

# **Metadatenintegration im UBS Data Warehouse Programm**

**Bernard Mayer**

UBS

**Florian Melchert**

Universität St.Gallen

*Im Rahmen ihres Data Warehouse Programms hat die UBS AG in den letzten Jahren ein unternehmensweites Data Warehouse sowohl technisch als auch organisatorisch erfolgreich umgesetzt. Dabei erkannte man, dass sich durch die Integration und systematische Nutzung der während Entwicklung, Betrieb und Nutzung des DWH anfallenden Metadaten die Effektivität und die Effizienz des Data Warehousing signifikant steigern lassen würden. Man initiierte ein Folgeprojekt zur Realisierung eines integrierten Metadaten Repository. Der vorliegende Beitrag beschreibt im ersten Teil das dabei gewählte Vorgehen sowie das bisher realisierte System für ein durchgängiges Metadaten Reporting. Im zweiten Teil wird gezeigt, wie sich die Integration der Metadaten durch die Erstellung eines konzeptionellen Modells des gesamten Metadatenbestandes weiter verbessern lässt und inwieweit es sinnvoll ist, auf das Common Warehouse Metamodel der OMG als Referenzmodell zurückzugreifen.*

## **1 Metadaten und ihr Nutzen im Data Warehouse**

Seit einiger Zeit hat sich das Data Warehousing als fester Bestandteil der IT-Infrastruktur grosser Unternehmen etabliert und ermöglicht heute meist schon eine gute Versorgung der Fach- und Führungskräfte mit entscheidungsrelevanten Informationen. Die Basis für das Funktionieren dieser Informationsversorgung bildet eine Vielzahl unterschiedlichster Metadaten, die im Rahmen von Erstellung, Betrieb und Nutzung des Data Warehouse von allen beteiligten Systemkomponenten fortlaufend produziert und konsumiert werden (vgl. auch zum Folgenden Bauer, Günzel 2001, S. 326 ff.). Über ihren eigentlichen Verwendungszweck hinaus wird Metadaten aber auch erhebliches Potenzial zugesprochen, das Data Warehousing sowohl im Hinblick auf die Effizienz in der Systementwicklung als auch hinsichtlich einer effektiven Nutzung entscheidend zu verbessern. Aus Nutzersicht kann die Effektivität erhöht werden, indem das Auffinden und Interpretieren relevanter Daten bei der Ana-

lyse durch Metadaten unterstützt werden. Effizienzsteigerungen lassen sich bei Entwicklung und Betrieb des Data-Warehouse-Systems erzielen, wenn Metadaten zur Wiederverwendung von Entwicklungsbausteinen, zur Automatisierung von Prozessen sowie zur konsistenten und integrierten Dokumentation des Data-Warehouse-Systems herangezogen werden und dadurch zu einer Reduzierung des Entwicklungsaufwandes beitragen.

Um das Potenzial der Metadaten in der beschriebenen Weise ausnutzen zu können, ist es erforderlich, die meist lokal genutzten und verwalteten Metadatenbestände einzelner Komponenten zu integrieren, d. h. sie zueinander in Beziehung zu setzen und allen potenziellen Nutzern über eine einheitliche Schnittstelle zur Verfügung zu stellen. Voraussetzung für eine solche Metadatenintegration ist das Verständnis der Beziehungen zwischen den Metadaten. Der konzeptionellen Modellierung des Metadatenbestandes kommt daher wesentliche Bedeutung zu.

Der vorliegende Beitrag zeigt die Entwicklung eines integrierten Metadaten Repository im Data Warehouse Programm der UBS AG auf und diskutiert den Beitrag, den der Metadatenstandard Common Warehouse Metamodel im Rahmen der Weiterentwicklung des Systems leisten kann. In Kapitel 2 werden hierzu zunächst die Rahmenbedingungen des UBS Metadatenprojektes geschildert. Kapitel 3 umreißt Entwicklungsgeschichte und Funktionalität des bisher realisierten Metadaten Repository und motiviert über die identifizierten Probleme den Bedarf an der Erstellung eines konzeptionellen Metadatenmodells. Im vierten Kapitel wird anschliessend an zwei Beispielszenarios überprüft, inwieweit es im konkreten Anwendungsfall sinnvoll erscheint, sich bei der konzeptionellen Modellierung der Metadaten am Metadatenstandard Common Warehouse Metamodel der Object Management Group zu orientieren. Die Schilderung der dabei gewonnenen Erfahrungen sowie ein Ausblick auf das weitere Vorgehen und zu klärende Fragen beschliessen diesen Beitrag.

