

Vollständige Implementierung des ISO-Standards 19115 in einer Metadatenbank

Kristin Thamm und Frank Schwarzbach

Zusammenfassung

Der Beitrag beschreibt die Erfahrungen, die bei der vollständigen Implementierung des ISO-Standards 19115 »Geographic information – Metadata« in einer Metadatenbank gewonnen wurden. Das Datenbankschema wird zur freien Verwendung zur Verfügung gestellt.

Summary

The experience gained in connection with the complete implementation of the International Standard ISO 19115 »Geographic information – Metadata« in a metadatabase are described. The schema of the database is placed in the public domain.

1 Einleitung

In letzter Zeit gibt es kaum eine größere geodatenführende Einrichtung, die nicht den Aufbau einer Metadatenbank realisiert hat oder plant. Den inhaltlichen Rahmen für diese Vorhaben vermittelt der 2003 verabschiedete ISO-Standard 19115 »Geographic information – Metadata« (ISO 19115:2003). Dieser Standard liegt als deutscher Norm-Entwurf (Englische Fassung prEN ISO 19115:2004) vor und wird voraussichtlich vom Comité Européen de Normalisation (CEN) übernommen und damit innerhalb der EU zur (verbindlichen) Norm.

Aus der Sicht einer ISO-/OpenGIS-basierten (also Dienste-)Geodateninfrastruktur ist das logische Schema dieser Metadatenbank nicht relevant, wenn nur die auf die Datenbank zugreifenden standardisierten Interfaces (Catalog-Services) realisiert sind. Ungeachtet dessen muss die Datenbank die im Standard definierten Metadaten-elemente bereitstellen können. Aus dem Umfang und der Komplexität des Standards erwachsen eine Reihe von Problemen, die im Folgenden ansatzweise diskutiert werden sollen.

2 Definition von Profilen oder Implementierung des gesamten Standards?

Der ISO-Standard 19115 stellt im sog. »comprehensive metadata profile« einen umfangreichen Vorrat an Metadaten-elementen bereit. Anwendergruppen sind aufgefordert, sich auf dieser Grundlage eigene Profile zu definieren, wobei der ISO-Kern (core metadata for geographic datasets) enthalten sein soll und Erweiterungen möglich sind (Abb. 1).

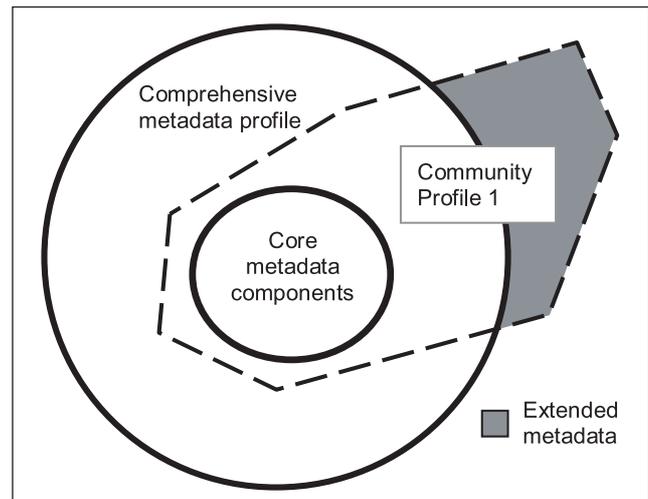


Abb. 1: Metadata community profile (ISO 19115:2003)

Von der Möglichkeit einer Profildefinition ist bereits vielfach Gebrauch gemacht worden. Entsprechende Implementierungen decken dann immer nur einen Teil der durch ISO 19115 vorgehaltenen Metadaten-elemente ab. Die Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) hat für ihre Geodaten ein vergleichsweise umfangreiches Profil nach ISO 19115 definiert (AdV 2003). Eine Realisierung des AdV-Profiles würde eine Implementierung von ca. 2/3 des ISO-Standards bedeuten.

Die Beschränkung auf ein eigenes Profil macht die Menge an Metadaten überschaubar und entlastet eine Implementierung im Rahmen eines Metainformationssystems (MIS). Allerdings setzt die Profildefinition einen möglicherweise langwierigen Abstimmungsprozess innerhalb der Anwendergruppe voraus. Auch ist es für die potentiellen Anwender aufgrund der Komplexität von ISO 19115 nicht immer leicht, ihre Anforderungen an die zu führenden Metadaten im Vorfeld ISO-basiert zu spezifizieren. In anderen Fällen werden Metainformationssysteme als offene Angebote konzipiert, so dass spätere Nutzer keinen Einfluss auf das implementierte Profil besitzen.

