurture the means and environment for spatial data sharing and development ((McLaughlin and Nichols, 1992), (Coleman and McLaughlin, 1998), (Rajabifard et al, 1999), (Rajabifard et al, 2000)). The ultimate objective of these initiatives, as summarized by Masser (Masser, 1998), is to promote economic development, to stimulate better government and to foster environment sustainability."

GIS tools are ideal to support the enabling platform of the SDI. Currently, the main processing and analysis of spatial data is done in GIS on desktop systems. In future, this might be transferred to Web services and Web GIS which are offered by SDI initiatives, online.

The wider use of internet affects GIS in three major areas: GIS data access, spatial information dissemination, and GIS modeling/processing (Peng and Tsou, 2003). SDIs support the spatial data search and exchange in the Web for experts and laymen. Currently, Web GIS enables certain processing and visualizing of spatial data in the Internet. If the functionalities are extended in future, Web GIS tools will be dynamically reusable: processing components might be added interactively if necessary and will contain a broader scope of functionalities. Thus, GIS might be considered a component of a Web service.

Both SDI and GIS, thus rely on the analysis of spatial data to solve a specific spatial problem and both need metadata to be able to find the adequate data.

The development of metadata for spatial data can be compared to metadata for books. Library systems, which formerly used only index cards to describe their books are now offering digital catalogue systems which create large networks of different libraries at various locations. The metadata information of one library has to be compatible and of similar quality

compared to others. The output of the search needs to offer standardized and consistent output so that the results might be compared. Thus, the Internet offers more possibilities for search but also creates the need for better cataloguing and adequate metadata.

2.3 Components of Spatial Data Infrastructures

In order to understand the existing processes in SDIs better, certain relevant components and their processes are analyzed. This is necessary for the further progress of this thesis to understand how integrated metadata-spatial data can support certain parts of the workflow and existing components.

SDI development is a long term process which needs long term investment and the consideration of organizational and technical issues (Najar et al, 2004b). The technical aspects are of greater importance in this thesis.

As shown in (Chan et al, 2001), there are various definitions of which components belong to an SDI and how these are defined. Some are more technical, others are oriented towards organizational issues. For example (McLaughlin and Nichols, 1992):

The components of a spatial data infrastructure should include sources of spatial data, databases and metadata, data networks, technology (dealing with data collection, management and representation), institutional arrangements, policies and standards and end-users.

Whereas, the American FGDC (FGDC, 1997) prefers to define the following:

National SDI is an umbrella of policies, standards, and procedures under which organizations and technologies interact to foster more efficient use, management, and production of geospatial data.

The definition used here is based on technical and organizational components (Golay et al, 2003): The technical elements comprise the Clearinghouse, web services, metadata as well as spatial data. The organizational components cover aspects like agreements or contracts, coordination, stakeholders, responsibilities, political support, economic issues (e.g. financing, fees), legal aspects (e.g. juridical foundation, data protection) and education.

The following sections will shortly describe the relevant components Web Services/Clearing-house, spatial data and standards as far as they are relevant for further elaboration of the concepts of this thesis.

2.3.1 Web Services and Clearinghouses

In general, a clearinghouse is an agency for collection, classification, and distribution especially of information or other matter or items requiring wide dissemination (Babcook Gove, 1993).

In the context of SDIs, "the Clearinghouse network is the means by which the fundamental datasets are made accessible to the community, in accordance with policy determined within the institutional framework and to agreed technical standards {Williamson et al, 2003 #137}". From the technical point of view, Clearinghouses are therefore specialized search engines for spatial data, which assist searching, viewing, transferring, purchasing, advertising, and disseminating spatial data from numerous sources via Internet.

They thus improve the exchange and sharing of spatial data between suppliers and users ((Crompvoets et al, 2004), (Crompvoets and Bregt, 2003)).

Here, a Clearinghouse is hence considered a specialized, complex and key Web service for an SDI. Current Clearinghouses exist on different levels from different initiatives and should be linked to each other in the ideal case. This might mean that a search query in one Clearinghouse can be transferred to other coupled search engines.

In general, Web services can fulfill many different kinds of functions in the Internet and help to process and share data. The currently best known are the e-commerce services which provide the possibility to shop and pay by credit card via Internet.

Moreover, Web services might fulfill either one specific task or a set of tasks. This means that they can be used alone or with other Web services to carry out a complex aggregation or a business transaction (Kreger, 2001). The task of linking different services to complete a complicated, composite problem is called 'chaining'. Different technical requirements need to be fulfilled for this, e.g. the existence of metadata for Web services.

So, technically speaking "Web services are distributed component systems that take advantage of XML as a standard for communication between components. The web services infrastructure provides an XML vocabulary, Simple Object Access Protocol (SOAP), for communication between components. Two other XML vocabularies, Web Services Description Language (WSDL) and Universal Description, Discovery, and Integration (UDDI) protocol, are used as the basis of a registry to help to make web services easier to find on the Internet (Worboys and Duckham, 2004):"

Geo Web Services are a specialized category of Web Services, which work with all aspects of spatial data (Najar et al, 2004a). SDIs need Geo Web Services as well as general Web services, e.g. for payment for a fulfilled process.

In order to be able to chain, for example a Geo Web Service for analyzing spatial data to work together with another service for visualizing and layouting the map, a common interface has to exist. The ideal situation would be if different services could be interactively connected. There have been different attempts to characterize and describe services in a sufficient way.

WSDL (Language Version 2.0 (WSDL 2.0)) is an example for one of the languages used to describe Web services based on an abstract model of what the service offers and an XML format. WSDL 2.0 enables to separate the description of the abstract functionality offered by a service from concrete details of a service description such as "how" and "where" that functionality is offered (Chinnici et al, 2005).

ISO 19119 is a specialized Web service standard for spatial data (ISO/TC211 International Organization for Standardization ISO, 2005), which also defines metadata for spatial services. The idea is to set certain attributes which identify the services precisely enough so the functionality and purpose is sufficiently described enabling a Clearinghouse to find the right service for a certain geospatial problem within the SDI. Thus, a Clearinghouse can search not only for spatial data but also for Geo Web Services with the user's input keyword. The future is that the SDI is considered rather a service infrastructure so that laymen do not have to deal with structures of spatial data or complicated GIS.

From the point of view of system architecture, multi-tier client-server architectures are needed (Figure 2-1). The middle tier can be supplied with software called Middleware, which allows two different technologies to work together. This increases the capability of allocating specific tasks to certain tiers. For example, a client web browser requests information from a data server via certain services which are situated on the middle tier.
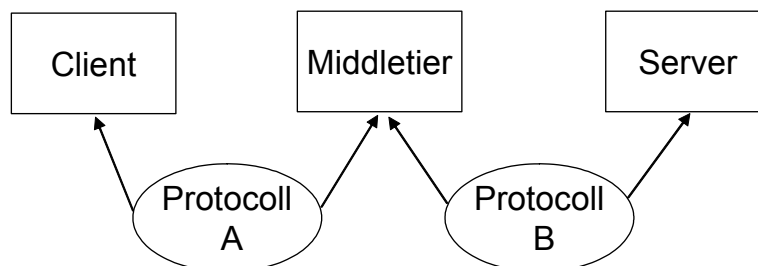
Figure 2-1: Multi-tier client-server architecture (Worboys and Duckham, 2004)

The role of the communication protocols in Figure 2-1 is to describe the information.

"A data communication network provides the physical infrastructure for sharing information in a GDI. Communication of data simply involves the transfer of binary digits (bits). For devices and software to make sense of these bits, however, they must be encoded in a specific format. This is the role of a particular communication "protocol"- to describe how the bits are arranged so that they can convey all the information necessary for effective communication. Protocols describe the format of transmission (Crosswell , 2000)".

One of the first Clearinghouses in 1994, hosted by the American Federal Geographic Data Committee used the client-server-architecture concept with Z39.50 protocols and servers (Nebert and Fullton, 1995). Therefore, for many years the Z39.50 protocol (also known as ISO 23950, (ISO/TC46 International Organization for Standardization ISO, 1998) was the state-of-art in SDI architecture.

This section has given a brief introduction to technical issues of web services. Certain aspects are closely linked to metadata management. For example, the fact that one of the most important web service components of an SDI, the Clearinghouse, relies on metadata for finding spatial data on decentralized data servers. Depending on how detailed and up-to-date the metadata are, the response to a user's query will give more or less adequate answers. Also, in this search and retrieval process, the metadata and spatial data work closely together.

2.3.2 Spatial Data for SDI

The situation of metadata and spatial data are closely linked. In the following the issues concerning spatial data for SDIs are analyzed.

Every SDI initiative needs to define its reference data. These have various names in different SDIs, e.g. framework, basic data, fundamental data etc., but mean in principle the same. These reference data are considered with highest priority when it comes to realizing a common data quality level. In literature, data quality elements with quantitative information are completeness, logical consistency, positional accuracy, temporal accuracy and thematic accuracy. Whereas qualitative information comes from the elements purpose, usage, lineage (Kresse and Fadaie, 2004).

Therefore, spatial data must be equipped with a model and metadata to be comparable to similar information of other data sets. Moreover, in this comparison a decision is taken, whether they are useful and their quality complies to a certain task.

Another issue in recent days is data licensing. A suggestion emerging from the US is that all spatial data should be donated in a common repository, thus renouncing all rights on ownership of the supplied data. The idea is to make the data provider or user create metadata and open licenses commonly on a web site that automatically generates certain part of the license's text and facilitates the process of legal metadata creation (Onsrud et al, 2004).

Nevertheless, the discussion among many countries is, that this approach might interfere with the goal of homogeneous and good data quality. The data provider would no longer hold interest or responsibility for keeping his or her data up to date when it is free of charge.

2.3.3 Standards for SDI

The political decision to create an SDI component and the technical development are usually realized on national, state or local level. By complying to guidelines offered by international standards, the compatibility to other, similar components in the world should be provided.

Since this thesis deals with metadata management, the international standard ISO 19115 is of major importance. Its structure and the way it deals with national profiles is essential for further elaboration of the common, integrated metadata-spatial data management. Therefore, this section will shortly introduce the most important organizations which are involved with standardization, normalization or which help coordinate SDI issues. Then, ISO 19115 for metadata and its structure as well as the Swiss profile will be introduced.

Standards have both an organizational or institutional as well as a technical point of view for discussion. At the moment there is much confusion concerning standards for geo-processes. There are many standards from different initiatives which are partly contradicting each other. Relevant international organizations for SDIs are:

- OGC (Open Geospatial Consortium, formerly known as Open GIS Consortium) (Open (GIS) Geospatial Consortium (OGC), 1998)
- ISO/TC 211 (ISO/TC211 International Organization for Standardization ISO, 2005)
- W3C (World Wide Web Consortium) (World Wide Web Consortium (W3C), 2004)
- International Cartographic Association (ICA) (International Cartographic Association (ICA), 1989)

There are also regional organizations like the European standardization committee, CEN, and topic related organizations such as the European Umbrella Organization for Geographic Information (EUROGI, (European Umbrella Organisation for Geographic Information (EUROGI), 1998) International Federation of Surveyors (FIG) (International Federation of

Surveyors (FIG), 2003) which are involved in guidelines and standards. Plus, there are recommendations like the SDI Cookbook (Nebert and Fullton, 2001) which give necessary background information to evaluate and implement existing components of SDI.

The Open Geospatial Consortium

The OGC promotes and harmonizes open standards and architectures which enable the integration of geospatial data and services into user applications and advance the formation of related market opportunities. At the moment, there are 291 (as of Nov. 2005) members of OGC registered. These belong to different organizations: commercial, consulting, development laboratories, government, NGO, research institution, university. Its approach is therefore to find consensus among their members and to develop, review, and release specifications. It tries to harmonize its geo-processing standards with other international IT standards. Also, OGC employs test beds, pilot projects, planning studies etc. to rapidly and efficiently develop, test, validate, and document vendor-neutral specifications based on user requirements.

The abstract specifications concerning metadata were developed and tested technically by OGC. Later OGC Abstract Specifications were adopted by ISO and replaced by ISO 19115. Other specifications which are relevant in this context are Geo Web Service interface definitions (e.g. Web Map Service (WMS), Web Feature Service (WFS) and Web Coverage Service Implementation Specification (Open Geospatial Consortium (OGC), 1994-2005)).

International Standardization Organization

ISO/TC211 is in charge of international standardization in the field of digital geographic information. Its members are from 52 countries. There are working groups for geospatial services, imagery, information communities, location based services, information management. The goals are to understand an use geographic information, increase the availability, access, integration and sharing of geographic information, enable interoperability of geospatial computer systems as well as ease the establishment of geospatial infrastructures. ISO has a cooperative agreement with OGC since 1998.

The so called base standards were the original focus of ISO/TC211: reference model, feature definition, spatial and temporal schema, coordinate reference system, portrayal, encoding, quality, and metadata. Now that these are completed the cooperation with the OGC is of importance for standardizing implementation issues, e.g. the Geography Markup Language (GML) (Kresse and Fadaie, 2004).

The most important standard for this thesis is ISO 19115. Within the standard a large amount of metadata attributes are defined. Usually only a part of the extensive list is used. In order to give the user guidance which metadata attributes are important a group of basic metadata are called core metadata and are offered as the minimum description for a dataset. (ISO/TC211 International Organization for Standardization ISO, 2005). Furthermore, the core metadata are distinguished between optional, conditional and mandatory attributes. The mandatory elements are essential for the description, whereas by adding the optional attributes will increase interoperability and the conditional elements categorized might be important for certain applications or topics.

Profiles of ISO 19115

Annex C of the ISO 19115 standard discusses metadata extensions and profiles. It describes the possibility of creating specific profiles or extensions of the standard. This might be interesting for fulfilling specialized purposes for a certain information as well as user community or nation.

Figure 2-2 illustrates the relationship between the Core Metadata components, the comprehensive metadata application profile and national, regional, domain specific or organizational profiles.
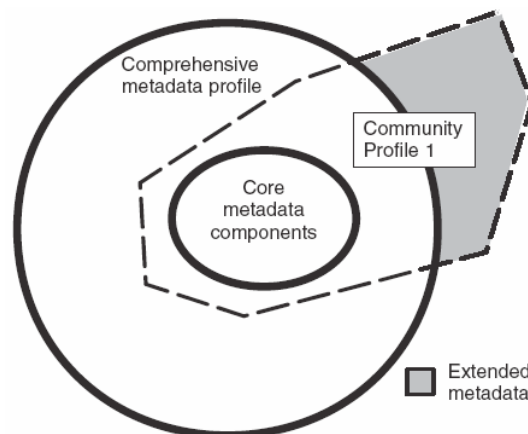
Figure 2-2: Metadata profile
(ISO/TC211 International Organization for Standardization ISO, 2001)

Many countries are currently creating their own national metadata profile of ISO 19115 in order to implement their specialized needs, e.g. the Swiss profile called GM03 adds an INTERLIS 2 description to the UML of ISO 19115. It thus specifies in some areas the exact way the model should be interpreted and implemented. When it comes to changing the ISO model certain rules are prescribed by Annex C (e.g. newly defined metadata attributes must

have unique names). Existing metadata elements may be changed in the profile to carry a more stringent obligation, like going from conditional to mandatory. The same is valid for the data type: initially, an attribute contained free text, but the profile chooses to define a list of given values. A code list can be expanded by new values. The general tendency of the rules is, that the profile permits at minimum the same the standard does. The profile can thus be stricter in the model but not more lenient than ISO 19115. Also, the profiles always include the ISO core, but can extend it to carry more elements.

Swiss Profile GM03

The working process for the Swiss ISO 19115 profile started in 2001. It was partly accelerated by the decision of the CEN to adopt all ISO standards. Since Switzerland is member of CEN, all CEN standards are compulsory. Two existing metadata catalogues were analyzed and compared to ISO 19115 and those attributes missing in ISO were added to extended metadata list (Figure 2-2). The standard SN 612050 was ratified by an expert working group in June 2004, by the Swiss standardization organization in March 2005 and published in May 2005. Some of the examples for differences of GM03 is that the class 'ResponsibleParty' is modeled as an independent class in order to avoid multiple acquisition of the same information. The multilingual support, which is optional in ISO, is implemented in GM03 and placed in the core.

2.4 Current SDI Initiatives

From an organizational point of view, there are many different SDIs on different hierarchy levels which should be coordinated and interconnected in the ideal case. This kind of interoperability is not only achievable by standards and technical specifications, but also by political, juridical and organizational incentives. In order to understand and discuss this complex system of different stakeholders (e.g. national, local governments or other organizations) a framework which supports hierarchical reasoning was introduced by Rajabifard et al, 2000. According to the concept of Hierarchical Spatial Reasoning (HSR) for SDIs an SDI hierarchy is composed of inter-connected SDIs at corporate, local, state/provincial, national, regional (multi-national) and global levels (Figure 2-3). It is possible to reason in a similar way in SDIs as in other hierarchical structures and therefore certain properties and principles apply (part-whole-property, Janus-Effect, Near Decomposability). This concept is used to understand and discuss the different levels of interoperability which are provided by SDIs and need to supported by spatial data and metadata management.

Many governments and organizations have recognized the importance of SDIs as a means to maximize economic, social, and environmental benefits. Therefore, there are many SDI initiatives in different hierarchical levels; according to (Borrero, 2002) and (Crompvoets and Bregt, 2003) in more than half the countries of the world. Only a few years later, in 2005, (Warnest et al, 2005) stated that most countries have national SDI initiatives. There is clearly a massive development of SDIs on the national level. It is important to develop all four levels, the global, regional, national and local level equally in order to facilitate data sharing and the coordination of data activities within the vertical relationships Figure 2-3. It also introduces the horizontal level as an important way of interconnecting different stakeholders on the same level in an SDI.

Vertical Relationships

Horizontal Relationships

- Global SDI
- Regional SDIs
- National SDIs
- State SDIs
- Local SDIs
- Corporate SDI
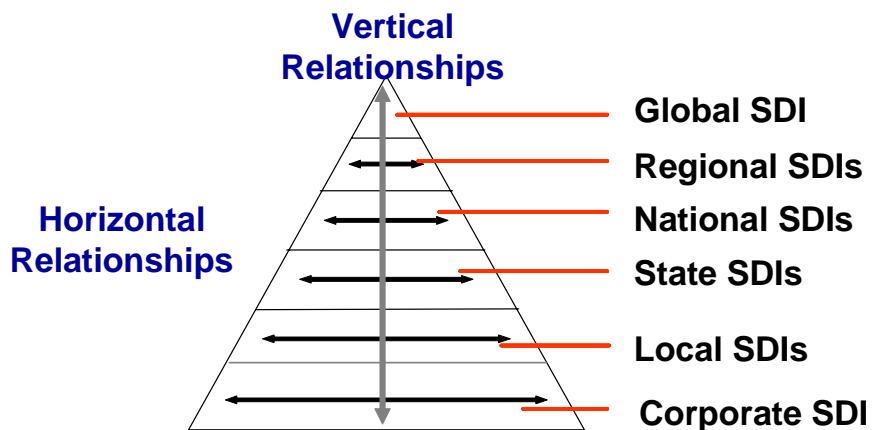
Figure 2-3: Levels for horizontal and vertical relationships (Rajabifard, 2002)

Global Level (GSDI)

The Global Spatial Data Infrastructure goal is to promote awareness and implementation of complementary policies, common standards and effective mechanisms for the development and availability of interoperable digital geographic data and technologies to support decision-making at all scales for multiple purposes (GSDI 5, 2001). The steering committee was created in 1999 and is comprised of representatives from the regions of the world and a cross section of the GSDI community in government, academia and private sector. The GSDI also host annual conference, which is a platform for the exchange of ideas and promotion of new results in the SDI environment.

A well known contribution to the international community of SDIs, is the SDI Implementation Guide, also called SDI Cookbook (Nebert, 2001), which is aimed at geographic information providers and users and provides the necessary background information to evaluate and implement existing components of SDI. Plus, it facilitate participation within the growing digital geographic information community of the GSDI.

Regional Level (PGCIAP, INSPIRE)

In the last decade there has been an emergence of regional Spatial Data Infrastructures organizations in Europe, Asia and the Americas. Africa and the Middle East are creating a Committee on Development Information under the auspices of the UN Economic Commission for Africa (Williamson et al (2003))

In 1997, at the Sixth UNRCC Americas, the delegates recommended the establishment of PCIDEA, reporting to the following meetings of the UNRCC-Americas.

The Permanent Committee on SDI for the Americas (PCIDEA) has the role to promote a better communication among the currently 24 member nations (Permanent committee on SDI for the Americas (PCIDEA), 2004).

For Asia, the Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP) documented information about an Asian Pacific (APSDI) in 1998: a network of databases, located throughout the region, that together provide the fundamental data. This network of databases are decentrally located by administrative and technical frameworks and standards throughout the region. A draft metadata profile complies with ISO 19115 profile (Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP), 1998).

INSPIRE - Infrastructure for Spatial Information in Europe

INSPIRE is an initiative and a legal framework launched by the European Commission and developed in collaboration with officials and experts in Member States and accession countries from the national, regional and local levels. It aims at making relevant, harmonized and quality geographic information available to support formulation, implementation, monitoring and evaluation of Community policies with spatial relevance. It will be implemented throughout the European Union (EU) from 2006/7 onwards with different types of geographical information gradually harmonized and integrated, resulting in a European Spatial Data Infrastructure. The ultimate vision is to establish integrated spatial information services, based upon a distributed network of databases, linked by common standards and protocols to ensure compatibility and inter-operability.

INSPIRE launched a prototype version of the EU Geo-Portal in order to better understand user requirements and to define the necessary standard-based technical specifications of the future operational systems (Infrastructure for Spatial Information in Europe (INSPIRE), 2005).

Guidelines for the development of implanting rules have been published and currently, a call for tender for development of prototype components for a European Spatial Data Infrastructure is in the decision process.

National Level (ASDI, e-geo.ch)

The national level is of significance to SDIs as this is often where juridical, political and administrative decisions are taken for a country and thus guidance and framework is given for the local levels. On a national level, there are many different organizations (private and public sector involved in producing, maintaining and offering spatial data sets (Najar and Giger, in process). Williamson et al (2003) suggest that countries with an efficient National SDI will be able to contribute to the regional and global SDI initiatives.

Here, two national SDIs (NSDIs) are shortly characterized and thus show what the situation is in two developed countries, one with a long SDI tradition and the other with a recent developed NSDI (Australian and Swiss). They mirror the situation of connectivity which has been achieved on national level. Both developed countries have federal systems. This makes the levels and hierarchy of integration in an SDI play an important role. The interoperability depends on the successful integration of different levels. The state and local levels have considerable meaning and must be integrated in a national SDI.

One of the earliest examples for a national SDI initiative is the **Australian SDI (ASDI)**. On the national level ANZLIC (Australia and New Zealand Information Council (ANZLIC), 1986), the Spatial Information Council is the key stakeholder and was established in 1986 in order to achieve a better coordination of land information management. It has developed a national spatial data infrastructure that comprises four components: Institutional Framework, Technical Standards, Fundamental Datasets and Clearinghouse Network.

Looking at the metadata issues, the ANZLIC metadata profile (Version2) is now being changed to a profile of ISO 19115 by the Metadata Working Group. The current ANZLIC metadata profile (version2.0, 2001) does not comply with the international metadata standard (ISO 19115). Several Australian agencies and jurisdictions are at varying stages of developing their own metadata profile based on the new standard. These agencies and jurisdictions agreed that efforts should be harmonized to produce an Australasian profile (WALIS News). The Australian Spatial Data Directory (ASDD), a metadata catalogue, was

launched in 1998 (ANZLIC - the Spatial Information Council, 2005). An ASDD audit from 2001 shows that one of the main problems is the fact that the metadata catalogue is almost exclusively known by the organizations who contribute to it (Williamson et al, 2003).

In Switzerland, the NSDI initiative is called **e-geo.ch**. It is headed by a steering committee and was initiated by the Swiss federal strategy for geo-information which was passed by the federal council in June 2001. The federal representative of the Swiss NSDI is KOGIS (Coordination of geographic information and geographic information systems (KOGIS), 2005). A network of components are defined which work together: basic spatial data, technical infrastructure, guidelines and standards, metadata, legal basis, basic geo services, education and training, research as well as strategy for distribution and pricing. The component metadata is enforced by a metadata catalogue called 'geocat' (Coordination of geographic information and geographic information systems (KOGIS), 2005) which supports the data acquisition with Web tools as well as the metadata storage in centralized or decentralized manner. It is also possible for already existing, local catalogues to connect to the national geocat catalogue.

Both the Australian and the Swiss metadata catalogue are not connected to international projects on the same level. This situation persists in many national initiatives in is therefore another sign of missing vertical interoperability (Najar and Giger, in process).

State Level

As mentioned for a national SDI, the state level is of importance especially in federal systems, in which several SDI initiatives are founded on the level of state (or canton in Switzerland). The various state SDIs must be integrated for a common national SDI. Common framework and standardized interfaces are needed. The SDI on state level is characterized by medium to large scale spatial data and the proximity to state administration.

Local Level

The local level of the SDI is often closely connected to the executing organizations of spatial data and thus of high relevance to a functioning, well-structured SDI. The national level of SDI which usually sets the guidelines and framework for the local level in the country must at the same time consider the specific needs and specialties of the local level. A functional SDI achieves interaction between the different levels.

For example the local **SDI for the region Rhein-Main** in Germany which is an industrial and well developed area, tries to first coordinate the different administrations and integrates other organizations by realizing test-beds and prototypes with them. At the same time the SDI Rhein-Main is part of the framework of the German SDI which sets certain strategic, political and technical goals which have to be followed (Giger et al, 2005).

There is evident potential for local initiatives to feed to higher level SDI projects, which in turn link up to create national portals (Beaumont et al, 2005).

It is altogether necessary to find a political and a technical compatibility on the vertical and the horizontal level in SDIs. It is necessary to have a sufficient metadata description of at least the spatial data and the existing metadata must be of high quality. Furthermore, a strict regulations have to be enforced for data custodians who want to participate in SDIs.

2.5 Spatial Data Modeling

This section gives an overview of current issues concerning spatial data modeling and the different concepts, systems and languages used. The insight in some of the modeling languages might seem tedious, but is essential in order to understand the case study in Section 5.1.

In general, "a model is a simplified representation of an object of investigation for purposes of description, explanation, forecasting or planning. A spatial model is a model of an object of investigation in bispace (space, attribute)…. (Fotheringham and Wegener, 2000)".

 "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 (Longely et al, 2001)".

The way the reality is represented in spatial data and later in a GIS, and which information it contains depends on the point of view of the end-user. A typical example is the question "What is a street?". A pedestrian might characterize a street as a location where there is a lot of noise and traffic. According to him or her, the geometry for a map would be a line. On the other hand a navigation system considers a street a means to calculate distances and routes. Therefore in this case, the topology and polyline segments need to be explicitly saved. Yet, the official administration for roads and infrastructure considers a street a 3 dimensional object with different layers of material and pavement which needs to be

maintained as well as signs. Here, the geometry would have to specify an area of the street (Najar and Giger, 2004). Therefore, models are important to capture and characterize the design decisions as well as to generate usable work products. Models are always an excerpt of reality and interpret by organizing, filtering and editing certain facts. They offer a contribution to understand the semantics of spatial data and therefore assist exchange and interoperability of spatial data. Nevertheless, the data model itself is not enough; the ideal combination are metadata and data model which complete each other in giving information about the spatial data.

In the context of database management and GIS implementation, literature often refers to five levels of abstraction, sometimes called levels of generalization or simplification: reality, informal, conceptual, logical and physical level ((Reinhardt et al, 2004), (Longely et al, 2001), (Worboys and Duckham, 2004), (Bill, 1999), (McFadden and Hoffer, 1994)). Currently, the state-of-the-art in spatial data modeling is confined to a static model. In future, dynamic developments might be part of the model, e.g. a temporal development of an object. Figure 2-4 shows how the static process of spatial modeling is related to the modeling of dynamic processes and on which level the steps of implementation are located.

The principal modeling steps of spatial data modeling are analog to the Figure 2-4 and involve firstly a selection of real world objects that are relevant and need to be characterized in the model. These objects are modeled with a conceptual schema, taking the following issues into account:

(1) Geometric and topological structure of the objects.
(2) The objects' characteristics which are defined by the attributes.
(3) The grouping of objects to hierarchical entities/classes for better overview.
(4) Relationships between objects and group of objects.

Then, this conceptual model needs to be implemented in a database or information system. Therefore, a logical schema has to be generated. The physical schema finally describes how the data is stored on the hard disk. Here, the optimization involves adequate performance such as response time and security.

	Spatial Data (static)	Process (dynamic)	Level
I			Reality
II	Informal description of objects	List of functions and activities	Informal Description
III	Data Model Defines groups and relationships	Workflow	Conceptual Model
IV	Datebase Schema	Process specifications	Logical Model
V	Data Format File Structure	Program code and control blocks	Physical Model
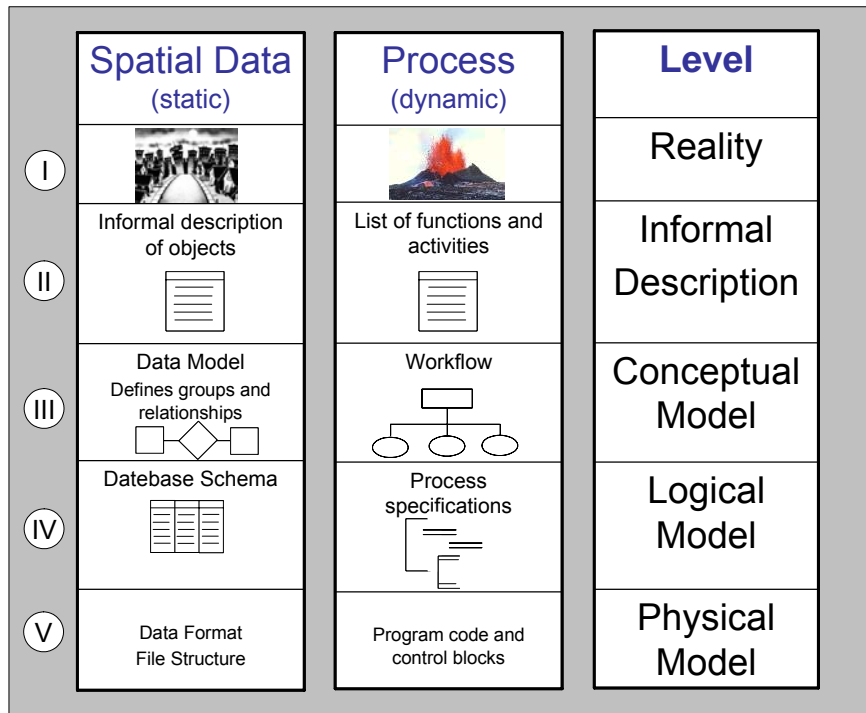
Figure 2-4: Five levels of abstraction and implementation for static and dynamic models

In order to be able to describe the models on the different levels, many modeling languages have been developed. **Figure 2-5** attempts to characterize some of the most important modeling languages for the GIS Community (see also the following sections).

*examples for tools which are used for spatial data

Legend:
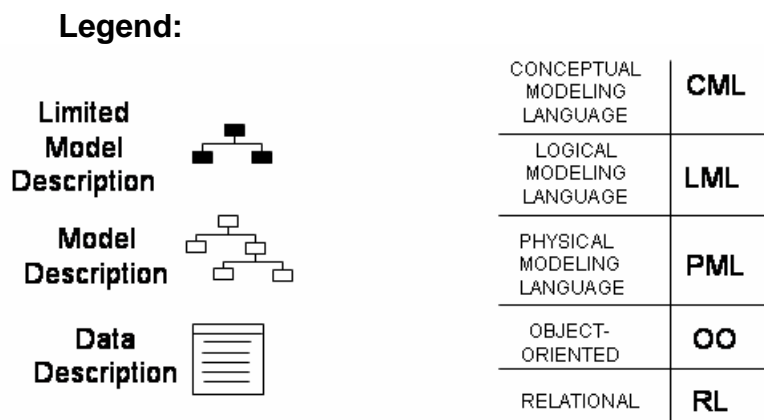
Figure 2-5: Overview of modeling languages

First of all the modeling languages differ in level of abstraction (conceptual, logical and physical) and in method they use for characterizing structures (object-oriented, relational). Some modeling languages, like UML give a graphical characterization of the model, but do not describe the data. Whereas, SQL can create and thus describe the data structures (e.g. by using the command "CREATE TABLE…", and query the information that a table contains, but does not offer any description of the model. The possibility of characterizing spatial models with GML and XML Schema is limited, e.g. inheritance is not possible and associations are complicated to define in GML schema (Gronno et al, 2002). Hence, GML

and XML are placed on the physical model, because their strength lies in the description of spatial data in a certain format.

Tools for supporting modeling languages are important as they make sure that the modeling is according to a standard and check the quality and make the modeling languages usable. This overview shows only a small, selected group of non-commercial tools as examples. In the following sections will describe the modeling languages of **Figure 2-5** in greater detail.

2.5.1 Conceptual Data Modeling

Conceptual data modeling belongs to the highest level of abstraction meaning that the grouping and structure of the data are described in a system-independent, general way. There are two main principles how to model conceptually: object-oriented and relational model. These principles each have modeling languages that support the main structures or principles. For object-oriented modeling, the languages UML, GeoUML and INTERLIS are introduced as they are relevant in the further progress of this thesis. For relational modeling, the Entity Relationship Model (ERM) is introduced with its graphical representation entity-relationship-diagram (ERD) as well as EXPRESS (ISO 10303-11), which can be used as a modeling language.

Object-oriented Modeling

Object-oriented modeling deals with objects which are grouped according to similar characteristics (attributes, behavior) and are represented in classes.
In this thesis, the following concepts are of importance for object-oriented models (Rumbaugh et al, 2005):

- **Encapsulation**: The ability to hide from a user the internal structure of an object. Instead, each object has a published interface that is made known to the outside world .
- **Inheritance**: The property that, when classes are arranged in a hierarchy, each class assumes (inherits) the attributes and methods of its ancestors, which are those higher up in the hierarchy.
- **Polymorphism**: Is the property whether the implementation of an operation may be overridden by descendant classes.

In order to realize these concepts with spatial data, modeling languages are needed to describe the data in a certain structure.

 "*A conceptual schema language is based upon a conceptual formalism that provides rules, constraints, inheritance mechanisms, events, functions, processes and other elements that make up a conceptual schema language. For the ISO 19100 family of standards the applicable conceptual formalism is the object-ortiented modeling as described by OMG* (Object Management Group (OMG), 2003) (Kresse and Fadaie, 2004)".

UML

Unified Modeling Language (UML) is one of the most broadly used conceptual, object-oriented modeling languages in various areas: software engineering, workflow for business processes, spatial data modeling.

It is a general-purpose visual modeling language that is used to specify, visualize, construct, and document the artifacts of an object-oriented system. In general, it offers static, dynamic, environmental, and organizational parts. Furthermore, it is intended to be supported by interactive visual modeling tools that have code generators and report writer. UML was developed in an effort to simplify and consolidate the large number of object-oriented development methods that had emerged (Rumbaugh et al, 2005).

UML is specified and supported by the OMG (Object Management Group (OMG), 2003).

Generally, it is possible to extend UML, e.g. by using stereotypes for different areas of application.

For geographic information, the ISO/TC 19103 (Conceptual schema language) defines a UML profile (Kresse and Fadaie, 2004).

The graphical representation of the model by using pictograms, e.g. for modeling databases, has the advantage of making complex relations easy to survey and to explain to non-experts ((Bédard and Paquette, 1989), (Bédard et al, 1992)). Nevertheless, this visual modeling is as such not computer-processable. It needs to be transferred in a machine-readable format.
Figure 2-6 gives a basic overview of important features of possible relationships in UML which are needed for this thesis.
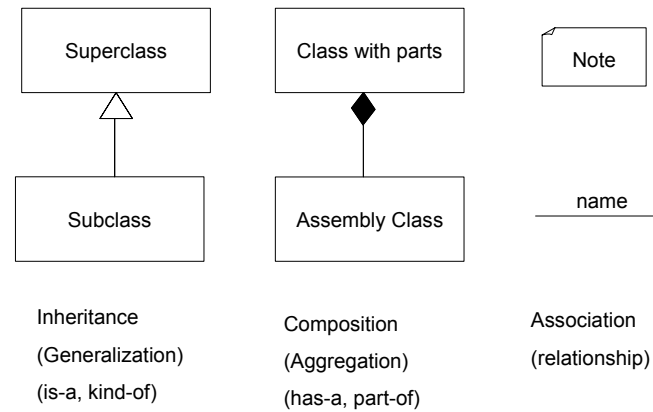
Figure 2-6: Basic class diagrams in UML

An association is the semantic relationship between two or more classifiers that involves connections among instances.

Aggregations and compositions are a form of association which specify a whole-part relationship between an aggregation (a whole) and a constituent part. A composition is a form of aggregation association with strong ownership and coincident lifetime of parts by the whole. This means that objects of, e.g. the class 'points' are necessary for the class 'lines'. In a composition, if one class is deleted, the other class is deleted, too.

An inheritance enforces that all the objects of the subclass have the same characteristics as the superclass, plus whatever is defined in the subclass itself.

The problem with UML, even in ISO/TC 19103, is that it is not created with spatial data in mind. Consequently, it lacks possibilities which allow the content, geometry types and spatial integrity constraints of a geographic database to be formally specified. Therefore, GeoUML was created for the Italian IntesaGIS project (Amadio et al, 2004).

GeoUML

GeoUML is a conceptual model using UML-based terminology which allows the formal specification of the content and the spatial integrity constraints of a geographic database. The design of GeoUML is ISO compliant, it is a specialization of ISO/TC 211 standards (Belussi et al, 2004a). GeoUML uses UML as conceptual language, but adds some conceptual constructs, such as a set of predefined spatial types, a set of rules for representing geographical data using UML, templates for the specification of integrity constraints, schema templates for the representation of structures shared by most geographic applications (Belussi et al, 2004b). Yet, what is missing for GeoUML are tools for using and processing GeoUML for spatial data.

INTERLIS

INTERLIS is a conceptual schema language, an interface and a transfer format. It is designed specifically for spatial data and its use is obliged by law for the cadastral surveying in Switzerland.

There are two different versions of INTERLIS: INTERLIS1 and INTERLIS2. Both are based on the model-driven approach and therefore separate the data from the model. Plus, they are conceptual modeling languages and have firstly a graphical representation, secondly a textual description of the model and thirdly a transfer file. Finally, both support multilingual models.

The graphical description of the model in INTERLIS1 is based on Entity-Relationship-Model (Table 2-1). The textual version of the model is offered as *.ili and the data file has the extension *.itf is ASCII. The graphical representation for INTERLIS2 is UML and the object-oriented concepts were thus added. The textual description of the model is still the *.ili format, as in INTERLIS1, but the data file uses XML-Schema (*.xsd) to describe the data instead of *.itf.