## **2 Motivation und Projektauftrag für ein Metadatenmanagement bei der UBS**

Das Fehlen einer einheitlichen unternehmensweiten Sicht auf entscheidungsunterstützende Informationen – nicht zuletzt als Resultat der Fusion von Schweizerischem Bankverein und Schweizerischer Bankgesellschaft im Jahre 1998 – veranlasste die UBS AG gegen Ende der 1990er Jahre dazu, das UBS Data Warehouse Programm (DWP) zu initiieren, das zum Ziel hatte, ein unternehmensweites Data-Warehouse-System zu realisieren (vgl. Meyer 2000, S. 46 ff.). Seitdem ist es gelungen, ein funktionierendes Data Warehouse zu etablieren und die zuvor genutzten bereichsspezifischen Management Support Systeme bereits zu grossen Teilen durch auf dem Data Warehouse basierende analytische Informationssysteme zu ersetzen. Die strategische Reichweite des Projekts und die Grösse des zu betreuenden DWH-

Systems bedingen den relativ grossen personellen Umfang des DWP. Als eigene Abteilung innerhalb des UBS IT Entwicklungsressorts etabliert, umfasst es etwa 50 Stellen, die sich auf drei Sektionen aufteilen. Die Sektion *Consulting and Integration Services* stellt dabei die Schnittstelle des Data Warehouse Projekts zu den Fachabteilungen dar, berät diese in Fragen der Anbindung von OLTP-Quellsystemen an das DWH-System und ist für das Management von Softwareversionen und Entwicklungsreleases des DWH-Systems zuständig. Die technische Entwicklung der DWH-Bewirtschaftungsmechanismen sowie die Entwicklung im Bereich der analytischen Applikationen werden von der Sektion *Build Services* getragen.<sup>1</sup> Zudem hatte man bereits früh die Bedeutung von Standards für Entwicklung und Pflege eines so grossen Systems erkannt und in Konsequenz eine separate Sektion *Infrastructure* aufgebaut, die sich mit der Entwicklung von Ordnungsrahmen für die Datenbewirtschaftung sowie für die Datenauswertung beschäftigt.

Ebenfalls sehr früh gelangte man zu der Einsicht, dass durch die grösstenteils aufgaben- und/oder personenspezifisch erfolgende Verwaltung von Metadaten entscheidende Nutzenpotenziale des Data Warehousing unerschlossen bleiben. Insbesondere technische Metadaten der zum Einsatz kommenden Systeme (z. B. Datenstrukturdefinitionen, Transformationsregeln, Kommentare), aber auch fachliche Informationen zu den Daten im DWH wie z. B. Fachbegriffsdefinitionen lagen bereits in grosser Menge vor, konnten den potenziellen Nutzern innerhalb des DWP aber nur in Ausschnitten zur Verfügung gestellt werden. Eine Verteilung der Metadaten an Nutzer ausserhalb des Projekts war zunächst gar nicht möglich. Da man sich von der Herstellung einer integrierten Metadatenbasis die Erschliessung von Synergiepotenzialen in Entwicklung und Betrieb des DWH als auch eine weitere Verbesserung im Bereich der Entscheidungsunterstützung versprach, führte man 1999 eine Anforderungsanalyse für ein Metadatenmanagement durch. Im Rahmen dieser Anforderungsanalyse konnten zunächst sechs Benutzergruppen identifiziert werden, die von einer integrierten Metadatenverwaltung in verschiedener Weise profitieren würden. Während die DWH-Endbenutzer insbesondere an fachlichen Erklärungen sowie der Herkunftsanalyse von in Berichten dargestellten Daten interessiert waren, artikulierten die Fachabteilungen in IT-Fragen unterstützenden Business Technology Centers darüber hinaus einen Bedarf an Möglichkeiten zur Analyse der Datenqualität und der Auswirkungen von Änderungswünschen der Fachabteilungen sowie an einem einfachen Zugang zu der Dokumentation des DWP. Weitere Nutzenpotenziale sahen insbesondere die Abteilungen innerhalb des DWP, wobei speziell der Abgleich der Arbeitsfortschritte zwischen den Stadien Entwicklung, Integration und Produktion als sehr wichtig angesehen wurde. Zusätzlich wurde das Reporting über technische Metadaten sowie die Nutzung der Metadaten zur Erkennung von Redundanzen als Vorteil einer integrierten Sicht auf die DWH-

---

<sup>1</sup> Aufgrund einer strikten Trennung zwischen den Ressorts *Architektur, Entwicklung* und *Betrieb* des Geschäftsbereiches IT der UBS AG wird die operative Durchführung der Datenbewirtschaftung nicht von der Abteilung DWP abgedeckt, sondern fällt vielmehr in die Zuständigkeit des Ressorts Betrieb.