Im Ergebnis dieser Überlegungen wurde am Geodatenzentrum des Bundesamtes für Kartographie und Geodäsie (BKG) entschieden, für das MIS der AdV alle Klassen des ISO-Standards in der zugrunde liegenden Datenbank zu implementieren und damit gleichzeitig die Skalierbarkeit bezüglich kommender Anforderungen sicherzustellen. Die Profile »ISO-Kern«, »AdV« und »ISO vollständig« werden durch die Applikationen (Eingabetool, Rechercheclient, Importschnittstelle, Catalog-Service) sichergestellt.

3 ISO 19115 – konforme Metadatenhaltung

Spätestens mit der Verabschiedung des ISO-Standards 19115 werden Metadatenbanken zumeist als »ISO-konform« deklariert. Die ISO 19115 enthält jedoch lediglich die konzeptuelle Sicht auf den Anwendungsgegenstand, die in über 100 Klassen in UML-Diagrammen festgeschrieben ist. Ein zugehöriges Datenbankschema ist nicht spezifiziert. Von ISO werden lediglich externe Sichten, jedoch außerhalb von ISO 19115, standardisiert. Nachdem in einem früheren Stand des ISO-Prozesses auch eine aus den UML-Klassendiagrammen folgende Document Type Definition (DTD) verbreitet wurde, soll nun ISO 19139 ein aus ISO 19115 abgeleitetes XML-Schema für den Datenaustausch enthalten. Abgesehen davon, dass sich ISO 19139 gegenwärtig noch im Entwurfsstadium befindet, hat das XML-Schema nach ISO 19139 wenig Einfluss auf das Datenbankschema (ausgenommen für eine Datenhaltung auf der Basis von XML). Aussagen, eine Datenbank sei »ISO 19115 - konform« oder gar eine »ISO 19115 - Datenbank«, sind daher zu relativieren. Das Datenbankschema wird in jedem Fall mit den individuellen Freiheiten eines logischen Entwurfsprozesses behaftet sein.

4 Grundsätze des Datenbankentwurfs

Auch um gute Voraussetzungen für eine mögliche Nachnutzung der Datenbank zu schaffen, wurden zunächst Grundsätze für den Datenbankentwurf definiert:

- Es sollte eine rein relationale Datenbank unter Verwendung von SQL92-Datentypen auf der Basis von Oracle9i aufgebaut werden. Daraus folgt, dass beim Aufsetzen eines Catalog-Services ein Mapping vom relationalen auf das XML-Schema erfolgen muss.
- Die Grundsätze der Abbildung auf das relationale Schema wurden unter weitgehender Berücksichtigung allgemein anerkannter Regeln vorab definiert und dokumentiert, nachdem Versuche mit einem CASE-Tool nicht die gewünschten Ergebnisse geliefert hatten. Wenn im Ausnahmefall von den festgelegten Grundsätzen abgewichen wurde, so wurde auch dies nachgewiesen.
- Die Abbildungsgrundsätze wurden auch auf die extern referenzierten Entitäten (vgl. (ISO 19115:2003), Punkt B.4) angewandt. Damit sind Teile der Standards ISO 19103, 19107, 19108, 19109 und 19111 bezeichnet, die innerhalb von ISO 19115 Verwendung finden.
- Das relationale Schema sollte sich so eng wie möglich an die UML-Klassendiagramme anlehnen. Im Zweifelsfall wurde diesem Prinzip aus Gründen der Nachnutzbarkeit der Vorrang gegenüber einer Optimierung des relationalen Schemas gegeben. So wurde beispielsweise jeder Klasse, jedem Datentyp, jeder Codeliste und jeder Enumeration immer eine Relation zugeordnet.