	Graphical Representation of Model	Textual Description of *Model*	Textual Description of *Data*	Transfer Format
INTERLIS1	ERM	*.ili (ASCII File)	*.itf (ASCII File)	*.fmt (ASCII File)
INTERLIS2	UML	*.ili (ASCII File)	*.xsd (ASCII File, XML-Schema)	*.xtf (ASCII File)

Table 2-1: Overview of INTERLIS formats

Figure 2-7 shows how the different object-oriented concepts have been implemented in INTERLIS2 and how the relationships have changed in comparison to INTERLIS 1. In INTERLIS 2 there are classes instead of tables. The object-oriented concepts support the possibility of offering incremental updates, hierarchical relationships, layout elements, etc. More information is given in the reference manual (Coordination of geographic information and geographic information systems (KOGIS), 2003).

INTERLIS 2 UML	Textual INTERLIS2	INTERLIS1 Entity-Block-Diagram	Textual INTERLIS1
	CLASS A= ..... END A;		TABLE A= ..... END A;
1		1	
0..1		c	
0..*		m	
1..*		mc	
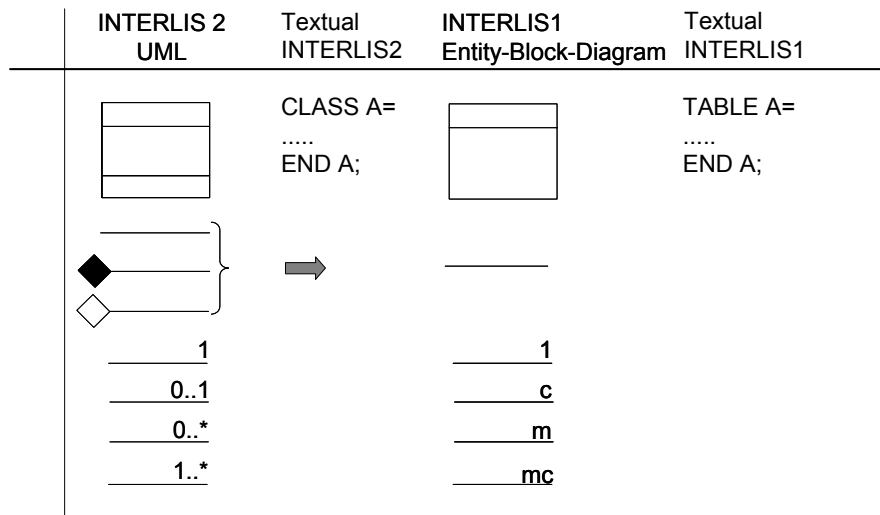
Figure 2-7: INTERLIS2 and INTERLIS 1 notation (modified, (Eisenhut et al, 2003))

Since INTERLIS will be used in the case studies, here an excerpt from a model description (*.ili) is offered.

Notation in INTERLIS 2 for metadata from GM03

!! are comments

…

CLASS MD_Distribution =

END MD_Distribution;

CLASS MD_Metadata = !! root entity which defines metadata about a resource or resources

 fileIdentifier : CharacterString; !! unique identifier for this metadata file

 language : LanguageCodeISO; !! language used for documenting metadata

 characterSet : MD_CharacterSetCode; !! full name of the character coding standard used for the metadata set

 dateStamp : MANDATORY Date; !! date when the metadata was created

 metadataStandardName : CharacterString; !! name of the metadata standard (including profile name) used

 metadataStandardVersion : CharacterString; !! version (profile) of the metadata standard used

 hierarchyLevel : BAG {0..} OF MD_ScopeCode_; !! scope to which the metadata applies (see Annex H for more information about metadata hierarchy levels)*

 hierarchyLevelName : BAG {0..} OF CharacterString_; !! name of the hierarchy levels for which the metadata is provided*

 dataSetURI : URL;

END MD_Metadata;

```
ASSOCIATION distributionInfoMD_Metadata =
        distributionInfo -- {0..1} MD_Distribution; !! provides information about the distributor       of
        and options for obtaining the resource(s)
        MD_Metadata -<#> MD_Metadata;
END distributionInfoMD_Metadata;
```

CLASS defines a class in INTERLIS 2 in which attributes and their data type are declared, e.g. the attribute 'dateStamp' in MD_Metadata is compulsory (*MANDATORY*) and has to comply to the format *Date*, which was defined explicitly at the beginning of the data description file of the data model.

ASSOCIATION defines the relationship between the class MD-Metadata and MD_Distribution via the attribute 'distributionInfo'.

An advantage of INTERLIS is that many tools exist which allow to process and link the data to existing GIS software or edit and work with them (INTERLIS, 1991). Also, INTERLIS is compatible to other languages (e.g. GML and Java). The Compiler can transfer a model in other description languages. The disadvantage of INTERLIS tools is that they are not well documented and many of the tools are difficult to handle without insider information by the programmer.

Relational Modeling

In literature the name relational or object-relational modeling can be found. Relational modeling on the conceptual level should not be confused with a relational model on the logical level. The latter has a flat hierarchy and does not support relationships as described in the following.

Entity-Relationship Modeling (ERM)

The main features of an Entity-Relationship model are entities, relationships and attributes. Each entity is defined by its component attributes; the affiliations, or relationships between entities, are identified (together, attributes and relationships are known as properties).

 "The entity-relationship model (ERM) is used to construct a conceptual data model, which is a representation of the structure of a database that is independent of the software (such as a database management system) that will be used to implement the database…The ERM is expressed in terms of entities, relationships among those entities, and the attributes (or properties) of both the entities and relationships. A ERM is normally expressed as an entity-

relationship-diagram (ERD), which is a graphical representation of an ERM. …There is no standard notation for ER modeling (McFadden and Hoffer, 1994)". ERM was not originally specialized for spatial data. It is also used in different organizational or business applications.

EXPRESS

EXPRESS is a relational data modeling language which is used for industrial design, e.g. for CAD design in mechanical engineering. It is standardized as ISO 10303-11 and is used as a modeling language by ISO 10303 known as STEP (Standard for the Exchange of Product model data (STEP)). An EXPRESS model can be defined in two ways: textually and graphically. The textual representation within an ASCII file is used for the formal verification and as an input. The graphical representation, called EXPRESS-G, does not describe all details, but is advantageous when explaining the model to someone else or to get a quick overview. Within a schema various datatypes can be defined together with structural constraints and algorithmic rules. As in INTERLIS there are tools to model and formally validate a population of datatypes- this is to check for all structural and algorithmic rules (Denno and Sanderson, 2000).

The disadvantage of EXPRESS is that it is not specialized on spatial data and therefore does not define any geometric datatypes. Also, the single parts in mechanical engineering are more standardized than in the GIS community. Therefore, the transfer tools do not have to work for every model, but only well-defined, specific models.

Altogether, most conceptual modeling languages offer a graphical representation and a computer-processable ASCII file. The latter is important for a network infrastructure, e.g. in an SDI. The choice which modeling language to use for spatial data depends on the features offered to describe geographical constraints (e.g. topological or datatype constraints) as well as geometry types. Furthermore, a modeling language needs tools to support the transfer to and from GIS software as well as managing tools. In general, conceptual schema languages give a system-independent, generic overview of the spatial data.

2.5.2 Logical Data Modeling

During the logical design process, a conceptual model is transferred to a logical model. If this is done with the purpose of managing data in a database, the goal is to translate and implement the logical database model on a chosen database management system (DBMS). Within the GIS community, spatial data are more and more stored by using DBMS, such as Oracle, which offers specialized spatial options and features. These DBMS are often

connected to a GIS desktop software in order to use the management utilities together with the functionalities of visualizing and processing of a GIS.

In a relational model, the data is represented in tables and relations. The tables are connected to each other in a flat hierarchy by primary and foreign keys. A primary key is the field or column in a database table that is indexed and maintains the main sequence of the table. The foreign key is an attribute that appears as a nonkey attribute in one table and as a primary key in another table. An Object Identifer (OID) is generally an implementation-specific integer or pointer that uniquely identifies a record.

Data manipulation is preformed with operations which are incorporated in the structured query language (SQL). SQL offers the possibility to define the data base scheme, and then insert, maniputlate, and retrieve data from the data base. Furthermore, it is possible to embed SQL in a general-purpose programming language, e.g. in order to extend its capabilities in computations.

2.5.3 Physical Data Modeling

The major objective of physical design is to implement the data as asset of stored records, fields, indexes, and other data structures that will provide adequate performance and ensure database integrity, security, and recoverability. This means that the logical database structures are mapped into an internal model, or in other words a set of physical data structures.

For spatial data, this can be database structures or single file structures. For the dominating IT world it is important to have computer-processable data which can be indexed by a search engine and thus be findable, exchangeable and retrievable by Internet. Therefore, markup languages like XML and GML seem to be a good solution for describing spatial data.

XML

In general, Extensible Markup Language (XML) is a subset of ISO 8879: 1986 Standard Generalized Markup Language (SGML). An XML document is composed of pairs of tags , each pair consisting of a beginning and an end tag, like <table> and </table>. A pair of tags is called element. The rules concerning how the tags are structured, are defined by the creator in specialized DTDs (Document Type Declaration). In contrast, HTML has predefined DTDs and is, therefore, inflexible and not useful for data description purposes. Since DTDs lacked semantic expressiveness and the possibility of checking, validating spatial data, XML Schema (XSD) has been developed. It defines the allowed elements structures in a XML file.

When using XML for spatial data, the data model is defined by XML Schema and the data itself is described in XML (Kresse and Fadaie, 2004).

GML

A dialect of XML is specialized as a transfer format for spatial data: the Geography Markup Language (GML) which is defined in ISO 19136. GML defines 42 basic geometry types. These are implemented in different versions 2.x and 3.x. Different vendors of GIS have implemented their own profile of GML. Consequently integration of data from different systems is not yet provided. Also, the versions of GML are not compatible. Therefore, the interoperability is confined.

RDF

RDF (Resource Description Framework) is a metadata transfer format which was developed by the W3C. Its goal is to define a mechanism for describing resources. It offers a representation model for metadata, as well as a syntax for encoding and transporting metadata using XML.

2.6 Semantic Processing of Spatial Data Sets

Processing spatial data includes exchanging, analyzing and restructuring the data. Analyzing and restructuring spatial data is usually done in a GIS, which provides the necessary functionalities. The next step is to exchange spatial data between different GIS. This section analyses the possibilities of transferring spatial data from one GIS to another.

As mentioned in Section 2.5, models are needed for the characterization of spatial data structures and for the interpretation of semantics. The latter depends on the point of view and application of the data set which was chosen for a certain topic. The example of "what is a street?" was given in Section 2.5 and the different interpretations depending on the use of the data. Another example is forest: if the forest warden point of view is chosen, there might be attributes with the name "tree-type", while when the surveyor acquires and models the data there might be attributes containing the area of the forest.

Thus the situation of semantic processing as a form of interoperability between two given spatial data sets of different semantics and stored in different commercial GIS is analyzed in this section.

For the exchange of data there are three generations of development (Shi, 2004):

The system-oriented data exchange is the oldest and affords a unified format (Figure 2-8). The users cannot create their models. Instead the exchange formats have to be created and comply with the different system. (example for mediating software like FME - Feature Manipulation Engine (Safe Software, 2005)
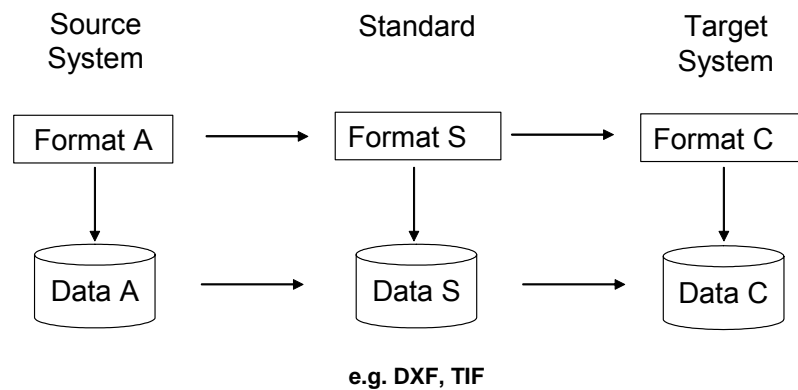
Figure 2-8: System-oriented data exchange (modified from (Shi, 2004))

For the first generation only data in certain formats are transferred. They are therefore specific for the producer and program-dependent. Examples are DXF and ESRI Shape files.

(1) Static model-driven exchange means that source and target system have the same, standardized model. Data is organized according to this common model and exchange is realized according to this interface. This is the case e.g. in Germany for topographic-cartographic Information system (ATKIS) and its interface called EDBS.
The second generation has a fixed model and different formats. Its field of application is specialized on one organization, but it is system-independent. Examples are : DIGEST, SDTS, ATKIS

(2) Dynamic model-driven exchange is based on dynamic application models. This means that data are delivered together with their model by the source system. An initial model from a GIS system A is transferred first to a standardized model S. This intermediate model is used to transfer the information in the target GIS system (Figure 2-9). Since the target system is able read and interpret the standard model, no information is lost. The translation between Model A and the target model C needs encoding rules for mapping. This procedure is used for INTERLIS.
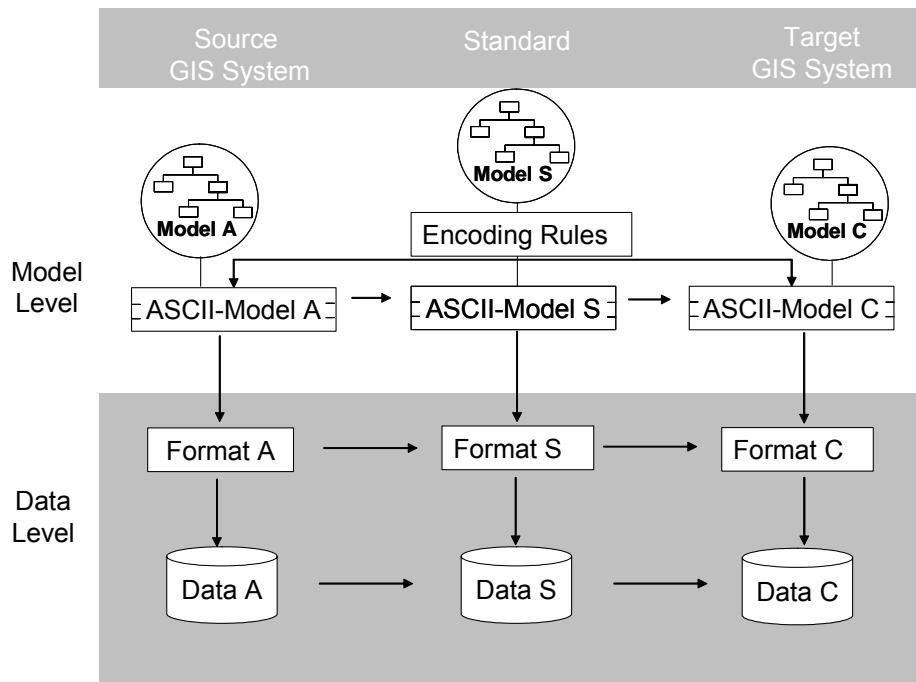
Figure 2-9: Dynamic model driven exchange (adapted (Eisenhut et al, 2003) and (Shi, 2004))

2.6.1 Model Driven Architectures

In the IT area of software design, a new methodology has been proposed and supported by the Object Management Group (OMG) (Object Management Group (OMG)). The main goal of the so called Model-Driven Architecture (MDA) is to build software that will be relatively easy to adapt to new technologies and is platform independent.

The MDA is an approach to system development, which increases the power of models. It is model-driven because it provides a means for using models to direct the course of understanding, design, construction, deployment, operation, maintenance and modification (Miller and Mukerji, 2003).

The Model-Driven Architecture defines models on different levels of abstraction (Figure 2-10). The Computation Independent Model (CIM) describes the problem space. It is comparable to the informal model in Figure 2-4 and does not contain technical details or the structure of the system. A CIM is for example used to describe a business system. The Platform Independent Model (PIM) defines the system functionality and models business aspects of the software system, using an appropriate specification language like UML. The PIM is derived from the CIM by manual translation. The PIM uses annotations to define business aspects and does not contain any elements that are specific to a certain platform. Platform specific elements are introduced in the transformation to one or more Platform Specific Models (PSM) for the actual implementation. In order to accomplish this goal, the

MDA defines an architecture that provides a set of guidelines for structuring specifications expressed as models. The PSM in turn is the basis for code generation for a specific target platform. The transformation and generation processes use the annotations in their respective source models, and templates that define the platform specific aspects and transformation rules (Seifert et al, 2004).
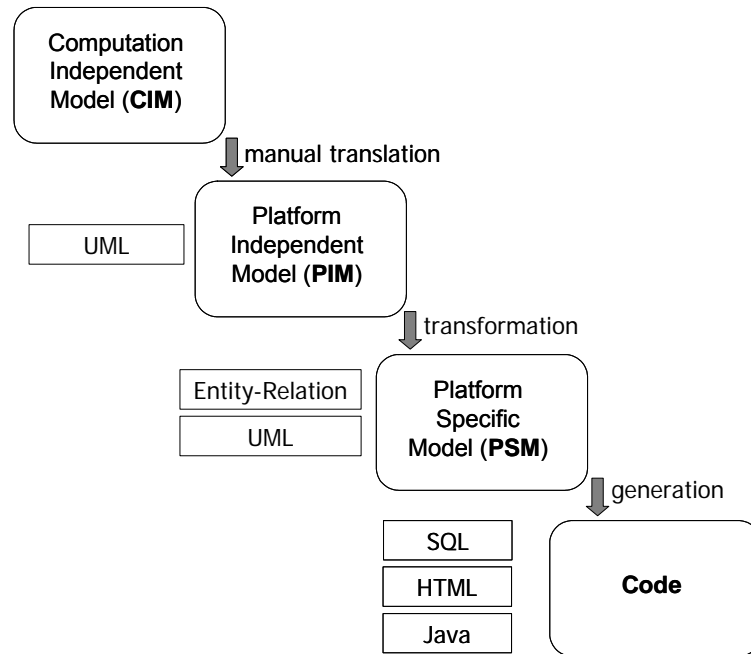
Figure 2-10: How does MDA work?

The MDA standard defines the terms PIM and PSM. The OMG documentation describes this distinction as if this is a clear black-and-white issue. In reality it is difficult to draw a line between platform independent and platform specific. Is a model written in UML specific for the Java platform because one of the class diagrams defines one or more interfaces? (Kleppe et al, 2003)

For the transformation from PIM to PSM and from PSM to Code, different transformation tools exist. (e.g. Tefkat an Eclipse Plugin based on Eclipse Modeling Framework). In future, the PIM should stay the same and only the steps of transformation need to be executed for a new platform. This saves time and increases the quality of the system.

The MDA relies on different standards, e.g. UML, XML Metadata Interchange (XMI), which is the transfer format for UML, MetaObject Facility (MOF), which provides the constructs for modeling and interchange and the Common Warehouse Metamodel (CWM), which covers the full life cycle of designing, building and managing data warehouse applications and life cycle (Miller and Mukerji, 2001).

For MDA to be accepted it is necessary to offer MDA tools, modeling wizards, support for model consistency checks, PIM-PSM translations, code generation and support for systems engineering.

Critical voices have argued that MDA, UML and other base technologies are still under development and therefore the necessary tools will change considerably. The standards are written in such a way that is subject to interpretation which means implementation tools differ depending on the vendor's perceptions (e.g. what exactly to include in a PIM and how to translate a PIM to a PSM and then to a compilable, deployable code. Thus, the status of being "MDA compliant " does not mean that a PIM created with one tool can be used with another tool. Also, UML does not allow business logic to be completely formulated. Experts therefore suggest to use MDA only for the initial generation of code and/or database scheme (Seifert et al, 2004), (Fowler, 2004).

2.6.2 Model Driven Architectures for the GIS community

In the GIS Information Community, the concept of the model-driven approach starting to be introduced and adopted for spatial data (van den Berg et al, 2005). Modeling languages support UML as a standardized, internationally used modeling language and extend it for spatial data purposes. Furthermore, the describing model is separated from the actual, technical data, similar to MDA where business aspects are divided from technical issues as well as implementation details.

As an example how the model driven approach could be implemented, Figure 2-11 shows how the MDA for spatial data exchange is organized In Switzerland with the conceptual schema language INTERLIS.

The advantage of such a system is that the different organizations which are involved can use arbitrary GIS software and define their models according to their own needs, but following given rules, similar to possibility of defining profiles for ISO standards. The ideas is that the different administrative regions in a federal system which are managed by local levels define the same data layers and are thus connectable to one homogenous map.
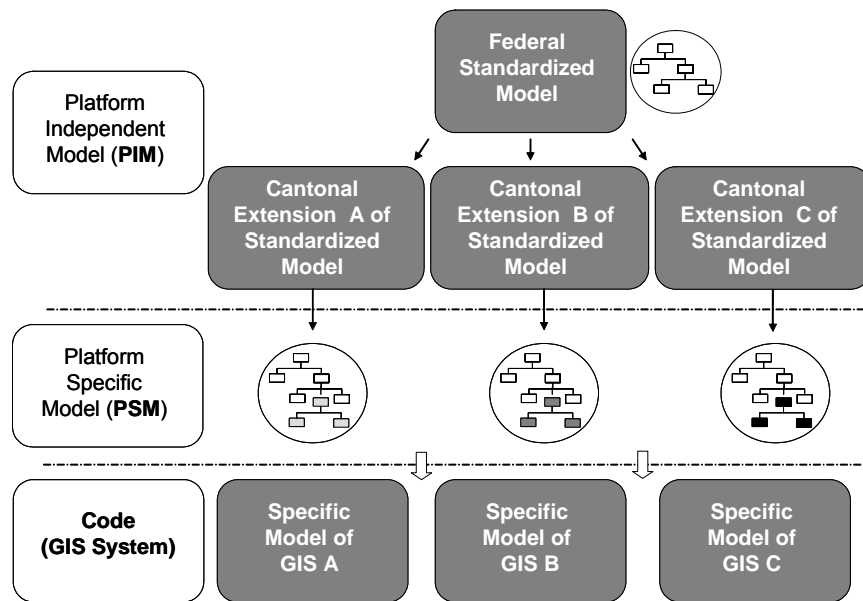
Figure 2-11: MDA for Switzerland using standardized models as PIM

2.6.3 Standardized Models

Ever more application areas and information communities in GIS ask for standardized models. For example, in the area of utilities such as gas, water and electricity, it has been recognized that standardized models result in significant savings during upgrade of the system. The flexibility and the access to the implemented data model of a GIS system depends on the policy of the vendor (Hesse, 2005).

A well known approach is the development of the Core Cadastral model by the International Federation of Surveyors (FIG) at the 2002 Congress in Washington (Oosterom and Lemmen, 2002). According to (Oosterom et al, 2004) it is based on the common denominator of functions of cadastral models:

- Keeping the contents of the relationship between persons and land up-to-date (based on transactions).
- Providing information on this registration.

A standardized core cadastral domain model, covering land registration and cadastre in a broad sense (multipurpose cadastre), will serve at least two important goals: avoid reinventing and re-implementing the same functionality over and over again, but provide an extensible basis for efficient and effective cadastral system development based on a model driven architecture.

It could also encourage and support the flow of information relating to land property between different government agencies and the public (Lemmen et al, 2005).

2.6.4 Model-Driven Services for Spatial Data

In the ideal situation services are model-driven and therefore first read and process the spatial data model and then the spatial data itself. For example OGC services (e.g. WFS, WMS) could be extended to be model-driven so they can read any model and then the corresponding data. The data can then be processed in the usual way and then transferred to GML, if necessary.

Currently only the data transfer is a working model-driven service. Therefore, the examples and requirements are focused on this service.

The model-driven exchange of data is only possible if certain prerequisites are fulfilled. The target system must be able to interpret and reconstruct the model from the source system. This implies that the quality of the schema has to be secured.

The innovation of creating model-driven services is that the user is not anymore dependent on the of the goodwill system producers and whether they offer interfaces for transfer. Examples for model-driven services (Gnägi et al, 2002):

- Data transfer service: the data model is first transferred from one system to the other and then the data in a format, which was automatically derived from the data model.
- Checking programs are capable to compare whether the model and the data correspond to each other. It is possible to check whether numbers of the data are in the defined interval of the model and whether relationships were realized, whether zones of an area classification create a consistent surface.
- Geoportals (geoshop) offer the presentation, geographic and thematic selection and the exchange of spatial data, which come from different systems. This service can be completed by an e-commerce functionality.

First the Model…

1. Step

….then the corresponding data

2. Step
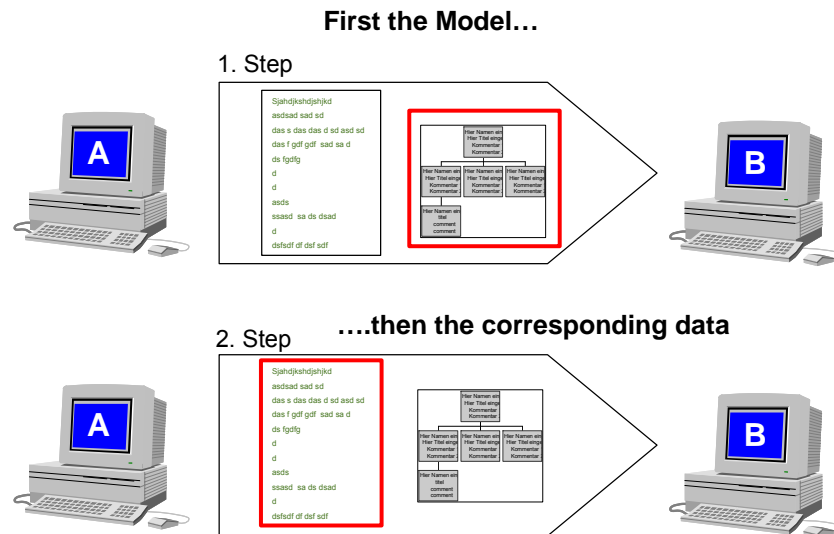
Figure 2-12: Two steps of the model-based service

for transferring data from system A to system B

First, any model is transferred from system A to system B and read by system B (Figure 2-12). Then, the corresponding data are sent and can be interpreted with the help of the model. Thus, two separate files are transferred: data and model. This exchange is format independent as model-driven, conceptual schema languages (e.g. INTERLIS) are used to describe the model and the data.

2.6.5 Ontology Based Interoperability

As mentioned in Section 1.1, ontologies are considered a more generic, application-independent description of spatial data. They can be placed on the level between the conceptual model and reality (Figure 2-13) in the abstraction model.
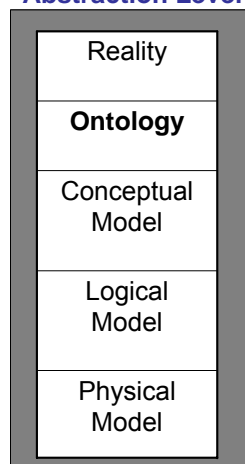
Figure 2-13: Levels of abstraction including ontologies, based on Figure 2-4

With the advance of Internet and the wish for seamless sharing and exchange of spatial information between systems and Information Communities, ontology has been looked into by different areas of research. Ontologies claim to provide a more cognitive foundation of information modeling. They try to build common "single" and "true" models of the reality and agree on a common understanding of the reality by specifying languages that describe these ontologies. In comparison, conceptual modeling is a specification of a single possible perspective of the world, whereas ontologies describe the same situation in a more general way and thus catch one or many state of affairs of the real world (Bishr and Kuhn, 2000) , (Hakimpour, 2003).

Hence, the example "What is a street?" (Section 2.5) is captured in models from different points of view. The perspective of the models depend on the purpose of the spatial data. Whereas, there is only one ontology corresponding to the different models, which represents the street in a generic way.

There have been many theoretical categorizations for classifying and designing ontologies in research. Here, a brief overview of some approaches:

There is, for example, the suggestion by (Guarino, 1997) to classify according to their dependence on a specific task or point of view:

(1) **Top-level ontologies** describe general concepts like space, time, matter, object, event, action, etc., which are independent of a particular problem or domain: it seems therefore reasonable, at least in theory, to have unified top-level ontologies for large communities of users.

(2) **Domain ontologies and task ontologies** describe, respectively, the vocabulary related to a generic domain (like medicine, or automobiles) or a generic task or activity (like diagnosing or selling), by specializing the terms introduced in the top-level ontology.

(3) **Application ontologies** describe concepts depending both on a particular domain and task, which are often specializations of both the related ontologies. These concepts often correspond to roles played by domain entities while performing a certain activity, like replaceable unit or spare component.

Based on this classification an ontology-driven geographic information system has been proposed by (Fonseca et al, 2002) which acts as a system integrator.

Also, (Fonseca, 2001) suggest to use the terms low-level ontologies which are for fine ontologies with detailed information and high-level ontologies. The latter are used for coarse

ontologies, and they represent more general information which is refined by low-level ontologies.

(Frank, 2001) suggest five tiers of ontology for GIS, in which different approaches are used on each tier:

Tier 0: human-independent reality

Tier 1: observation of physical world

Tier 2: objects with their properties

Tier 3: social reality

Tier 4: subjective knowledge

Another approach by (Kuhn, 2003), introduces semantic reference systems which strive to provide means for any pair of information communities to define their concepts and translate between them. It aims at developing tools for transformations among semantic spaces and projections to subspaces. A transformation may occur within or between information communities and involves a change to the reference system (for example, adding a new axiom to an ontology). A projection occurs typically within a community and reduces the complexity of a semantic space, e.g. by generalizing two entity classes to a super-class. A semantic reference system consists of ontologies that specify concepts as well as mappings between them, embedded in a formalism that supports the computation of these mappings.

In order to formalize ontologies, description languages that support the concepts are needed. Analogue to data modeling languages, there are many more than the ones listed here: ONTOLINGUA (Farquhar et al, 1997) and (Gruber, 1993), low-level and high-level LOGIC (Fonseca, 2001), Description and Frame-based Logic (Hakimpour, 2003), OWL (Web Ontology Language) (World Wide Web Consortium (W3C), 2004).

Many of these ontological approaches are theoretical structures and are not used beyond research. Altogether, ontologies are not used directly for this thesis, but are important to understand in comparison to models. The aim is to describe spatial data in an extensive way with models and metadata.

2.7 Summary

Chapter 2 offers definitions for the most important terms of this thesis and introduces basic notions and understanding of the current situation in relevant areas of research.

Spatial Data Infrastructures (SDIs) are enabling platforms which facilitate the access to spatial data and comprise organizational, political, cultural and technical aspects. The traditional spatial data processing tool GIS is growing together with Web applications and ideally supports SDIs as a tool for processing and visualizing spatial data.

The relevant components of SDIs (Web Services/Clearinghouses, spatial data and standards) and their processes within an SDI have been analyzed with regards to metadata and in order to understand how integrated metadata-spatial data can support certain parts of the workflow and existing components.

As metadata describe the contents of spatial data, spatial data models define the structure. Usually, the data is grouped according to certain characteristics or topic and relationships are defined. Different modeling languages can describe a model. They differ in level of abstraction (conceptual, logical and physical) and in method they use for characterizing structures (object-oriented, relational). In context of modeling and spatial data management, five levels of abstraction are referred to. They are sometimes called levels of generalization or simplification: reality, informal, conceptual, logical and physical level. Different examples of languages on different abstraction levels are provided.

In the IT area of software design, a new methodology has been proposed called Model-Driven Architecture (MDA). Its goal is to build software that will be relatively easy to adapt to new technologies and is platform independent by using models. These models are generic and thus can be transferred to any platform and then to the necessary code.

In the GIS Information Community, the concept of the model-driven approach has been adopted in a specialized way for spatial data. Modeling languages support UML as a standardized, internationally used modeling language and extend it for spatial data purposes. Furthermore, the describing model is separated from the actual, technical data, similar to MDA where business aspects are divided from technical issues as well as implementation details. Thus, the transfer from one system to another can be realized.

Other approaches to achieve interoperability between different systems are provided (e.g. ontologies).

Chapter 3   State of the Art in Metadata

The previous chapter reviewed the current situation in relevant areas which concern this thesis indirectly. This chapter looks into the metadata themselves and the topics of research which directly concern this work. It is therefore focused on concepts which relate to the topic of model-driven approach of integrated metadata.

Metadata have been defined in Section 1.1 and Section 2.1 to be data about data and therefore describe spatial data in content, use, purpose, source as well as quality (what?, where?, when?, who?).

Since spatial metadata are inherently, by their nature of description, linked to spatial data, it is necessary to look into certain aspects of spatial data handling. On the one hand metadata records would not exist without spatial data. Spatial data need to be described for interoperability reasons. On the other hand, one of the points of investigation in this thesis is whether the character of metadata is close enough to spatial data so that tools and methods which are applied on spatial data can be used for metadata as well.

Spatial Data Infrastructures are the application environment for the future use of concept described in this thesis. Therefore, only the SDI-relevant issues of metadata are mentioned here.

3.1 Current Situation in Research of Metadata Management

The classical definition is that metadata are "data about data". Today, the amount of data is growing quickly and metadata must fulfill ever higher requirements and tasks. The functional definition by (Turner, 2002) (quoted in (Calhoun, 2002)) is taken from library research and reflects its ambivalent nature. Metadata:

(1) help find and access information

(2) serve particular purposes

(3) can be used by people and/or machines

(4) often have structure and/or content rules

(5) can be created by people or machines

Metadata are used by many information communities and for different purposes (Point 2). While library cataloguing use metadata for describing books, the GIS community needs metadata to describe spatial data and services. Both objects of description have their own

particularities and need special metadata attributes to describe them sufficiently. When metadata is the result of a query by Clearinghouse, it is viewed by people. It is then important to support the user by making it easier to understand with aids such as graphical representation. In this case, metadata are used by people. Nevertheless, in order for the Clearinghouse to be able to provide an answer to the query it needs metadata catalogues which contain machine-readable metadata (Point 3).

In the following sections, the points of the functional description are addressed. Metadata created by people or machines (Point 5) is discussed in the Section 3.1.1 Metadata Acquisition. Metadata management, which is related to spatial data management is elaborated in Section 3.1.2 Metadata Management. The fact that metadata helps to search for information (Point 1) is discussed in Section 3.1.3 Metadata and Interoperability.

3.1.1 Metadata Acquisition

Until today, the majority of metadata are generated manually after the spatial data has been collected. The time period between the acquisition of spatial data and metadata can be crucial for the quality of metadata. As early as 1996, it was stated "…that retroactive compilation of metadata is an error prone and sometimes even impossible task (Beard, 1996)". The metadata quality is essential for retrieving and accessing spatial data in an SDI environment.

Using integrated approach of commonly managing metadata and spatial data, metadata are acquired at the same time as spatial data and are saved in the same file or database. Therefore, the metadata quality should be improved.

There have been some efforts in metadata acquisition research. Different methods have been developed which attempt to generate metadata automatically from spatial data ((Balfanz, 2002b), (Hill et al, 1999), (Manso et al, 2004)).

Hill et al (1999) attempts to develop a general model of digital library collections. Two classes of metadata are used: inherent metadata which can be derived through computer analysis of the contents of any collection (e.g. temporal coverage, types of items, formats), and contextual metadata: information which is supplied by the collection provider or collection maintainer that cannot otherwise be derived from the collections's contents (e.g. title, responsible party, scope and purpose, type of collection, date, update frequency, terms of conditions of use for the collection).

(Manso et al, 2004) tries to identify metadata information which can be acquired from standardized spatial data files, e.g. raster files like GIF, PNG, JPG, TIFF, GeoTIFF, Bmp, GRD ESRI, IMG erdas, DOC (idrisi) etc.

(Balfanz, 2002a) presents an integrated approach that uses metadata technology to guide users through data and method selection in a network environment, e.g. a data warehouse. The visualization part aims to make the user aware of the goal-related spatial data qualities. The underlying metadata is provided by a repository-based generator, which creates descriptive metadata by analysis and interpretation of the original spatial data.

All the mentioned approaches require some manual work and long waiting because the processing takes a lot of time. Regardless, a small group of metadata can be derived.

The relevance of metadata acquisition for this thesis is not the automatic acquisition as such. The concept of integrating metadata and spatial data (Section 4.2) uses existing, already acquired metadata and the main focus of research concerns the optimization of metadata management. Nevertheless, commonly managed metadata and spatial data can have an impact on metadata acquisition: using the idea that certain metadata attributes already exist as spatial data in the data set, they do not need to be acquired a second time. Therefore, less metadata need to be generated.

There are many commercial and non-commercial tools for metadata acquisition on the market. There have been attempts to categorize these by ((Moellering, 1998), (Federal Geographic Comittee FGDC, 2001)). Usually, these tools support the acquisition by using templates and masks according to a certain metadata standard (Balfanz, 2002b). Different standards offer different metadata models and therefore the resulting metadata sets differ stronly in structure and content (Hill et al, 1999).

3.1.2 Metadata Management

Here, a short overview of recent metadata management research is given in order to show the current development.

The importance of metadata management is stressed by the fact that "metadata are not worth much unless they are captured, stored and consistently managed in order to be uniformly accessible by users and software components (Vaduva and Dittrich, 2001)."

Furthermore, without effective metadata management, the corresponding spatial data is less valuable. (Nebert, 2001) calls data, plus the context for its use (documentation, metadata) information.

There are significant benefits to such asset management of metadata:

- Metadata helps organize and maintain an organization's investment in data and provides information about an organization's data holdings in catalogue form.

- Coordinated metadata development avoids duplication of effort by ensuring the organization is aware of the existence of data sets.

- Users can locate all available geospatial and associated data relevant to an area of interest.

- Collection of metadata builds upon and enhances the data management procedures of the geospatial community.

- Reporting of descriptive metadata promotes the availability of geospatial data beyond the traditional geospatial community.

- Data providers are able to advertise and promote the availability of their data and potentially link to on line services (e.g. text reports, images, web mapping and e-commerce) that relate to their specific data sets (Nebert, 2001).

Metadata management provides an infrastructure for storage, update of metadata according to certain metadata standards and thus strives to sustain high metadata quality. Today, repositories for metadata are used which are either file structured or databases and which are created, managed and updated separately from the spatial data.

Often, metadata management systems have evolved historically and the set of metadata attributes have changed over time. This results in redundancies and inconsistencies in discovering which information belongs where, e.g. does the reference system belong to the spatial data or metadata or to both?

It is, therefore, widely acknowledged, that new concepts for metadata management are necessary ((Bulterman, 2004),(Onsrud et al, 2004),(Balfanz, 2002b)).

Current approaches for new metadata management in research have suggested to enhance HTML documents in the Web, e.g. with geo-tags so that they carry their metadata with them and are, therefore, findable by a search-engine ((Schwering et al, 2003), (Brinkhoff, 2003)). This might be done with XML/HTML tags or by using already given spatial information such as the URL country code as well as information in the Web text. A disadvantage of this approach is, that it creates a lot of work and affords effort. Also, this information is inserted only in new internet sites, it will take a long time to make this system fully functioning.

Another attempt which is provided by some organizations or companies (e.g. ISO, ESRI, GML) is the option of putting an URL link in the metadata which connects it to the spatial data set. This does not provide any support in creating high quality metadata, as still two data sets have to be updated. Furthermore, the link has to be kept up to date.