Metadaten genannt. Im Rahmen der DWH-Entwicklung sah man die Möglichkeit, die Konformität zu Modellen und Standards besser gewährleisten zu können.

Abbildung 1 zeigt die Zuordnung der zehn identifizierten Nutzenpotenziale zu den sechs als mögliche Nutzer einer integrierten Metadatenmanagementlösung gesehenen Adressatengruppen. Die aus der dargestellten Situation hervorgehenden Nutzenpotenziale wurden als ausreichender Beweggrund erachtet, ein Projekt für die Entwicklung und Einführung eines Metadatenmanagementsystems zu initiieren. Um die Komplexität der Problemstellung beherrschbar zu halten, wählte das Projektteam eine evolutionäre Vorgehensweise, bei der die Herstellung einer integrierten Sicht auf die existierenden Metadaten als erstes Ziel gesetzt wurde. Das folgende Kapitel beschreibt das aus dieser Initiative entstandene Metadatenmanagement, das ebenfalls in den Zuständigkeitsbereich der Sektion Infrastructure fällt.

	DWH-Endbenutzer	Business Technology Centers	DWP Consulting Services	DWP Build Services	DWP Datenbankadministration	DWP Infrastruktur-Team
Vergleich zwischen Development/Integration/Produktion			X	X	X	X
Business Reporting	X	X	X			
Technisches Reporting			X	X	X	
Redundanzerkennung			X	X	X	
Herkunftsnachweis	X	X				
Datenqualitätsanalyse		X		X		
Informationsbasis für die Entwicklung				X		X
Dokumentation		X		X		
Auswirkungsanalyse		X		X		
Modell-/Standards-Konformitätskontrolle				X		

**Abb. 1:** Benutzergruppenspezifische Nutzenpotenziale eines integrierten Metadatenmanagements im UBS DWP

### 3 Erste Ausbaustufe: Statisches Metadaten-Reporting

Einen ersten Schritt zur Schaffung einer integrierten Sicht auf die im Rahmen des DWP anfallenden Metadaten bildete die Identifikation der bestehenden Metadatenquellen. Während technische Metadaten wie Datenbankkataloge, ETL-Mappings oder OLAP-Berichtsdefinitionen vielfach in Repositories der genutzten Standardsoftwareprodukte vorlagen und mit vertretbarem Aufwand lokalisiert werden konn-

ten, offenbarte sich dem Metadatenteam insbesondere im Bereich der fachlichen Metadaten eine unübersichtlichere Vielfalt kleinerer, dezentraler Eigenlösungen, die spezialisierte Mitarbeiter in ihrem eigenen Arbeitsumfeld aufgebaut hatten und mehr oder weniger regelmässig pflegten. Aus Gründen der Komplexitätsreduktion beschränkte sich das Metadatenteam daher zunächst auf die Anbindung von dreizehn Quellen mit überwiegend technischen Metadaten.