- Jeglicher Versuch, die konzeptuelle Sicht von ISO 19115 zu verändern, wurde unterlassen. Beispielsweise wurde darauf verzichtet, die real existierenden Abhängigkeiten zwischen Umringspolygon, Bounding Box und geografischem Identifikator »nachzumodellieren«. Um die Konsistenz zu erhalten, muss hier ein identischer Fremdschlüsselwert in den Tabellen für die Subklassen verwendet werden.
- Die Definition der Abbildungsgrundsätze erstreckt sich auch auf Namenskonventionen für die unterschiedlichen Tabellen- und Attributtypen.
- Das Datenbankschema befindet sich vollständig in der zweiten Normalform. Höherwertige funktionale Abhängigkeiten, die insbesondere in den Datentypen *CI_Citation* und *CI_ResponsibleParty* auftreten, wurden zugunsten eines möglichst engen Bezugs zwischen konzeptuellem und relationalem Schema nicht aufgelöst.

5 Spezielle Abbildungsprobleme und Beispiele

Die allgemeine Vorgehensweise bei der relationalen Abbildung von UML-Klassendiagrammen soll hier nicht diskutiert werden. Vorgestellt wird dagegen die Behandlung einiger ausgewählter Probleme:

Multiple Attribute vom Wertebereich »Free text« unter Beachtung der Mehrsprachigkeit

Die Anlage J aus ISO 19115 enthält das Schema für die Behandlung freier Texte. Auf dieser Grundlage wurde zunächst die Tabelle *PT_FreeText* implementiert. *PT_FreeText* enthält als einziges Attribut *idFreeText*. Hierbei handelt es sich um einen Identifikator für genau einen freien Text, unabhängig davon, in wie vielen verschiedenen Sprachen der Text in *PT_Group* (siehe unten) geführt wird:

```
PT_FreeText {idFreeText primary key}
```

Weiterhin empfiehlt Anlage J die Klasse *PT_Group* mit den Attributen *languageCode*, *country*, *characterSetCode* und schließlich *plainText*. Da die tatsächlich auftretenden Kombinationen von Sprache, Land und Zeichensatz jedoch überschaubar sein dürften, würde eine 1:1-Implementierung eine erhebliche Redundanz bedeuten. Diese Attribute wurden daher in einer gesonderten Tabelle implementiert:

```
LanguageIdent {idLanguageIdent primary key,  
ptLangCode,  
ptCountry,  
ptCharCode}
```

Die Tabelle, welche die Klasse *PT_Group* abbildet, besteht dann nur noch aus den Fremdschlüsseln *idFreeText*

und *idLanguageIdent* sowie den verschiedenen Texten in den unterschiedlichen Sprachen als Attributwerte von *ptText*:

```
PT_Group {idFreeText foreign key references
    PT_FreeText,
    idLanguageIdent foreign key references
    LanguageIdent,
    ptText}
```

Der Vorteil dieser Abbildung besteht nun darin, dass das relationale Schema beim Hinzukommen einer neuen Sprache nicht geändert werden muss. Es müssen lediglich ein Tupel für die neue Sprache in *LanguageIdent* sowie dann die neuen Texte in *PT_Group* eingefügt werden.

Die Ausführungen sollen anhand der Klasse *MD_Metadata* verdeutlicht werden (Abb. 2).

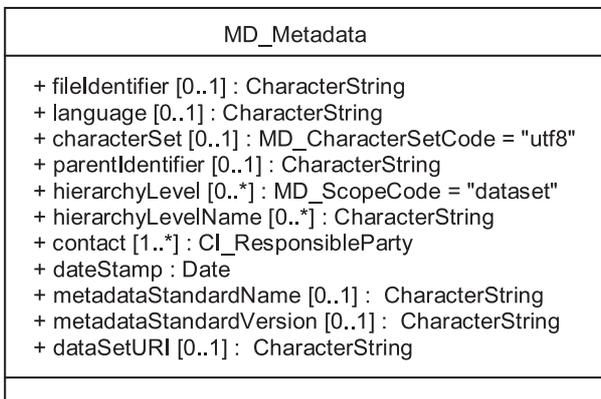


Abb. 2: Klasse *MD_Metadata* (ISO 19115:2003)

Im Falle einer Kardinalität von höchstens 1 wird das Freitext-Attribut (z. B. *metadataStandardName*) in der »Mutter-Tabelle« *Metadata_B21* als Fremdschlüssel, welcher *idFreeText* aus *PT_FreeText* referenziert, implementiert. Treten Freitext-Attribute, wie beispielsweise *hierarchyLevelName*, dagegen multipel auf, wird für jedes Freitext-Attribut eine Join-Tabelle zwischen der »Mutter-Tabelle«

und *PT_FreeText* benötigt. Somit ergibt sich für das Beispiel die in Abb. 3 dargestellte Tabellenstruktur.