3.1.3 Metadata and Interoperability

This section gives an overview of the relevant research in interoperability and links it to metadata management. In addition, the context between metadata standards and metadata interoperability is analyzed.

Metadata interoperability is composed by two different kinds of interoperability: spatial data interoperability and metadata standards interoperability.

Spatial Data Interoperability

The fact that heterogeneous spatial data from various sources need to be exchanged and used commonly for example in one GIS, has been discussed and identified as a important problem which must be solved by various authors and in different research areas in the following sequence of development:

 (1) Within various IT research areas, general data interoperability is relevant.

Integration of information from different heterogeneous data sources is an 'old' subject in computer science and database research. Before the advent of Internet, integration of heterogeneous data sources was already an issue in situations where data from several databases had to be combined to get the requested information (Hess et al, 2004), (Halevy, 2003).

This area in computer science covers research concerning Semantic Web, Description Logic as well Database Research and Description Logics as well as Ontologies. These topics deal with a great variety of data which can exist in general, whereas this thesis specializes on interoperability of spatial data, which are based on specific standards and interfaces for spatial data.

(2) In a GIS environment, the exchange of spatial data between different systems has been discussed.

Seamless integrating and interoperating of geo-spatial information between different GIS in a, is an essential requirement for such information to become a necessary tool and commodity in modern information based societies. Yet, this GIS interoperability is not yet operational. ((Fonseca et al, 2002), (Bishr and Kuhn, 2000), (Sheth et al, 1999), (Bishr, 1998)).

The exchange of spatial information which used to be done with paper maps has now evolved to interoperability between different data repositories as well as dissemination of GIS capabilities within an organization or with external institutions. The use of data models, transfer format techniques, standardized interfaces and metadata are important for this horizontal integration of spatial integration.

(3) Lastly, in a network or SDI environment the call for interoperability has found a new testbed.

The complexity and functions of the components involved in an SDI environment increases. Here, a shift from interoperability focusing on spatial data to chaining services and processing information, is realized.

With regards to interoperability, not only the horizontal (different data custodians offering spatial data) but also the vertical level of interoperability needs to be developed (local level SDIs connecting to national, regional or global levels).

Therefore, interoperability is a key issue for SDIs (Beaumont et al, 2005). On the one hand within one SDI, but also connecting to other initiatives on the same level. Lastly, the development of interoperability with other geoportals (Askew et al, 2005) on the horizontal and vertical level is a challenge.

The general definition of interoperability used in this thesis is:

"…the ability to develop conventions so as data exchange and integration becomes possible."

In accordance with this, it involves:…"making multiple information sources access, manipulate and share data across their boundaries (Nogueras-Iso et al., 2004)."

The need of interoperability for heterogeneous spatial data, whether in IT, distributed GIS or in network or SDI environment is one of the reasons why metadata are necessary. They assist together with spatial data models to make spatial data interconnectable. Therefore, improved metadata management improves spatial data interoperability.

Metadata Interoperability

Metadata need to have a standardized structure of their content in order to be comparable semantically and syntactically.

According to (Coulombre et al, 1998), a classification of metadata interoperability should be divided into three levels:

- Metadata that describe the semantics of a domain (or information community)
- Metadata describing the production specification chosen by the producer and the data organization
- Metadata that represents information on data contained in the file

This last level differentiates between metadata on the "set level", "category level" and "object level". For example projection information is needed on the level of the set . On the other hand position accuracy is needed on level of objects or category.

There is discussion on how detailed the metadata description should be and on which levels spatial data should contain metadata. Library collections have an obvious basic unit: a book. In ISO 19115 (ISO/TC211 International Organization for Standardization ISO, 2001) the notion of levels of metadata abstraction is discussed. This is relevant for metadata management and interoperability because it influences the size, type and structure of the metadata repository.

The typical level of abstraction used for metadata acquisition is the dataset, but the definition of a dataset depends on the institutional and software environment of the original organization. As a result, spatial data custodians define their own level of abstraction. Consequently, some data sets can be exchanged on object level and others merely on data series level.

In the concept of metadata-spatial data integration the discussion "what is a dataset" is not necessary because the metadata are stored and managed together with the spatial data and therefore where data exist there is also a metadata description. Accordingly, common metadata-spatial data management provides thus more interfaces and possibilities for interoperability for catalogue services and Clearinghouses.

Exactly which elements must be in the metadata set is defined in the metadata standard and its model. There are many metadata standards and they differ greatly in the level of information they support and the elements they define ((Onsrud et a, 2004), (Nogueras-Iso et al., 2004), (Balfanz, 2002b), (Balfanz and Göbel, 1999), (Hill et al, 1999), (Beard, 1996)). Until now, each Information Community and country offers their standards, e.g. DUBLIN Core for general metadata (Dublin Core Metadata Element Set, 2005), UDK (Umweltdaten-katalog (UDK), 2005) for environmental data in Germany, the Earth Observation System (EOS) which was initiated by the NASA for satellite systems (Earth Science Enterprise Program (ESE), 2005), ANZLIC metadata model in Australia an New Zealand (ANZLIC - the Spatial Information Council, 2005), the Federal Geographic Data Committee in the United States of America defined a standard called Content Standard for Digital Geospatial Metadata (CSDGM) (Federal Geographic Data Committee (FGDC), 2005), Alexandria Digital Library Metadata Schemas (Hill et al, 1999), etc. ISO 19115 is supposed to unify this situation for new metadata, but experience shows that the effort of changing the existing metadata records to a new model is too high for many organizations. One of the reasons is that the different metadata models which belong to the standards are semantically overlapping and difficult to relate to each other. Also, the standards are large, complex and difficult to understand, e.g. ISO 19115 has over 400 metadata elements. It is therefore interesting to create a support for the translation between different standards and thus facilitating the exchange of metadata in different standards.

In the ideal case, this would be a software which automatically transfers metadata to a certain standard. A prerequisite for this, are the mapping definitions: which field in standard X corresponds semantically to which other field in standard Y?

This translation is called "crosswalk" and is defined by the Dublin Core Metadata Glossary as a table that maps the relationships and equivalences between two or more metadata formats. Crosswalks or metadata mapping support the ability of search engines to search effectively across heterogeneous databases, i.e. crosswalks help promote interoperability (Dublin Core Metadata Initiative, 2001)."

There are many initiatives which provide crosswalks for mapping between different metadata standards ((Metadata Architecture and Application Team, 2005), (Day, 2002), (Digital Geographic Information Working Group (DGIWG), 2005)).

Crosswalks identify fields of the data model that correspond to fields described in current standards, requirements, and other data models. The crosswalks to these schemas identify similarities found between definitions for fields of the data model and definitions of in the schemas. At the same time, they represent interpretations of similarities and do not imply

compliance or conformance of the interesting data model with a particular schema (Center for International Earth Science Information Network (CIESIN), 2005).

Nougueras-Iso et al (2004) tested the possibility of using a repository of crosswalks for different standards and a software that can activate and process these crosswalks when needed has been tested. This automatic translation is complicated by the difficulty of constructing stable crosswalks. This error-prone task, requires deep knowledge and experience with the various standards, which use different terminology, specialized methods and processes.

In conclusion, crosswalks provide a one-to-one mapping between metadata, that are acquired in different standards. This is an important basis for creating views for metadata. Yet, the concept of views advances a step further in mapping not only between standards, but also formats, extracts of datasets etc.

The complexity of the existing standards can only be understood and handled by experts. Therefore, common models for spatial data and metadata, which take standards into account and can be reused by non-modeling experts who acquire the data, would lead to better cooperation and reduced mistakes.

3.2 Relevance of Concept of Integration in Research

The concept of metadata-spatial data integration (Section 4.2) can be used in the various areas of research to support and encourage existing approaches. The following sections elaborate on the relevant areas in which a new, integrated metadata management system would have the largest impact.

3.2.1 Metadata and SDIs

SDI is and will be in future, the application area for implementing and using new metadata management concepts.

It has been stated in various literature sources that metadata are an important piece in the puzzle to technical interoperability in SDI and in networks of distributed data repositories which use search engines or Web services ((Frank et al, 2004), (Bucher, 2003), (Masser, 1999), (Smith, 1997)).

The first level of importance is supporting the horizontal level and relationships with adequate metadata in an SDI (see also Section 2.4). Enhancing the horizontal dimension of the SDI

means creating the possibility of communication between different stakeholders who want to access or offer their data in a Clearinghouse.

The second level is providing for the vertical penetration and diffusion between the SDIs on the different hierarchy levels.

Metadata have to be readable and interpretable by humans, but also by software that indexes, searches and retrieves documents over the internet. From a technical point of view, metadata must be both processable, meaning machine-readable and connectable. (Nebert, 2001).

In an SDI, Web services such as catalogue services realize the functionality to publish meta-information on spatial data resources and to search and query this information, respectively (Bernard et al, 2005).

Catalogue services are also a discovery and access system to use metadata as target for query on raster, vector, and tabular geospatial information. Indexed and searchable metadata provide a disciplined vocabulary against which intelligent geospatial search can be performed within or among SDI communities.

Also, other Geo Services for visualizing or processing spatial data are described and published in a catalogue. This can include information about functionalities of the service and which spatial data is processed and offered by the service. Therefore, metadata play an essential role for finding services, via a catalogue, in a clearinghouse.

In order to fulfill these high requirements in an SDI environment, metadata must be of high quality, meaning up-to-date, reliable and consistent. This could be achieved by the common, integrated management of metadata and spatial data. As they are updated together and redundancies are less probable.

Moreover, catalogues need machine-readable metadata in different structures. This could be supported by the view concept, which offers a possibility of leaving the original, common data set untouched on the custodian's data server while views extract the data according to a certain structure.

It is also known that metadata for Web services are essential for chaining the operations (Lemmens et al, 2003) and making the services retrievable for Clearinghouses.

If services carry their own metadata with them, they can be linked to each other according to certain required characteristics. These characteristics are documented in the metadata and can be extracted by a mediating service.

3.2.2 Semantic Integration and Interoperability

The heterogeneity of spatial data is often mentioned in literature as one of the obstacles to interoperability ((Donaubauer, 2005), (Rodriguez and Egenhofer, 2004),(Kashyap et al, 1996)).

(Bishr, 1998) offers a categorization of types of heterogeneity in the context of database:

1. **Semantic heterogeneity:** a Real World fact with more than one description.
2. **Schematic heterogeneity:** object classes describing the same Real World facts can have different aggregation or generalization hierarchies.
3. **Syntactic heterogeneity**: each database may be implemented in a different DBMS of different paradigms (e.g. relation or object oriented models).

According to ((Brodeur et al, accepted for publication in March 2005), (Egenhofer, 1999), (Ouksel and Sheth, 1999)), there has been considerable progress in addressing syntactic and structural (schematic) heterogeneities, but it is necessary to also look into semantic heterogeneities, geometric and temporal heterogeneities.

In this thesis, especially semantic heterogeneity is addressed and analyzed by using metadata and conceptual models.

Semantic Interoperability is a special kind of interoperability which provides systems with the ability of access, consistently and coherently, to similar (though autonomously defined and managed) classes of digital data, objects and services distributed across heterogeneous repositories.

In the context of semantic interoperability between metadata, two main lines of approach can be found in literature. One deals with ontologies (Section 2.6.5) and the other uses crosswalks and/or one-to-one mapping (Nogueras-Iso et al., 2004).

The wider the targeted scope of interoperability, the more difficult it is to achieve accurate, precise mappings and machine understanding of metadata descriptions which conform to schemas from different domains (Hunter et al, 2001).

There have been different attempts to map two existing spatial data models, e.g. if there is no standardized model and two topographic data sets which have defined different layers and attributes should be visualized together in a map.

When it comes to integrating different geographic databases, matching schemas are needed in order to identify how the elements of the data model relate to each other. This is done by modifying the individual schemas in order to make them as close as possible. The result are rich and close schemas, which are expressed in the same language with the same

principles. The automation, or at least semi-automation of this task is highly difficult (Mustière et al, 2004). One of the main challenges during the integration process of two models is to identify the corresponding features and to transform them (Schwering and Hart, 2004).

This semantic translation process between two spatial data models is comparable to the creation of crosswalks for metadata standards, but the data models are not standardized in most cases (Section 2.6.3). Accordingly, research has explored semantic mapping between two spatial models ((Mustière et al, 2004), (Schwering et al, 2004)) and two metadata models ((Nogueras-Iso et al., 2004), (Hunter et al, 2001), (Metadata Architecture and Application Team, 2005), (Day, 2002), (Digital Geographic Information Working Group (DGIWG), 2005)) but not the integration of a standardized metadata model and a spatial data model.

In the concept of integrating spatial data and metadata, the task of semantic mapping is adapted for a metadata standard and a spatial data mode. It is tackled by utilizing conceptual models, instead of ontologies. The conceptual spatial data model and the spatial data are enhanced with metadata. Thus, the semantic interoperability of the system is promoted.

3.2.3 Workflow modeling

For metadata management, concepts from other disciplines are useful, e.g. spatial data modeling and modeling workflows. The workflow technology, has been proven to be a useful enabling technology in many areas, permitting improved productivity and process efficiency by facilitating, automating or controlling the corresponding business processes. This method has been already used to model distributed GIS data production environments (Li et al, 1998). Now, this research has been continued in greater depth by defining the data quality inspection process in a workflow management system (WMS), which is the primary software implementation of workflow technology. The idea is to increase efficiency by defining one high-level project-wide process and used for all workflow processes. This should allow a continuous flow-of project-wide activities over the project life span and provide important information for project management, such as internal work status. Plus, the project-wide workflow process should take into account the distributed function roles (Li et al, in press (2005)). The authors suggest to extend these geoprocessing workflows to automate the updating process of spatial data repositories and thus using appropriate scripts or applications which can invoke tasks to update distributed GIS data. The drawbacks named in the paper are the lack of specific target projects for which these models will serve.

Similarly, the common process of modeling spatial data and metadata could be made more effective and standardized by using workflow modeling. In the next step of SDI, the workflow modeling could help in the integrated metadata-spatial data environment to cope with

chaining of Web services. The workflow model might be used as guidance or metamodel to arrange which services are needed in which sequence for certain tasks. Extracting certain necessary metadata from the common data set according to a specific schema might ask for a planned combination of services which are chained in sequence.

3.2.4 Spatial Data Update

Spatial data management ensures, similar to metadata management, the data quality and actuality of spatial data. In this context, the process of the spatial data update is important. The possibility of "incremental update" is a keyword often used in the context of object-oriented spatial data management. It involves partially updating the data set. Only the modified elements of the data set is updated. This can be important in large data sets, e.g. land cover where not all information is dynamic and therefore only a subset of the data has to be sent for update purposes by acquisition organization to the client.

Spéry et al (2001) suggest in their paper that updates for cadastral databases can be facilitated and improved by using lineage metadata which document the changes within cadastral data. Their proposed model associates document sources, that generate cadastral update operations, with the description of changes and modified land parcels. They contend that the current situation in metadata management sets following obstacles for this procedure:

- Current metadata models give relatively limited information about individual object changes. The reason is that the metadata are often focused on the identification of geographical objects. These metadata are usually acquired once during the production process.
- Current applications do not take the temporal changes of cadastral data into account. The changes of cadastral data between the successive versions are not documented.
- The metadata lineage is created once and thus does not document the changes of cadastral data.

This approach would be facilitated and supported by managing metadata and spatial data commonly, as the lineage would be updated automatically together when the spatial data is updated. Nevertheless, this is an argument for setting the lineage for object-level and not on data set level as especially in cadastral surveying different objects can be updated in different cycles. The idea of using lineage for documenting the change is discussed again in this thesis for the prototype implementation when it comes to create views out of integrated data sets (Section 6.2.4).

3.3 Summary

Chapter 3 provides and insight in the current situation in research, especially concerning metadata management.

Different approaches have analyzed the possibility to **acquire metadata** automatically by deriving information from existing spatial data sets. Consequently, this metadata information does not need to be generated when metadata and spatial data are managed commonly.

Metadata acquisition tools are only partly capable of supporting this computation of metadata from spatial data as each supports different metadata schemas, without a common interface. **Metadata management** is widely acknowledged as an essential factor in storing and providing consistent, high quality metadata. Nevertheless, little research exists in the area of metadata management research. Some of the current research is focused on Web-based solutions, such as enhancing of HTML pages to carry metadata. These spatial metadata use georeferencing in the content which is hidden in tags. The heterogeneous and dense information in the Web makes this approach complex and extensive.

Other approaches offer URL links between existing metadata on the Web and spatial data which are stored on servers. This affords large effort and manual work in updating metadata, spatial data as well as the links. Currently, the solutions for new metadata management concepts do not improve metadata quality.

However, with the advance of Web-based solutions for spatial data search, exchange and processing, the requirements for metadata and efficient metadata management are growing. Vertical and horizontal **interoperability** in a network environment, between different spatial data sets is only possible with adequate metadata. The interoperability between different metadata sets is constricted by the heterogeneity of metadata standards in structure as well as semantics. Crosswalks link the fields in one specific standard to the corresponding in another standard (one-to-one mapping). The automatic translation between different metadata schemas by using crosswalks and a processing software is an error-prone task.

For the current area of **research concerning SDIs**, metadata need to be readable and interpretable by humans on the one hand and by Web services that index, search and access information on the other hand. In order to fulfill these high requirements, metadata must be up-to-date, reliable and consistent.

This affords concise and clearly defined processes for planning, managing and updating metadata and an organizational or technical link to similar workflows for the corresponding spatial data. There are approaches for implementing **workflow modeling** to support data quality inspection by using a workflow management system. This is supposed to be extended for updating processes in a distributed environment and might be useful for spatial data management processes.

Research in the area of **incremental update for object-oriented spatial data management** has identified the current metadata management as one of the obstacles for the approach. It does not enable current metadata to give sufficient information about changes on object level, the metadata lineage is not updated frequently and temporal changes in cadastral models are not taken into account.

Chapter 4   Concept of Integration and Views

The previously mentioned problems in Section 1.2 and 1.3 have shown that new concepts for metadata management are necessary. The ideal workflow suggested by this research to support an interoperable environment needs the consecutive realization of two concepts: First metadata and spatial data must be managed commonly according to the concept of integration. This involves creating common metadata-spatial data models and common metadata-data sets (Section 4.2). The second concept uses these extended data sets and models to support the flexible creation of user-defined views (Section 4.3). Views are virtual tables, that are linked to the real, physical database table by queries or mapping. Applied to the integrated metadata-spatial data set, the aim is to realize the liberty of creating metadata according to different standards and various excerpts of data formats without changing the original, integrated data set. In the ideal case this would happen on the conceptual level. Both concepts use semantic mapping as a common method. Therefore, Section 4.1 introduces the basic steps of semantic mapping and how it is related to creating views.

4.1 Introduction to Semantic Mapping

The process of semantic mapping is important for both the concepts of metadata-spatial data integration as well as views. Creating views is a special form of semantic mapping. It uses the same steps as semantic mapping, but focuses on restructuring an excerpt of model and data (**Figure 4-1**).
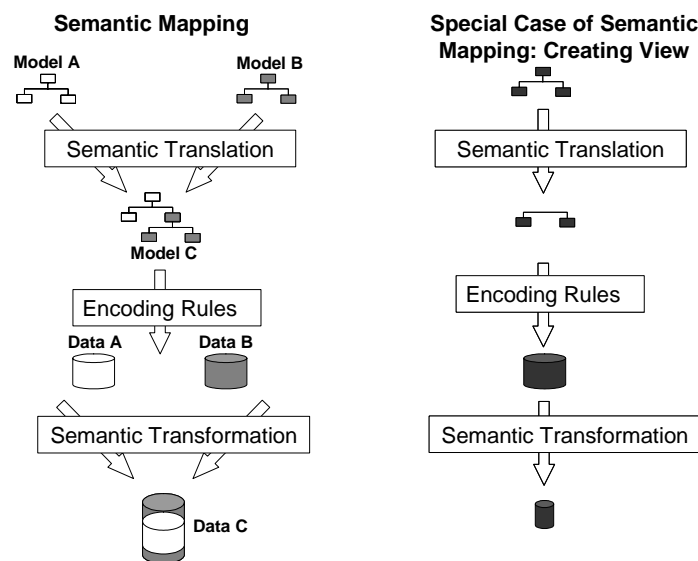

Figure 4-1: Common steps of semantic mapping and creating views

As mentioned in Section 1.1, semantics deal with inherent meaning of objects. If data sets from different information communities, countries, cultures or languages are integrated it is necessary to translate not only the name but also the meaning of certain topics.

Semantic mapping consists of two steps (**Figure 4-1**):

    (1) Semantic Translation and

    (2) Semantic Transformation

When two spatial data models are integrated, the corresponding tables/classes as well as attributes have to be identified and translated (**semantic translation**). The result are a common model and encoding rules, which define how to map one element from one model to another (**Figure 4-1**). Therefore, the precondition for integration of metadata and spatial data as well of creating rules is the existence of conceptual models which are described by formal description language (modeling language).

Then, the data is restructured according to the common model using the encoding rules (**semantic transformation**).

The encoding rules are generalized and broadened crosswalks. Crosswalks were defined in Section 3.1.3 as tables that map relationships and equivalences between two or more metadata schemas. Encoding rules are not specialized on metadata schemas but define how to map ay models for integration.

The use of semantic mapping can be demonstrated with Switzerland as an example. In Switzerland, according to the statistics of December 2004, there is no cadastral survey yet available for 13% of the country's territory (Swisstopo, 2005). In areas where there is no digital cadastral data coverage, topographic data is sometimes used. Nevertheless, since Switzerland has a nationally standardized cadastral model, this topographic data has to be restructured to fit to the cadastral structure and interface. For this, semantic mapping was used to combine the cadastral and the topographic model for certain regions and the topographic data was used as temporary information (Balanche, 2005). In order to do so, it was necessary to identify how the elements of the topographic data relate to the cadastral data and how the topographic model can be translated to the cadastral model. Lastly, the actual topographic data were transformed according to the new structure of the cadastral model, thereby filling in for the missing cadastral data.

The characteristic of semantic mapping is that two models and their corresponding data sets from the same extract of reality, in this case the same region of Switzerland, are combined.

Creating views is a special case of semantic mapping (Figure 4-2). In order to create views the same steps of semantic translation and transformation are used to change a given model and its data set to agree with another, new model. This new model describes an excerpt of the original data set in a restructured way. The resulting data view is therefore a manipulated subset of the original data.
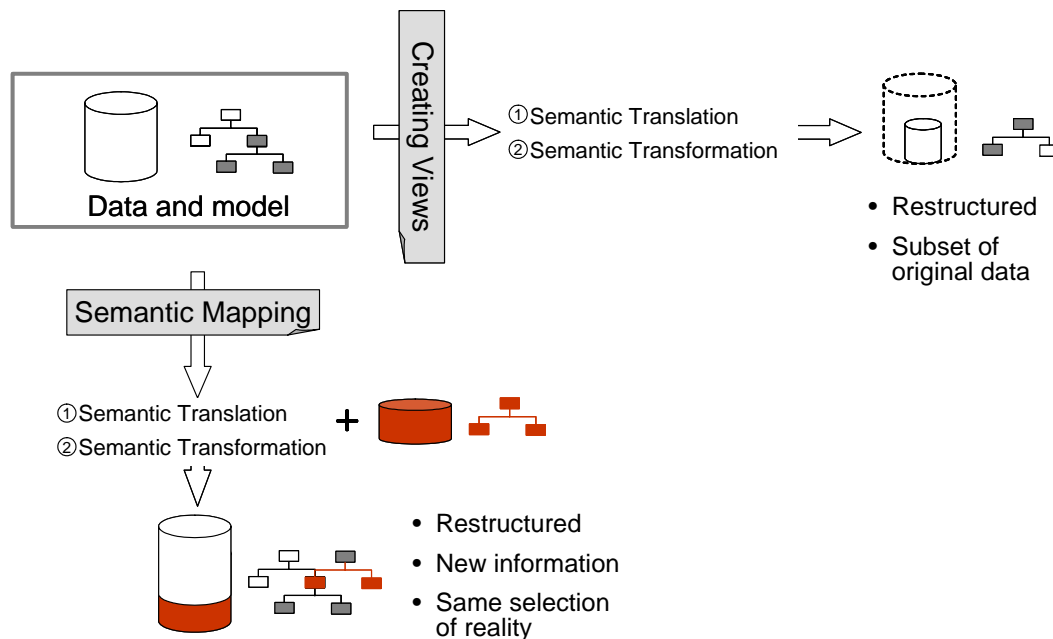
Figure 4-2: Comparison between semantic mapping and creating views

For common management of spatial data and metadata, the first task is to integrate the metadata and spatial model and data sets or databases. For this, semantic mapping is needed. The further use of the comprehensive data-metadata in an interoperable environment is supported by creating views.

4.2 Concept of Metadata-Spatial Data Integration

This section introduces the concept of common management of spatial data and its corresponding metadata in a single database or file. Consequently, spatial data will carry their own (metadata) description inherently with them.

The concept distinguishes between the approach of using already existing spatial data sets and newly planned, not yet existing spatial data.

The process and steps of integration are discussed in Section 4.2.1 using the method of semantic mapping between a given metadata model (prescribed by a metadata standard) and the spatial data model. Section 4.2.2 analyzes the structures and hierarchies of conceptual models and the consequence for the integration.

In order to achieve this 'self-describing spatial data set' there are two possibilities for the integration (Najar et al, in process):

1) Metadata are part of the data set from the beginning. This means the planning, modeling, acquiring of spatial data and metadata are a common process, e.g. in a workflow.

2) Metadata are added to existing data sets.

Modeling and acquiring metadata and spatial data commonly from the beginning, as suggested in point one, is the ideal goal that should be achieved in future. It creates a homogeneous data set carrying its own metadata. In order to find suitable strategies how to attain this ideal situation, option number two is used and the process as well as the consequences of metadata-spatial data integration for existing data sets (see also Section 5.1) is investigated.

One of the preconditions of the integration is to keep the data set as small, and its structure as simple as possible. In order to do so, only those metadata attributes are used for the integration, which provide information that is not yet present in any other way in the data set (Figure 4-3). These required metadata are chosen from a standardized metadata guideline, which is symbolized by the Christmas tree ornaments in the figure.
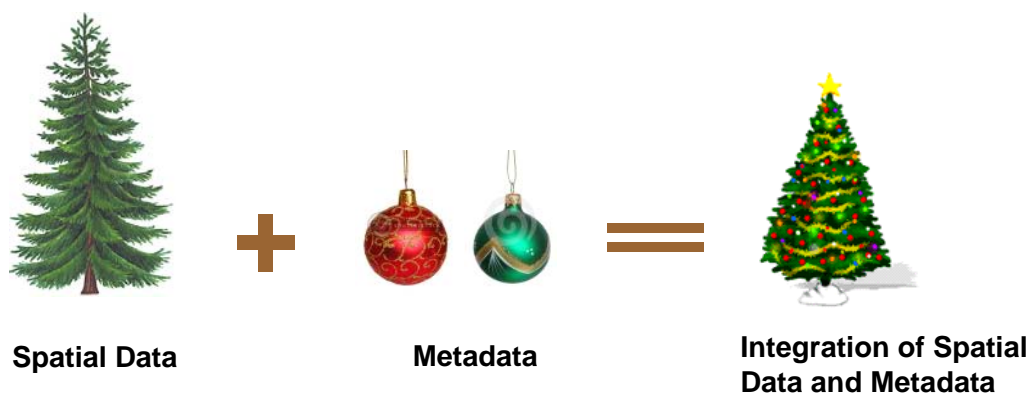

Spatial Data   **Metadata**   **Integration of Spatial Data and Metadata**

Figure 4-3: Integration of a certain, necessary group of metadata

In the case of integrating metadata in an existing spatial data set, different categories of metadata integration are considered:

1) Metadata derivable from the spatial data set

There are metadata attributes or values which may be derived explicitly or implicitly from spatial data ((Balfanz, 2002a), (Hill et al, 1999)).

Explicitly means that the value or attribute is not changed but can be copied or used directly, whereas the **implicit** method changes the model hierarchy or calculates new values for metadata attributes (e.g. the bounding box may be calculated implicitly by given coordinates and the date of actuality might be used directly, explicitly).

2) New Metadata attributes and their information

There are metadata attributes that do not exist in any form yet in the spatial data set. These must be newly inserted into the model and later in the data set (e.g. the information which metadata standard or profile has been used for the metadata acquisition)

4.2.1   Principles of Integration

It is important to be aware of two basic methods of integrating attributes into a model: the top-down and the bottom-up principle.

The Top-Down Principle (Figure 4-4)

Metadata are inserted as high in the model hierarchy as necessary and possible. This principle is applied to metadata of general relevance to the model. For instance, the name and contact address of the custodian, which is valid for the whole data set. It is assumed that such metadata are relevant for all levels of hierarchy of the model. Depending on the modeling paradigm this can be implemented in different ways. In object-oriented modeling this might be solved by inheritance: the general metadata are inherited to lower hierarchy structures.
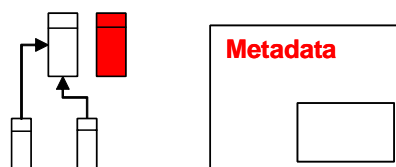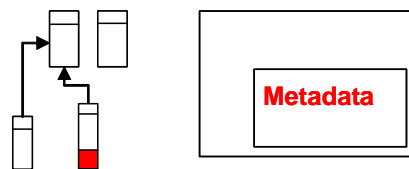

Figure 4-4: Top-down principle

The Bottom-Up Principle (Figure 4-5)

Certain metadata are placed at the lowest level possible so that e.g. information about accuracy, date of acquisition and completeness of the data describes specific classes/entities and not a whole series. Being able to describe small groups or layers of data may be of relevance especially in data sets with heterogeneous data. For example environmental measurements with different units and different accuracy, e.g. ozone in the air which is measured at different, discrete points in the city and is then interpolated over an area. In order to visualize the critical areas the ozone information is overlaid with an exact topographic map. Consequently, there are two layers in the data set with different accuracies and acquisitions methods. Another motivation for the bottom-up principle is given by (Lemmens et al, 2003) who stated for services that "… the level of connectivity that we can anticipate on, depends on the availability of meta information linked to a dataset and operation at those levels". ISO 19115 states in a similar context of Clearinghouses "….by creating several levels of abstraction, a linked hierarchy can assist in filtering or targeting user queries to the requested level of detail (ISO/TC211, 2003)". This way search engines and services could retrieve not only whole data sets but also specific groups or layers of spatial data and thus process only the relevant, interesting information.

Figure 4-5: Bottom-up principle

In summary, the top-down principle comprises the approach of seeking the "common-denominator". This includes finding groups of spatial data that have common characteristics and therefore need common metadata description. On the other hand, the bottom-up approach searches for the groups/layers/attributes of data that are 'different' and therefore need different metadata description than the general top-down metadata. The combination of both principles provides a balance: general, mandatory metadata attributes, e.g. the language of the data set, which applies to the whole model, must be inserted according to the top-down principle. Whereas, specific metadata attributes that only concern e.g. a class, or object, or attribute should be inserted according to the bottom-up principle (e.g. quality information).

4.2.2   Steps of Integration

The procedure needed for the realization of the concept 'integration of metadata and spatial data' requires a metadata standard which defines a metadata model and a data dictionary. A data dictionary is at least a table with the names of the existing attributes and their description. If the integration is applied on an existing data set, a spatial data model which is described in a formal description method needs to be provided. If the data set is going to be newly modeled, a catalogue of requirement is mandatory. The latter relates the content as well as which data and how the data are going to be acquired. Also, for what purpose the new data are going to be used and which queries are going to be possible.

For both cases of integration (new spatial data and existing spatial data), there are three steps of integration (Figure 4-6), of which the first differs depending whether one is dealing with new data or an existing data.

(1) Semantic Analysis and Translation (Step 1, Figure 4-6)

The semantic analysis and translation of the models answers the question "Which attributes are necessary?". The result is a group of required metadata attributes.

In order to answer this question, the element definitions, semantics and structure of both models are compared and similar notions or definitions are related to each other. Thus, the metadata model is translated to match the spatial data model or catalogue of requirement.
For an existing data set:

As mentioned, the first question is what metadata are needed to describe the spatial data set. Simultaneously, the challenge is met to analyze which spatial data can be used as metadata implicitly or explicitly and which information and algorithm is needed to derive certain attributes.

For newly modeled data set only the question which metadata are needed to describe the data set is relevant.

(2) Structural Analysis and Hierarchical Integration (Step 2, Figure 4-6)

The structural analysis and hierarchical integration answers the question where the selected metadata attributes should be inserted in the spatial data model. The result is a guideline (encoding rule) for every metadata attribute where it is placed in an extended metadata-spatial data model.

The chosen metadata attributes from step one are grouped thematically and placed strategically into the model. For this, a combination of the top-down and bottom-up principle is used: the general metadata which are valid for the whole data set are inserted according to the top-down principle. Metadata which label certain parts of the data set are inserted at the bottom. As a model is mostly grouped so that objects of the same topic are together, this provides a basis for the metadata labeling.

(3) Semantic Transformation (Step 3, Figure 4-6)

Step three moves from the model point of view to the actual data. The question is where to place the metadata in the concrete e.g. data base according to the extended model and encoding rules. The result is a database or spatial data file with spatial data and metadata.

The extended model is now used as pattern. The encoding rules are used as instructions which metadata information must be put where in the e.g. spatial database.
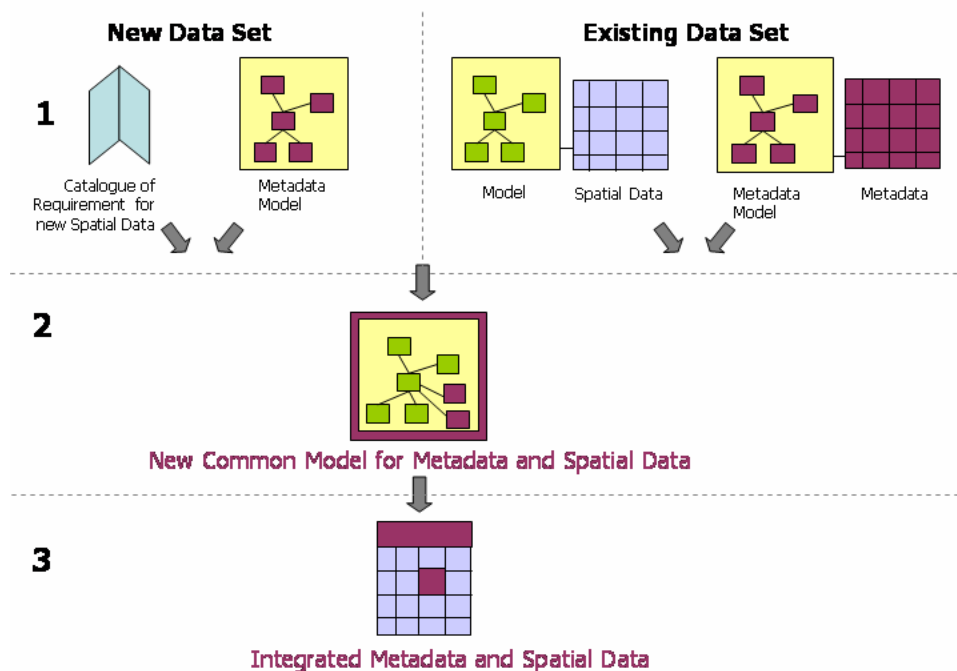
Figure 4-6: Steps for extending existing data sets to carry metadata

Altogether, the three steps of integration result in the semantic mapping between a metadata model and a spatial data model. The first step of semantic mapping (Semantic Translation) equals step one and two of the integration (Semantic Analysis and Translation + Structural Analysis and Hierarchical Integration). The second step of semantic mapping is equivalent to the Semantic Transformation (step two to three) of the integration process.

4.3 Concept of Views

There have been efforts to install specialized spatial search engines, called Clearinghouses, in an enabling environment in the Internet . This gives users the opportunity to search for spatial data. It has also been mentioned that current Clearinghouses usually give a list of metadata as a result to a specific search query. Sometimes, it would be interesting for the user to receive additional information on the spatial data, e.g. an excerpt of a certain data layer. Yet on the other hand, it is important for the custodian, with regard to user rights, license restrictions and authenticity to offer this service without giving the user access to the whole dataset and without changing the original data set. Therefore, it is important to introduce views, e.g. in an SDI environment.

In the following section the existing concept of views is introduced from a new perspective: with respect to metadata and metadata-spatial data integration.

In Section 4.3.1 the general idea of views is explained using metadata examples. It also stresses the technical, database aspect of views, the fact that the original data is not changed when views are created. Section 4.3.2 applies the 'metadata-spatial data view concept' on the integrated metadata-spatial data sets.

4.3.1   Introduction

From a database perspective, views are virtual classes. Their data instances do not originate from a physically existing structure or class in the database, but are derived mathematically from other objects.

A possibility to take a look at the spatial data without need to access the spatial data itself is offered by different database management systems (DBMS) like Oracle (Oracle, 2005) and PostgreSQL (PostgreSQL, 2005) in combination with SQL. The DBMS enables the management of spatial data and SQL offers the possibility of creating virtual tables, called views. These views define which attribute of which table of the spatial data set can be seen and which not.

Our approach looks at the view concept from the conceptual perspective. It advances one step further by trying to interpret metadata as a view on the original data and its model (Giger and Najar, 2003).

Metadata can be used for different purposes and therefore in different formats, too. The idea of having metadata for laymen, specialists and computers would be supported by the view concept.

Metadata are needed in a network environment (SDI) for different tasks and can be considered from different points of view according to the person who is dealing with them. Figure 4-7 shows this in an example:

A data provider, e.g. a national mapping agency (in Switzerland called swisstopo, (Swisstopo, 2005)) offers and sells topographic data. As the custodian wants this data set to be known to a wider audience of clients, he creates a leaflet in the Web. Therefore, the **first view** of metadata is an easily readable, commercial advertisement sheet for laymen in the internet. For a more advanced user, possibly someone who wants to use spatial data for a GIS application this would not suffice. He or she needs more information about the accuracy, method of acquisition, language or specific layers in the data set. This can be provided in a detailed html description according to a certain standard schema. In Switzerland, this would be provided by the metadata catalogue geocat (Geocat, 2005) according to the Swiss ISO 19115 profile and is shown in Figure 4-7 as **view two.**

→ Leaflet for Beginners

→ For Advanced and GIS Users

View 1: Metadata for Layman

View 2: Metadata for Spatial Data Users

→ For Data Providers, Custodians, etc.

→ For Web Services, Clearinghouses, etc.

View 3: Metadata for Experts

View 4: Computer Processable Metadata

Figure 4-7: Four views of metadata with different purposes

View three (Figure 4-7) changes the point of view from a user to the metadata custodian and expert. In order to be able to create a metadata description similar to view two, a metadata standard with model is necessary. A specialized person works with metadata schemas in different modeling languages and with various tools, e.g. according to ISO 19115.

View four is from the perspective of an SDI. Web Services (e.g. a Clearinghouse search engine) need to be able to process the metadata in order to find the interesting keywords. Therefore, the metadata have to also exist in an ASCII/text-based format which is readable by computers.

