

DISS. ETH NO. 1 64 74

**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-

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 be 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 Geodatenätze „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 Interoperabilität 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.