Die Abbildung der Codelisten und Enumerations aus ISO 19115 Anlage B5 (ISO 19115:2003) erfolgt mehrsprachig analog zu den freien Texten.

Vererbung

Für die relationale Abbildung von Vererbungshierarchien stehen verschiedene Möglichkeiten zur Verfügung, die hier nicht im Einzelnen diskutiert werden sollen. Im vorliegenden Fall wurde entschieden, jeder Ober- und jeder Subklasse zunächst genau eine Relation zuzuordnen. Der Primärschlüssel der Oberklasse migriert als Fremdschlüssel in die Subklassen. Die Primärschlüsselwerte der Oberklasse tauchen dann in keiner (lediglich die Attribute der Oberklasse werden verwendet), genau einer oder mehreren Subklassen auf. Letzterer Fall tritt beispielsweise ein, wenn die geografsche Referenz sowohl als Umringspolygon als auch durch einen geografischen Identifikator angegeben wird. Als Abbildungsbeispiel sei die Klasse *MD_Constraints* sowie deren beide Subklassen vorgestellt (Abb. 4). Implementiert wurde das Schema der Abb. 5.

Es muss darauf hingewiesen werden, dass durch diese Abbildung bei Recherchen über die Oberklasse vollständige Joins mit allen Subklassen erforderlich sind. Die Kardinalität 1 bei der Oberklasse zeigt, dass im konkreten Projekt der Eintrag eines Tupels in einer Subklasse in jedem Fall auch zu einem Eintrag in der Oberklasse führen soll. Andere Herangehensweisen wären denkbar, z. B. mit Hinblick auf die Oberklasse *Ex_GeographicExtent*, welche nur ein Attribut mit Wertebereich [0,1] besitzt. Das Datenbankschema ist von diesen Überlegungen nicht betroffen.

Keine Regel ohne Ausnahme: Im Package *Data quality information* sind 20 Subklassen von *DQ_Element* enthalten, die alle keine eigenen Attribute besitzen. Hier wurde in der zugehörigen Tabelle *DQElement_B243* ein Attribut *dqCategory* hinzugefügt. Dieses enthält, als Auswahl aus

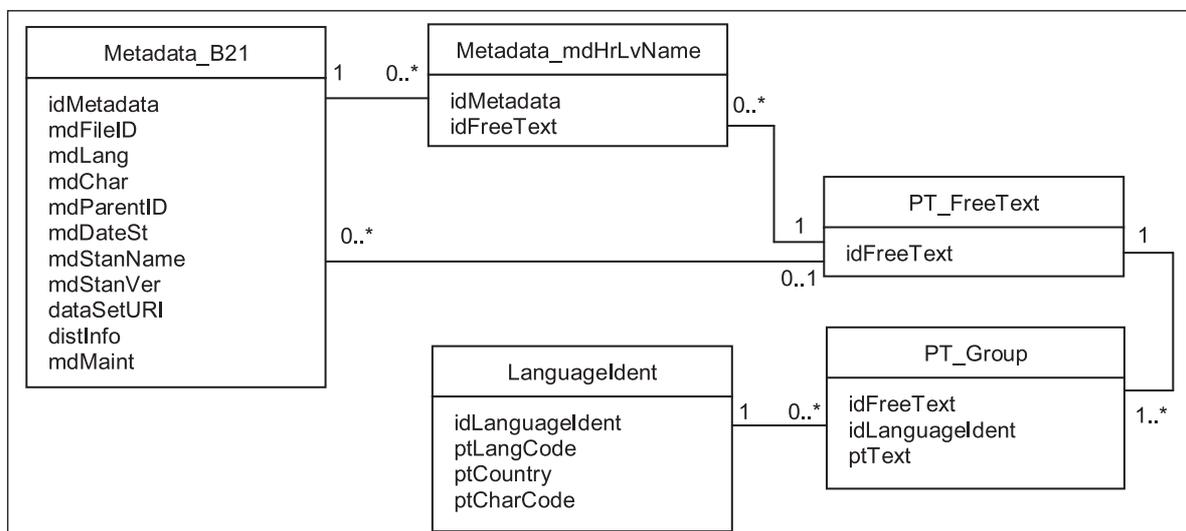


Abb. 3: Relationales Schema zur Abbildung der Mehrsprachigkeit (Beispiel)