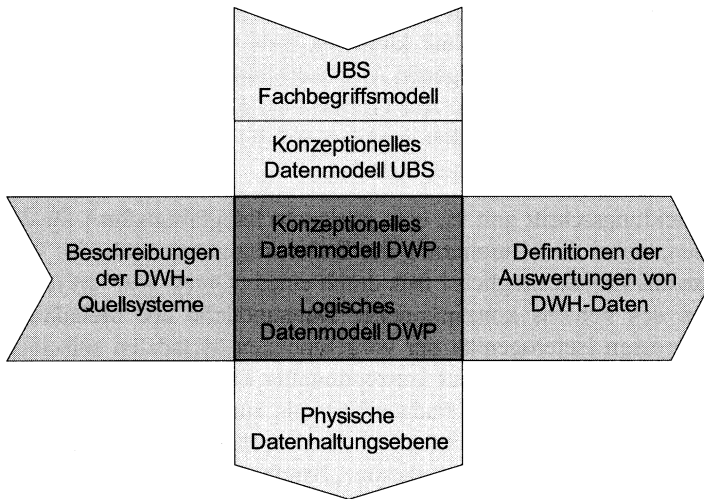
Als nächsten Entwicklungsschritt galt es, eine geeignete Integrationsform für die Metadaten zu finden, wobei die Minimierung des Entwicklungsaufwandes im Vordergrund stand, um das Projekt möglichst früh durch einen „Quick Win“ zu rechtefertigen. Aufgrund des Fehlens konzeptioneller Datenmodelle der Metadatenquellen und einer grossen Heterogenität der vorgefundenen Metadaten entschied man sich dagegen, mit der Integration auf konzeptioneller Ebene zu beginnen und begnügte sich stattdessen damit, die Metadatenbestände zunächst auf physischer Ebene in einer zentralen Datenbank unintegriert zusammenzuführen. Die so entstandene Metadatenammlung konnte nun als sog. „Sandbox“ genutzt werden, um in einem Bottom-up-Ansatz nach und nach Verbindungen zwischen einzelnen Metadaten auf physischer Ebene zu identifizieren und so schrittweise ein logisches Modell der Datenbank zu erstellen. Bei der Durchführung dieser Form der Modellkonstruktion wurden im Wesentlichen zwei Dimensionen in das Zentrum der Bemühungen gestellt (vgl. Abb. 2): Einerseits sollten die verschiedenen Datenbeschreibungen auf konzeptioneller, logischer und physischer Ebene miteinander verbunden, andererseits die Datenflüsse von den operativen Quellsystemen bis zu fertigen Berichtsbestandteilen auf der Auswertungsebene aufgezeigt werden. Das Verbindungsglied dieser beiden als orthogonal betrachteten Dimensionen stellen das konzeptionelle und das logische Datenmodell des Data Warehouse sowie der Data Marts dar, welche von der DWP-Abteilung selbst gepflegt werden<sup>2</sup>.

Um die auf Basis der Inhalte identifizierten Verbindungen zwischen Metadaten verschiedener Quellen für Endbenutzer sichtbar zu machen, wurden zunächst entsprechende Datenbankviews erstellt. Als Frontend für eine breitere Nutzergemeinde eignete sich diese Lösung allerdings nicht, da die vorgefertigten Views zu unflexibel waren, keine aktive Verknüpfung zu anderen Views boten und zudem freie Ad-hoc-Abfragen auf der Sandbox den Benutzern nicht zugemutet werden sollten.

Man entschloss sich daher, einen Generator zu implementieren, der die Metadaten in Form von HTML-Dokumenten im Intranet der UBS bereitstellt. Neben der Portierung der Sandbox-Metadaten in HTML-Dokumente wurden die identifizierten Verbindungen in Form von Hyperlinks umgesetzt, durch die es den Benutzern des HTML-Frontends ermöglicht wurde, sehr flexibel durch den Metadatenbestand zu navigieren. Abbildung 3 zeigt beispielhaft eine durch den Generator erstellte Seite

---

<sup>2</sup> Im Gegensatz zum DWP-Datenmodell wurden Fachbegriffe und Datenmodelle der operativen Quellsysteme als Vorgaben anderer Abteilungen übernommen und konnten nicht beeinflusst werden.

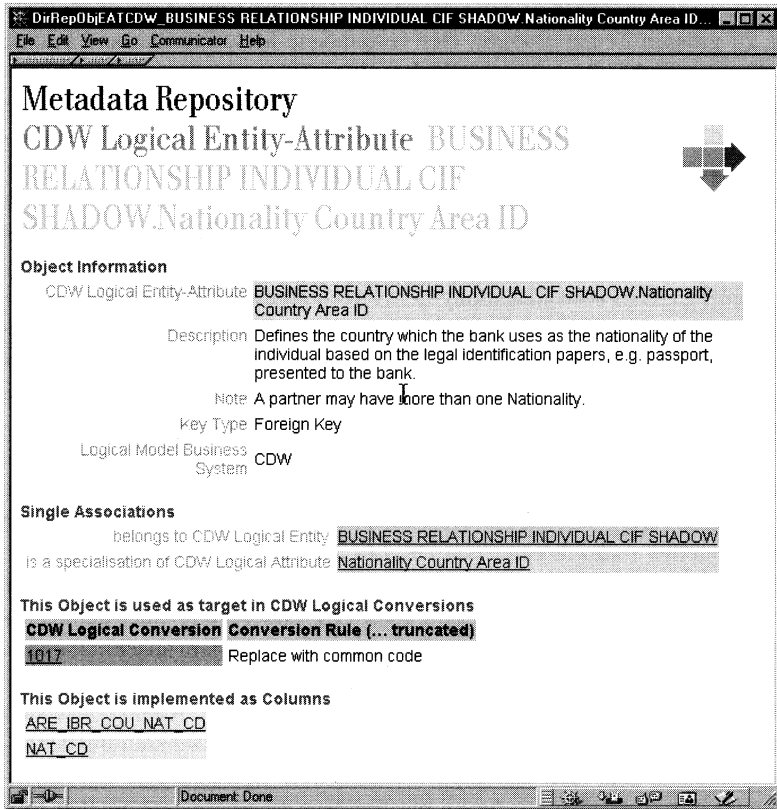


**Abb. 2:** Leitgedanken bei der Entwicklung des DWP Metadata Repository des HTML-Frontends, die alle verfügbaren Metadaten zu einem einzelnen Attribut im konzeptionellen Datenmodell des Data Warehouse enthält.