Altogether, the same original metadata are needed in four different views. This does not mean we have to acquire the metadata four different times, but the metadata should be acquired once according to some standard which provides a model and then it should be possible to create different views out of this file or database.

Thus, the "View-Concept" supports a transfer from one original (meta-)data set, that is not changed, to different, derived views. These views can then be used for various purposes, e.g. in a SDI environment.

4.3.2 Views for Integrated Metadata-Spatial Data

By applying the View Concept on the Integrated Metadata-Spatial Data Integration Concept (Section 4.2), the original, unchanged set is not either data or metadata but both: the common metadata-spatial data set. This offers the advantage of a larger variety of possible views: either certain metadata or certain spatial data might be extracted. To illustrate this with examples metadata, might be extracted according to a needed standard or parts of standards (Core of ISO 19115) and spatial data might be needed in specific layers including its metadata, certain formats or within a certain bounding box (one state, instead of the whole country).

Formally, the procedure of creating views is shown in Figure 4-8. A given data set *i* (in our case with integrated metadata) is described by its model *I* with a certain formal description method *M1*, e.g. a modeling language. The goal is to restructure the data set *i* in such a way so that it corresponds to the new schema *O* which is described with the formal description method *M2*. This restructured data set is then called *o* and is the view of the original data set *i*. In order to do so, certain mapping or encoding rules must exist or be generated which define how to get from model *I* to model *O*. These encoding rules are ideally integrated in a viewing tool or service which can execute and realize the rules on the data set *i* and thus produce the view data set o (Giger and Najar, 2003).

The square and octagonal boxes help to translate/map from one model to the other. In the example of Figure 4-8, Model *I* would be the extended data-metadata model and the model *O* could be the metadata schema of FGDC. The encoding rules provide the information of how algebraic, logic or language elements of the formal description methods are mapped to physical representations (e.g. metadata of FGDC or file formats like *.dxf) (Giger and Najar, 2003).
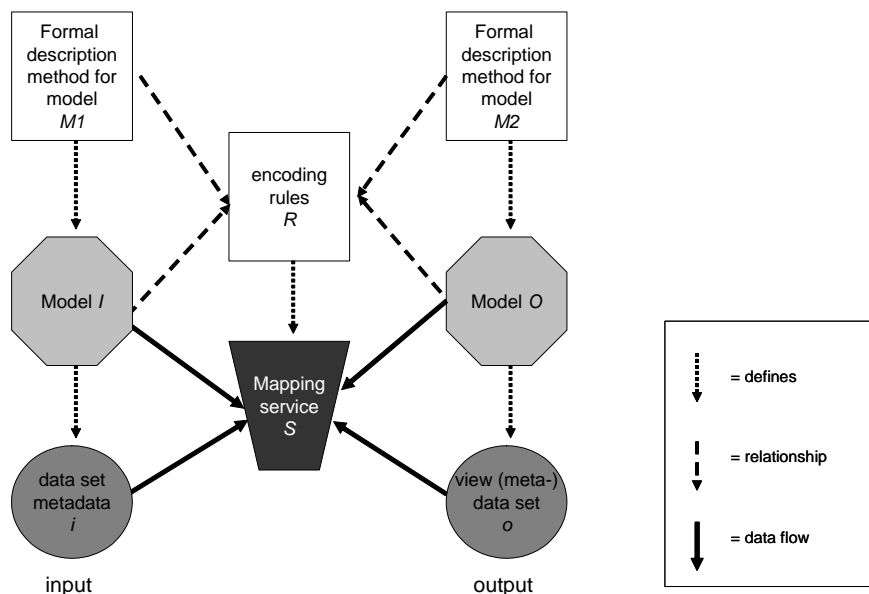
Figure 4-8: Theoretical process of producing a view for a given (meta-)data set

Figure 4-9 shows the fourth step of creating views, e.g. metadata according to FGDC standard or a certain excerpt of spatial data including its metadata for a certain bounding box. Step four follows the integration as described previously in Figure 4-6. It is possible that by generating views of different metadata schemas, certain spatial data are needed to
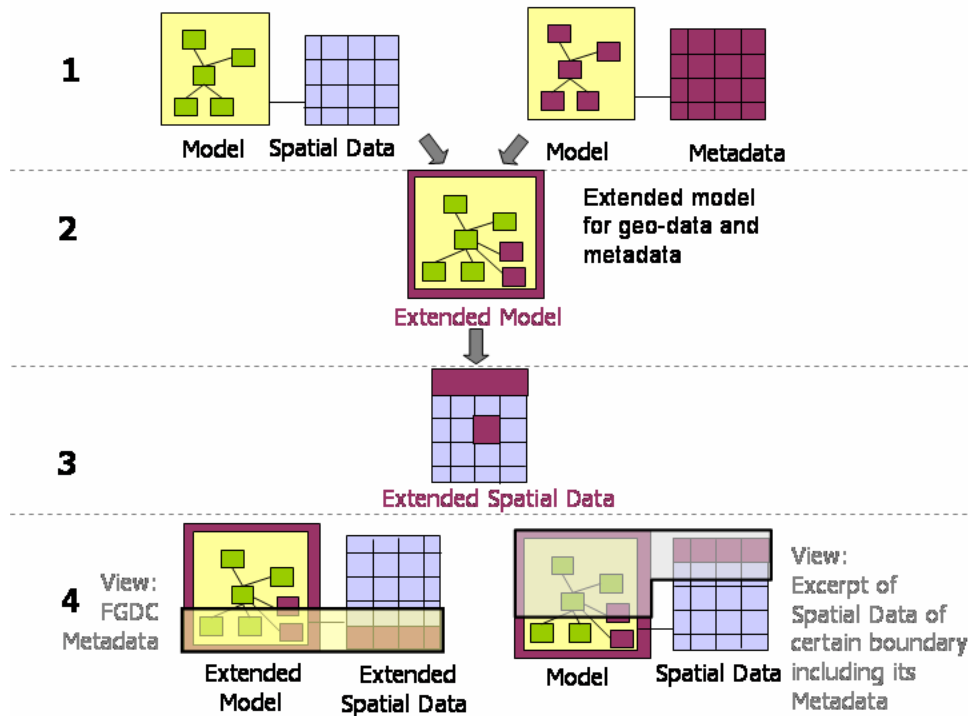
Figure 4-9: 4ᵗʰ Step to create views from integrated metadata-spatial data

In Section 6.3 the following, practical questions are going to be answered:

- Is it possible to create views according to standardized models?
- Is it possible to extract metadata attributes from the original data sets and structure these according to a certain metadata model?
- Is it possible to derive metadata attributes from spatial data in order to create views?

Altogether, the successive step after achieving the integration of data sets is to use the advantage of flexibility which comes with this management system. This is where the concept of views comes in.

The goal is to filter metadata according to different standards from an existing, integrated metadata and spatial-data set: Input one metadata standard and output flexible, various standards. The concept of views is used for this procedure. Different views according to different metadata schemas are defined. As mentioned in Section 3.1.3 the existing metadata standards originate from different distinct domains and thus cover different and structures. Accordingly, in a common metadata-spatial data set it might be necessary to use

spatial data attributes to create views for specific metadata schemas. This process is more complicated and would require an organizational effort if the metadata and spatial data set were managed separately.

4.4 Discussion in Context of SDI

In this section the possible implications of the concepts in an SDI environment of the future are discussed and whether certain implementations might change as a consequence.

In summary, the concept of common management of spatial data and metadata suggests that the metadata must be inherently connected to the spatial data. This would can provide flexible interpretations of what is data and what is metadata in future. It will also lead to only one process of spatial data and metadata acquisition, which would naturally produce consistent data sets. Consequently, a common metadata-spatial data set would achieve less redundancy, as only one update is needed. Another advantage of a common acquisition process is that metadata are generated as part of the spatial data process. As a consequence metadata will always exist and can be offered in an homogeneous way. Furthermore, the expert in charge of the spatial data is available to offer precise information for the metadata (e.g. about the acquisition method, accuracy, date etc.). A common metadata-spatial data management also enforces common update-cycles, which means that the metadata information is altered as soon as changes occur in the spatial data. It would no longer be necessary to keep track which information belongs to metadata and which to spatial data.

If metadata can be treated in the same data set as spatial data, the same tools should be applicable. Since special data processing tools are more advanced in development than metadata tools, which has been a consequence of higher value and priority of spatial data, the usability of metadata is increased if these same tools can be used for both.

An existing spatial data method, which has been used in DBMS up to now, is creating views. If this is applied on integrated data sets, the situation of interoperability can be improved in an SDI.

In Section 3.1.3 the use of crosswalks for one-to-one mapping between different metadata standards was elaborated on (Nogueras-Iso et al., 2004). Figure 4-10 shows how this approach is developed further for the view concept.
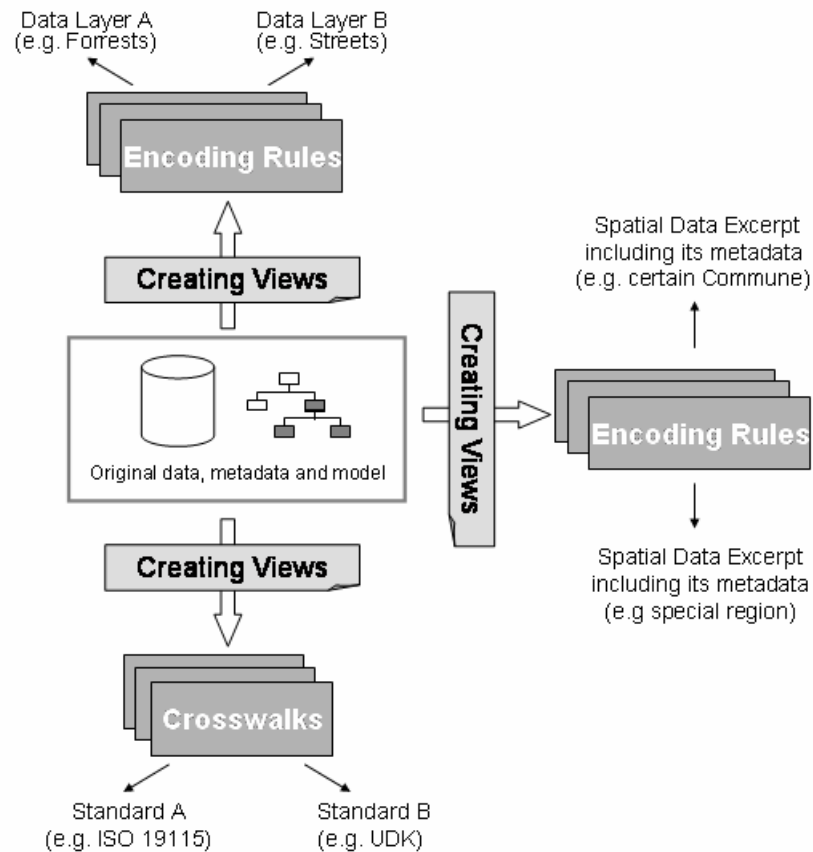
Figure 4-10: Context between crosswalks, views and metadata

The crosswalks correspond to one type of encoding rules in the view concept as they identify the correspond features and the structure between metadata standards. The other encoding rules create other views, e.g. data formats or extracts of the original data. The original data is not changed and the excerpts can be produced, e.g. according to specific user rights. A geoportal of an SDI might allocate these using different user profiles. A typical view would be metadata according to the standard which is required for a certain Clearinghouse.

As mentioned in Section 3.1.3 is currently a maze of different existing standards for different SDI components (Figure 4-11), which partly contradict each other (Golay et al, 2003).
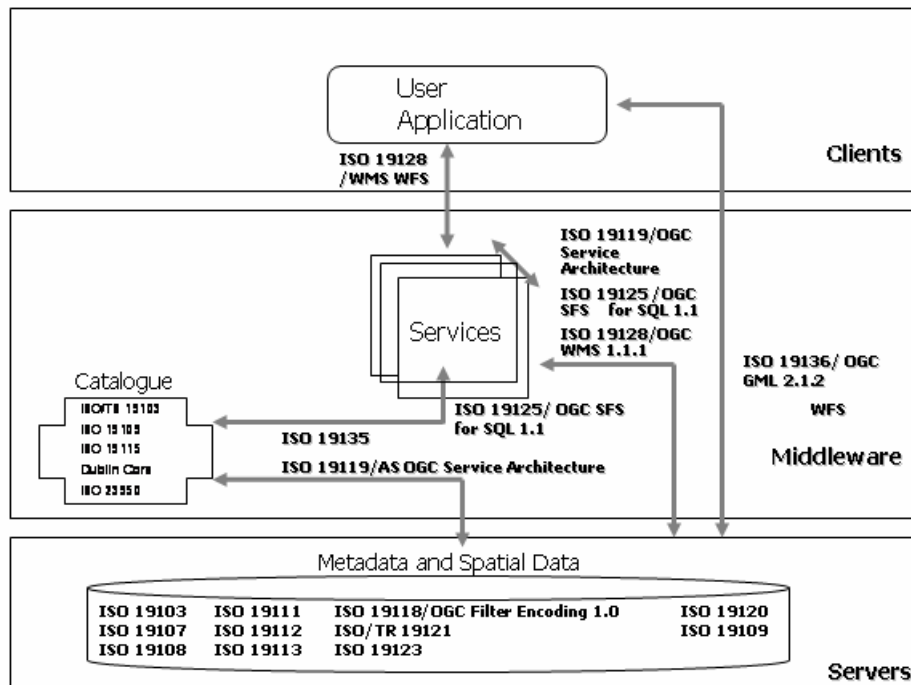
Figure 4-11: Overview of standards for different SDI components (Golay et al, 2003)

Current politics of SDI initiatives tend to oblige data custodians and service providers to use specific standards if they want to participate in a certain SDI. This standard often evolves and changes during time.

If interoperability between different metadata standards is achieved, it would be possible to liberalize the usage of different metadata standards, thereby considering them merely an interface for acquisition or mapping algorithm and thus offering view software to support the current application schema needed by a certain Clearinghouse.

This would mean that by implementing the view concept in a Web service, it might be possible to create for an Australian Clearinghouse the necessary ANZLIC standard and for a German, environmental Clearinghouse the UDK standard for metadata. Thus, the interoperability between the horizontal and the vertical level of SDIs might be facilitated.

4.5 Summary

Chapter 4 introduces the two relevant concepts of the research: the Concept of Integration and the Concept of Views for metadata-spatial data management.

The Concept of Integration involves creating common metadata-spatial data models and common metadata-data sets. The second concept uses these extended data sets and models to support the flexible creation of user-defined views.

By integrating, the first concept offers a common management of spatial data and its corresponding metadata in a single database or file. Consequently, spatial data will carry their own (metadata) description inherently with them. The integration distinguishes between the approach of using already existing spatial data sets and newly planned, not yet existing spatial data.

Views are virtual tables, that are linked to the real, physical database table by queries or mapping. Applied to the integrated metadata-spatial data set, the aim is to realize the flexibility of creating metadata according to different standards and various excerpts of data including their metadata without changing the original, integrated data set. To illustrate this with examples, metadata might be extracted according to a needed standard or parts of standards (Core of ISO 19115). In the ideal case this would happen on the conceptual level.

The process of semantic mapping is important for both the concepts of metadata-spatial data integration as well as views. Creating views is a special form of semantic mapping. It uses the same steps as semantic mapping, but focuses on restructuring an excerpt of model and data, whereas semantic mapping adds new information to model and data.

Chapter 5   Creating Principles for Metadata- Spatial Data Integration

The aim of this chapter is to review the feasibility of the Concept of Metadata-Spatial Data Model Integration that was introduced in the Section 4.2.

To illustrate, a case study (Section 5.1) focuses on three existing data sets and models, which are extended to carry their own metadata. The choice of the case study models is discussed (Section 5.1.1) and the procedure of the integration in general (for the models and the data, Section 5.1.2). Finally, the integration is analyzed for each single case study model (Section 5.1.3).

Section 5.2 discusses the overall result and feasibility of the integration procedure as well as outcome.

If the integration of metadata is possible, then the next step is to investigate whether general principles for any existing data set can be derived and whether these principles are also applicable in the workflow of newly modeled spatial data (Section 5.3). For this reason the three extended models from the case study are compared.[1]

5.1   Case Study with Existing Spatial Data Sets

In order to be able to test the hypothesis and the feasibility of the concept of integration of spatial data and metadata, the case study methodology is utilized. This enables an evaluation of the concept based on the following questions:

(1) Is it possible to integrate data that are acquired independently using a common model?

(2) Can the original data be extended in order to carry its own metadata on the basis of the common model.

(3) Is it possible to extend the data (values in the database or file) in an automatic way based on the common model?

---

[1] Certain information in the following sections is taken from a semester paper, which was realized in collaboration with and under the supervision of the author. For further details please check (Stopper, 2004)
.

5.1.1 Choice of Test Models

The scope of the case study is strongly dependent on the test data models. Here models of vector data are used. It is important to have a large variety of different models in complexity, size and which were created by different modeling experts for testing feasibility of concept. In order to attain a fair representation of diverse information communities the data sets should be of different topics. As shown in Section 2.5, there are many different modeling approaches and languages.

Table 5-1 characterizes the three chosen data models which are the Swiss cadastral model, the model for the water supply and the environmental model. The Swiss cadastral data model which is defined by law and therefore obliging for everyone dealing with cadastral data management in Switzerland, the model for the water supply of the city of Zurich, which is defined in a national standard and has the status of recommendation and the environmental model for Canton of St.Gallen, which is a local guideline.

These models differ in **complexity** (how many hierarchy levels exist) and **size**, which is measured with the amount of classes as well as attributes (Table 5-1). The "syntactic heterogeneity" is characterized according to (Bishr, 1998) by the different paradigms. Based on this syntactic heterogeneity, different modeling languages are used to describe the models which support relational or object-oriented paradigms (**modeling structure**). If national binding regulations exist for these models which provide a object catalog or a model description for a certain topic, than these models are considered **standardized**.

CRITERIA		DATA SETS		
Complexity		**Cadastral**	**Water Supply**	**Environmental**
	High	x		
	Medium		x	
	Low			x
Size				
	Large	x		
	Medium		x	
	Small			x
Modeling Structure				
	Relational	x	x	x
	Object-Oriented	x		
Modeling Language				
	INTERLIS1	x	x	x
	INTERLIS2	x		
	ERM	x	x	x
	UML	x		
Standardized				
	Yes	x	x	
	No			x

Table 5-1: Overview of the characteristics of the three test data models

These three existing test data sets with a wide spectrum of models are chosen to represent the procedures and difficulties that can occur during the integration process using any data set.

5.1.2 Procedure of Case Study

In Section 4.2, the general approach for the integration of spatial data and metadata is described. Three steps for the integration have been identified:

(1) Semantic Analysis and Translation which produces a group of necessary metadata attributes.

(2) Structural Analysis and Hierarchical Integration which defines where the metadata are placed in the data model and thus creates a common metadata– spatial data model and encoding rules.

(3) Semantic Mapping (Semantic Transformation) uses the common data model as a pattern and restructures the data and metadata information, e.g. in a database.

Now, these steps are used in the case study and thus tested practically on existing models. The test data sets have different models and separate metadata descriptions saved in individual files or databases. This metadata description consists of the metadata information itself and the standard, including model, according to which the metadata was acquired. In this case the Swiss Profile of ISO 19115, called GM03, has been used (Section 2.3.3).

Since two models exist, one for the spatial data and one for the metadata, it is necessary to decide which is selected to be the target model for the integration. Here, the spatial data model is utilized as basic, target model and the metadata model is altered to fit into it. Therefore the following steps, based on Section 4.2, have been executed:

Step 1: Semantic Analysis and Translation

First, the mandatory metadata attributes from the source model are added to the required group of metadata attributes for the target model. Then the conditional and optional attributes from, in our case GM03Core, are analyzed (step one in Figure 4-6) as well as the specialized metadata attributes for cadastre from the comprehensive section of the metadata standard. Therefore, the list of added metadata attributes consists of mandatory, optional and conditional attributes from GM03Core and specialized metadata attributes:

Group of Required Metadata Attributes $_{\text{target model}}$ =

$\text{GM03}^{\text{mandatory}} + f(\text{GM03}^{\text{optional}} + \text{GM03}^{\text{conditional}} + \text{Specialized Attributes})_{\text{target model}}$

This group of required metadata is then compared with the data model. The implicit and explicit attributes are identified.

Step 2: Structural Analysis and Hierarchical Integration

The group of required attributes are separated into metadata which are valid for the whole model, the 'top-down attributes', and metadata which can change with features within the model and groups, the 'bottom-up attributes'. This selection is not exclusive. A metadata attribute like 'title', which differs depending on the hierarchy level, whether the title of the whole data set or one of the classes/tables, can be used as a top-down and bottom-up attribute.

The difficulty of the structural analysis is the implicit assumption that all the attributes of the model are contained in the data set, too (Mustière et al, 2004). Often, the model expresses all the possibilities of a state of a data set. Yet, in some modeling languages, e.g. INTERLIS it is possible to define whether an attribute is mandatory or not. Those that are not marked with the keyword 'OPTIONAL' are compulsory for the data acquisition and therefore must exist in the data set.

Step3: Semantic Transformation

The ideal semantic transformation works on the conceptual level: it uses the common metadata-spatial data model as a pattern to restructure the information in the data files or database as well as add the metadata information in a specific place in the data set. For this step a tool for semantic reading and mapping would be necessary (Section 6.1). At the moment, semantic transformation can only be solved on the logical model level. Here, INTERLIS is the modeling language which is the easiest to work with as both the model and the data file are simple ASCII files which are easy to process and many useful tools for INTERLIS already exist. In the case study three possibilities were tested:

1. Manual manipulation of the INTERLIS data file (called INTERLIS Transfer (ITF) file (Section 2.5.1)

2. Import the data into a GIS software, in order to restructure and add the metadata on the database level.

3. Program the conversion in a specialized INTERLIS tool (INTERLIS Conversion System, (InfoGrips GmbH, 1994-2005))

The first possibility involves processing the data file in a text editor. In the case of the cadastral data set for whole Switzerland this file has 1774 pages (about 2800KB). Every tab-count has a meaning, so that programs like Microsoft Excel spreadsheet or similar software cannot be used for this processing.

When importing the data into a GIS software, e.g. GeoMedia by INTERGRAPH (Intergraph Corporation, 2005), the add-on tool for importing INTERLIS1 into internal Microsoft Access tables is crucial (GeosPro, (a/m/t software service AG, 2005)). The tables are stored in an internal structure so that altering the tables means working on the physical level. Another disadvantage is that GeosPro only exists for the relational INTERLIS1 model.

The third method uses the machine language iG/Script for configuring the program INTERLIS Conversion System (ICS). The restructuring of the data file is defined in two configuration files. The input is the old data model and the extended model as well as the data information files. The result is the restructured data file. The disadvantage is, that it is not possible to add new metadata into the newly restructured data set. Therefore, ICS fulfills the first step of restructuring the data file according to the common model, but the metadata have to be added manually (Figure 5-1) .
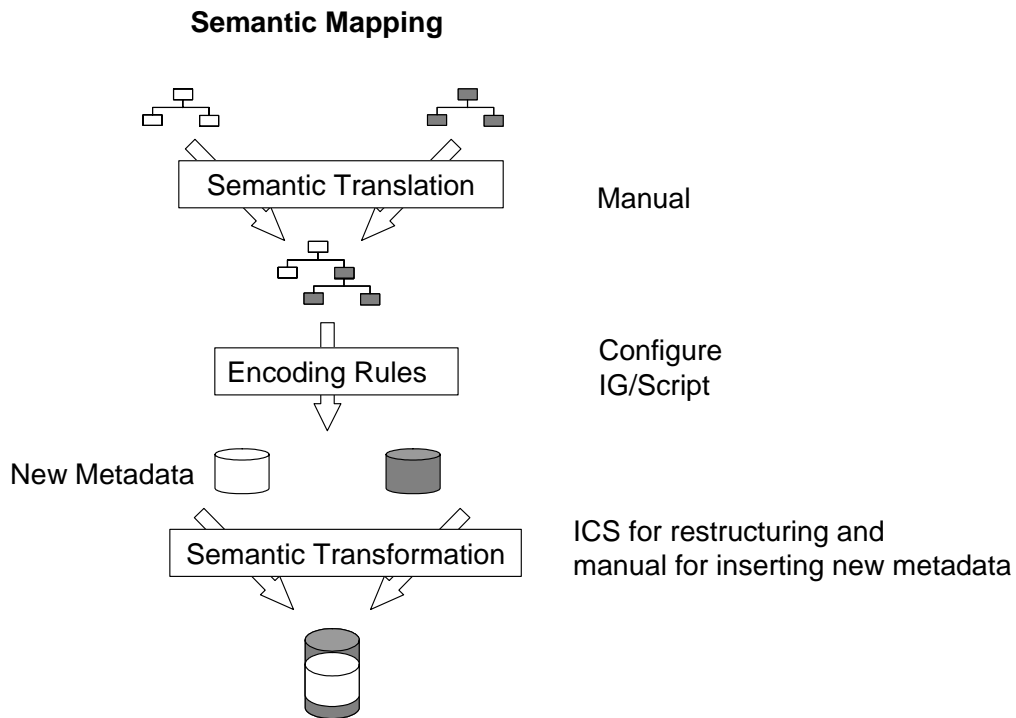
Semantic Mapping

Figure 5-1: Semantic mapping for combining existing spatial data with metadata

The following three sections will briefly introduce the three test data models and discuss the metadata integration.

The common metadata-spatial data models have been analyzed with experts in a workshop. Thereby, the usefulness of the integration was checked and the implicit as well as explicitly derived attributes have been verified. The workshop showed general approval towards the models.

5.1.3 Cadastral Data Model

The cadastral data model (called DM.01-AV-CH, version 23) and test data are freely accessible on the internet, (Cadastral Surveying, 2005) offered by the Swiss mapping agency (called swisstopo). The cadastral model is used at the moment only in relational INTERLIS 1 by the cantons and the federation. Notwithstanding, the unofficial (meaning not legally required model) can be found modeled in object-oriented INTERLIS 2 for testing and practice on the INTERLIS webpage {INTERLIS, 2005 #286}. Both sources offer the model in three languages: French, German and Italian. The content is basic information about the land coverage and cadastre. It consists of eight, independent thematic layers (Figure 5-2).

8 Information Layers:

Control points

Land cover

Miscellaneous objects

Heights

Names

Property

Pipelines

Administrative Boundaries
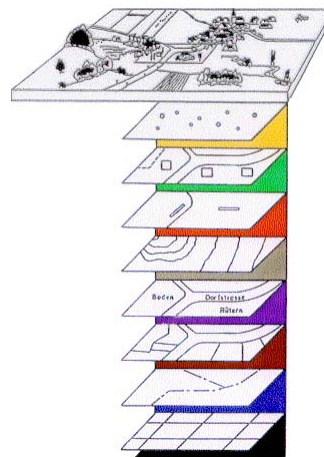
Figure 5-2: Eight independent information layers for the Swiss official cadastral surveying

As shown in Table 5-1 the Swiss cadastral model is a highly complex and large data model. It consists of more than 500 attributes in 19 INTERLIS1 TOPICS. These topics are on the one hand important for the modeling in the modeling language INTERLIS, but might also be considered as thematic groups.

Due to the size of the original model, **Figure 5-3** below only shows one of the INTERLIS topics concerning the control points (LFP).
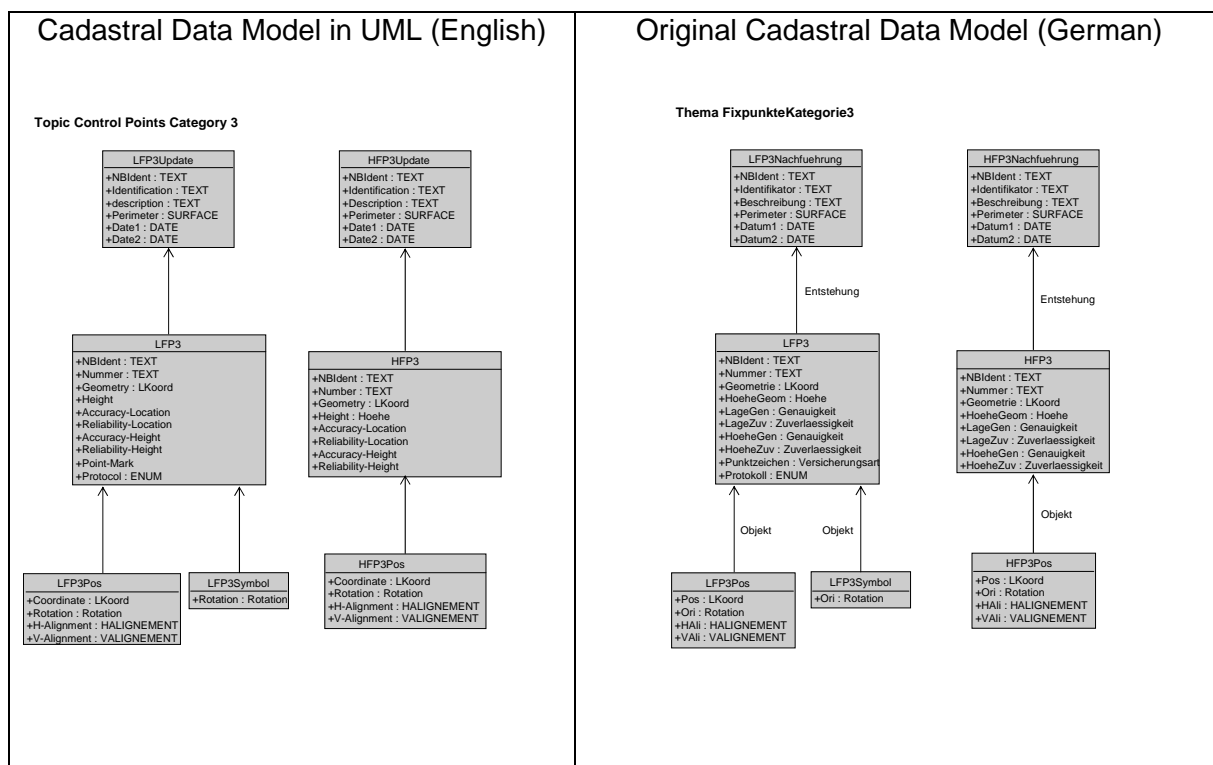

```
TOPIC FixpunkteKategorie3 =
  CLASS LFP3Nachfuehrung =
   ATTRIBUTE
     Identifikator: MANDATORY TEXT*12;
     Beschreibung: MANDATORY TEXT*30;
     Perimeter: SURFACE WITH (STRAIGHTS, ARCS) VERTEX LKoord
       WITHOUT OVERLAPS > 0.200;
     Datum1: MANDATORY INTERLIS.INTERLIS_1_DATE;
     Datum2: INTERLIS.INTERLIS_1_DATE;
  END LFP3Nachfuehrung;
  CLASS LFP3 =
   ATTRIBUTE
     Nummer: MANDATORY TEXT*12;
     Geometrie: MANDATORY LKoord;
     HoeheGeom: Hoehe;
     LageGen: MANDATORY Genauigkeit;
     LageZuv: MANDATORY Zuverlaessigkeit;
     HoeheGen: Genauigkeit;
     HoeheZuv: Zuverlaessigkeit;
     Begehbarkeit: MANDATORY (
       begehbar,
       nicht_begehbar);
  END LFP3;
  CLASS LFP3Pos =
   ATTRIBUTE
     Pos: MANDATORY LKoord;
     Ori: Rotation;
     HAli: INTERLIS.HALIGNMENT;
     VAli: INTERLIS.VALIGNMENT;
  END LFP3Pos;
  CLASS LFP3Symbol =
   ATTRIBUTE
     Ori: Rotation;
  END LFP3Symbol;
```

**Figure 5-3: Excerpt of original cadastral model (DM.01-AV-CH, Version 23);
TOPIC Control Points of 3rd Category (INTERLIS, 2005)
LFP3: control points of position / HFP3: control points in height**

As the case study has been realized with German models, the English names in the following sections are basic translations of the models. Figure 5-3 gives an impression of models in INTERLIS2 and UML as well as two natural languages German and English.

The UML description of the cadastral model in Figure 5-3 originates from the annotation for the cadastral model (Fachstelle AV-Datenmodellierung und -austausch (FADMA), 2001). It uses dependencies to connect certain classes to each other.

Figure 5-4 shows the top-down attributes which were chosen from GM03. The following areas are covered by the metadata description:

- Metadata Information
- Data Identification and Content Information
- Responsible Party
- Data Qualitiy
- Extent Information
- Identifier Information
- Reference System Information

The original structure of the GM03 model was preserved wherever possible. Some of the metadata can be derived from existing spatial data, such as the temporal extent, which is defined in GM03 as the time period covered by the content of the dataset: it is derived by using the attributes "date1" which exist in each of the fifteen tables called "*update" of the cadastral model. In order to not have fifteen couples of dates, a minimum-maximum algorithm that picks out the youngest and oldest date needs to be implemented.

Other implicit attributes were discussed with experts, but discarded for various reasons. For example the discussion whether the description of the spatial data "abstract" can be derived from the titles of all the tables belong to the model might be a good example. The disadvantage is that often abbreviations are used and therefore the description could be incoherent for non-experts.
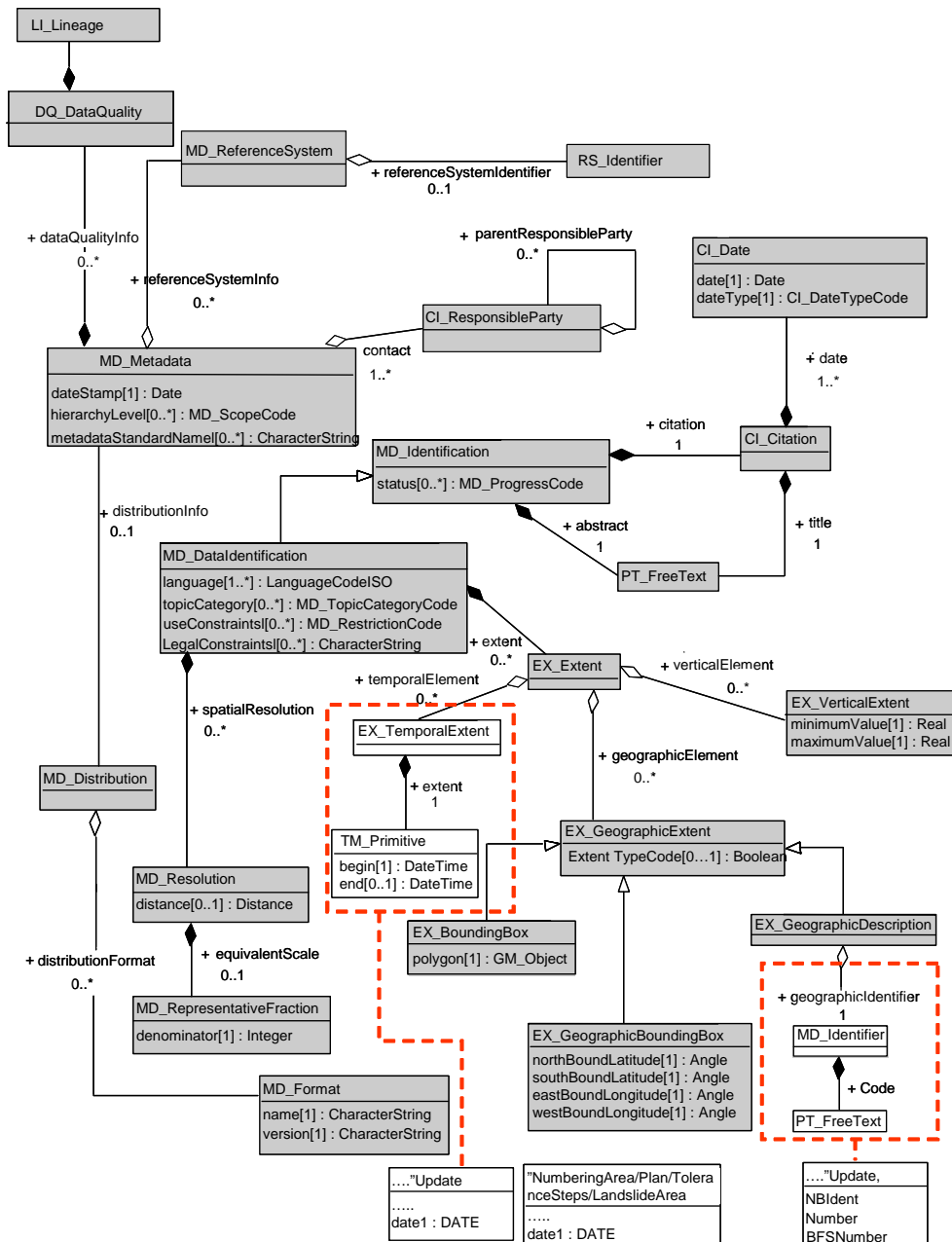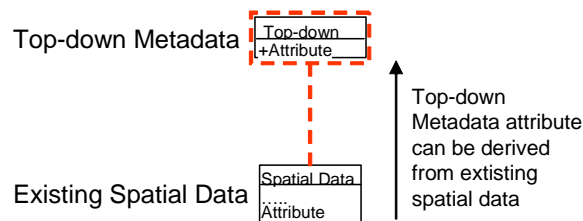
Figure 5-4: Top-down attributes for cadastral model with indication where implicit attributes can be used from the existing data model (according to GM03)

Even though the bounding box is one of the suggested metadata attributes which can be derived from spatial data according to (Balfanz, 2002b), there are examples that make the automatic calculation difficult. In the Swiss cadastral system, the enclaves and exclaves must be calculated separately.
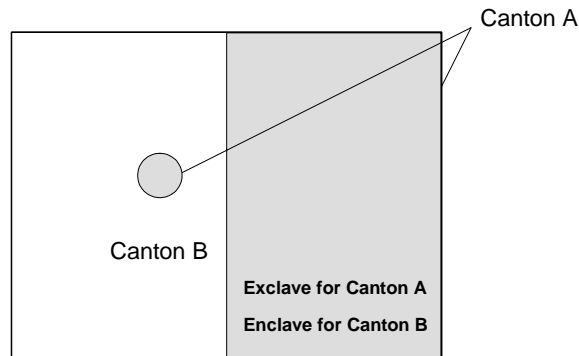
**Figure 5-5: Exclave and enclave in Swiss cadastral surveying
makes bounding boxes difficult to derive automatically**

The Swiss administrative borders have some particularities: enclaves and exclaves. If Canton B has an island of another Canton A within its borders, it is called enclave. From point of view of Canton A the same situation is called an exclave. The bounding box of canton A would be miscalculated (Figure 5-5).

If the maximum and minimum coordinates of a data set are computed and the bounding box is derived by using the extreme coordinate pairs, exclaves and enclaves must be excluded or a wrong result will be provided for the bounding box. A semi-automatic calculation by processing enclaves and exclaves manually might be possible.

Figure 5-6 shows the bottom-up attributes that were chosen from GM03 for the cadastral model on the example of the class of control points.