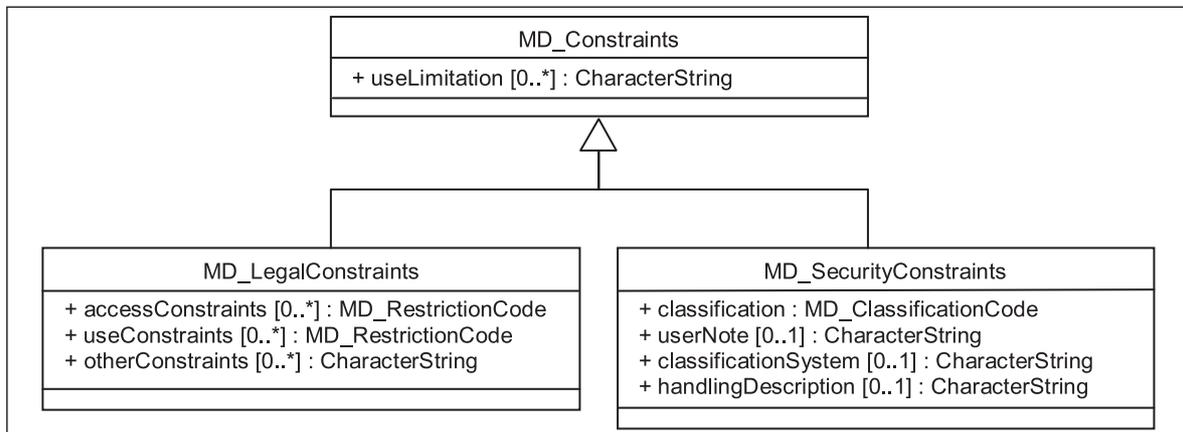


Abb. 4: Vererbungshierarchie aus *Constraint information* (ISO 19115:2003)

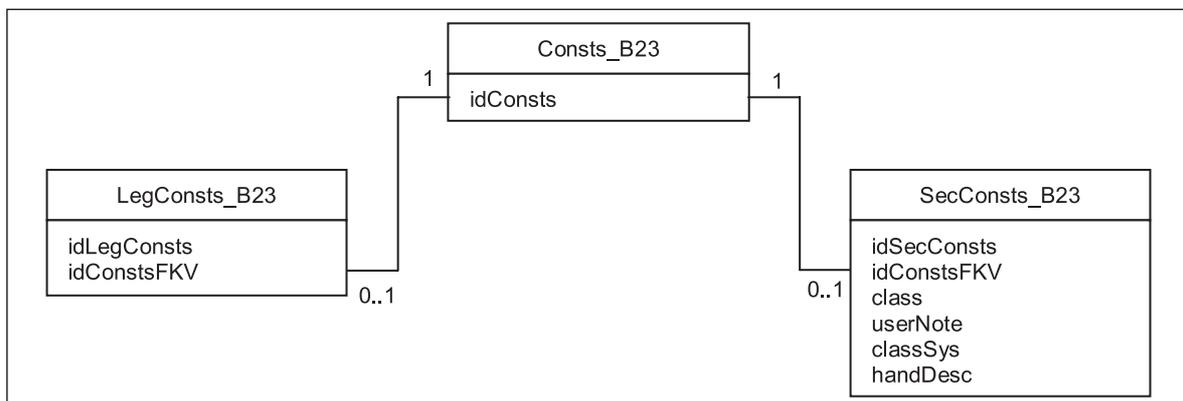


Abb. 5: Relationales Schema zur Abbildung der Vererbung (Beispiel)