Neben der textuellen Definition, einer Notiz, dem Schlüsseltyp und dem System, dem das Attribut zuzuordnen ist, werden Informationen über den Entitytypen und dessen Generalisierung, über Konvertierungsoperationen vom konzeptionellen zum logischen Modell sowie über die dem Attribut entsprechenden Tabellenspalten des logischen Modells gegeben. Mithilfe der hinterlegten Hyperlinks kann der Benutzer jeweils detaillierte Informationen zu den Entitytypen, den Konvertierungsoperationen oder den Tabellenspalten abrufen.

Über die im Beispiel dargestellten Beziehungen hinaus bot das Frontend auch Zugang zu Fachbegriffsdefinitionen, Notizen und Datentypen, sodass bereits in einem frühen Entwicklungsstadium eine recht umfassende Informationsquelle für Metadaten aus verschiedenen Quellen genutzt werden konnte.

Eine Frage, die das Metadatenteam von Projektbeginn an beschäftigte, war, ob sich Kosten und Entwicklungszeit für die angestrebte Lösung mithilfe am Markt verfügbarer Standardsoftware für Metadatenmanagement gegenüber einer Individuallösung reduzieren lassen würden. Parallel zur Entwicklung des ersten Prototyps wurde daher eine Evaluation möglicher Standardsoftwarepakete durchgeführt. Aufgrund der relativ hohen Lizenzkosten der in Frage kommenden Kandidaten war die Identifikation mehrerer UBS-interner Nutzergruppen für die Standardsoftware Voraussetzung. Die restriktive Software-Strategie der UBS stellte zudem hohe Anforderungen an die Seriosität des Software-Anbieters. Schon nach relativ kurzer Zeit identifizierte man einen Kandidaten, der als einziger die gestellten Anforderungen an Funktionalität, Benutzerfreundlichkeit, Seriosität des Anbieters und Lizenzierungskosten erfüllen konnte. Nachdem diese Software jedoch an einen kleineren,



**Abb. 3:** Screenshot des HTML-Metadaten-Frontends der UBS

nicht mehr als ausreichend stabil eingestuftem Hersteller verkauft worden war, verwarf man die Pläne zum Einsatz von Standardsoftware für Metadatenmanagement und konzentrierte sich in der Folge auf die Weiterentwicklung der eigenen Individuallösung, wobei ausschliesslich die in der UBS etablierten Softwarekomponenten zum Einsatz kommen sollten. In einem ersten Arbeitspaket wurden die bereits gewonnenen Erkenntnisse über die strukturellen Zusammenhänge der vorliegenden Metadaten genutzt, um die bisherige Metadaten-Sandbox, die lediglich 1:1-Kopien der Metadatenquellen beinhaltete, durch ein *Metadata Repository* (MDR) zu ersetzen, dessen Struktur besser an die späteren Verwendungszwecke der Metadaten angepasst sein sollte. Neben den als Basistabellen bezeichneten Kopien der Quell-Metadaten enthielt das MDR zusätzlich eine Reihe sog. Mart-Tabellen, in denen Quell-Metadaten mithilfe von ETL-Prozessen rekombiniert und denormalisiert abgelegt wurden, um so die Generierung des HTML-Frontends zu erleichtern. Zudem wurden die Prozesse zum Extrahieren, Transformieren und Laden der Quellmetadaten in das Repository automatisiert und in die bereits existierende Standardinfrastruktur zur Befüllung des UBS Data Warehouse integriert.