Legend:

Bottom-up Attribute: *HierarchyLevel: MD_ScopeCode*
Explicitly Derivable Attribute: **Date1 : DATE**
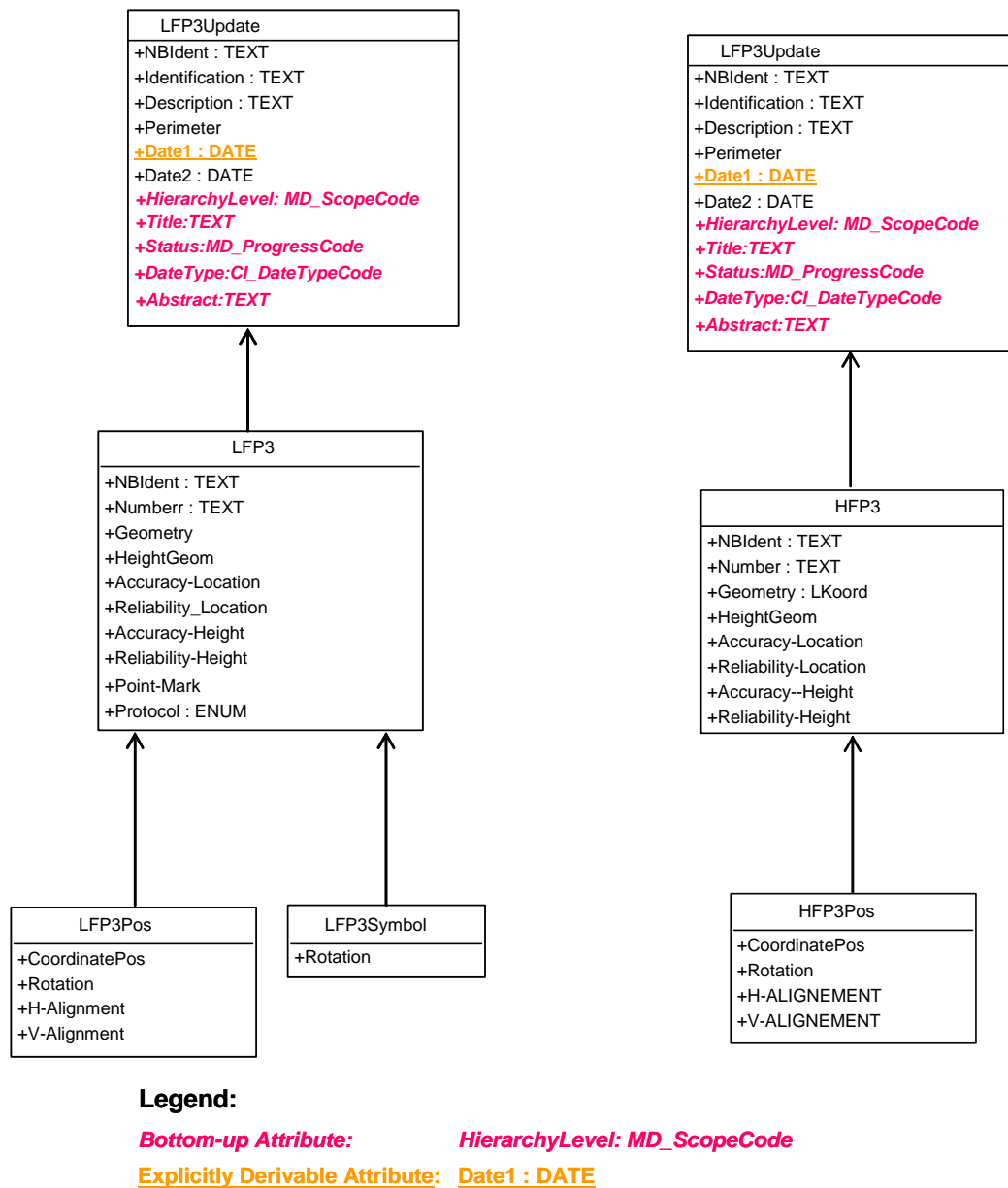
Figure 5-6: Bottom-up integration for cadastral model, object-oriented

The attribute date does not have be inserted as metadata attribute, as it exists already. It is used as the actuality of the spatial data.

The class 'LFPupdate' (LFP are control points of position) has even two attributes called date: there has been a lot of discussion with experts about date1 and date2 and which one is the correct date that should be used as metadata. The one is defined by the law to be entry in the technical dossier, while the other is defined by every canton individually and therefore varies in meaning. It was decided upon date1 to be the representative date of actuality.

In Swiss cadastral mapping there are two kinds of updates. Some objects are continuously updated and others periodically. For examples streets and buildings are updated continuously when changes occur, but the forest boundaries are only updated periodically. Therefore, the date depends on the instance of the object which is described. Accordingly, the metadata attribute must be placed on the lowest hierarchy level as a bottom-up attribute.

5.1.4 Water Management Data Model

The water supply model for the city of Zurich is modeled according to a Swiss standard ('SIA bulletin 2016 GEO 45 Data Exchange').[2] The model is described in ERM and INTERLIS1 and is considered a medium sized model with about 50 attributes in 8 INTERLIS1 TOPICs. Figure 5-7 shows the original model which is structured in three main component groups: the fresh water pipeline, the pipeline nodes and the facilities. Each of these main tables is provided with a labeling table which is called *Text (e.g. Pipeline_Text).
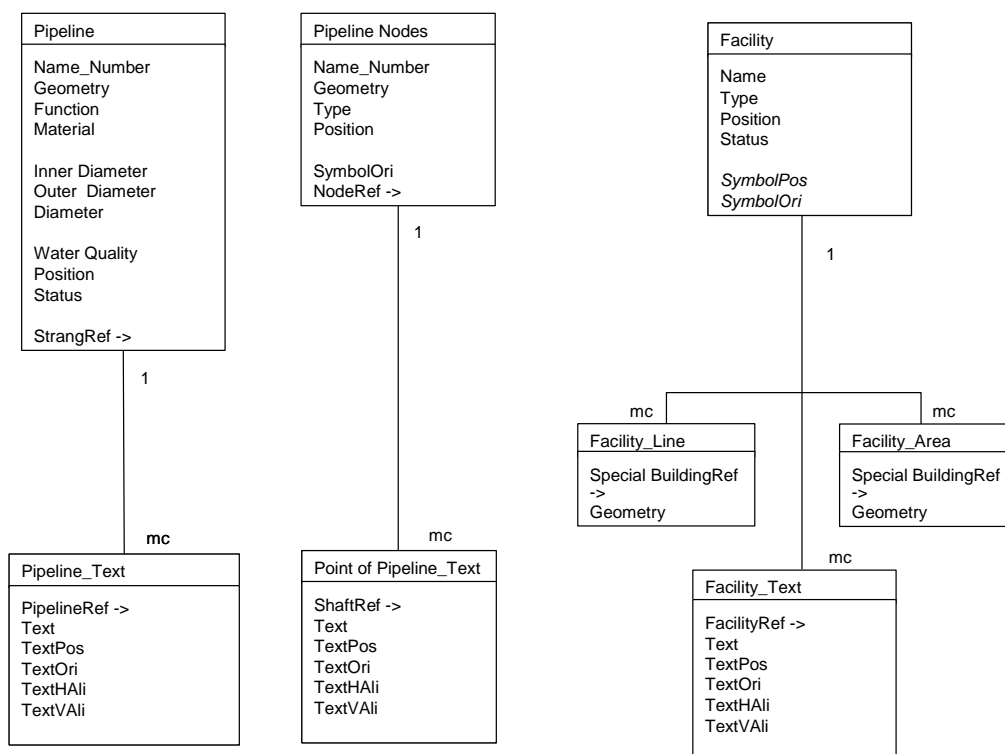
Figure 5-7: Original model of water supply Zurich (according to SN 500405) in entity-relationship model

---

[2] The water management model is used thanks to the permission of the administration of the city of Zurich

The top-down metadata are depicted in Figure 5-8 and are divided into information about the spatial data ('SpatialDataInfo'), the metadata ('MetadataInfo') as well as the contact details for the spatial data set ('Organisation' and 'ResponsibleParty') and the special information ('SpatialDataSpecInfo'), e.g. about the use constraints of the spatial data and update cycle. Since the titles of the original tables are easy to understand, plus the attribute 'Function' gives the information what kind of a pipeline is used, this information can be used for the abstract
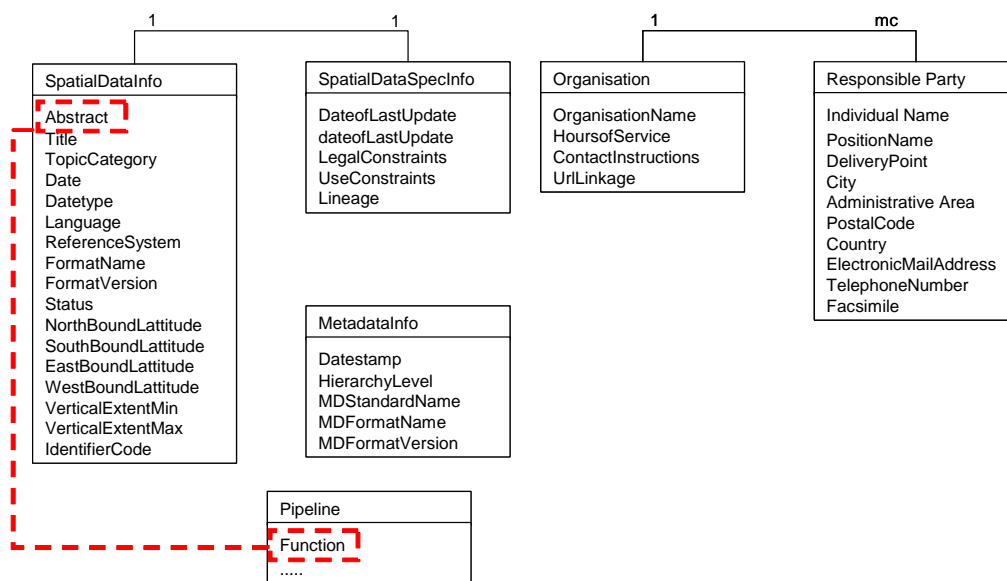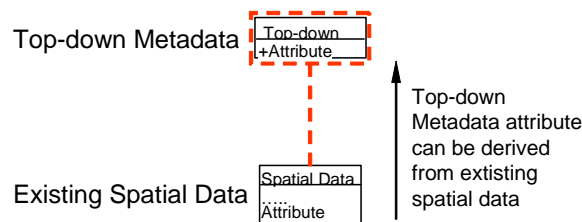
Figure 5-8: Top-down metadata attributes, entity relationship model

Furthermore, it is important for this model to have an attribute with legal and use constraints as the model is not prescribed by law as the cadastral model, but standardized only by a guideline.

The bottom-up attributes are shown in Figure 5-9. The red attributes denote the bottom-up metadata attributes and those spatial data attributes which can be used as metadata description.
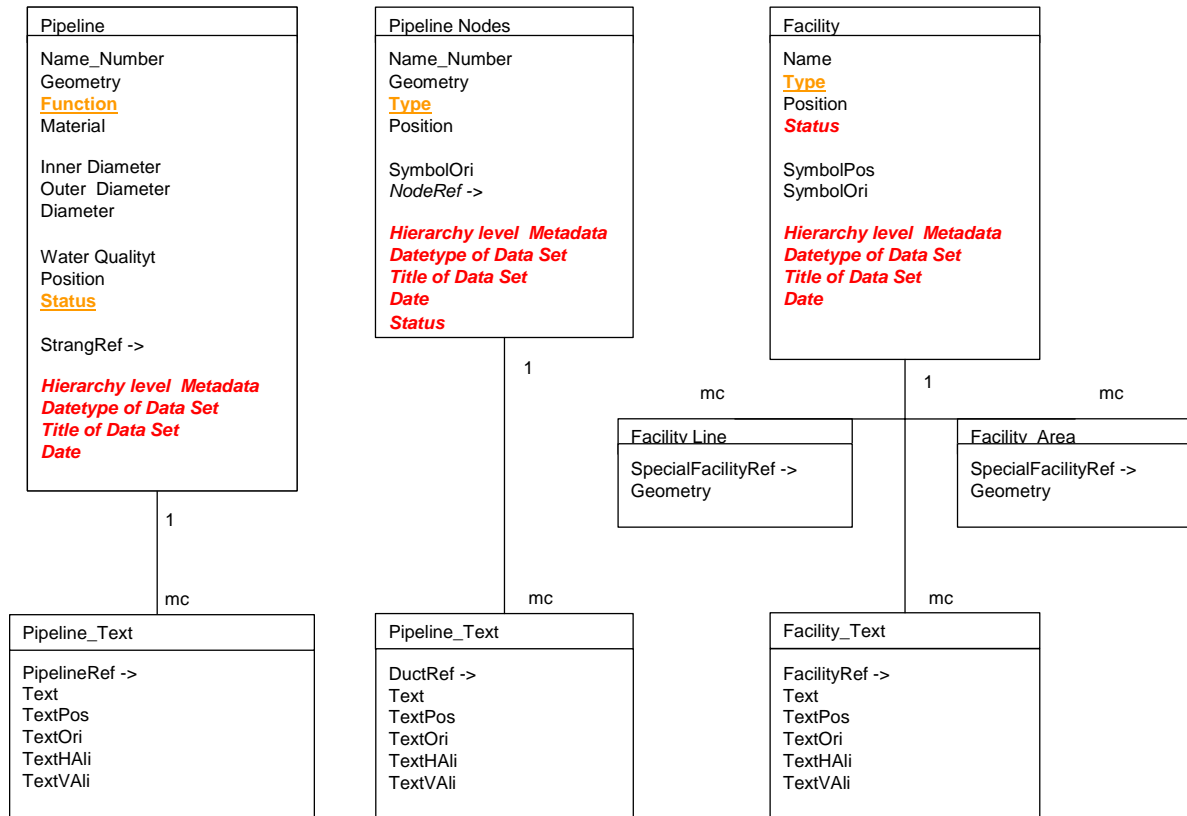
Legend:

Bottom-up Attribute: *Date*

Explicitly Derivable Attribute: Status

Figure 5-9: Bottom-up attributes for water management model of city of Zurich

The attribute 'status' (table Pipeline) is explicitly used as status of the spatial data set. The other bottom-up attributes are equivalent to those in the cadastral model. The attribute function (table Pipeline) is used as a short abstract for the entity as it contains the description which pipeline is used.

5.1.5 Environmental Data Model

The relational, environmental data model is the least complex as well as the smallest of the three models.[3] The model is not based on any standard but on information from the cadastral register and the Swiss Water Information System (Federal Office for Water and Geology (FOWG), 2005) which is administered by the Federal Office for Water and Geology.

The body of water is structured in segments, which have common ecological, political and morphological characteristics. The main table, called "WaterBodySegement", merely defines starting and ending distances through which it flows (in kilometers). The geometry and topology are determined by the table "Topology_Node" (Figure 5-10).
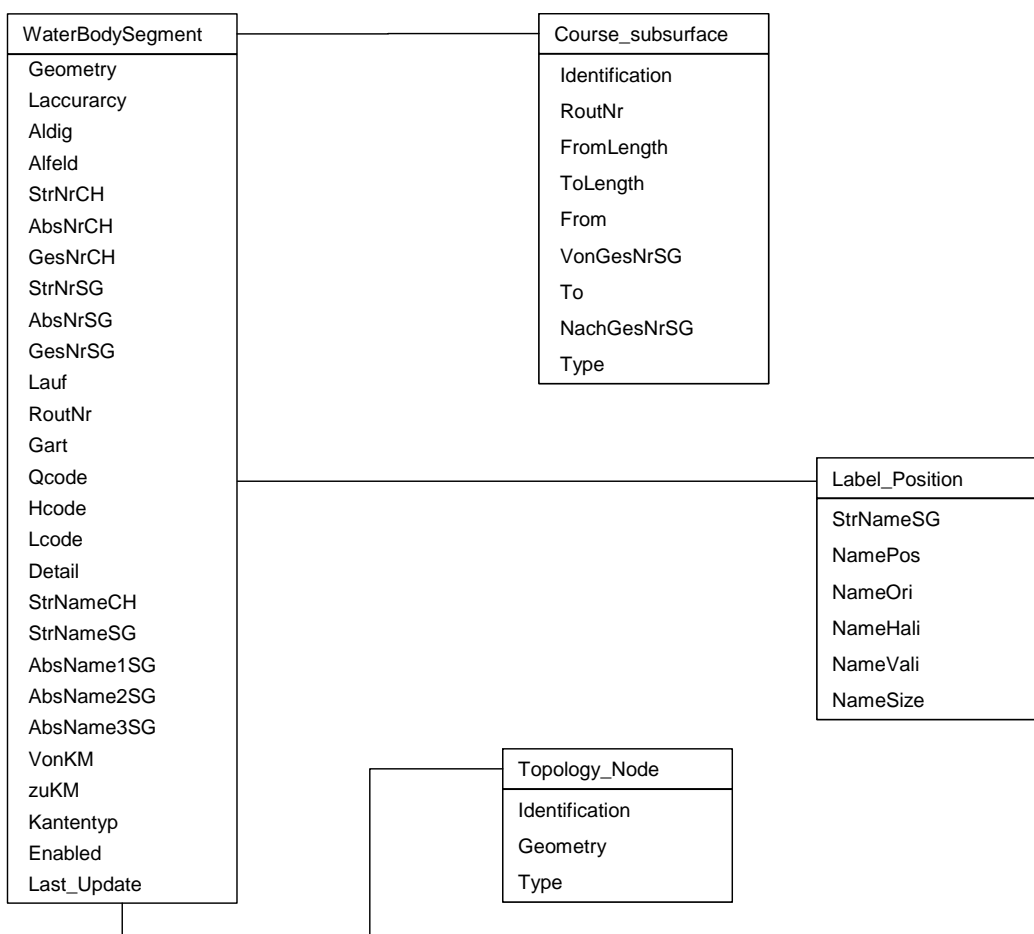
Figure 5-10: Original model for environmental data of Canton St.Gallen

---

[3] The environmental data model was provided by the Canton St.Gallen

The top-down metadata are similar to those used for the water management model, but the derived metadata attributes differ: The title of the tables are not abbreviated and understandable and self-explanatory. Therefore, they can be used together with the two attributes 'Gart' (typecode: creek, river, canal etc) 'StrNameCH' (name of section) to describe the content of the data set. So, in this case the top-down attribute 'abstract' (Figure 5-11) can be created by using the titles of the different tables. Also, the bottom-up attribute 'title' uses the information of the name of the table.
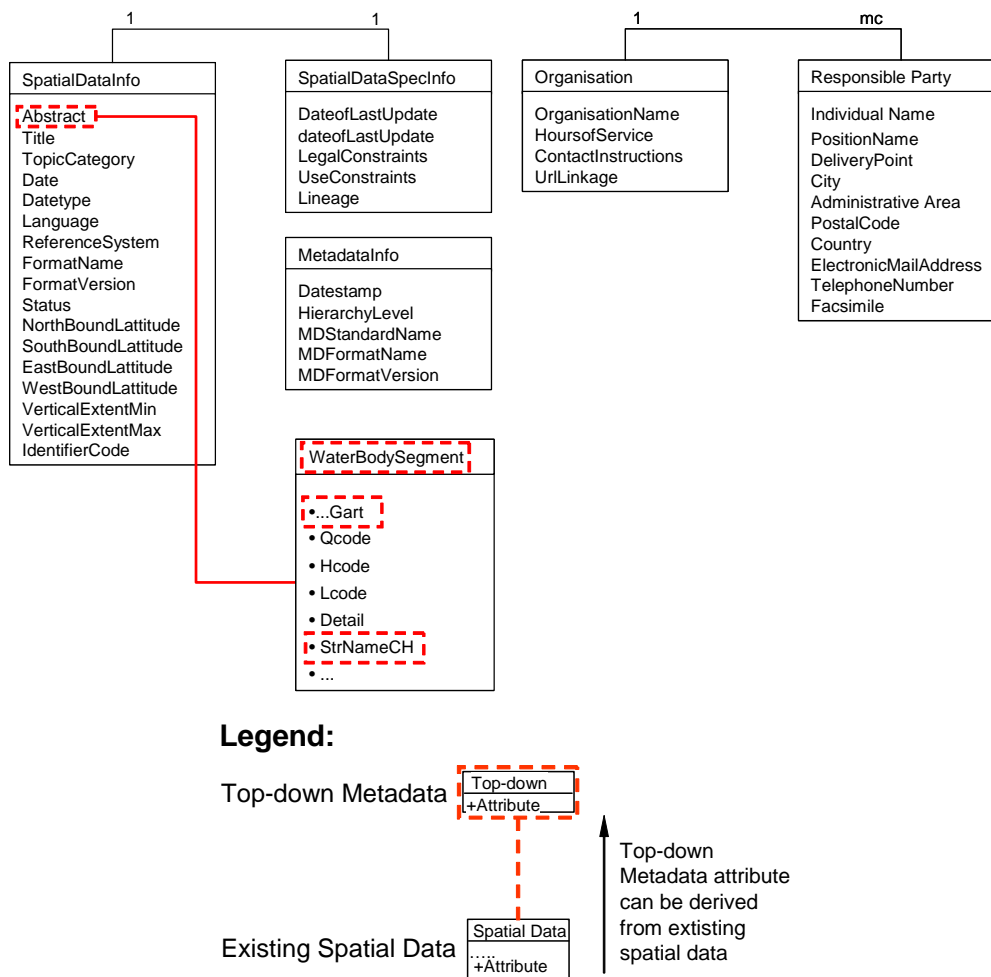
Figure 5-11: Top-down metadata derived from data set

In an expert workshop the possibility of using the attributes 'Geometry' for the bounding box and the attribute 'LastUpdate' for the date were discussed with the experts. It was agreed upon that both are not useful in this context. This shows once again how necessary the expert knowledge for understanding the semantics of a model is.

5.2   Discussion of Integration

In conclusion, it was possible to integrate metadata in one object-oriented cadastral model which was modeled in the languages INTERLIS2 and UML and in two relational models (in INTERLIS1 and ERM). The examples showed that the integration is feasible. In a workshop with experts who knew the original models well, the common models were verified and accepted with minor corrections.

Yet, different problems occurred during the case study and were solved in different ways. This will be discussed in the following:

(1)   **The result of the metadata-spatial data integration**
          depends on the characteristics of the modeling language.

The modeling languages have different properties (e.g. one uses inheritance and the other does not) and use different keywords to describe the data structure. The result of the integration, e.g. the level of detail of metadata description, depends on the interpretation and possibilities of the modeling language. Furthermore, it is important to understand the semantics and structure of the model in detail.

(2)   **The decision which metadata can be derived automatically**
          should take the characteristics of the model into account

It is possible to define in some modeling languages whether an attribute is optional or mandatory. Looking at only the model, it is not always clear whether this element actually exists in the database e.g. in the case of spatial attributes modeled to be "optional" (unless these attributes are explicitly checked in the database). Therefore, in process of the decision whether an metadata attribute is categorized as "automatically derived", preference should be given to mandatory spatial data attributes.

(3)   **The feasibility of the integration does not depend on the quality of a model,**
          nevertheless the resulting metadata description differs.

There is a great heterogeneity between the different models, depending whether a group of experienced modelers created them for a standard or whether the models are merely used for internal purposes.

It is not task of the integration process to improve the quality of the data model. Also, the process of integration and the result, managing metadata and spatial data commonly, is not obstructed by the different models.

Nevertheless, a strongly thematically structured model supports detailed metadata on the lower hierarchy levels in a better way: In contrary, the environmental data model offers a entity called "WaterBodySegment" with attributes of different nature (such as different names, length from km to km, quality, geometry etc.). The goal is to divide for example a river in different segments and each part is thoroughly described with this table. Yet, the consequence for the metadata description is that the metadata must describe a heterogeneous table in e.g. abstract and title. The metadata will therefore be less accurate than if the large table was divided into various simpler tables which offer different perspectives of the description of the same segment, e.g. geometry_segment (containing geometry, length, starting km and ending km, border type etc.), name_segment (containing the five different attributes for the name), identification_segment (route number, source description etc.) and waterbody_characteristics_segment (describing the watermanagement characteristics).

The object-oriented modeling, by definition, uses classes to describe objects with common characteristics. It tends to support the goal of a detailed metadata description in a more concise way. The reason is that the danger of large heterogeneous groups of spatial data is rather uncommon in object-oriented modeling.

(4)   **Standardized models support the automatic metadata integration and make the integration reproducible**.

A standardized model, which has been enforced by law for some time, like the Swiss cadastral model, is well thought-through and tested in practice. It is stable and well documented. This assists the automatic metadata integration, as a specified, common cadastral metadata-spatial data model can be reused. Also, the derived attributes remain the same. This makes the common model reproducible for other cadastral data sets, e.g. on cantonal level.

Nevertheless, the fact that the Swiss cadastral model is at the moment in the transition phase between relational INTERLIS1 and the object-oriented INTERLIS2 makes the automatic integration more complicated. The documentation (Fachstelle AV-Datenmodellierung und -austausch (FADMA), 2001), which is needed for semantic distinction of certain attributes, uses only some object-oriented methods, but the possibilities of the object-oriented concept are not yet fully implemented (for the model version 23).

(5)   **There is no common denominator between the three case study models with regards to the decision which metadata can be explicitly or implicitly derived**.

The actual decision which spatial data attribute can be explicitly used as a metadata description and which meta-information might be implicitly derived from certain attributes, requires exact knowledge of the model and the spatial data as well an understanding of the semantics. There are no spatial data attributes that can always be used explicitly or implicitly. In the case study the date could be explicitly used in the cadastral model, but in the water management and environmental model it had to be inserted as a new metadata attribute.

When dealing with semantics, human analysis is necessary.

5.3   The Design of Principles

The last sections have discussed the integration of specific data models in detail. The next aim is to derive generally applicable rules from the case study for the integration of any metadata and spatial data. These should be phrased in a general way in order to support and to be able to handle any spatial data set and enrich it with metadata.

In order to define what "general" means for spatial data, it is necessary to consider the basic structure of a spatial data model. What do all spatial data models have in common?

Spatial data models are set of constructs that represent objects and processes in a digital environment. Depending on the modeling language, they therefore consist of fields/entities/classes/tables, attributes/properties and relationships. The modeling language formalizes the metadata- spatial data models on conceptual level and describes these characteristics by using its own specific structure. Depending on which modeling language is utilized to describe the model, the outcome varies in keywords, size and concept (object-oriented or relational concepts are used).

Altogether, it is necessary to use a similar framework, in order to be able to compare the three extended models of the case study under the same preconditions and thus make the result reproducible. Therefore, the spatial data set is considered a set of structured data which is grouped according to either the object-oriented or relational modeling principle. Accordingly, two types of principles for metadata integration are considered in the following sections: for object-oriented and relational structured data.

The next question is whether it is possible to automate certain parts of the integration. The goal is to find certain metadata attributes from a given standard which can be always placed in the same part or place in the common model.

5.3.1 Evaluation of the Case Study

In order to be able to find a common denominator and create general principles which can possibly be used in an automatic way, it is necessary to compare and analyze the results from the case study.

There are on the one hand certain communalities between the integrated models, but on the other hand the integration is to some extent model-specific. Every model has to be handled individually when it comes to the decision which spatial data could be used as metadata and which metadata can explicitly and implicitly derived. Also, concerning the judgment for which tables/classes it is necessary and useful to add the bottom-up attributes. Yet, similar top-down and bottom-up attributes were found for all three test models.

The discussion is ambivalent: how to make the general principles as flexible and adaptable to specific model structures as possible while simultaneously being automatically executable? In this context the idea of defining different bottom-up attributes according to the type of table has arisen, but was finally discarded. It is discussed here anyway to show different approaches to define general rules for integration. In the case of the cadastral model it was possible to distinguish between bottom-up attributes for thematic, geometry and layout classes. Figure 5-12 gives examples for the three categories of classes.

Difficulties arise, when different models might not differentiate between geometry and thematic tables or have separate layout tables. Also, the next question is how to identify the different categories automatically. Keywords that characterize, e.g. the geometry table in INTERLIS are the datatypes AREA, SURFACE and POLYLINE. Yet, this is a specific trait of only one modeling language. Regardless, the layout and thematic tables are more difficult to recognize. Altogether, the studied spatial data models are too diverse to be able to categorize their tables according to Figure 5-12.
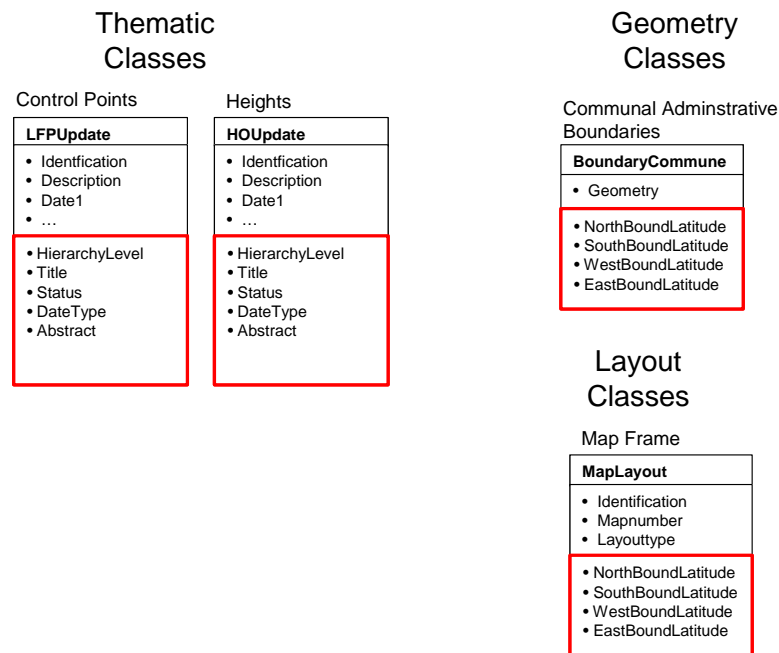
Figure 5-12: Different categories of bottom-up attributes

The case study showed that semantics of the model attributes are not always clear, even if the model is standardized and well documented as in cadastral surveying. For example, there were three spatial data attributes named "date1", "date2" that could have different meanings:

- Date of spatial data acquisition
- Date of acceptance of the spatial data set by the official administration after checking
- Date for entry in land register

An expert must decide which of the three versions is closest to the metadata "date" (MD_Identification.citation>CI_Citation.date>CI_Date.date) definition of ISO 19115: "a reference date for the cited resource". Therefore, it is not possible to automate every decision for the metadata integration.

Also, a detailed analysis for every single data set is too much effort. Therefore, general rules were defined based on two groups of metadata:

a) general, automatically inserted metadata
b) specific, manually inserted metadata by an expert

The **group of general metadata** constitutes the basic description necessary for a typical spatial data set. These metadata are inserted automatically in all spatial data models in the same way. They consist on the one hand of **general top-down attributes**, placed in separate metadata classes and valid for the whole model and on the other hand of **general bottom-up attributes** that fill in the missing meta-description in lower hierarchy levels of the model.

As only the general metadata can be used for every day set and be inserted automatically, they are the topic of discussion in the following sections.

5.3.2 Object-Oriented Principles

As mentioned in Section 2.5.1, one of the fundamental activities of an object-oriented design is establishing relationships between classes by using the two fundamental methods: inheritance and composition. Inheritance expresses the subclass/superclass-relationship and answers the question "is-a?". Whereas composition expresses a 'part-of' or 'has-a relationship'.

Using inheritance to connect the top-down metadata attribute classes with the rest of the spatial data classes would allow the latter to inherit the general metadata attributes. This would make bottom-up attributes unnecessary on the one hand, but on the other hand this would enlarge the actual data file considerably as this means that a lot of the general metadata which are valid for the whole data set needed to be copied to every single object. Various object-oriented guidelines warn to use inheritance for the sake of code re-use ((Oestereich, 2001), (Venners, 2005))

In terms of the case study, only the cadastral model is relevant for this discussion, because the two other models are relational. The former has a flat structure since dependencies and no inheritance have been implemented. Therefore, there are only two levels: the top-down metadata attributes constitute the upper level and the lower level are the existing classes. The latter are described by the bottom-up attributes, if changes in the data make it necessary. In the case study, the bottom-up attributes are only added to selected classes. The two levels are connected by a composition rather than inheritance, because the top-down, general metadata attributes should not be passed on to every class in the lower level. The integration was thus adapted to the specific needs of the model.

For new metadata-spatial data modeling this situation has to be analyzed again. Here, the structure of the data model is not known yet and for the sake of automatization, the bottom-up attributes are either inserted in all the classes of one level or none. Therefore, a general recommendation is necessary.

In general three possibilities are realistic. Figure 5-13 shows the different possibilities of connecting spatial data and metadata in a object-oriented environment schematically.
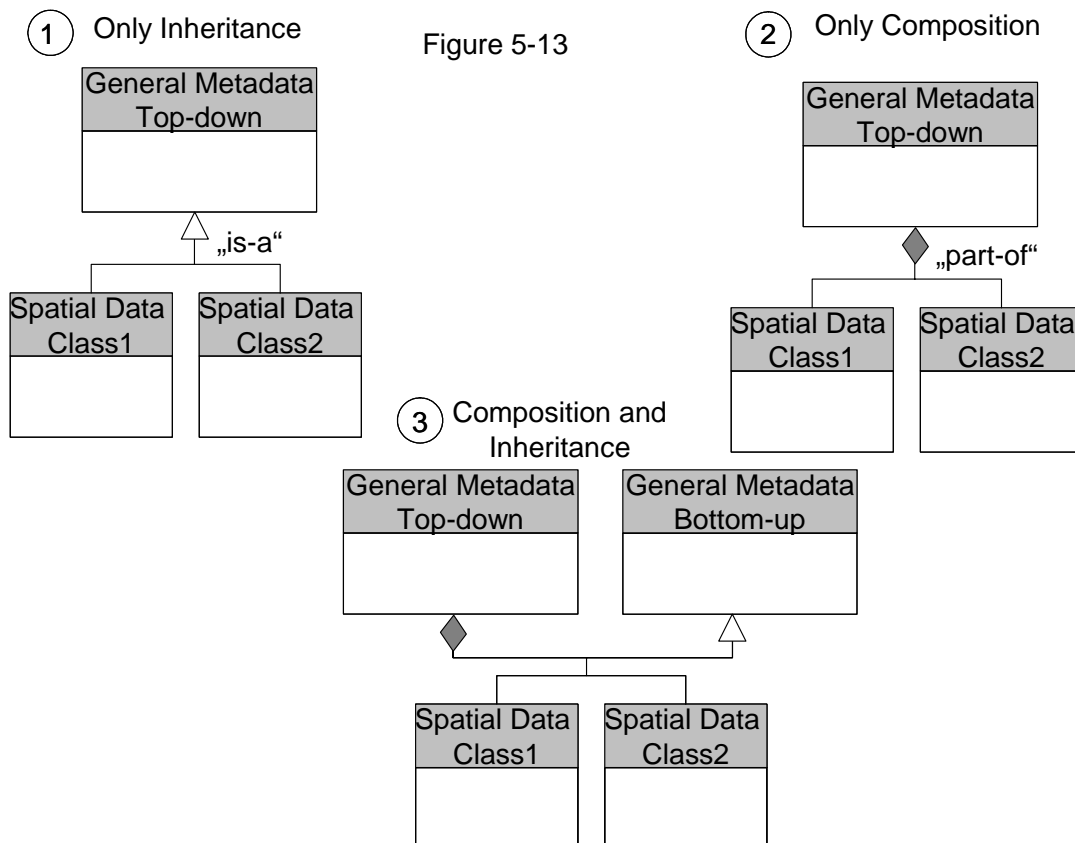
Figure 5-13: Different ways of connecting the metadata and the spatial data in the object-oriented environment.

In possibility number one only inheritance is utilized. Consequently, every object inherits all the general metadata. Now, information e.g. about the responsible party or the name of the metadata standard are not only saved on the hierarchy level of the data set but in all lower levels, too. This makes it unnecessary to add bottom-up attributes and all classes are described in the same way. Also, the existing structures of the data models are strongly modified.

The second possibility connects the top-down metadata by a composition to the spatial data classes. The original data structure is thus kept and legally obliging models like the Swiss cadastral model would not have to be changed. Yet, the metadata description is inflexible and only general top-down metadata are used. Changes in the data set on lower levels are not taken into account.

The third possibility is a combination of inheritance and composition. The top-down metadata are connected by a composition and the bottom-up metadata are inherited. This creates the necessary flexibility for the metadata description while not blowing up the data set unnecessarily.

It was decided that the combination of inheritance and composition is the closest to the goals of integration and follows the advice of object-oriented modeling: on the one hand it is flexible enough and on the other hand it does not blow up the data set too much.

In summary, the object-oriented principles utilize a combination of composition and inheritance (Number 3 in Figure 5-13**) for connecting the general bottom-up metadata (**Table 5-2**) and the general top-down metadata (**Figure 5-4**) to the spatial data.**

5.3.3 Relational Principles

Since concepts like inheritance and composition do not exist for relational models, the integration has to be solved with a combination of top-down and bottom-up attributes which are connected by foreign keys.

The general top-down metadata attributes are shown in Figure 5-14.
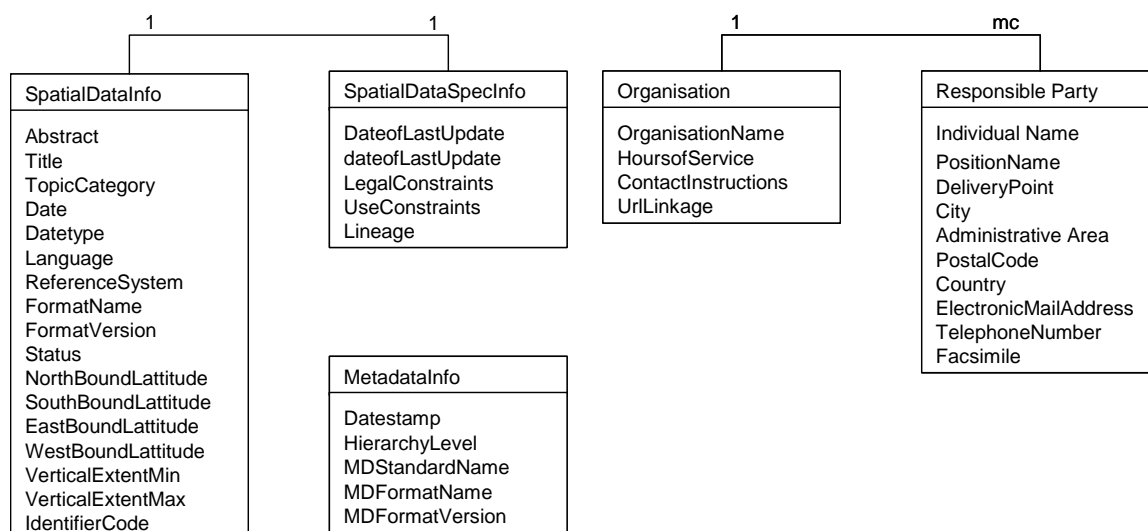
Figure 5-14: General top-down metadata attributes for relational models

In order to describe all levels of spatial data and thus enable interoperability in the sense of ISO 19115, it is necessary to add bottom-up metadata attributes to the relational data model. By analyzing the case study examples and following the changes of metadata from the level of the general spatial data set to the table level, a list of bottom-up metadata could be derived (Table 5-2).