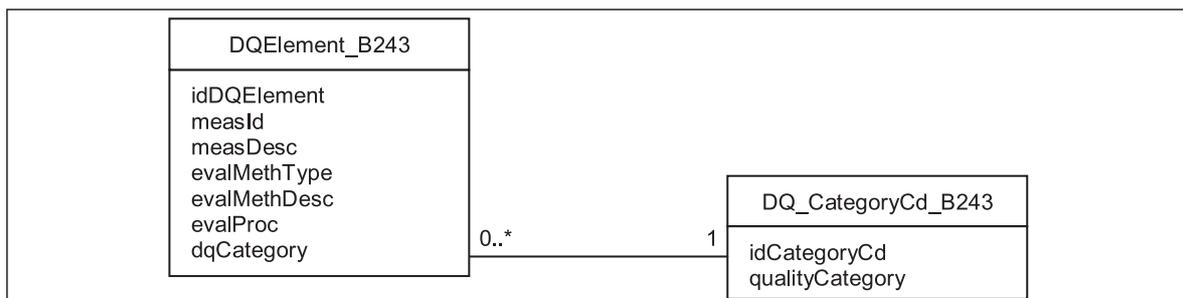


Abb. 6: Relationales Schema zur Abbildung von Subklassen ohne eigene Attribute (Beispiel)

einer Enumeration, die Namen aller instanzierbaren Subklassen. Ausnahmsweise wurden also im Zuge der Abbildung Klassen zu Attributen »degradiert«. Die Implementierung von *DQ_Element* zeigt Abb. 6.

Bei dieser Abbildung wird davon ausgegangen, dass durch Auswahl aus der Enumeration einem Tupel in *DQ_Element* nur genau eine *DQ_Category* zugeordnet werden kann. Das ist aus inhaltlicher Sicht der Attribute von *DQ_Element* möglich. Eine Multiplizität ist durch mehrfache Verwendung von *DQ_Element* gegeben.

Union-Klassen

Im ISO-Standard 19115 werden insgesamt zwei Union-Klassen beschrieben: *MD_Resolution* (Abb. 7) im Package *Identifikation information* und *MD_ScopeDescription* im Package *Maintenance information*.

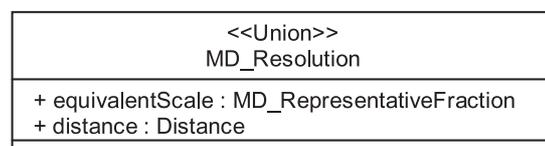


Abb. 7: Union-Klasse *MD_Resolution* (ISO 19115:2003)

Von den Attributen einer Union-Klasse soll jeweils genau eines verwendet werden, wobei die Attribute von verschiedenen Datentypen sind.

Die relationale Abbildung von Union-Klassen kann verschieden erfolgen. Zur Vermeidung zahlreicher Nullwerte in den Attributspalten der Union-Tabelle wurde die Abbildung jedes Attributs in einer eigenen Tabelle bevorzugt. Die Union-Tabelle selbst enthält somit nur die eigene Primärschlüsselspalte. Alle Attribute werden über jeweils eigene Join-Tabellen der Union-Tabelle zugeordnet.

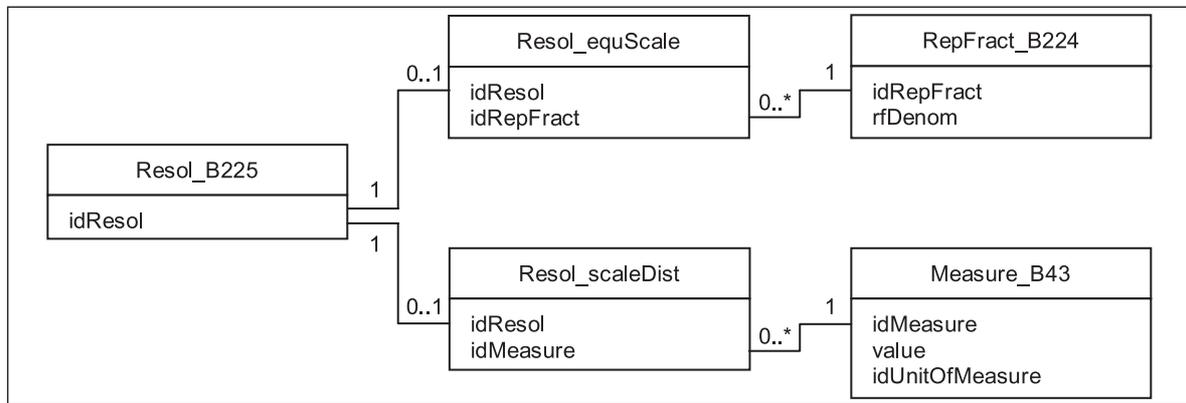


Abb. 8: Relationales Schema zur Abbildung von Union-Klassen (Beispiel)

Durch einen Trigger oder die Applikationen muss sichergestellt werden, dass zu einem Tupel der Union-Tabelle genau ein Eintrag in einer der zugehörigen Join-Tabellen erfolgt. Die Darstellung (Abb. 8) zeigt die Abbildung der Klasse *MD_Resolution*.