In dieser Ausbaustufe des Systems hatte das Metadatenmanagement-Team das erste Ziel bereits zu grossen Teilen erreicht. Der vorhandene Metadatenbestand konnte über eine übersichtliche Benutzeroberfläche eingesehen werden, und die hergestellten Bezüge zwischen Metadaten unterschiedlicher Quellen ermöglichten eine einfache Navigation von Elementen der konzeptionellen Ebene bis zu korrespondierenden Elementen der physischen Ebene. Der so gewonnene Überblick offenbarte jedoch auch bisher nicht erkennbare Probleme aus inhaltlicher Sicht: Die aus sehr heterogenen Quellen stammenden Metadaten basierten jeweils auf eigenen Begriffssystemen, die nicht aufeinander abgestimmt waren. Durch die Zusammenführung wurde deutlich, dass sowohl Synonym- als auch Homonymbeziehungen existierten. Darüber hinaus war kein einheitliches System zur Schlüsselvergabe vorhanden, wodurch die Beziehungen zwischen Metadaten verschiedener Quellen nur teilweise identifiziert werden konnten. Insgesamt musste man erkennen, dass die blossе Zusammenführung der Metadaten sowie ihre Aufbereitung für die Auswertung nicht ausreichten, um ein integriertes Metadatenmanagement zu realisieren. Zur Verbesserung der Situation orientierte man sich bewusst an den Erfahrungen aus der Konstruktion des Data Warehouse. Hier hatte man die Übersicht über die Daten verschiedener Quellen mithilfe einer von DEVLIN als Reconciled Data Layer (vgl. Devlin 1997, S. 69-77) bezeichneten integrierten Datenbasis hergestellt, in der die Quelldaten zunächst in normalisierter Form unter Zugrundelegung eines gemeinsamen konzeptionellen Datenmodells abgelegt werden. Erst in einem anschliessenden zweiten Aufbereitungsschritt werden die normalisierten Daten dann in an der Auswertung orientierte multidimensionale Datenstrukturen überführt. Analog dazu kam man auch für das Metadaten Repository zu der Einsicht, dass zur Herstellung der gewünschten Übersicht über den Metadatenbestand eine normalisierte Zwischenschicht benötigt wird, welches wiederum die Existenz eines konzeptionellen Modells der Metadaten voraussetzt. Im Folgenden Abschnitt wird dargestellt, welchen Weg das UBS Metadatenteam beschritten hat, um ein solches Metadatenmodell zu konstruieren.

## **4 Zweite Ausbaustufe: Konzeption eines integrierten Metadaten Repository**

Als Grundlage für die Erstellung einer integrierten Metadatenbasis wird die Konstruktion eines konzeptionellen Modells des MDR gesehen. Dieses Modell soll nicht nur alle existierenden Elemente und Beziehungen zwischen den Metadaten abbilden, sondern auch die Einbindung neuer Metadatenquellen in das Repository erleichtern. Hierzu sollen klare Modellierungsstandards vorgegeben werden, mit deren Hilfe sich neue Modellbestandteile – etwa bei der Aufnahme neuer Metadatenquellen in das Repository – einfach integrieren lassen. Als Ansatz zur Reduktion des Modellierungsaufwands bei gleichzeitiger Sicherstellung einer hohen Modell-



qualität erschien es dem Metadatenteam der UBS sinnvoll, der Modellierung ein existierendes Referenzmodell zugrunde zu legen. Hier konnte man auch auf Vorarbeiten zurückgreifen, an denen sich das UBS DWP unabhängig vom beschriebenen Metadatenprojekt bereits seit 1998 aktiv beteiligt hat: Im Rahmen einer Initiative der Object Management Group (OMG) hatte man zusammen mit anderen Grossunternehmen an der Etablierung eines Industriestandards für Data-Warehouse-Metadaten mitgewirkt. Das aus dieser Initiative entstandene Common Warehouse Metamodel (CWM) bietet heute ein von der OMG anerkanntes Standardmodell für technische und fachliche Metadaten im Umfeld des Data Warehousing und liefert so eine geeignete Ausgangsbasis für die Entwicklung unternehmensspezifischer und dennoch standardkonformer Metadatenmodelle (vgl. OMG 2001a; Poole et al. 2002; Melchert 2003). Für die Auswahl und nähere Untersuchung des CWM als möglichem Referenzmodell sprachen vor allem folgende Gründe:

- Durch die aktive Beteiligung der UBS AG sind bereits einige der unternehmensspezifischen Anforderungen während der Entwicklung des CWM berücksichtigt worden.
- Das CWM ist ein Modell, das speziell für Metadaten im Data-Warehousing-Kontext entwickelt wurde und somit genau auf die existierende Anwendungsdomäne zugeschnitten ist. Vergleichbare Modelle liefern zudem nicht die Breite an durch das Modell explizit abgedeckten Metadatenkategorien. (vgl. Do, Rahm 2000, S. 6)
- Das CWM ist herstellerneutral, da es aus einer Kooperation mehrerer Hersteller und Anwender entwickelt und durch die Object Management Group als weltweiter Standard anerkannt wurde. Dies begünstigt die weitgehende Unterstützung des Standards durch die Softwareindustrie, da einzelne Hersteller die Ausgestaltung des Standards zu ihren Gunsten nur in sehr engen Grenzen beeinflussen werden können.
- Durch Verwendung der standardisierten Modellierungssprache UML bei der Konstruktion sowie durch die Einbindung in die als Metametamodell propagierte Meta Object Facility erreicht das CWM eine hohe Konformität mit den übrigen Standardisierungsbestrebungen der OMG (vgl. Melchert 2003, S. 94; OMG 2001a, S. 1-2ff.). So lassen sich die auf Basis des CWM modellierten Metadaten z. B. direkt über den ebenfalls MOF-kompatiblen Schnittstellenstandard XMI in Form von XML-Dokumenten austauschen. (vgl. Poole et al. 2002, S. 69 ff.) Der hohe Grad an Standardkompatibilität soll die Akzeptanz des CWM bei Anwendern und Herstellern von Standardsoftware gleichermaßen fördern und helfen, die Möglichkeiten zur inner- und zwischenbetrieblichen Metadatenintegration zu verbessern.

Nachdem man sich aus den genannten Gründen für das CWM als zweckmässigstes Referenzmodell entschieden hatte, galt es zunächst, die theoretischen Vorteile in der praktischen Umsetzung zu überprüfen und den genannten Vorteilen den zusätzlichen Aufwand für die Einarbeitung in das CWM und seine Modellierungsprinzi-

pien sowie für die Anpassung des Modells an die Unternehmensanforderungen gegenüberzustellen. Das UBS-Metadatenteam entschied sich daher zunächst dafür, in Zusammenarbeit mit dem Kompetenzzentrum Data Warehousing 2 die Anwendung des CWM als Referenzmodell in zwei unterschiedlichen Szenarien zu evaluieren: In einem ersten Szenario galt es, die Möglichkeiten des CWM im Bereich der Modellierung technischer Metadaten zu überprüfen, wobei ein bereits recht vollständiges konzeptionelles Modell der Recordstrukturen im UBS DWP als Informationsbasis diente. In einem zweiten Szenario sollte dann die Eignung des CWM für die Modellierung von Terminologien näher beleuchtet werden – ein Bereich, für den bisher noch kein einheitliches konzeptionelles Modell existierte. Im Folgenden werden die beiden Testfälle in Bezug auf das Vorgehen und die gewonnenen Erfahrungen näher erläutert und in einem anschließenden Fazit zusammengefasst.

## **4.1 Konzeptionelle Modellierung technischer Metadaten auf Basis des CWM**

Ein wesentlicher Teil des UBS Metadaten Repository beinhaltet die technischen Metadaten der als Quellen an das Data Warehouse angeschlossenen operativen Legacy-Systeme der UBS. Hierbei handelt es sich hauptsächlich um Definitionen recordbasierter Datenstrukturen, wobei sowohl die liefernden Systeme, die genutzten Recordtypen als auch die darin enthaltenen Recordfeldtypen festgelegt werden. Die Heterogenität der Legacy-Systeme spiegelt sich auch in den Metadaten wider: Während zu einigen Systemen wenig mehr als die nötigsten Metadaten wie ID, Name und Speicherort einer Datenstruktur vorliegen, umfassen andere Systeme einen sehr umfangreichen Katalog von Attributen zu verschiedenen Sachverhalten. Das konzeptionelle Modell der Recordstrukturen wurde bereits während der Entwicklung der ersten Ausbaustufe des MDR erstellt und umfasst insgesamt 18 Klassen und ca. 116 Attribute.

Bei der Überführung des Modells in eine CWM-konforme Form diente das von SCHÜTTE vorgeschlagene Vorgehen zur unternehmensspezifischen Anpassung von Referenzmodellen als Orientierungshilfe, das die Schritte Modellauswahl, Modellmodifikation, Modellintegration und Qualitätssicherung umfasst (vgl. Schütte 1998, S. 316ff.). Die Phasen Modellintegration und Qualitätssicherung werden im Rahmen dieses ersten Modellierungsanlaufes zunächst vernachlässigt, da zum einen die Integration des CWM-basierten Modellteils in das restliche MDR-Modell nicht Gegenstand der Untersuchung ist und zum anderen die Qualitätssicherung erst im Rahmen eines produktiven Einsatzes des CWM-konformen Modells sinnvoll durchgeführt werden kann.