Name	Path in Swiss ISO19115 Profile (GM03)
Title of data set	MD_Identification.citation>CI_Citation.title>PT_FreeText
Spatial resolution of the dataset	MD_Resolution.equitvalentScale>MD_RepresentativeFraction.denominator
Metadata hierarchy level	MD_Metadata.hierarchyLevel=”dataset
Creation date of data set	MD_Metadata.citation>CI_Citation.date>CI_Date.date
Creation data of metadata	MD_Metadata.dateStamp
Lineage	DQ_Dataquality.lineage>LI_Lineage.statement>PT_FreeText
Content of data set	MD_Identification.abstract>PT_FreeText

Table 5-2: Bottom-up attributes for relational models

5.4   Summary

The aim of Chapter 5 was to review the feasibility of the Concept of Metadata-Spatial Data Integration that was introduced in Section 4.2. This was realized by utilizing the case study methodology.

To illustrate, the case study focuses on three existing data sets and models, which are extended to carry their own metadata. The choice of the case study models is discussed. It is important to have a large variety of different models in complexity, size and which were created by different modeling experts for testing feasibility of concept. The three chosen models and data sets are the Swiss cadastre, the water management of the City of Zurich and environmental data. The integration is executed in three steps: firstly semantic analysis and translation, secondly structural analysis and hierarchical integration of the models as well as thirdly semantic transformation of the data sets.

Finally, the integration is analyzed for every single case study model. For each of the chosen models certain top-down and bottom-up metadata attributes are defined. Furthermore, the implicitly and explicitly derivable attributes are ascertained. The common models were verified in a workshop with experts who know the original models well.

Section 5.2 discusses the overall result and feasibility of the integration procedure as well as outcome. Altogether, the concept of integrating metadata and spatial data is feasible, as shown in the case studies.

In order to be able to use the Concept of Integration on any data set with model in a similar way, rules for the integration are necessary. Therefore, general principles were derived for object-oriented and relational modeling languages by comparing the results from the case study and abstracting them to a general case of any spatial data set. A group of general, automatic principles to insert certain metadata at a specific place in the model has been defined. These automatic principles consist of general top-down metadata that are valid for the whole model and bottom-up metadata that mirror the changes and heterogeneity of data within the model. Notwithstanding, it is also necessary for the modeling expert who knows the spatial data well to choose which metadata can be derived implicitly and explicitly.

Chapter 6   Prototype Implementation

In the previous chapters, the workflow was to first integrate metadata in existing spatial data sets in order to check whether the concept of integration is feasible. Then, imagining that this integration is utilized in future for every new data set, metadata will consequently be modeled and acquired together with the data from the start. This is where the view concept becomes important. Assuming that integration will be the standard procedure and most spatial data carry their own metadata, it will be challenging to utilize the view concept to create metadata extractions according to different standards and specific needs, automatically. Therefore, certain tools must be provided to support these processes.

This chapter introduces two prototype implementations which support the mentioned concepts: the UML Editor implements the principles for metadata-spatial data integration and thus realizes the common modeling of spatial data and metadata (Section 6.2). The second prototype implementation explores the possibilities of creating views and functionalities of views in the relational database management system Oracle 9i (Oracle, 2005) (Section 6.3). Finally, a demonstrator-program visualizes how the concept of integration of metadata-spatial data and the view concept will facilitate certain processes for SDIs in the future (Section 6.4).

6.1 Introduction of General Tools

Various software tools are needed to support the concept of metadata-spatial data integration (Giger and Najar, 2003). Some of them are related to managing and enforcing the concept and others are for planning or modeling spatial data Figure 6-1. Again others support the data quality management. Also, many of these services already exist for certain purposes or specific modeling languages. Nevertheless, they are not helpful, because they are often stand-alone solutions which cannot be used in combination, e.g. the acquisition tool does not work together with a data/model quality verification software and the modeling tool. These compatibility problems can stem from a lack of standardized interfaces or transfer formats, e.g. a spatial data set is modeled according to a certain modeling tool, but the quality verification tool cannot import and read the data and model because it needs another format. Another difficulty is the missing documentation, especially among the open-source and freeware software.
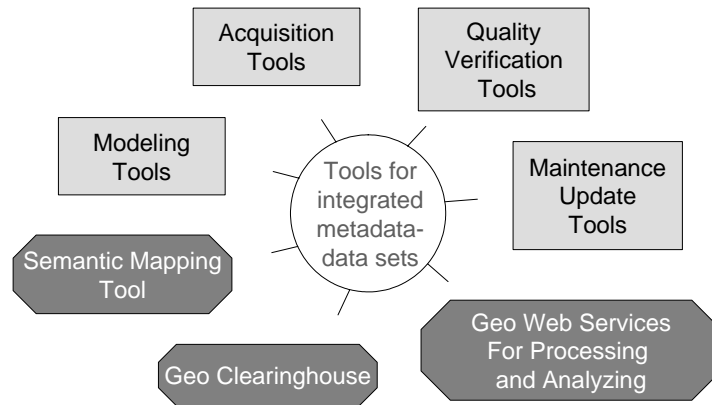
Figure 6-1: Categorization of important software tools to support the integrated management

Here, the most important types of services that are needed to support an ideal workflow are listed and described shortly (Figure 6-1):

Maintenance and Acquisition Tools

- Tools for quality verification: a given data set must correspond to its model. Other quality aspects, such as topologic correctness, completeness, consistency are also checked.

- Modeling tools to support new data modeling and existing data models: tools to extend existing models and to create common new models according to the metadata principles of Section 5.3. are necessary.

- Maintenance tools for processes such as (incremental) update.

- Metadata acquisition tools which support the acquisition according to given standards.

Processing and Management Tools

- Tools for semantic reading and mapping: dealing with different models, ontologies and semantics in order to be flexible, compatible and transferable. This could be a view service which not only restructures existing data sets but also adds the new data (e.g. metadata) in the right place in the model.

- Web services for SDIs for processing and analyzing services for spatial data (Geo Web Services).

In many cases, the effort of extending existing tools is less than redesigning a new tool. Some of the functionalities from existing tools can be reused and adapted. There have been various modeling tools developed for spatial data modeling on conceptual level , e.g. the Perceptory tool by the University of LAVAL (Bedard et al, 2004) or the Add-on tool for GML which works with the software Rational Rose by Clemens Portele. The communality is that

UML is used as basic modeling language to model spatial data. At the same time it is necessary to be able to export UML to a computer-processable language like XML or INTERLIS in order to make the model readable for Web services in SDIs.

6.2 Implementation of the Modeling Prototype

6.2.1 Introduction to INTERLIS/UML-Editor

The INTERLIS/UML-Editor (Eisenhut and Hirzel, 2003) is an open source modeling software which was created to support the use of the model-driven approach and thus combining the text-based INTERLIS with the graphical UML (Section 2.5.1).

Therefore, the INTERLIS/UML-Editor allows to model in two standardized modeling languages UML and INTERLIS2 simultaneously. Furthermore, it is possible to derive different formats, such as XML, automatically. This modeling tool simplifies and supports the use of the complex UML and INTERLIS languages and thus open them to a large user-community (Hirzel, 2004))

The reasons for choosing the INTERLIS/UML-Editor to implement the general principles for integrating metadata and spatial data are various:

- It is designed to support the spatial data modeling process.
- It supports two major object-oriented modeling languages for spatial data.
- The open source code allows an easy modification.
- The test data sets are described precisely in the supported modeling languages UML and INTERLIS.

In order to understand the possibilities which are provided originally and how these can be extended in the next step, this section will shortly introduce the INTERLIS/UML-Editor with its original functionalities and its specialties for modeling in UML and INTERLIS 2.

The INTERLIS/UML-Editor is based on the UML Metamodel as defined by the Object Management Group (Object Management Group (OMG), 2005). Therefore, all models which are created by the INTERLIS/UML-Editor are compatible with the UML 1.4 specification (International Organization for Standardization ISO, 2005).If the UML-metamodel is updated by OMG, the tool can be adapted to the new metamodel with little effort. Furthermore, an INTERLIS-heavyweight extension specializes the UML-metamodel with INTERLIS elements and makes it usable for INTERLIS and UML modelers.

The tool is programmed in Java (SUN Developer Network (SDN), 2005).The major part of the software model is designed in the program Rational Rose. Although Rational Rose is a commercial software which is used for general UML modeling and not specialized on spatial data modeling. Here, only the functionalities of the UML Editor are designed in Rational Rose and the code is generated by a specifically extended Java-Code-Generator This way the design always stays up-to-date with the source code (Hirzel, 2004)).

There are consecutive versions of the INTERLIS/UML-Editor. Here, the functionalities of the INTERLIS/UML-Editor 2.3.0 are described.
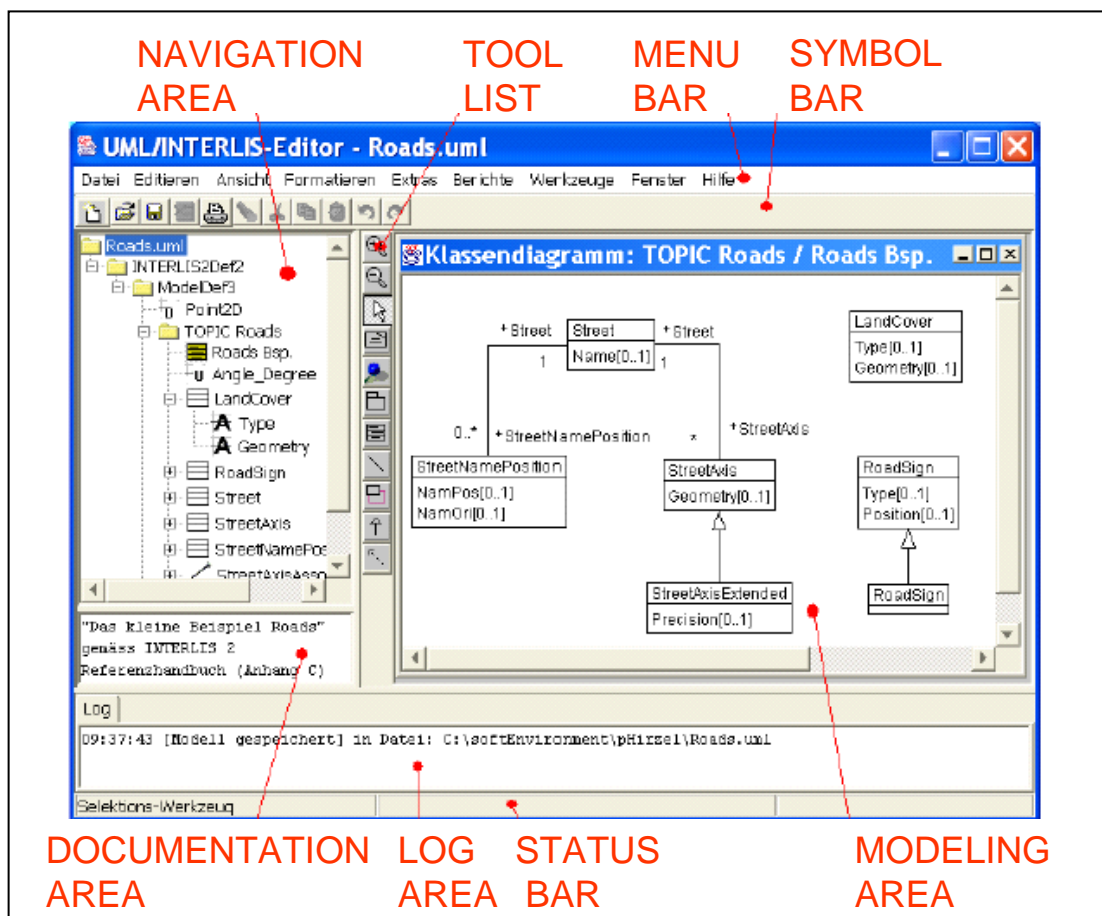
Figure 6-2: Modeling in the INTERLIS/UML Editor

The main part of the modeling process takes place in the navigation area and in the modeling area (Figure 6-2). The navigation area on left hand side of the window offers a quick possibility to navigate between the different elements of the model and edit them. Equally, the modeling area shows the elements and their relationships graphically in order to process the different elements. Plus, the documentation area offers a window for entering a description of the chosen element.

When the model is finished, it is possible to generate an automatic data dictionary in a html file and check whether the model complies with the INTERLIS 2 rules concerning language syntax and semantics.

There is an import function for UML (XMI/Rose) and INTERLIS 2 model files and an export possibility which converts the current model to an INTERLIS textual model. Furthermore, XML- Schema (XSD) can be exported.

6.2.2 Implementation of Prototype for Common Modeling

Using the information of the original functionalities and the technical details, the UML/INTERLIS-Editor is extended so that it enforces the general principles of metadata-spatial data integration. This is realized with Java 1.3.1_01 in the programming environment Eclipse 2.0.

The first task of the implementation is to understand the existing program and to know which class must be modified to achieve the necessary changes.

Corresponding to its building blocks the INTERLIS/UML-Editor Java program consists of three parts with more than 600 Java classes:

(1) UMLEditor: defines how the application itself works and how it is presented on the screen. The design and functionalities are related to the program Rational Rose.

(2) UML1.4:   implements UML 1.4 according to ISO/IEC 19501

(3) INTERLIS2: implements INTERLIS 2, e.g. defines the INTERLIS 2 elements, constraints as well as graphical description and implements the INTERLIS 2 modeling rules according to the INTERLIS 2.2 version.

In order to be able to work with the Java code, it is necessary to know in which classes the main()-method exists, which is the entry point into a Java program (Flanagan, 2002). In this case, there is only one main()-method and it is located in the class "LauncherView.java" in the 'UMLEditor' part of the program.

For implementing the principles for metadata integration (Section 5.3) three steps are necessary (Figure 6-3).
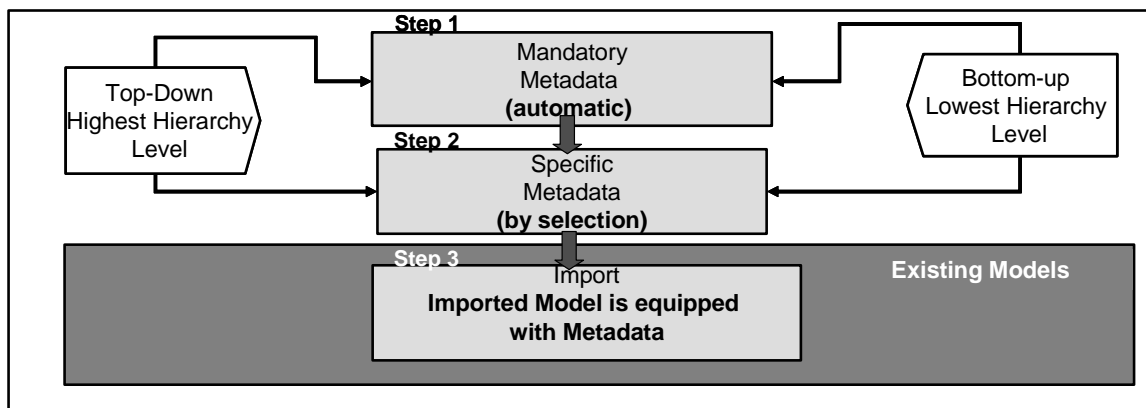
Figure 6-3: Steps for extending the INTERLIS/UML Editor

According to the principles for object-oriented models (Section 5.3.2), there are two kinds of attributes automatically inserted: the general top-down attributes, which are connected to the rest of the spatial data by a composition, and the general bottom-up metadata attributes which all spatial data classes inherit (Figure 5-13). Both are therefore mandatory metadata groups and are realized in step 1. In order to realize this in the Java code, two classes in the 'UML Editor/application' part of the code are modified: ElementFactory.java and LauncherView.java. In ElementFactory.java, the method createClassDef(Element) was extended to carry all the bottom-up attributes and insert them automatically if a new class is drawn (Figure 6-4).

```
public static ClassDef createClassDef(Element parent)
Throws ClassNotFoundException,InstantiationException,IllegalAccessException {
// generic initialization
ClassDef classDef =
(ClassDef) createOwnedElement(ClassDef.class, parent);

// default initialization
classDef.setKind(ClassDefKind.CLASS);

//new, extra, automatic Metadata attribute
ch.ehi.interlis.attributes.AttributeDef attributeDef21 =
ElementFactory.createAttributeDef(classDef);
attributeDef21.setDefLangName("MD_Hieararchy level metadata");

ch.ehi.interlis.attributes.AttributeDef attributeDef22 =
ElementFactory.createAttributeDef(classDef);
attributeDef22.setDefLangName("MD_Title of data group)");
.....
```

Figure 6-4: Excerpt of java code: adding two new bottom-up metadata attributes

Thus, the bottom-up attributes are generated as soon as the user draws a new class in the model and the top-down classes are added automatically to a new model when the modeling is started.

This automatic integration is completed by specific metadata (step 2) which are selected by the expert. The former are offered to the user by a pop-up window which opens automatically and contains various panels with thematically structured attributes that need to be marked in a check-box, if wanted in the model. This can be interesting for altering the automatic metadata by completing or deleting both the bottom-up as well as the top-down attributes manually.

In the code, a new class called "MetadataJTabbedPane.java" is added to the 'UmlEditor\ application' part of the program. This new class defines index cards with thematically structured metadata (at the moment according to GM03).

Step 3 ensures that imported models, such as the case study models which don't have any metadata attributes inserted yet, receive the automatic metadata while importing and the specific metadata by selection, too.

Altogether, the newly created models are now supplied with the necessary metadata automatically during the modeling process. As soon as a new class is drawn in the Navigation or Modeling Area of the INTERLIS/UML-Editor. Simultaneously, the specialist is offered the choice to add supplementary metadata attributes by a pop-up window and can add certain extra attributes, if needed, or delete existing metadata attributes.

6.2.3 Verification of General Integration Principles

In order to be able to verify the general principles and also the prototype implementation, one of models from the case study is opened in the INTERLIS/UML-Editor and supplied with automatic bottom-up metadata attributes. Figure 6-5 shows the same control points of the cadastral model as in Figure 5-3 and 5-6, but this time in the INTERLIS/UML-Editor environment and with the automatic metadata integration.
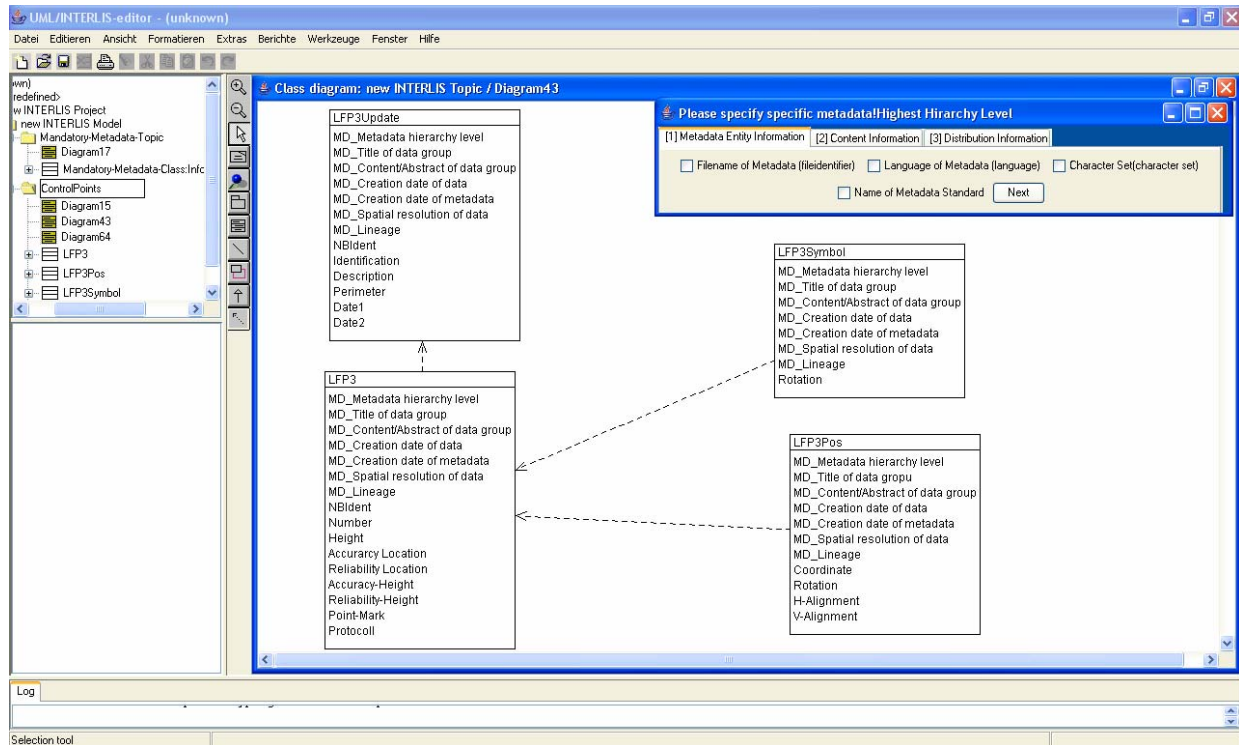
Figure 6-5: Control points from cadastral model are supplied with automatic metadata attributes by the UML Editor

The main difference is that the explicit attribute "Date1" was inserted by the automatic metadata integration of the INTERLIS/UML-Editor a second time, even though it already exists in the spatial data model. The capability of semantic comparison is reserved to humans. In this case, the specialist would therefore delete either the attribute "MD_Creation date of data" from the class "LFP3Update" or "Date1".

Another difference is that new bottom-up attributes about the quality were added in the automatic integration: lineage and spatial resolution. This was the result of discussion with experts.

The Swiss Federal Directorate of Cadastral Surveying (Swisstopo, 2005) launched a project to find the essential metadata for the Swiss official cadastre surveying within the GM03 comprehensive (Eidgenössische Vermessungsdirektion + ITV Geomatik AG, 2004). If those suggested cadastral metadata attributes are compared with the result of the automatic integration according to the general principles, some metadata are missing. These metadata are the specialized attributes which have to be inserted manually, e.g. by an expert using a software wizard. They are flexible, topic-related metadata which can be used according to the model's characteristics.

Among the missing and therefore model-specific metadata, there are attributes relating to the package of maintenance information and legal aspects, e.g. maintenance update frequency, note for maintenance, date of last update, contact address for maintenance, information about legislation type and information. This is typical for cadastral surveying data sets, because cadastral surveying is usually inherently connected to extensive legislation. Rules, regulations and restrictions are important for a commodity like land and property which needs to be regulated in a society with growing population. The frequency of maintenance and note for maintenance might be interesting if method of maintenance is complex and depends on the object, e.g. whether it is updated periodically or continuously.

At the moment the specialized metadata are grouped according to thematic packages of GM03 in the pop-up window. In the future, they are maybe offered in groups according to model topics (e.g. cadastral) and their characteristics.

In summary, the verification shows that the automatic integration offers the basic, neutral metadata description without adding unnecessary, topic-related attributes. Thus, the model is not enlarged needlessly. The flexibility of the integration relies on the wizard for specialized attributes to make the metadata description adapted to the characteristics of the model. Also, the decision which spatial data attribute can be used as a metadata attribute has to be taken by a human and is not replaceable by a computer in an automated process.

6.2.4 Discussion of Modeling Prototype

The basic functionalities of the extended INTERLIS/UML-Editor for metadata have been introduced in the previous sections. There are many further possibilities of expanding the INTERLIS/UML-Editor to facilitate the general usability and the workflow of common metadata-spatial data management.

One of the main constraints of the current modeling tool is the missing link to other metadata tools. Therefore, the next step for future implementation would be an external link or interface to a metadata acquisition program. The ideal support would enable a user to first model the spatial data and metadata commonly in the INTERLIS/UML-Editor and then transfer this common model to the metadata acquisition tool, which supports the user in entering only the required metadata, according to the common model. Another example for external linkage would be connection to the spatial data, e.g. in order be able to derive specific attributes.

Furthermore, it would be good to offer a math-assistant that lets the expert choose the right algorithm (e.g. Maximum and Minimum function) and the corresponding attributes in the model as well as in the spatial data base.

A wizard which assists users in modeling their new spatial data and metadata would be useful, too. Observing the different test data used, it is evident that there are many different styles of models. Even though each modeling language has best practice examples, the actual modeling is diverse and individual. In order to make the modeling more uniform and facilitate the process, an assistant could be created which offers certain topics components, similar to a code-library. In the case of e.g. the contact to the data custodian which is needed quite often, the package in GM03 could be offered. If the modeling expert is interested in reusing this bit of code, he or she could recycle it from the code-library and insert it into the new model in the INTERLIS/UML-Editor.

During the prototype implementation it has been discussed whether it is useful to add a lineage attribute from bottom-up to every class. This might be interesting when it comes to supporting other applications, e.g. creating views (Section 6.3). The general principles allow a certain pattern of metadata integration, which are documented and traceable. This means, the metadata attributes can be retrieved, if needed, by retracking the general principles. Nevertheless, the specialized metadata attributes are added individually and therefore not documented. They are not identifiable as newly inserted attributes. In some cases it might be important to retrack this, e.g. for licensing or legal reasons. If the lineage attribute saves all the changes it would give the necessary information. The lineage collection could be implemented in an automatic way, e.g. derived automatically from inserted metadata attributes which are chosen from the list.

6.3 Implementation of the View Process

The processes of semantic mapping as well as its special case "creating views" are needed in different steps within the workflow of common metadata-spatial data management. In order to receive common metadata-spatial data sets, the metadata must be integrated with a semantic mapping tool into the existing data set. The view tool is used to create various specialized file formats, excerpts, or in our case metadata according to different standards from the common metadata-spatial data set.

The overall goal is to be able to apply semantic mapping on the conceptual level, because more information about the model, its overall structure and data is available. The conceptual level is also independent of any database management system or other implementation. A prerequisite is to cover semantic mapping and creating views with the same tool. Both processes have common steps: semantic translation and semantic transformation, but semantic mapping requests added functionalities. For example it is necessary have the possibility to add new data in a restructured data set, whereas a view tool only has to restructure data without adding new information (see also Figure 4-1).

There have been various attempts to create semantic mapping tools on different levels (Section 5.1.2, step 3: semantic transformation). The main functionality missing is the adding of new data in the restructured data file. Also, generic scripting languages are not available on conceptual level with the necessary functionalities. Nevertheless, here the process is essential and the implementation is executed on the logical level with the DBMS Oracle.

Figure 6-6 shows the workflow on the different levels. First the models are integrated on conceptual level with the INTERLIS/UML-Editor(1). In order to be able to add the metadata to the spatial data set, according to the pattern of the integrated model, it is necessary to work on the logical level. This is possible in for example a DBMS like Oracle. The disadvantage is that during the translation, the conceptual structures and relationships are flattened and simplified to the relational database schema. In order to preserve as much of the information of the conceptual level as possible, the integrated model is exported with a special tool (ILI2ORA, Section 6.3.2) which saves the original relationships, structures and names of classes in an external table. The result is the integrated metadata- spatial data in the database (2). Returning to the conceptual level the view model, e.g. the semantics of the metadata standard model which should be created, is manually translated. Encoding rules how to map the integrated model to the view model are the result (3). These are used on the logical level to restructure an excerpt of the integrated data. This semantic transformation results in the view data (4).
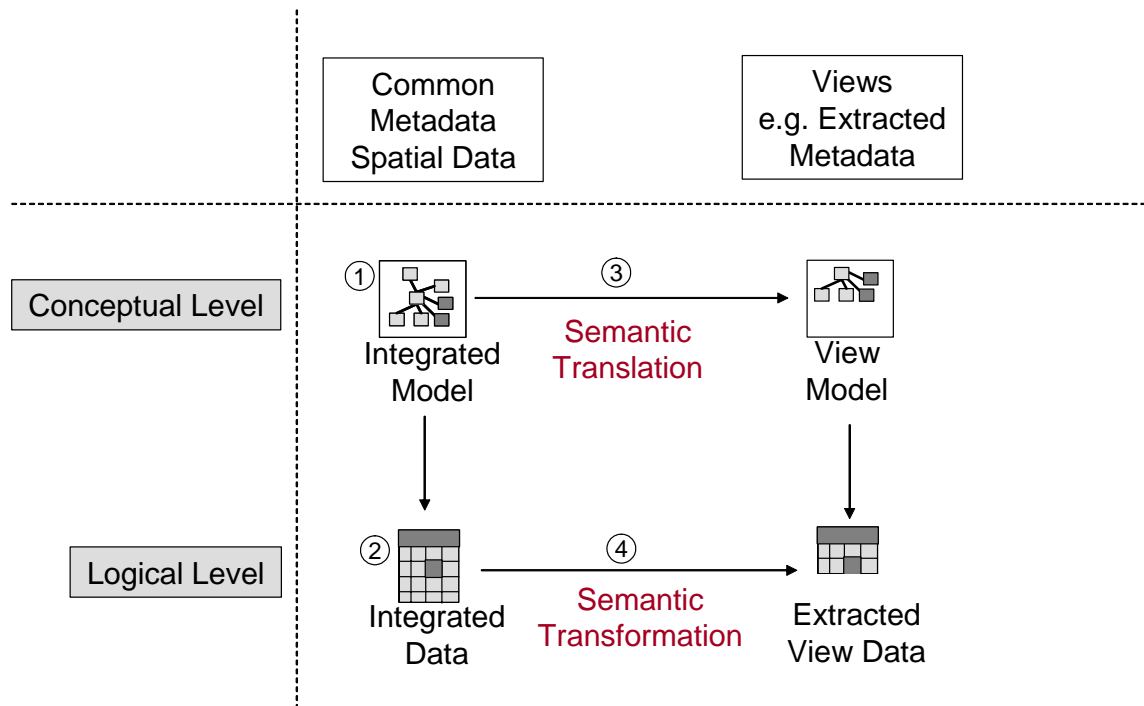
Figure 6-6: Conceptual versus logical level for the workflow for semantic mapping

The goal of the following prototype implementation is to check the feasibility of creating metadata views according to different schemas from an integrated data set. This would be useful, e.g. in a SDI where the extended spatial data might be stored on a data server and the search engine requests the updated metadata for indexing purposes. For data protection reasons, not all the spatial data should be indexed, so only a view of the metadata are created for the Clearinghouse and the original, common data set is not changed.

6.3.1 Introduction to Oracle Views

"A view is a virtual table in the relational data model in which data from real (base) tables are combined so programmers can work with just one (virtual) table instead of the several or more complete base tables…. A view gives a current, but restricted, "window" into the database; that is, a view is not a temporary table that has to be rebuilt when changes are made to the original data. Because its contents are derived (by the DBMS) as needed from base tables, a view always reflects the latest changes made in any of its base tables. The advantage of an SQL view is that we can use it just like a table for subsequent data retrieval (McFadden and Hoffer, 1994)."

A view can be dependent on multiple tables or other views. These tables or views must exist in the database, and the columns must be defined. The user who uses the command 'CREATE OR REPLACE VIEW' must have certain privileges to execute it and must the direct access and necessary privileges on the tables in the view.

If a table, the view depends on, is dropped, the view is not dropped, it is merely marked 'invalid' and any access attempts will result in an error.

Views are generally used to focus, simplify, and customize the perception each user has of the database. They can also be used as security mechanisms by allowing users to access data through view, without granting the user permission to directly access the underlying base tables of the view.

Views have been part of the functionality of any larger and sophisticated Database Management Systems (DBMS) like Oracle and PostgreSQL for a while. It is therefore interesting to explore the functionalities of creating views in Oracle, because it is possible to understand the advantages and limitations of working on the logical level, with a powerful software. Also, the feasibility of creating views with integrated spatial data and metadata are tested. Nevertheless, the long-term goal for the view and semantic mapping tool is to be able to work on the conceptual level.

Even though INTERLIS2 offers the possibility to define views, there are functionalities missing. Views in INTERLIS 2 can describe simple classes but no associations or complex structures. So, a view can bundle or divide certain classes, but no associations can be defined. In order to extract metadata according to certain standards, associations are absolutely needed. Therefore, INTERLIS 2 cannot be used at the moment for this purpose.

In order to be able to use Oracle for metadata views, certain discussions and tools were needed. Coming from an object oriented INTERLIS 2 model to a relational model where there are no relationships between tables, it is necessary to define how the certain model structures are mapped. *1-m* relationships are defined by object identifier using primary keys between the associated tables. In the case of m-n relationships (Figure 6-7) an intermediate table needs to be inserted to dissolve the multiple relationship to simple *1-m* relationships, which are handled as mentioned above.
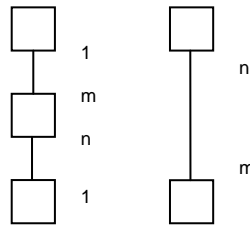
Figure 6-7: Intermediate table for n-m relationships in Oracle

Altogether, working with a DBMS that supports views with specific, well tested functionalities, helps creating views for metadata. Yet, by coming from a conceptual, object-oriented level to database level a lot of information is lost by translation.

6.3.2 Choice of Test Data Sets

In order to have a variety of different data sets with different metadata integration, two different topographic data sets in different scales are chosen: VECTOR25 and VECTOR200 (Swisstopo, 2005) with their metadata from geocat application (Geocat, 2005) are imported into Oracle and integrated.[1] Furthermore, the integrated cadastral data model and data sets are used, which were mapped with the INTERLIS Conversion Software (Section 5.1.2, Step3). Thus, the chosen test data sets have two different Semantic Transformation methods.

Since the topographic data sets are only offered in ESRI shape files, a specialized tool , called "shp2sdo", for importing shape files into oracle is used. Figure 6-8 shows the schematic workflow of the tool (Oracle9i Spatial, 2005).
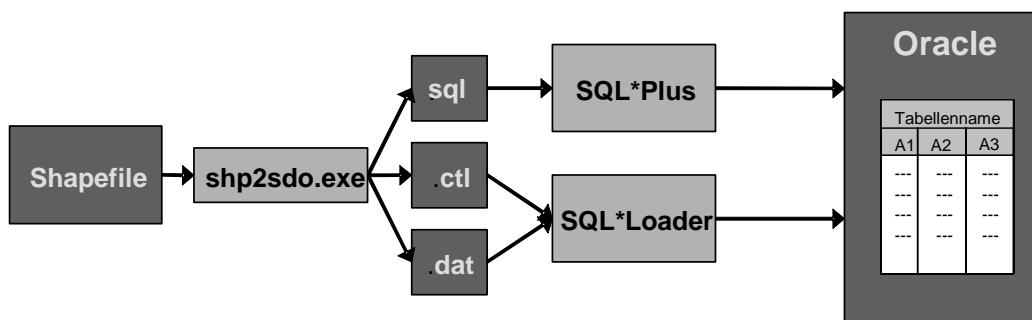
Figure 6-8: Data import from shape file to Oracle

---

[1] Certain information in the following sections is taken from a diploma paper, which was realized in collaboration with and under the supervision of the author. For details please check (Filli, 2005)

Furthermore, the INTERLIS 2 files are imported with another specialized tool, called ili2ora. This tool was defined, tested, discussed and improved together with the programmer specifically for the semantic mapping prototype implementation. It consists of a compiler which first checks syntax of the model and only if this is executed without major problems, it proceeds to insert the data into new Oracle tables.
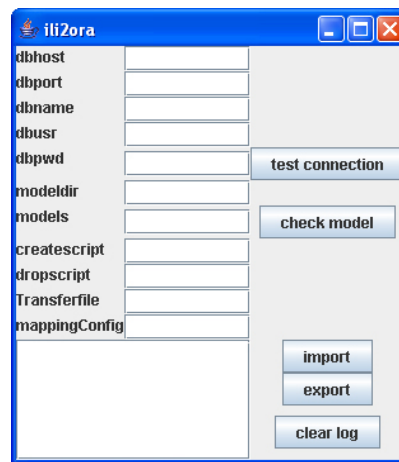
Figure 6-9 shows the mask which has to be filled in with the necessary information before the import can be done.

Figure 6-9: Data input window ili2ora (INTERLIS to Oracle)

During the import of the INTERLIS2 files the ili2ora tool creates the translation tables for class names and relationships in addition to the actual data tables. The content of the extra tables are the original name of the classes and the current name in Oracle as well as the original relationships between the classes. Saving the names is necessary, because Oracle only allows thirty characters to be used for a name of a table and has its own algorithm for abbreviating the original name which can be incomprehensible: it uses the first and the last two characters of the original name.

6.3.3 Implementation of Views

After having chosen the test metadata and spatial data, they are imported in Oracle using "shp2sdo" for the topographic data and the INTERLIS 2 import tool ili2ora for the rest of the data. The next step is to understand and create views in Oracle according to three mapping schemas: the regional, Belgian Core profile METAWAL (Wallonne, 2005), the Swiss Core GM03 (Coordination of geograhic information and geographic information systems (KOGIS), 2003) and the ISO 19115 Core (ISO/TC211 International Organization for Standardization ISO, 2001).

Thereby, it is interesting to look at the feasibility and assess the differences as well as the effort it takes to create views from different profiles of the same metadata standard. Furthermore, the comparison between the views and the question whether it is possible to create generic views is analyzed. By generic views, the reuse of the same (SQL query) code for different data sets is possible. Another point of investigation is the calculation of derived metadata attributes from spatial data.

Certain characteristics of the Oracle DBMS have to be taken into account while working on views:

It is necessary to make unique allocations between all tables, from which entries are extracted for views (symbolized as V1 in Figure 6-10). This is realized by using unique Object Identifiers (OID) and primary keys. Otherwise, all the entries are automatically combined by a cartesian product when extracted for views. Consequently, the OIDs must be unique over the whole data set.