6 Freie Verfügbarkeit des Datenbankschemas

Das beschriebene ISO 19115-basierte Datenbankschema ist in Zusammenarbeit zwischen dem Geodatenzentrum des BKG und der Hochschule für Technik und Wirtschaft Dresden (FH) entstanden. Am BKG dient es der Weiterentwicklung des vorhandenen AdV-Metainformationssystems im Rahmen des Aufbaus einer nationalen Geodateninfrastruktur in Deutschland (GDI-DE).

Als Beitrag zur Anwendung des ISO-Standards und zur Herausbildung einer GDI-DE wird das abgeleitete Datenbankschema als SQL-Create-Script unter www.geodatenzentrum.de (Software) zur freien Benutzung zur Verfügung gestellt. Ein SQL-Insert-Script legt in der gesamten Datenbank einen leeren Datensatz an, womit das Füllen der Datenbank wesentlich erleichtert wird, da bereits alle Fremdschlüsselbeziehungen realisiert sind. Die Grundsätze des Datenbankdesigns werden in einer Begleitdokumentation umfassend dargestellt.

Mit dieser Offenlegung des Datenbankschemas soll insbesondere auch eine künftige Interoperabilität zwischen dem AdV-Metainformationssystem und in Konzeption befindlichen ISO-basierten Systemen der Bundesländer und Dritter unterstützt werden.

7 Zusammenfassung und Wertung

Im Ergebnis der Arbeiten entstand eine relationale Datenbank, die den gesamten ISO Standard 19115 implementiert und sich dabei eng an diesen anlehnt. Die Datenbank umfasst ca. 250 Tabellen. Dabei muss beachtet werden, dass ISO 19115 und damit die vorgestellte Implementierung eine sehr detaillierte Beschreibung von Geodaten zulässt. Die Lösung ist somit eher für größere Anwendungen, etwa zum Nachweis der Geodaten eines

Bundeslandes, geeignet. Für eine Übersicht über die Geodatenbestände eines überschaubaren Anwendungsbereichs, z. B. auf der Grundlage der Pflichtfelder des ISO-Kerns, wäre die Datenbank überdimensioniert.

Die Datenbank befindet sich beim BKG seit ca. einem Jahr im praktischen Einsatz und wurde bereits durch die Migration der Daten aus der bisherigen CEN-basierten Datenbank des MIS der AdV gefüllt. Zugehörige Applikationen (Eingabetool, Rechercheclient, Importschnittstelle und Catalog-Service) sind in Arbeit bzw. fertig gestellt. Durch die enge Anlehnung an den ISO-Standard und die Implementierung eines Catalog-Service (Realisierung durch die Fa. GIStec GmbH, Darmstadt) ist die Integration der vorgestellten Lösung in eine Geodateninfrastruktur technisch möglich.

Dank

Die vorgestellten Arbeiten wurden durch Herrn Frank Schnabel, Diplomand am Fachbereich Vermessungswesen/Kartographie der Hochschule für Technik und Wirtschaft Dresden (FH), wesentlich unterstützt.

Literatur

AdV 2003: Gemeinsamer Metadatenkatalog der AdV. 2003.
ISO 19115:2003: International Standard 19115: Geographic information – Metadata. 2003.

Anschrift der Autoren

Dipl.-Phys. Kristin Thamm
Bundesamt für Kartographie und Geodäsie
Außenstelle Leipzig Geodatenzentrum
Karl-Rothe-Straße 10-14, 04105 Leipzig
kristin.thamm@bkg.bund.de

Prof. Dr.-Ing. Frank Schwarzbach
Hochschule für Wirtschaft und Technik Dresden (FH)
Fachbereich Vermessungswesen/Kartographie
Friedrich-List-Platz 1, 01069 Dresden
schwarzbach@htw-dresden.de