Creating views out of linked tables, ideal solution

Automatic combination of table entries by Oracle: Cartesian product
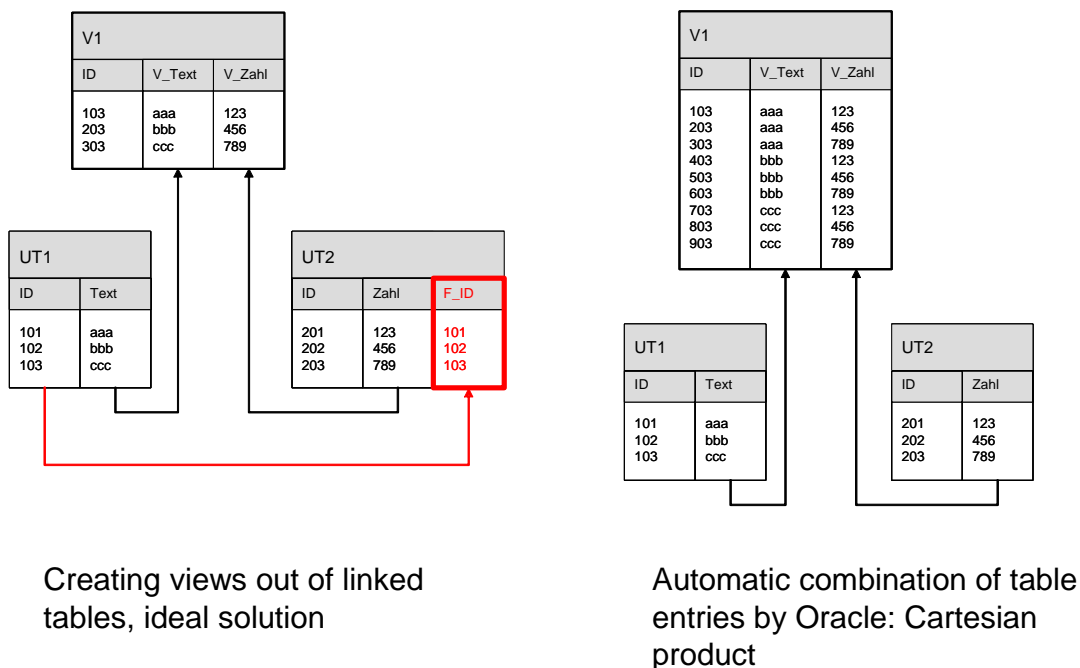
Figure 6-10: Creating views from tables

Another precondition for creating views in Oracle is that the original spatial data tables should not have empty entries. If a view is created from a table with no attributes it is empty too. This is a problem if the view obtains attributes from different sources and only one provides empty entries. Then the resulting view is empty as well. This should be avoided by first using a query which checks the tables and excludes empty tables or gives single entries the value zero.

Another implementation detail is that views need to create their own primary key for their entries. The reason is that if entries from different tables are combined in a view, a conflict between primary keys may occur. Therefore, it is the best to generate the primary keys automatically and in a global way.

When it comes to assessing the effort of creating views the amount of object entries is important. The more entries there are in the original table, the more unique allocations have to be done.

Furthermore, unique OIDs are essential for SDI applications in order to combine different geo-objects from different sources. Therefore, the call for coordinated, overall unique OIDs is a point of discussion in SDI initiatives which want to solve this with a Web service.

In comparison between the different profiles of ISO 19115 it was found that ISO Core is compatible for all profiles. This does nevertheless does not apply for the comprehensive part.

6.3.4 Results and Discussion of View Implementation

It was found possible to create views for different profiles of metadata standards, guidelines and in extension for different standards. In order to be able to derive metadata from spatial data, e.g. a bounding box, mathematical functions like minimum and maximum are needed. Profiles of the same standard are essentially compatible, but details like optional classes have to be take into account and mapped, possibly manually, adequately.

A view cannot be expressed in a generic way in SQL so that it is applicable for any data set and model. It is necessary to at least to embed SQL in a more powerful programming language.

- **It is possible to calculate metadata from spatial data.**

SQL offers this comparison to find the smallest and largest coordinates, but it takes about 30min for one thematic layer of the Swiss topographic data set (about 300 MB, ESRI Shape file). It is thus a time consuming operation.

- **The feasibility to create views from different profiles of the same metadata standard.**

The "Multilinguism package" in ISO 19115 Core (ISO/TC211 International Organization for Standardization ISO, 2001) was a point of discussion during the implementation in Oracle. It is interesting to compare different ISO 19115 profiles to see how similar they really are and whether profiles therefore facilitate the process of creating views. In the case of the package

"Mulitlinguism information", which is offered in ISO 19115 as an option, and is modeled with two main classes: PT_FreeText and PT_Group. If the mulitlinguism option is used, PT_Group contains the textual description, like 'abstract', 'purpose', 'keyword', 'title', etc. (Figure 6-11). If the multilinguism option is not used, the textual information is saved in a decentralized way in the different classes, e.g EX_Extent, CI_Citation, MD_Identifier etc. The class PT_FreeText contains the allocation of primary keys from all the connecting classes.

The Swiss profile GM03 uses the multilinguism package in the Core metadata (Figure 6-11). This makes sense for a country with multiple official language like Switzerland or a large community like the European Union. Other countries do not have this necessity and therefore do not implement the package "Multilinguism Information" in their national metadata profile. The resulting question is how to deal with this situation when creating views. For an SDI it would be interesting to create views with metadata of all the offered languages. Yet, using metadata in all languages means to take over the class PT_Group in a view (Figure 6-11). This involves creating new foreign keys as the original keys which were saved in the class PT_FreeText are not transferable. As the allocation in PT_FreeText cannot be used for every class, it is not possible to formulate in a general way.
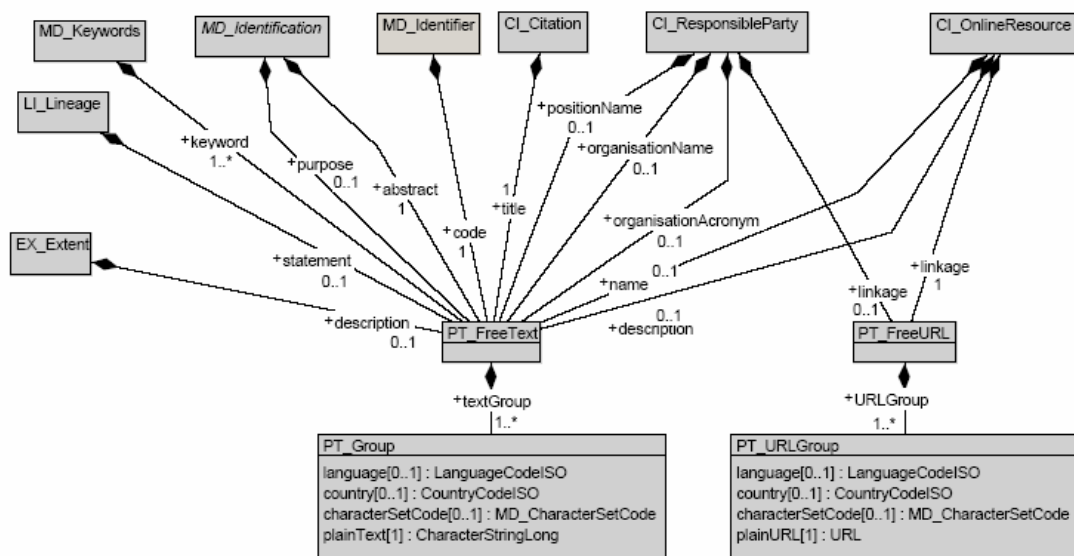
Figure 6-11: Multilinguism information according to GM03

The other possibility of creating views is using only one language which means that the classes PT_Group and PT_Text are dropped. Therefore, the information has to be collected from the single tables. The disadvantage is, that the multi-lingual information is lost.

This discussion is especially interesting in the context of global or regional SDIs, in which many countries with different languages and different metadata standards and profiles meet.

It also shows that profiles from the same standard are essentially compatible, but details have to be taken into account and mapped adequately.

- **Another important issue is the question whether it is possible to create generic views.**

In the ideal case, a view for a certain metadata standard can be reused for different metadata-spatial data sets without the need to repeat the whole semantic mapping. The feasibility of creating generic views depends on different factors:

Are the original metadata-spatial data sets are all acquired according the same metadata standard?

Were the same principles of metadata integration used on the data sets and models?

How much in common do the profiles of the standard have with each other?

In the case of this implementation, all the metadata-spatial data sets were acquired according to the same metadata standard profile GM03. The topographic data set was integrated according to the same, automatic principles, whereas the cadastral data had been integrated in the case study manually. The output standards were all profiles of ISO 19115. Therefore, the results can only based on this situation and further view implementations are necessary.

It appears that working on the logical level with SQL, generic views are only possible if an external, e.g. programming language supports the SQL commands. This could be done for example by using variables which can replace the table names so that the views are applied to different models by filling in the new table names for the variables at the beginning. When differences between different profiles occur, e.g. one set uses the multilinguism package and the other does not, the most basic solution for the reusable view has to applied, even if this means loss of important information.

For future view implementations, it isimportant to be able to work on the conceptual level.

6.4 Demonstrator for Web Services using Integrated Metadata Spatial Data Sets

The mapping tool and the INTERLIS/UML-Editor, which were introduced as prototypes in the previous sections could be realized as a standard software, but in the future they might be used as a Web Service for SDIs.

Imagining that future spatial data carries its own metadata, this section gives an outlook in future implementations for Web Services in SDIs by using a demonstrator to visualize the functionalities and possibilities of future Geo Web Services[2]. The advantages of the concept of metadata-spatial data integration and concept of views to facilitate different processes and workflows in SDIs are discussed.

The modeling tool might be important for data providers whereas the mapping tool is needed by users and data custodians. In SDIs, the mapping tool needs to be usable for laymen, too. Possibly the first question when starting the service could be whether the basic or the advanced user interface is requested. The advanced user interface offers more functionalities concerning the modeling and the schemas of standards needed.

This would make the SDI independent e.g. of incompatible data formats. The interesting layer or part of the data set could be extracted.

6.4.1 Introduction to Geo Web Services

As mentioned in Section 2.3.1 Geo Web Services support the user within the enabling platform of SDIs to deal, process, visualize or process spatial data. There are technical and organizational issues revolving around the provision of Geo Web Services. It is useful to distinguish between basic service and end-user service for SDI (Najar et al, 2004a):

The former provide fundamental services and ensure the basic functioning of an SDI. This might be a search engine for retrieving spatial data within the infrastructure, maintaining databases and data quality. End-user services are usually needed by spatial data providers, companies or institutions that offer spatial data on the Web and end-users, laymen or experts that use the infrastructure via Web. Basic Services should be provided by a neutral organization.

---

[2] Certain information in the following section is taken from a diploma (master) thesis and a semester paper, which were realized under the supervision of and in collaboration with the author. For further details please check ((Henrich 2003), (Hübner 2004))

On the other hand end-user services are more complex in functionality. They can serve different purposes and may chain various basic services to a complex service. They are often focused on users, possibly laymen, who are looking for information in the SDI and need answers to spatial questions. Therefore, end-user services may be provided by companies or different institutions.

6.4.2 Demonstrator of Future Web Services

In general, it is known with regards to search engines that the design of user interfaces for searching and browsing for certain information needs improvement in usability. Basically, the users are not interested in search for its own sake but in answering a specific question (Hearst, 2002).

Also, the currently working Clearinghouses are not easy to handle and often do not answer the user's questions. A comparison of different Clearinghouses ((Crompvoets, 2004), (Henrich, 2003)) shows that there is a lack of interoperability between the different Web Services. This is confirmed by practical examples of trying to chain two standardized Web Services WMS and WFS (Hübner, 2004). It is therefore known that Web Services need to carry their own metadata to be able to identify themselves and be chainable with other services (Reference ISO 19119) On the other hand the result of a search, in which keywords are used, is a long list of metadata links which have to be opened and analyzed one by one. The data quality is heterogeneous and no samples of spatial data are provided. One of the limiting factors is identified to be the lack of consistent spatial data and metadata. Therefore, it is interesting to analyze how the integration of spatial-metadata and the concept of views as well as the prototypes contribute to the ideal situation.

The demonstrator in Figure 6-12 shows a possible Clearinghouse of the future from a user perspective and visualizes the possibilities and limits. It suggests a Clearinghouse that consists of three chained services and so fulfills the user's needs of flexibility: Navigator, Selector and Inspector. The Navigator is used for the rough search by defining the spatial parameter with different input methods. The Selector supports choosing the interesting spatial data from a list of visualized sample data with metadata description offered by the Navigator. The Inspector supervises the analyzing and processing of the spatial data.
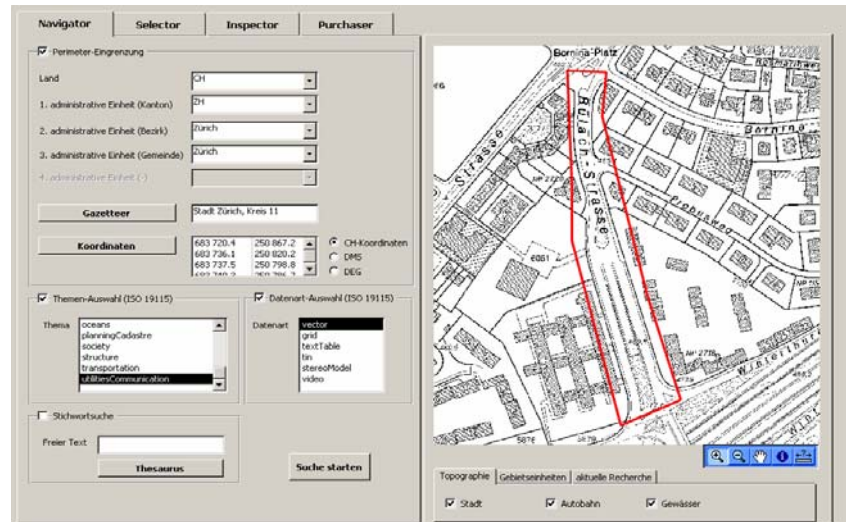
**Figure 6-12: Demonstrator for Clearinghouse as a Basic service
which can be chained to End-User service**

In a scenario for the future, the mapping and modeling tools would be basic services which work together with the Clearinghouse. The processes of modeling and updating and handling spatial data in an SDI could be supported by workflow modeling. The proper chaining of the different Geo Web services could happen by workflow modeling similar to (Li, in press (2005)).

The mapping service would extract views of metadata of good quality as updated together with the spatial data from integrated spatial data -metadata sets for the custodian and data provider of SDIs. The user of SDI has the benefit of spatial data samples during the Clearinghouse search, which are extracted from the integrated data by the mapping services. When the user decides for one of the spatial data sets he or she can than specify in which data format it is needed and whether only certain layers or the whole set is of interest, this might work similar to the demonstrator in Figure 6-13.

Figure 6-13: Demonstrator for mapping tool, different possibilities to select extract of data set

The modeling tool could support the Clearinghouse by modeling not only the metadata-spatial data commonly but also the service metadata which can then be exported as XML.

6.5 Summary

Chapter 6 implements two prototypes which support the Concepts of Integration and Views and demonstrators which discuss future possibilities of Web services using integrated metadata-spatial data sets.

The first prototype is an existing open source modeling software called INTERLIS/UML Editor, which was created to support the use of the model-driven-approach. It combines the text-based INTERLIS with the graphical UML. The extension of the INTERLIS/UML Editor implements the principles for metadata-spatial data integration from Section 5.3. To illustrate, one new functionality is that certain metadata are added automatically when a new model is generated. Consequently common modeling for spatial data and metadata is supported in a harmonized way.

The second prototype implementation explores the possibilities of creating views and functionalities of views in the relational database management system Oracle 9i (Oracle, 2005). Views according to different profiles of ISO 19115 are extracted from integrated data sets.

Finally, a demonstrator-program visualizes how common metadata-spatial data management and the full implementation of the two concepts will facilitate certain processes for SDIs in the future.

Chapter 7 Discussion of Results

7.1 Summary of Results

Altogether, the objectives of the thesis have been fulfilled and the concept of commonly managing metadata and spatial data has been proved to be possible. Moreover, this concepts offers advantages by achieving better metadata and interoperability between metadata as well as spatial data. Therefore, the following statements summarize the results of this thesis and are discussed from the point of view of concept (contribution to research), implementation (proof and verification of concept) as well as future research:

- **The feasibility of integrating, managing and modeling <u>existing</u> metadata and spatial data commonly is realistic.**

Existing spatial data sets and models have been completed with their metadata description in the case study (Section 5.1). A common management for spatial data and metadata is supported by common models. These are generated in the modeling tool prototype, for example by importing an existing spatial data model into the program, which is automatically completed with general metadata (Section 6.2). For flexible handling of an integrated metadata-spatial data set, views with different excerpts of the comprehensive data set are needed. For this reason, the prototype for the data mapping process was created (Section 6.3).

The interface or link between the prototype modeling tool and a metadata acquisition tool remains to be implemented. The advantage of the coupling would be that the special requirements of the respective model and data set can be taken into account in the metadata acquisition, i.e. by transferring the needed elements to the acquisition tool. Also, there is so far only an indirect connection (using several conversion tools) between the model in the prototype and the database with the actual information. A direct interface or chaining of processes, would be important for deriving certain metadata attributes from spatial data.

- **It is possible to model and produce <u>new</u> metadata and spatial data in the same working process.**

A common framework for planning and structuring metadata and spatial data with a common modeling tool has been created for new data sets, too (Section 4.2). Models treat metadata and spatial data equally and the user does not have to differentiate between metadata and spatial data.

This consequently promotes simultaneous acquisition of metadata and spatial data as well as the possibility of automatically deriving implicitly and explicitly metadata from spatial data during the acquisition process.

New data sets are planned, modeled and updated with metadata in the INTERLIS-UML prototype from the beginning (Section 6.2). The prototype allows a modeler to start a new conceptual model and create spatial data classes, according to his or her external demands and duties. The general metadata is added automatically during the modeling work process. The modeler has to merely change or add specialized metadata and specify which metadata can be derived from spatial data.

In future, a wizard software could assist the user in modeling new spatial data and metadata in the prototype. If certain libraries and rules for modeling were implemented, this would not only support common management but also consistent modeling and the general modeling of spatial data as such. At the moment, it is not possible to rely on the existence of explicit formal descriptions of conceptual models for every spatial data set. As a result, the explicit formal description of data models is promoted.

- **The flexible creation of views is possible. It enables the possibility of supporting different metadata schema (e.g. ISO 19115 profiles), without changing the original data.**

The concept of generating views for comprehensive (metadata-spatial) data sets without changing the original structure is of advantage for certain environments (such as SDI and distributed GIS) in which multiple applications or users (laymen as well as experts) need different versions or excerpts of the same data for various purposes (Section 4.3).

The implementation of the mapping process to create views from the integrated metadata-spatial data set has shown, that in general, the flexible creation of views is possible (Section 6.3). The semantic mapping between two models on conceptual level is feasible, but has to be solved manually by an expert. Once this mapping pattern is created, e.g. by using the modeling prototype from Section 6.2, data can be mapped, too (Section 6.3).

The prototype implementation of the mapping process has been executed on the logical database level, which was enhanced with information from the conceptual modeling level . This was possible, by using a specialized tool (ILI2ORA) which was specifically developed for this application. It saves the information from the conceptual model in separate tables, which can be used for retranslating the data back to the conceptual level (Section 6.3.3).

Certain problems, which are encountered when creating views for different profiles of one standard were discussed in Section 6.3.4. In summary the result is, that profiles of ISO 19115 are in general more compatible to each other, than crosswalks between different standards. Nevertheless, certain implementation details, e.g. 'how is the package multi-linguism used in a certain profile?', have to be verified before views can be created. Thus, the statement by (Nogueras-Iso et al., 2004) that crosswalks between different metadata standards are instable, is also true for certain packages of ISO profiles.

Since working on the logical level means loosing important structure information, a semantic mapping tool which works mainly on the conceptual level and is able to add new, restructured data is missing. It should be developed for further implementation of SDI interoperability, in future.

- **It is possible to integrate metadata in different spatial data models and repositories according to a set of general principles. They define where to put which metadata.**

Section 5.3 offers a set of principles for metadata integration, which were derived from the experience of the case study and discussion with experts. The rules are categorized in object-oriented and relational paradigms (Section 5.3.2 and 5.3.3).

In order to be able to integrate metadata in a time- and size-efficient way, the author suggests in Section 5.3.1 that two categories of metadata are defined and inserted:

– the *general, automatic metadata* which describe any spatial data set in general terms

– the *specific, manual metadata* description which has to be created by an expert who knows the characteristics of the data set well and can e.g. therefore decide which metadata can be replaced by spatial data.

Both groups are implemented in the modeling prototype (Section 6.2). The general group of metadata are assorted elements from the Swiss ISO 19115 profile. The specific metadata are offered in a pop-up assistant window and are also thematically grouped attributes from ISO 19115.

In future, the specific metadata attributes could be visualized and grouped according to type of model. This effort would be supported by the existence of more standardized models. At the moment the selection window is not user friendly and for people, who are not used to the thematic structure of ISO, difficult to work with.

This thesis looks at spatial data and metadata and the rules apply to mainly spatial vector data. It might be a research topic of the future to check these principles on other, non-spatial, data (e.g. for library catalogues) and perhaps another metadata standard (e.g. Dublin Core).

7.2   Discussion of Results and Conclusion

The following sections discuss various topics which are related to the concept of common metadata-spatial data management or views and which might be influenced in a relevant way by the concepts and results of this thesis.

7.2.1   Discussion about the Automatic Integration of Metadata

The thesis has shown that the automatic integration of general metadata combined with a manual integration of specialized metadata covers the necessary metadata description for a certain topic and is thus sensible (Section 6.2.3) and feasible (Section 6.2.2).

The degree of automatization of the integration depends on the aim of the integration. In the case of this thesis, the prerequisite was to not enlarge the model and data set unnecessarily and to provide adequate metadata description for interoperability in a networked environment (Section 1.4). Accordingly, the metadata description penetrates until object-level, and, if possible, metadata are replaced or derived from spatial data.

Figure 7-1 discusses in a qualitative overview the different approaches for metadata integration and visualizes the advantages and disadvantages for each solution. The **horizontal axis represents the degree of differentiation specialization of metadata**: it answers the question how specialized the metadata description is to the characteristics of a certain data set and with which level of detail the semantic translation and analysis of metadata attributes has been executed. The **first orange vertical axis represents the degree of integration effort**: it shows how much manual labor for metadata integration is necessary or whether the integration can be programmed to work automatically. This is related to the effort invested in structural analysis and hierarchical integration. It also has an effect on the effort for processing spatial data. This is shown on the **blue dashed vertical axis**: If metadata description for the spatial objects are general and automatic, certain steps in e.g. geospatial analyzing steps might process the wrong information. Consequently, more effort needs to be invested in exploitation. The more specific and thus significant the metadata describe the spatial data, the less costs and time needs to be placed in the exploitation.

Point 1 stands for the approach "place all metadata attributes on the highest hierarchy level of the model". In object-oriented terminology it would mean creating a single, large "metadata package/class" from which all the metadata might be inherited to lower hierarchy levels. In this case, the metadata description would not change for special or modified parts of the model (for example a certain part of the data set varies in its data quality). Consequently, model and data would be enlarged equally and there would be a higher risk of misinterpretation by the user.

This integration approach is almost equivalent in terms of organizational effort and quality of description to the current situation of strictly separating spatial data and metadata management.
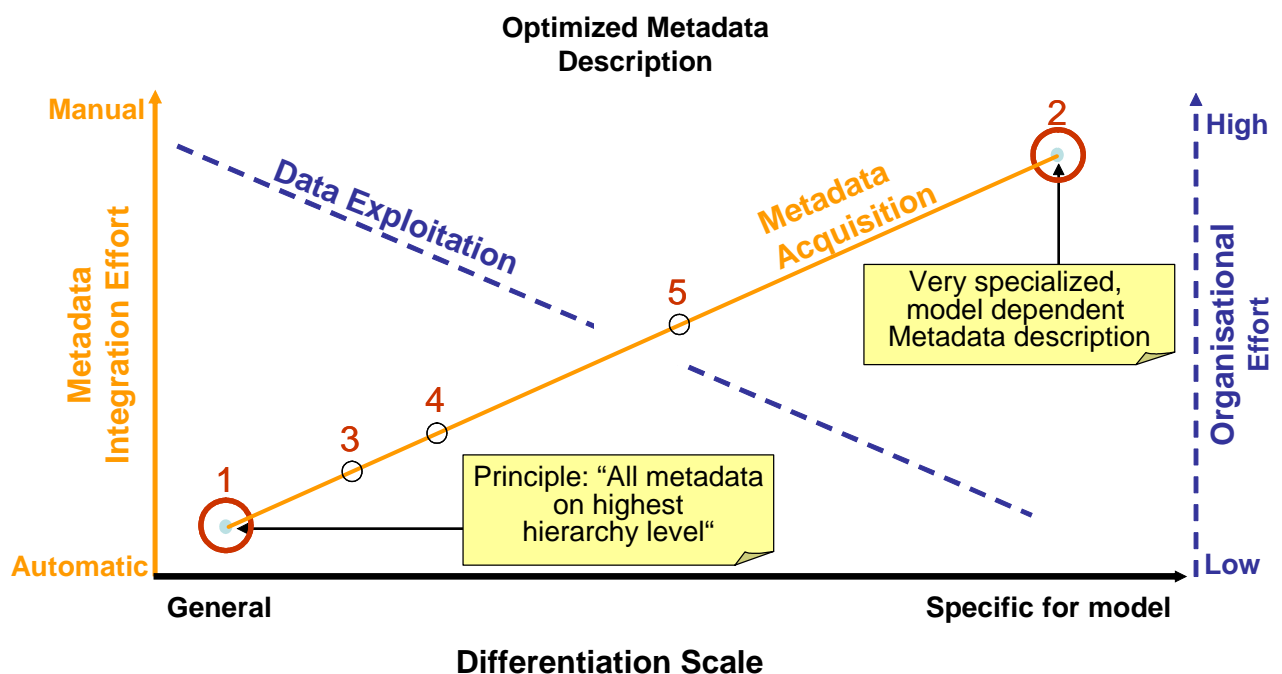
Figure 7-1: Qualitative discussion of the optimized metadata description for spatial data
(Najar and Giger, in process)

Point 2 reflects the other extreme: every metadata attribute from a certain spatial data model is semantically and structurally compared to the metadata standard and an expert decides which metadata attributes need to be newly inserted and which existing spatial data attributes can be used as metadata.

This is the procedure that was used for the case study. Thus, only the necessary metadata attributes are inserted at the ideal place in the model and the metadata description is highly specialized on the individual data set. This causes much effort and takes a lot of time.

In terms of spatial data processing and interoperability, the precise, specialized metadata description in Point 2 optimally supports the integration of spatial information from different sources and the geospatial analysis on object level. If the original data set is subdivided, by e.g. geospatial analysis and only one layer is needed, then this layer can be provided with metadata without reformatting. A specialized metadata description down to the spatial data objects increases the flexibility of processing, e.g. by use of Web services.

Point 3, Point 4 and Point 5 show different stages of the metadata integration in a certain data set model. The higher the points are defined on the graph of metadata acquisition, the more specialized the metadata are and the more effort was invested in semantic and structural analysis of the model.

The **optimized metadata description** depends on the interests, politics, investment and goals of the integration. Accordingly, the optimum will lie somewhere on the orange graph of metadata acquisition. In the case of this thesis an optimized description would be close to a virtual intersection of the two graphs. This means a moderately precise metadata description which can be partly inserted automatically. Furthermore, this choice suggests an automatically inserted group of metadata, which are generally applicable and a specific, manually inserted group of metadata which are chosen by experts for the integration.

The advantage of the optimized metadata description is that it finds a balance between a precise metadata description by manual integration and the sensible degree of automatization which is needed for leveling organizational and processing efforts.

7.2.2 Organizational Aspects of Metadata Acquisition

With regards to the metadata acquisition, Section 1.2.1 mentions, that large organizations currently have the problem that it is not always clear who is in charge for the metadata acquisition. The possibilities of a decentralized acquisition in which every unit is in charge of its own metadata as well as a centralized organization where a central IT or GIS unit creates metadata for the other units have been discussed.

The integrated management of spatial data and metadata which uses clear general rules for the integration help the management and acquisition of metadata in large organizations or companies.

The acquisition and update responsibilities are divided according to the separation of general and specific metadata integration. In the case of commonly management metadata and spatial data, the expert in the specialized unit is in charge for the specific, data set-dependent metadata, which need a lot of expert knowledge, whereas the general metadata

can be provided by templates and automatic tools by the general, centralized unit. As a result, the metadata is managed by two different instances in an organization, but the repository is a common metadata- spatial data database.

As ISO 19115 suggests the levels series, data set, feature etc., a confusion was created on which hierarchy level to acquire metadata. This would be solved by dividing the responsibility of acquisition and inheriting the general metadata to the lower levels, whereas the specific metadata are already acquired on the necessary level.

This is supported by the linkage between modeling tool and acquisition tool (Section 7.1). The idea is that the metadata acquisition tool is connected to the modeling tool and the model is read, like a step-by-step instruction recipe. The common model defines what metadata are required for the existing data set and where to put them in the data file or data base. Thus, the metadata acquisition tool will ask the expert to acquire the right metadata information and put the typed in information in the right place in the database.

7.2.3 Contribution to Interoperability in SDIs

The task of semantic mapping has been identified in literature as one of the main challenges of integrating different geographic databases and their schema ((Mustière et al, 2004), (Schwering and Hart, 2004), (Section 3.2.2). In this thesis, semantic mapping has been extended by using it on the integration of metadata and standardized spatial data models (Section 5.1).

The following mapping combinations have been executed:

- Standardized spatial data model (cadastral)⇔Standardized Metadata Model

- Semi-standardized spatial data model (water supply) ⇔ Standardized Metadata Model

- Not-standardized spatial data model (environmental) ⇔ Standardized Metadata Model

- Comprehensive metadata-spatial data set (views)⇒ Profile of Metadata Model

The method of semantic mapping has thus been tested on metadata models and the feasibility of creating views.

Two steps of the process of semantic mapping have been identified (Section 4): **semantic translation** and **semantic transformation**. The former translate elements of one model to corresponding elements of the second model on conceptual level, resulting in encoding rules. The latter step realizes the encoding rules by transforming the model and data structures according to the translation.

The semantic transformation has been realized for all examples on the logical or semi-conceptual level (Section 5.1.2 and Section 6.3 in the prototype for mapping process). Due to missing semantic information, the current mapping prototype requires the integrated metadata to be marked (e.g. by a character *, so that a parsing tool can find the metadata).

Possibly this extra documentation could be realized with the modeling prototype by implementing a simple lineage recorder. As the metadata inserted into the model and data by using the general principles are standardized, the changes in the models need to be recorded only for the specific metadata. Whenever the expert modifies metadata in the modeling prototype it should be documented automatically by the program and saved in the metadata element "lineage".

By creating appropriate tools for semantic transformation, which can work with a conceptual modeling language and thus support, read and interpret the semantic translation, the actual transformation can be automated.

Hence, the **technical interoperability** between two models, in this case metadata model and spatial data model can be solved with appropriate tools or Web services.

In future, existing tools can be extended in functionality and improved in usability for this purpose. Also tutorials or instructions need to be provided so that the user knows in which workflow they can be applied. So, technical interoperability can be solved on conceptual level, and, provided that the right tools are developed, in an automatic way and for a broad use in the public or SDI.

On the other hand the semantic translation, which generates encoding rules between two models needs interpretation of the semantics and human knowledge. Even if a model is standardized, it does not give enough semantic guidance: In the Swiss cadastral model it is for example defined that the location of a (water) source has to be mapped, but it is not declared whether it is mapped as a point or an area.

Therefore, **semantic interoperability** needs human intelligence and interaction for semantic mapping. The reason is that the model description usually does not give enough semantic information and the information is only partly computer processable. Also, it depends on the modeling language used.

In future, (Web) services are going to have more functionality and the user is going to need less experience in dealing with spatial data and GIS. Good metadata management, which in this case means metadata-spatial data integration, is the key to this kind of interoperability as it provides the possibility for different services to be found for a certain task by a Clearinghouse and provides the capability of describing spatial data in an adequate way so

that the Geo Web Services can process them. Thereby, a spatial question of a non-expert can be answered without him or her dealing with the actual spatial data or their current file formats. Also, the normal users does not need to bother according to which standard the metadata is acquired.

7.2.4 Flexibility by Means of View Concept

At the moment, views are manually created on the conceptual model level and realized on the logical level (Section 5.1 and 6.3). The goal for the future is to be able to support the conceptual level, so that the mapping can be realized in a graphical, semi-automatic way. This could happen by visualizing the models and supplying the possibility of linking the respective elements by arrows graphically. It should be possible to translate this graphical information (like in UML) in encoding rules with an interface to the data. Finally, the data restructuring and completion on the logical level should be initiated, managed and guided by the conceptual level.

In order to be able to implement this view concept-tool, the basic precondition is that data models exist for spatial data and that current conceptual modeling languages support the view on conceptual level. Currently, there are some modeling languages, e.g. INTERLIS 2, that provide some simple functionalities for describing views as part of the model, but they are not powerful enough to support for example metadata crosswalks. These functionalities in modeling languages should be extended as suggested in Morf et al (2004).

The Model Driven Architecture (MDA, Section 2.6.1) works partly along the same lines as the view concept, as it forces the custodian to model the data and requires checking and transfer tools for conversion between the models on the different levels.

The **view concept** can be applied to integrated metadata-spatial data and is important for SDIs. Metadata for different purposes (machine-processible, human-readible, different (Web) formats) can be created without acquiring them several times. Flexibility of metadata schema: it will not be necessary to convert all the old metadata catalogues in manual way, instead one script will be able to convert the metadata schemas in an ad-hoc way when needed in another schema. This also means that the information communities don't have to all agree on one metadata schema. It is better to equip the view- tool with sufficient conversion scripts to cope with different metadata schemas.

In the a further development of the concept, issues of security can be solved: for example if a data provider wants to regulate certain steps of progressive availability for different clients. Possibly an administration would like to offer other administrations full access to all the spatial data, but external users from private companies should be able to access only certain layers which are of no concern to privacy issues.

7.2.5 Contribution to Quality Issues

The integration concept improves data quality as only one data repository needs to be updated and managed and consistency will be improved.

Also, due to the bottom-up method, there will be an information about data quality on every hierarchy level and thus for every object.

Therefore, the comprehensive management of metadata and spatial data makes the objects more interoperable and usable for the right purposes.

In general, object-oriented concepts support the metadata integration in a better way than relational modeling paradigms. The reason is that classes are organized by definition into groups with common characteristics. It is possible to find large tables with a collection of miscellaneous attributes in relational modeling. These are difficult to provide with a precise metadata description, e.g. give a title or abstract, as they are too heterogeneous in content.

Object-oriented metadata-spatial data integration might be of advantage for institutions that cannot afford metadata acquisition down to the object: inheritance in object oriented, integrated metadata-spatial model will take care of this. The integrated data-metadata set inherently provides a metadata description down to object level.

For example, if someone wants to overlay or intersect spatial data of high quality with spatial data of low quality, metadata information on object level can be passed on after e.g. the merge and is shown in the lineage (part of the core data set in GM03) or shown as a range of data quality.

When the quality of spatial data changes within objects of the same kind (e.g. a river in the environment data of the case study is constituted by different segments which have each a different quality) the integrated metadata change with the object.

Since part of the integrated metadata gives information about the data quality, the view concept can help when data of different quality is combined, overlaid or intersected.

7.3 Outlook

With respect to metadata handling, the next goal is to develop tools and services that allow software to access and deal with metadata without a non-expert user needing to understand what metadata are.

Metadata will be processed in the future by tools, e.g. Web Services and a few experts, who e.g. know how to deal with ISO standards. These experts are in charge for the interpretation and correct use of the relevant metadata schemas. The tools provide an efficient user interface for acquiring metadata in a correct way, e.g. scroll down menus, templates and inheritance of meta-information to lower levels of the data set. In future only the rendered, readable metadata will be accessed by normal users, e.g. on Web pages.

This nevertheless does not mean that metadata standards loose importance, but instead shift in purpose. In future, they will be interpreted as an interface definition, only.

Another field of research is the coupling and modeling of metadata-spatial data workflow and management technologies This could use UML modeling tools to improve the interoperability between different services which can read and interpret the workflow an thus be coupled to support the management process is another field of future research. It would refine and improve the common management workflow and link it to the necessary services. This could be the basis for different transaction in an SDI of the future. Furthermore, organizational elements (such as responsibilities who is in charge for which update and acquisition of which data set) as well as financial, economic or responsibilities (how much do certain data sets cost and statistics of how much have been sold) could be included in the workflow and would thus support internal coordination within an organization.

Standardized models are important as long as the model driven approach has not been fully implemented, meaning the necessary tools do not exist in an interoperable way. Notwithstanding, object-oriented, standardized models support the reproducibility of metadata integration. Metadata, which are derived implicitly or explicitly from spatial data can be determined once and then reused. Therefore, future standardization of models for certain topics is important. Linked to this discussion the question arises which modeling language should be used. This thesis has shown, that the main criteria of decision is the existence of tools to support the usability and interoperability of the modeling language. In the area of developing tools and interfaces for modeling languages, there is a lot of future research and development necessary.

Integrated metadata for Web services is also a topic for future discussion. The question is whether it makes sense for Web services to carry their metadata in a "self describing way", similar to the integrated spatial data. This might support the chaining of different services. Also, it would be interesting to investigate whether modeling tools, e.g. UML Editor can support integrated Web services to manage their metadata description commonly with their XML code-description.

References

AMADIO, G., CANNAFOGLIA, C., CORONGIU, M., DESIDERI, M., ROSSI, M. (2004). IntesaGIS: the basis for a National Spatial Data Infrastructure. 10th EC-GI& GIS Workshop, ESDI: The State of the Art, Warsaw, Poland.

A/M/T SOFTWARE SERVICE AG, Software GEOS Pro, (Accessed: Dec. 2005), http://www.amt.ch/

ANZLIC - THE SPATIAL INFORMATION COUNCIL, Australian Spatial Data Directory (ASDD), (Accessed: Dec. 2005), http://asdd.ga.gov.au/asdd/

AUSTRALIA AND NEW ZEALAND INFORMATION COUNCIL (ANZLIC), The Spatial Information Council, (Accessed: Dec 2005), http://www.anzlic.org.au/

ASKEW, D., EVANS, S., MATTHEWS, R., SWANTON, P. (2005). "MAGIC: a geoportal for the English countryside." Computer, Environment ad Urban Systems 29: pp.71-85.

BABCOOK GOVE, P., Ed. (1993). Webster's Third New International Dictionary, unabridged. Springfield, Massachusetts, Miriam-Webster Inc.

BALANCHE, R. (2005). Integration der Daten der Landeskarte 1:25'000 (VECTOR25) ins Datenmodell der Amtlichen Vermessung (DM.01-AV-CH). Interoperabilität für die breite Nutzung von Geoinformation, Zurich, Switzerland.

BALFANZ, D. (2002a). Automated geodata analysis and metadata generation. SPIE conference on visualization and data analysis (VDA), San Jose, USA.

BALFANZ, D. (2002b). Visuell unterstützte, automatisierte Generierung und Nutzung von Geo-Metadaten, Thesis. Darmstadt, Fraunhofer IRB Verlag.

BALFANZ, D., GÖBEL, S. (1999). Bridging Geospatial Metadata Standards towards Distributed Metadata Information Systems. IEEE Computer Society: 3rd IEEE Metadata Conference, Bethesda, Maryland.

BEARD, K. (1996). A structure for Organizing Metadata Collection. 3rd International Conference/Workshop on Integrating GIS ans Environmental Modelling.

BEAUMONT, P., LONGELY, P.A., MAGUIRE, D.J. (2005). "Geographic information portals- a UK perspective." Computer, Environment ad Urban Systems 29: pp. 49-69.

BEDARD, Y., LARIVEE, S., PROULX, M-J., NADEAU (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. COMOGIS Workshops ER2004, Shanghai, China.

BÉDARD, Y., PAGEAU, J., CARON, C., (1992). Spatial Data Modeling: The Modul-R Formalism and Case Technology. ISPRS Symposium, Washington, United States.

BÉDARD, Y., PAQUETTE, F., (1989). Extending Entity/Relationship formalism for spatial information systems. AutoCarto9, Baltimore.

BELUSSI, A., NEGRI,M., PELAGATTI, G. (2004a). GeoUML: A Geographic Conceptual Model defined through Specialization of ISO TC 211 Standards. Proceedings of 10th ECGI & GIS Workshop "ESDI-State of the Art", Warsaw, Poland.

BELUSSI, A., NEGRI,M., PELAGATTI, G. (2004b). GeoUML: an ISO TC211 compatible data model for the conceptual design of geographical databases. internal report, Dipartimento di Informatica, Università di Verona and Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy.

BERNARD, L., KANELLOPOULOS,I., ANNONI, A., SMITS, P. (2005). "The European geoportal- one step towards the establishment of a European Spatial Data Infrastructure." Computers, Environment ad Urban Systems 29: pp.15-31.

BILL, R. (1999). Grundlagen der Geo-Informationssysteme, Band 1: Hardware, Software und Daten. Heidelberg, Germany, Herbert Wichmann Verlag.

BISHR, Y. (1998). "Overcoming the semantic and other barriers to GIS interoperability." IJGIS 12(4): 299-314.

BISHR, Y., KUHN, W (2000). Ontology Based Modelling of Geospatial Information. 3rd AGILE Conference on Geographic Information Science, Helsinki, Finland.

BRINKHOFF, T. (2003). Geodienste im Rahmen der Architektur einer Geo-Suchmaschine. Münsteraner GI-Tage, Münster.

BORRERO, S. (2002). The GSDI Association: State of the Art. 6th GSDI Conference, Budapest, Hungary.

BRODEUR, J., BÉDARD, Y., MOULIN, B. (accepted for publication in March 2005). "A Geosemantic Proximity-Based Prototype for the Interoperability of Geospatial Data." Computers, Environment ad Urban Systems.

BUCHER, B.(2003). Translating User Needs for Geographic Information into Metadata Queries. AGILE, Lyon.

BULTERMAN, D. C. A. (2004). "Is it Time for a Moratorium on Metadata?" Multimedia, IEEE 11(04): 10-17.

CALHOUN, K. (2002). "Special section: metadata." Library Collections, Acquisitions & Technical Services 26: 195-197.

CADASTRAL SURVEYING UNIT OF SWISSTOPO, Data model for the official cadastral survey in Switzerland, (Accessed: Dec. 2005), http://www.cadastre.ch/en/target/interlis/

CENTER FOR INTERNATIONAL EARTH SCIENCE INFORMATION NETWORK (CIESIN) (2005). Data Model for Managing and Preserving Geospatial Electronic Records, Version 1.00. CIESIN's Guide to FGDC Compliant Metadata. Palisades, New York, Center for International Earth Science Information Network (CIESIN), Socioeconomic Data and Applications Center (SEDAC), Columbia University.

CHAN, T. O., FEENEY, M., RAJABIFARD, A., WILLIAMSON, I.P. (2001). "The Dynamic Nature of Spatial Data Infrastructures: A Method of Descriptive Classification." Geomatica 55(1): 451-462.

CHINNICI, R., MOREAU, J.-J., RYMAN, A., WEERAWARANA, S., Web Services Description Language (WSDL), Version 2.0 Part 1, Core Language, W3C Working Draft 3, (Accessed: Dec. 2005), http://www.w3.org/TR/wsdl20/

COLEMAN, D. J., MCLAUGHLIN, J.D. (1998). "Defining Global Geospatial Data Infrastructure (GGDI): components, stakeholders and interfaces." Geomatica 52(2): 129-144.

COORDINATION OF GEOGRAPHIC INFORMATION AND GEOGRAPHIC INFORMATION SYSTEMS (KOGIS) (Accessed: Dec. 2005), http://www.kogis.ch/,

COORDINATION OF GEOGRAPHIC INFORMATION AND GEOGRAPHIC INFORMATION SYSTEMS (KOGIS) (2004). GM03 - Metadatenmodell Ein Schweizer Metadatenmodell für Geodaten, Version 1.3, Final Draft

COULOMBRE, S., LIBOUREL, T., SPÉRY, L. (1998). Metadata and GIS, A classification of Metadata for GIS. 1st International Conference and Exhibition on Geographic Information, Lisbon, Portugal.

CROMPVOETS, J., BREGT, A., RAJABIFARD, A., WILLIAMSON, I. (2004). "Assessing the worldwide developments of national spatial clearinghouses." Int.J. Geographical Information Science 18(7): 665-689.

CROMPVOETS, J., BREGT, A. (2003). "World Status of National Spatial Data Clearinghouses." URISA Journal 15(Access and Participatory Approaches (APA).

CROSSWELL, P. L. (2000). Chapter 5: The role of standards in support of GDI. Geospatial Data Infrastructure, Concepts, Cases and Good Practice. R. Groot, McLaughlin, J. Oxford, Oxford University Press.

DAY, M., Mapping between metadata formats, (Accessed: November 2005), http://www.ukoln.ac.uk/metadata/interoperability/

DENNO, P. O., SANDERSON, D.B. (2000). "Structural Information Mapping with EXPRESS-X." To appear in: ACM Special Interest Group Management Data.

DIGITAL GEOGRAPHIC INFORMATION WORKING GROUP (DGIWG), Metadata Development Efforts, (Accessed: November 2005), http://metadata.dgiwg.org/

DONAUBAUER, A. (2005). A Multi-Vendor Spatial Data Infrastructure for Local Governments Based on OGC Web Services. FIG Working Week 2005 and GSDI 8, Kairo, Egypt.

DUBLIN CORE METADATA ELEMENT SET, Dublin Core Metadata Initiative, (Accessed: Dec. 2005), http://dublincore.org/

DUBLIN CORE METADATA INITIATIVE, Dublin Core Metadata Glossary, (Accessed: November 2005), http://library.csun.edu/mwoodley/dublincoreglossary.html

EARTH SCIENCE ENTERPRISE PROGRAM (ESE), Earth Observing System Data and Information System (EOSDIS) Core System Metadata, (Accessed: Dec. 2005), http://ecsinfo.gsfc.nasa.gov/

EIDGENÖSSISCHE VERMESSUNGSDIREKTION + ITV GEOMATIK AG, Metadatenmodell
GM03_Profi_AV, Anwendungs-Handbuch, (Accessed: Dec. 2005),
www.cadastre.ch/pub/down/publications/reports/manualGM03Profil_AV_V10_de.pdf

EGENHOFER, M. J. (1999). Introduction: Theory and Concepts. Interoperating Geographic
Information Systems. M. Goodchild, Egenhofer, M., Fegeas, R., Kottman, C. Boston,
Massachusetts, Kluwer Academic Publisher.

EISENHUT, C., GNÄGI, H.R., KELLER, S.F., SPÄNI, B. (2003). INTERLIS Grundkurs. Zürich.

EISENHUT, C., HIRZEL, P. UML/INTERLIS-Editor, (Accessed: Dec. 2005),
http://prdownloads.sourceforge.net/umleditor/umleditor-2.3.0.src.zip?download,

EUROPEAN UMBRELLA ORGANISATION FOR GEOGRAPHIC INFORMATION (EUROGI) (Accessed:
November 2005), http://www.eurogi.org/index_1024.html,

FACHSTELLE AV-DATENMODELLIERUNG UND -AUSTAUSCH (FADMA) (2001). Datenmodell 2001
der Amtlichen Vermessung - Erläuterungen. Wabern, Bundesamt für
Landestopographie, Eidg. Vermessungsdirektion: 1-43.

FARQUHAR, A., FIKES, R., RICE, J. (1997). "The ontolingua server: a tool for collaborative
ontology construction." International Journal of Human-Computer Studies 46: pp.707-
727.

FLANAGAN, D. (2002). Java in a Nutshell, Fourth Edition. Sebastopol, CA, O'Reilly Media, Inc.

FEDERAL GEOGRAPHIC COMMITTEE FGDC, Metadata Tool Surveys, (Accessed: Dec. 2005),
www.fgdc.gov/metadata/metatool.html

FEDERAL GEOGRAPHIC DATA COMMITTEE (FGDC), FGDC Content Standards for Digital
Geospatial Metadata (CSDGM), (Accessed: Dec. 2005),
http://www.fgdc.gov/metadata/meta_stand.html

FEDERAL OFFICE FOR WATER AND GEOLOGY (FOWG), Swiss water information system
GEWISS, (Accessed: Dec. 2005),
http://www.bwg.admin.ch/bwg/kontakt/zustaendig/e/gewaesserinformationssystem_g
ewiss.htm

FGDC Content Standards for Digital Geospatial Metadata (Accessed: 2005
www.fgdc.gov/metadata.contstan.html

FGDC (1997). Framework, introduction and guide. F. G. D. Committee. Washington: pp 106.

FILLI, R. (2005). Viewbasierte Erzeugung von Metadaten aus einem integrierten Geo- und
Metadatenbestand. Zürich, Swiss Federal Institute for Technology, diploma paper,
supervisors: C. Giger, C. Najar

FLANAGAN, D. (2002). Java in a Nutshell, Fourth Edition. Sebastopol, CA, O'Reilly Media, Inc.

FONSECA, F., EGENHOFER, M., DAVIS, C., CAMARA, G. (2002). "Semantic Granularity in
Ontology-Driven Geographic Information Systems." Annals of Mathematics and
Artificial Intelligence 36(1-2): 131-151.

FONSECA, F., EGENHOFER, AGOURIS, P. (2002). "Using Ontologies for Integrated Geographic Information Systems." Transactions in GIS 6(3).

FONSECA, F. T. (2001). Ontology-Driven Geographic Information Systems. The Graduate School,, The University of Maine.

FOTHERINGHAM, S., WEGENER, M. (2000). Spatial Models and GIS: New Potential and New Models, Taylor & Francis.

FOWLER, M. (2004). Model Driven Architecture.

FRANK, A. U., GRUM, E., VASSEUR, B. (2004). Procedure to Select the Best Dataset for a Task. GIScience 2004, Maryland, USA, Springer-Verlag.

FRANK, A. U. (2001). "Tiers of ontology and consistency constraints in geographic information system." IJGIS 15(7): 667-678.

GEOCAT, Swiss Metadata Catalogue by Coordination of geographic information and geographic information systems (KOGIS), (Accessed: Oct. 2005), www.geocat.ch

GIGER, C., NAJAR, C., LOIDOLD, M., KOLLMANN, T., FEEG, P. (2005). Rahmenkonzeption für die Geoinformationsinfrastruktur Hessen (GII Hessen). H. Z. f. D. Hessisches Landesamt für Bodenmanagement und Geoinformation.

GIGER, C., NAJAR, C. (2003). Ontology-Based Integration of Data and Metadata. 6th AGILE Conference on Geographic Information Science, Lyon, France.

GNÄGI, H. R., SPÄNI, B. (2002). Systemunabhängige Daten-und Anwendungs-Modellierung. Kommunale Geo-Informationssysteme, Basiswissen, Praxisberichte und Trends. Heidelberg, Herbert Wichmann Verlag.

GOLAY, F., GIGER, C., MORENI, C., NAJAR, C. (2003). Etude Préliminaire au Projet e-geo.ch, Ecole Polytechnique de Lausanne and Eidgenössisch Technische Hochschule Zürich.

GRONNO, R., SOLHEIM, I., SKOGAN, D. (2002). Experiences of UML-to-GML Encoding. Proceedings of the 5th AGILE Conference on Geographic Information Science, PALMA.

GRUBER, T. R. (1993). "A translation approach to portable ontologies." Knowledge Acquisition 5(2): 199-220.

GSDI 5 (2001). Resolutions of the 5th GSDI Conference, Cartagena, Columbia.

GUARINO, N. (1997). Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. Summer School on Information Extraction, Frascati, Springer Verlag.

HAKIMPOUR, F. (2003). Using Ontologies to Resolve Semantic Heterogeneity for Integrating Spatial Database Schemata. Mathematisch-naturwissenschaftliche Fakultät. Zürich, Switzerland, University of Zurich.

HALEVY, A. (2003). Data Integration: a Status Report. Proceedings of the German Database Conference, BTW, Leipzig, Germany.

HEARST, M., ELLIOTT, A., ENGLISH, J., SINHA, R., SWEARINGEN, K., YEE, K.-P. (2002). "Finding the Flow in Web Site Search." Communications 45: pp. 42-49.

HENRICH, S. (2003). Entwicklung eines Geodaten-Portals für eine Standortevaluation, ETH Zürich, diploma thesis, supervisors: C. Najar, C. Giger,

HESS, C., DE VRIES, M. (2004). From Models to Data: a Prototype Query Translator for the Cadastral Domain. Joint 'FIG Comission 7' and 'COST Action G9' Workshop on Standardization in the Cadastral Domain, Bamber, Germany.

HESSE, W. (2005). "Standard Data Models." Position 17(June/July).

HILL, L., JANÉE, G., DOLIN, R., FREW, J., LARSGAARD, M. (1999). "Collection Metadata Solutions for Digital Library Applications." Journal of the Amercan Society for Information Science 50 (13): 1169-1181.

HIRZEL, P. (2004). UML-Editor Referenzhandbuch the Art of modelling. INTERLIS2-UML-Editor.

HÜBNER, K. (2004). Geo Web Services, ETH Zürich, semester thesis, supervisors: C. Najar, M. Loidold, C. Giger, Zürich,

HUNTER, J. (2001). "Metanet-a metadata term thesaurus to enable semantic interoperability between metadata domains." Journal of Digital Information 1(8).

INFOGRIPS GMBH, infoGrips Conversion Development System - ICS, (Accessed: Dec 2005), http://www.infogrips.ch/

INFRASTRUCTURE FOR SPATIAL INFORMATION IN EUROPE (INSPIRE), INSPIRE, European Geo-Portal, (Accessed: Dec. 2005), http://eu-geoportal.jrc.it/

INTERGRAPH CORPORATION, GeoMedia, (Accessed: Dec.2005), http://www.intergraph.com/geomedia/

INTERLIS, INTERLIS, the Geo Language, (Accessed: Dec. 2005), www.interlis.ch

INTERNATIONAL CARTOGRAPHIC ASSOCIATION (ICA) (Accessed: Jan. 2006), http://ncl.sbs.ohio-state.edu/ica/home.html,

INTERNATIONAL FEDERATION OF SURVEYORS (FIG) (Accessed: Dec 2005), http://www.fig.net/,

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ISO, Information technology -- Open Distributed Processing -- Unified Modeling Language (UML) Version 1.4.2, ISO/IEC 19501, (Accessed: Dec 2005), http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=32620&scopelist=

ISO/TC46 INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ISO (1998). Information and documentation -- Information retrieval (Z39.50) -- Application service definition and protocol specification

ISO/TC211 INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ISO (2005). Geographic information — Services, ISO 19119

ISO/TC211 INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ISO (2001). Geographic information — Metadata, ISO 19115

ISO/TC211 INTERNATIONAL ORGANIZATION FOR STANDARDIZATION ISO, Technical Committee 211 Geographic Information - Geomatics, (Accessed: November 2005), www.isotc211.org

KASHYAP, V., SHETH, A. (1996). Semantic Heterogeneity in Global Information Systems: The Role of Metadata, Context and Ontologies. Cooperative Information Systems. M. Papazoglou, Schlageter, G.

KLEPPE, A., WARMER,J., BAST, W. (2003). MDA Explained, The Model Driven Architecture: Practice and Promise. Boston, USA, Addison-Wesley.

KOGIS (2003). INTERLIS 2- Reference Manual. KOGIS, (Accessed: Dec. 2005) http://www.interlis.ch/interlis2/download_e.php

KREGER, H. (2001). Web Services Conceptual Architecture (WSCA 1.0), IBM Software Group.

KRESSE, W., FADAIE, K. (2004). ISO Standards for Geographic Information, Springer-Verlag Berlin.

KUHN, W. (2003). "Semantic Reference Systems." Int. J. Geographical Information Science 17(5): 405-409.

LEMMEN, C., VAN OOSTEROM, P., ZEVENBERGEN, J., QUAK, W., VAN DER MOLEN, P. (2005). Further Progress in the Development of the Core Cadastral Domain Model. FIG Working Week 2005 and GSDI-8, Cairo, Egypt.

Lemmens, R., de Vries, M., Aditya, T. (2003). Semantic Extension of Geo Web Service Descriptions with Ontology Languages. Proceedings of the 6th AGILE, Lyon.

LI, S., COLEMAN, D. (2005). "Modeling distributed GIS data production workflow." Computers, Environment and Urban Systems 29 (4):pp. 401-424

LI, S., BOETTCHER, R., COLEMAN D.J. (1998). Geomatics workflow management requirements in a collaborative production environment. Proceedings of the American congress on surveying and mapping.

LONGELY, P. A., GOODCHILD, M.F., MAGUIRE, D.J., RHIND, D.W. (2001). Geographic Information Systems and Science. Chichester, John Wiley & Sons, Ltd.: pp. 154-161.

MANSO, M. A., NOGUERAS-ISO, J., BERNABÉ, M.A., ZARAZAGA-SORIA, F.J. (2004). Automatic Metadata Extraction from Geographic Information. 7th AGILE Conference on Geographic Information Science, Heraklion, Greece.

MASSER, I. (1998). The first generation of national geographic information strategies. 3rd Global Spatial Data Infrastructure Conference, Canberra, Australia.

MASSER, I. (1999). "All shapes and sizes: the first generation of national spatial data infrastructures." International Journal of GIS(13): 67-84.

MCFADDEN, F. R., HOFFER, J.A. (1994). Modern Database Management.

MCLAUGHLIN, J. D., NICHOLS, S.E. (1992). "Building a national spatial data infrastructure." Computing Canada 24.

METADATA ARCHITECTURE AND APPLICATION TEAM, Internaional Metadata System Tools, (Accessed: November 2005), http://www.sinica.edu.tw/~metadata/tool/tool-eng-frame.html

MILLER, J., MUKERJI, J. (2001). Model Driven Architecture (MDA)- A technical perspective.

MILLER, J., MUKERJI, J. (June 2003). MDA Guide Version 1.0.1. OMG.

MOELLERING, H. (1998). Evaluating spatial metadata processing software. OGRIP Metadata Working Group, Columbus, USA, Dept. of Geography, Ohio State University.

MUSTIÈRE, S., CLEACH, A., FORT, J. (2004). Unifying Databases: Experiments with a Hydrographic Database. 7th AGILE Conference on Geographic Information Science, Heraklion, Greece.

NAJAR, C., GIGER, C., GOLAY, F., MORENI, C., RIEDO, M. (2004a). Geo Web Services: An NSDI-embedded approach. 7th AGILE Conference on Geographic Information Science, Heraklion, Greece.

NAJAR, C., GIGER, C., GOLAY, F., MORENI, C., RIEDO, M. (2004b). Geodata does not Always Equal Geodata: Meet the Challenge of a GDI in the Swiss Federal System. ESDI State of the Art, 10th EC-GI&GIS Workshop, Warsaw, Poland.

NAJAR, C. GIGER, C. (2004). Studie zur Integration von Geodaten und Metadaten, 1.Zwischenbericht. Zurich, Swiss Federal Institute of Technology Zurich.

NAJAR, C., GIGER, C. (in process). "Spatial data and Metadata Integration, a Model Driven Approach." Computers, Environment and Urban Systems.

NAJAR, C., RAJABIFARD, A., WILLIAMSON, I.P., GIGER, C. (in process). "A Framework for Comparing Spatial Data Infrastructures on the basis of Web Services and Metadata Management, An Australian-Swiss Case Study."

NEBERT, D. (2001). Developing Spatial Data Infrastructures: The SDI Cookbook, V1.1.

NEBERT, D., FULLTON, J. (1995). Use of Z39.50 to search and retrieve geospatial data. Digital Libraries Conference.

NOGUERAS-ISO, J., F. J. ZARAZAGA-SORIA, J. LACASTA (2004). "Metadata standard interoperability: application in the geographic information domain." Computer, Environment ad Urban Systems 28(6): pp. 611-634.

OBJECT MANAGEMENT GROUP (OMG), Website of Object Management Group (OMG), (Accessed: November 2005), www.omg.org

OBJECT MANAGEMENT GROUP (OMG) (2003). UML Resource Package.

OESTEREICH, B. (2001). Objektorientierte Softwareentwicklung. München, Oldenbourg Verlag.

ONSRUD, H., CAMARA, G.,CAMPBELL,J., SHARAD CHAKRAVARTHY, N. (2004). Public Commons of Geographic Data: Research and Development Challenges. Geographic Information Science, Springer-Verlag.

ORACLE9I SPATIAL, Shape to SDO Converter, Documentation, (Accessed: Dec. 2005), http://www.oracle.com/technology/software/products/spatial/files/text_files/using_shp2sdo.txt

ORACLE, Website of Oracle DBMS, (Accessed: Dec. 2005), http://www.oracle.com/index.html

OUKSEL, A. M., SHETH, A. (1999). "Semantic Interoperability in Global Information Systems: A Brief Introduction to the Research Area and the Special Section." Sigmond Record 28.

OOSTEROM, V., P., LEMMEN, C. (2002). Impact Analysis of Recent Geo-ICT developments on Cadastral Systems. FIG XXII Congress, Washington, USA.

OOSTEROM, V., P., LEMMEN, C., VAN DER MOLEN, P. (2004). Remarks and Observations related to the further development of Core Cadastral Domain Model. Joint 'FIG Commision 7' and 'COST G9' Workshop on Standardization of the Cadastral Domain, Bamberg, Germany.

OPEN GEOSPATIAL CONSORTIUM (OGC) OpenGIS Guide, (Accessed: Dec. 2005), www.opengis.org, today: www.opengeospatial.org,

OPEN GEOSPATIAL CONSORTIUM (OGC), OpenGIS® Specifications, (Accessed: Dec 2005), http://www.opengeospatial.org/specs/?page=specs

PENG, Z.-R., TSOU, M.-H. (2003). Internet GIS. Hoboken, New Jersey, John Wiley&Sons Inc.

PERMANENT COMMITTEE ON GIS INFRASTRUCTURE FOR ASIA AND THE PACIFIC (PCGIAP), A Spatial Data Infrastructure for the Asia and the Pacific Region, No.1, (Accessed: Dec. 2005), http://www.gsi.go.jp/PCGIAP/tech_paprs/apsdi_cnts.htm

PERMANENT COMMITTEE ON SDI FOR THE AMERICAS (PCIDEA), The 10th Meeting on the PCGIAP, Report from PCIDEA, (Accessed: Jan.2006), http://www.gsi.go.jp/PCGIAP/bangalore/bangalore_pcidearep.htm

POSTGRESQL, Website of DBMS PostgreSQL, (Accessed: Dec. 2005), www.postgresql.org

RAJABIFARD, A. (2002). Diffusion of Regional Spatial Data Infrastructures: with particular reference to Asia and the Pacific. Department of Geomatics. Melbourne, University of Melbourne.

RAJABIFARD, A., FEENEY, M-E., WILLIAMSON, I. (2002). "Future directions for SDI development." International Journal of Applied Earth Observation and Geoinformation 4: 11-22.

RAJABIFARD, A., ESCOBAR, F. AND WILLIAMSON, I.P. (2000). "Hierarchical Reasoning Applied to Spatial Data Infrastructures." Cartography Journal 29(2).

RAJABIFARD, A., WILLIAMSON, I.P., HOLLAND, P., JOHNSTONE, G (2000). From Local to Global
SDI initiatives: a pyramid building blocks. Proceedings of the 4th Global Spatial Data
Infrastructures Conferences, Cape Town, South Africa.

RAJABIFARD, A., CHAN, T. O., WILLIAMSON, I. P. (1999). The Nature of Regional Spatial Data
Infrastructure. Proceedings of the AURISA 99, Blue Mountains, NSW, Australia.

RÉGION WALLONNE, PORTAIL CARTOGRAPHIQUE DE LA RÉGION WALLONNE, (Accessed: Dec.
2005),
http://cartographie.wallonie.be/PortailCarto/Maquette/admin_prof/document.php

REED, C. (2005). "Interoperability, standards and the geospatial industry." GIS Development
9(5): 16-17.

REINHARDT, W., JOOS, G., KUHLMANN, H., MÜLLER-HERMES, E., SCHMITT, B., SEEBERGER, S.,
WALTER, V. (2004). Raumbezogene Informationssysteme. Heidelberg, Herbert
Wichmann Verlag.

RODRIGUEZ, M. A., EGENHOFER, M.J. (2004). "Comparing geospatial entity classes: an
asymmetric and context-dependent similiarity measure." IJGIS 18(3): pp.229-256.

RUMBAUGH, J., JACOBSON, I., BOOCH, G. (2005). The Unified Modeling Language Reference
Manual. Boston, Addison-Wesley.

SAFE SOFTWARE, FME SUITE (The Feature Manipulation Engine), (Accessed: Dec. 2005),
http://www.safe.com/

SCHWERING, A., HART, G. (2004). A Case Study for Semantic Translation of the Water
Framework Directive and a Topographic Database. 7th AGILE Conference on
Geographic Information Science, Heraklion, Greece.

SCHWERING, A., WAGNER, R.M., SCHNEIDER, B. (2003). A Geographic Search Ways to
Structure Information on the Web. Proceedings of the 6th AGILE, Lyon.

SEIFERT, T., BENEKEN, G., BAEHR, N. (2004). Engineering Long-Lived Applications using
MDA. IASTED International Conference on Software Engineering and Applications,
Cambridge, MA.

SHETH, A. (1999). Changing Focus on Interoperability in Information Systems: from System
Syntax, structure to Semantics. Interoperating Geographic Information Systems.
Norwell, MA, Kluwer Academic Publishers.

SHI, W. (2004). Zum Modellbasierten Austausch von Geodaten auf Basis XML. Fakultät für
Bauingnenieur- und Vermessungswesen. Munich, Germany, Universität der
Bundeswehr München.

SMITH, T. R. (1997). "Meta-information in digital libraries." International Journal on Digital
Libraries 1: 105-107.

SPÉRY, L., CLARAMUNT, C., LIBOUREL, T. (2001). "A Spatio-Temporal Model for the
Manipulation of Lineage Metadata." GeoInformatica 5(1): 51-70.

SPYNS, P., MEERSMAN, R., JARRAR, M. (2002). "Data modelling versus Ontology engineering." ACM SIGMOD Record, SPECIAL ISSUE: Special section on semantic web and data management 31(4): pp.12-17.

STOPPER, R. (2004). Demonstrator für einen Geodatendienst und Untersuchung zur Integration von Geodaten und Metadaten. semester thesis, supervisors: C. Najar, C. Giger, Zürich, ETH Zürich.

SUN DEVELOPER NETWORK (SDN), Java Technology, (Accessed: Dec. 2005), http://java.sun.com/

SWISSTOPO, Statistics of cadastral data coverage in Switzerland, (Accessed: October 2005), www.cadastre.ch/en/av/stat.html

SWISSTOPO, Web Page of the Swiss National Mapping Agency, (Accessed: Dec. 2005), www.swisstopo.ch

TURNER, T., What is metadata? (Accessed: Nov. 2005), http://www.library.cornell.edu/staffweb/kaleidoscope/volume10/feb02.pdf

UMWELTDATENKATALOG (UDK) (Accessed: Dez. 2005), http://www.umweltdatenkatalog.de/wwwudk/V-UDKServlet,

VADUVA, A., DITTRICH, K.R. (2001). Metadata Management for Data Warehousing: Between Vision and Reality. Database Engineering & Applications, International Symposium on., Grenoble, France, IEEE.

VAN DEN BERG, B., HULSHOF, B., TIJSSEN, T., VAN OOSTEROM, P. (2005). Harmonisation of distributed geographic datasets - A model driven approach for geotechnical & footprint data. 8th AGILE Conference on Geographic Information Science, Lissabon, Portugal.

VENNERS, B., Inheritance versus composition: Which one should you choose? (Accessed: Dec. 2005), http://www.javaworld.com/javaworld/jw-11-1998/jw-11-techniques.html

WORLD WIDE WEB CONSORTIUM (W3C) (Accessed: Nov. 2005), http://www.w3.org/TR/owl-ref/, OWL Web Ontology Language

WORLD WIDE WEB CONSORTIUM (W3C) (Accessed: Dec. 2005), http://www.w3.org/,

WARNEST, M., RAJABIFARD, A., WILLIAMSON, I. (2005). A Collaborative Approach to Building National SDI in Federated State Systems: Case Study of Australia. GSDI 8, Kairo, Egypt.

WIEDERHOLD, G. (1994). Interoperation, Meditation and Ontologies. International Symposium on Fifth Generation Computer Systems (FGCS94), Tokyo, Japan.

WILLIAMSON, I., RAJABIFARD, A., FEENEY, M.-E. F. (2003). Developing Spatial Data Infrastructures, from Concept to Reality. New York, Taylor&Francis.

WORBOYS, M., DUCKHAM, M. (2004). GIS, A Computing Perspective. Boca Raton, CRC Press LLC.

Student Papers in Collaboration and Supervision by the Author of this Thesis

Filli, R. (2005). Viewbasierte Erzeugung von Metadaten aus einem integrierten Geo- und Metadatenbestand. Zürich, Swiss Federal Institute for Technology, diploma thesis

Henrich, S. (2003). Entwicklung eines Geodaten-Portals für eine Standortevaluation, ETH Zürich, diploma thesis

Hübner, K. (2004). Geo Web Services, ETH Zürich, semester thesis

Stopper, R. (2004). Demonstrator für einen Geodatendienst und Untersuchung zur Integration von Geodaten und Metadaten, ETH Zürich, semester thesis

Glossary

(Network) Protocol	A data communication network provides the physical infrastructure for sharing information in a GDI. Communication of data simply involves the transfer of binary digits (bits). For devices and software to make sense of these bits, however, they must be encoded in a specific format. This is the role of a particular communication "protocol"- to describe how the bits are arranged so that they can convey all the information necessary for effective communication. Protocols describe the format of transmission (Crosswell, 2000).
ANZLIC	Australian and New Zealand Information Council
ASCII	Acronym for the American Standard Code for Information Interchange. ASCII is a code for representing English characters as numbers, with each letter assigned a number from 0 to 127.
Catalogue	Spatial data catalogues are discovery and access systems that use metadata as the target for query on raster, vector, and tabular geospatial information. Indexed and searchable metadata provide a disciplined vocabulary against which intelligent geospatial search can be performed within or among SDI communities (Nebert, 2001).
Clearinghouse	From a technical point of view, it is considered to be a specialized, complex and important Web service which assists in offering the exchange and sharing of spatial data between different users and suppliers in an SDI.
COGIS/KOGIS	Coordination of geographic information an geographic information systems Active since the 3rd January 2000, the objective of GIS coordination is to benefit as much as possible from the existing potential of synergy in the processing, data acquisition, management, use and distribution of geographical information within the Swiss federal administration.
Crosswalks	Crosswalks are tables that map relationships and equivalence between two metadata schemas. These schemas originate from different metadata standards.
Data Dictionary	A data dictionary is a collection of descriptions of the data objects or items in a data model. It helps clarify the semantics of attributes, especially when the name of the attribute is not clear or misleading. In the simplest case, a data dictionary is a table with name of attribute and description.
Data Quality	Fitness of use, e.g. for a certain task

DBMS	Data base management system
DEM	Digital Elevation Model
DM01.AV	Standardized, official, federal cadastral model of Switzerland.
Encoding rules	Encoding rules define how two map and integrate any two models. They provide relationships and equivalences between the models.
ERM	Entity relationship model describes an excerpt of reality on the conceptual level and provides a graphical notation in so called entity-relationship diagrams.
ESRI shape files	Commercial GIS format
FGDC	Federal Geographic Data Committee
FIG	International Federation of Surveyors
Format	The pattern into which data are systematically arranged for use on a computer. A file format is the specific design of how information is organized in the file. For example, ARC/INFO has specific, proprietary formats used to store coverages.
Geoportal	A geoportal is the access point of an SDI on the Internet.
GIS	Geographic Information System is a tool for acquisition, management, processing, analyzing and visualizing spatial data.
GM03	Swiss profile of ISO 19115 for metadata, consists of a core and a comprehensive attribute list. Also, the attributes are categorized as mandatory, optional or conditional.
HSR	Hierarchical Spatial Reasoning, Concept for SDI hierarchy reasoning developed by (Rajabifard, 2000).
HTML	Hypertext Markup Language
Integration	Data and their models from heterogeneous sources and systems are combined for the purpose of common use
INTERLIS	Conceptual modeling language, interface and transfer format for spatial data. Two versions exist. INTERLIS1 is based on Entity-Relationship Models and INTERLIS2 on object-oriented UML models.

Interoperability	Interoperability has many aspects. This thesis focuses on the technical exchange between spatial and metadata as well as integration of data from heterogeneous sources. In order to achieve interoperability between various spatial data and functions from different computer systems, metadata and standards are essential for a defined interface and semantic mapping.
Java	Programming language
LFP1	First category (highest importance), horizontal control points in Swiss cadastral system
Metadata	Data about data. Here, spatial metadata are considered. They describe spatial data by giving information for identification, extent, quality, spatial reference and distribution.
Model	Models describe selected aspects of reality in a computer; thereby, utilizing relationships and hierarchical structures as well as groupings. Spatial models describe the reality in geo-referenced context by including the position of the described objects. A model is written in a modeling language.
Modeling Language	Modeling languages describe specific excerpts of reality (systems or models) in a systematic way (e.g. UML, INTERLIS)
Ontology	An ontology is an explicit specification of a conceptualization to which the vocabulary of one information community refers (Gruber, 1993). Ontologies are considered a more generic description of spatial data than models.
Schema	Synonym to model (see also model)
SOAP	Simple Object Access Protocol SOAP is a protocol for exchanging XML-based messages over a computer network.
Spatial Data	Are also called geo-data or geospatial data. Spatial Data are data about objects which have a reference to location (spatial reference).
Spatial Data Model	Spatial Data Models describe the structure of spatial data. Usually, the data is grouped according to certain characteristics or topic and relationships are defined.
swisstopo	Swiss Federal Institute of Topography (Swiss Mapping Agency)

TOPIC

Keyword used in INTERLIS. It groups spatial data in the same thematic area. Several tables/classes are clustered in one TOPIC.

UDDI

Universal Description, Discovery, and Integration
A platform-independent, XML-based registry for businesses worldwide to list themselves on the Internet. UDDI is an open industry initiative (sponsored by OASIS) enabling businesses to publish service listings and discover each other and define how the services or software applications interact over the Internet.

UDK

Umweltdatenkatalog, metadata schema for environmental data in Germany

UML

Unified Modeling Language
An object modeling and specification language

URL

Abbreviation of Uniform Resource Locator, the global address of documents and other resources on the World Wide Web. The first part of the address indicates what protocol to use, and the second part specifies the IP address or the domain name where the resource is located.

View

Virtual tables, that are linked to the real, physical data table by queries or mapping.

Web Service

Distributed component system that takes advantage of XML as a standard for communication between components in the Internet. A Web service might fulfil either one specific task or a set of tasks.

WFS

Web Feature Service is an interface allowing requests for geographical features across the web. It uses the XML-based GML for data exchange. The OGC defines the WFS specification.

WMS

Web Map Service is an interface specification which comprises a definition for Internet Map Servers, i.e. how to describe geographic data to be published on the web or how to compose and render a map-image to be published on the web. The Open Geospatial Consortium (OGC) defines WMS.

WSDL

Web Services Description Language
The Web Services Description Language (WSDL) is an XML format published for describing Web services. Version V 1.1 has not been endorsed by the World Wide Web Consortium (W3C), however it has released a draft for version 2.0 on May 11, 2005, that will be a recommendation (an official standard), and thus endorsed by the W3C.

XML

Extensible Markup Language

XSD

XML Schema Definition
An XML Schema Definition (XSD) is an instance of an XML schema written in the W3C's XML Schema language. An XSD defines a type of XML document in terms of constraints upon what elements and attributes may appear, their relationship to each other, what types of data may be in them, and other things. It can be used with validation software in order to ascertain whether a particular XML document is of that type, and to produce a Post-Schema Validation Infoset.

Z39.50

An information search and retrieve protocol which enables distributed digital libraries and sophisticated search and discovery. Often used in SDIs

Curriculum Vitae

Christine Najar

Title: Dipl.-Ing.

Group for GeoInformation Technologies
Institute for Geodesy and Photogrammetry
ETH Hönggerberg
8093 Zürich

Date of Birth: 15. December 1976
Nationalities: Polish, American

Education
2002- 2006   Doctoral student at Group for GeoInformation Technology at ETH
1996-2001   Studies of Rural Engineering at the University of Rostock, Germany and the Federal Institute of Technology Zurich (ETH), Switzerland

Diploma Paper (Master Thesis) (2001): „Lösungsansätze zur Verknüpfung von Flächendaten zur Biomassenbewirtschaftung in GIS mit der Methode der Stoffflussanalyse für den regionalen Ressourcenhaushalt- dargestellt an Waldparzellen der Stadt Zürich"
(Approaches for linking georeferenced biomass data in a GIS using the method of Material-Flux-Analysis for a regional ressource management- applied on forest parcels in the city of Zurich)

1987-1996   Asam Gymnasium, Munich, Germany

Professional Experience:
1996-2001   Various projects in IT and engineering for different engineering offices and universities
2002-2005   Scientific projects for the Group for GeoInformation Technologies at the ETH Zurich
2005   Visiting Academic at the University of Melbourne, Australia

Acknowledgements

I would like to extend my special gratitude to my supervisor, professor and mentor Christine Giger. Her motivation, patience, open-mindedness and enthusiasm allowed me to fulfill my potential. She made it possible for me to realize this thesis and offered much advice and positive words during the process.

My gratitude also goes to my expert committee Alain Buogo, Francois Golay and Abbas Rajabifard for advising and supporting me each in their individual way.

During three years of my PhD I had important discussions and cooperation with the administration of the Canton of Zurich, St. Gallen and Aargau. Their input and advice was a valuable balance for my work and for me from the practical application point of view.

Gratitude is also due to Claude Eisenhut (Eisenhut Informatik AG) and Hans Rudolf Gnägi who took time and effort to discuss programming and modeling issues.

In the final phase of writing up my thesis, Anna Boin and Martin Tomko offered me invaluable support by reading my thesis; Karika Kunta and Rossella Nocera their friendship as well as loyalty.

My former students Romedi Filli and Regula Stopper have contributed to my thesis and I would like to credit their work.

Also, I would like to thank many people from the Institute for Geodesy and Photogrammetry, especially the members of the group for GIS and Theory of Errors and of the GeoIT group Manfred Loidold, Joachim Baldegger.

My former professor and mentor Prof. Ralf Bill from the University of Rostock deserves my kindest reverence. His energy and encouragement for his students has opened many doors for me.

My highest gratitude is due to my partner Matthias Kistler who supported me in all the ups and downs of my PhD. His patience, kindness and discussions were very valuable to me.
I owe much to my parents and family for accepting me the way I am and helping me.