### *Modellauswahl*

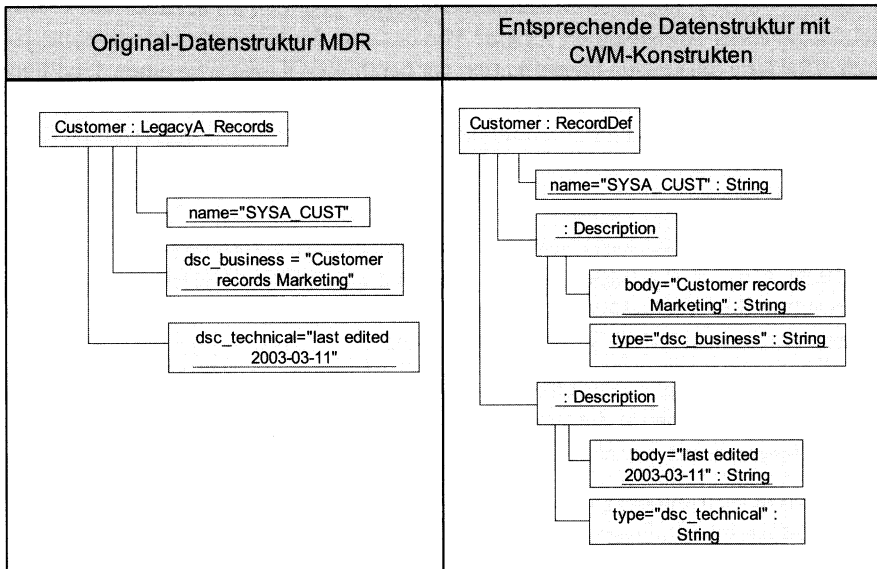
Im ersten Schritt galt es hierbei, die für das bestehende Modell relevanten Teile des CWM auszuwählen. Durch seine modulare Package-Struktur ermöglicht das CWM hierbei eine recht genaue Selektion der benötigten Modellkonstrukte (vgl. Poole et

al. 2002, S. 80 ff.). Die nach Metadatenkategorien gebildeten Packages des CWM dienten als erster Anhaltspunkt für eine Vorselektion der relevanten Modellteile, jedoch erforderte die genaue Identifikation der benötigten Konstrukte einen detaillierteren Abgleich auf Klassen- bzw. Attributebene. Hierzu wurde sukzessive für jedes Attribut des betrachteten MDR-Modellausschnitts versucht, ein äquivalentes Konstrukt im CWM zu identifizieren, wobei zwischen folgenden vier Fällen unterschieden werden musste:

1. Ein MDR-Attribut kann sowohl inhaltlich als auch terminologisch auf ein CWM-Attribut abgebildet werden. Dieser Fall entspricht einer identischen Konstruktion des unternehmensspezifischen Modells und des Referenzmodells und erfordert keine weitere Modellmodifikation.
2. Ein MDR-Attribut entspricht inhaltlich einem CWM-Attribut mit einer alternativen Bezeichnung. In diesem Fall ist die Abbildung möglich und die Terminologieunterschiede werden als Synonymbeziehung festgehalten.
3. Ein MDR-Attribut entspricht einer CWM-Klasse. In diesem Fall wird das Attribut auf ein inhaltlich möglichst passendes Attribut der Klasse abgebildet. Sofern mehrere MDR-Attribute derselben CWM-Klasse entsprechen, kann ein weiteres Attribut dieser Klasse genutzt werden, um das korrespondierende MDR-Attribut zu charakterisieren. Abbildung 4 verdeutlicht diesen Sachverhalt anhand eines Beispiels. Die beiden unterschiedlichen Beschreibungsattribute *dsc\_business* und *dsc\_technical* der Recorddefinition *Customer* werden auf zwei Instanzen der CWM-Klasse *Description* aus dem *BusinessInformation*-Package abgebildet, die zu der Instanz *Customer* der Klasse *RecordDef* in Beziehung stehen und die durch die beiden Instanzen „*dsc\_business*“ und „*dsc\_technical*“ des Attributs *type* unterschieden werden können.
4. Ein MDR-Attribut lässt sich nicht mithilfe von CWM-Konstrukten abbilden. In diesem Fall muss das betreffende Attribut als Kandidat einer möglichen Modellmodifikation vorgemerkt werden.

Als Ergebnis des attributweisen Modellvergleichs kann festgehalten werden, dass sich 27 Attribute des MDR-Modellausschnitts auf CWM-Attribute und 23 MDR-Attribute auf CWM-Klassen abbilden lassen, wobei Konstrukte der nachfolgend aufgelisteten CWM-Packages zur Nutzung ausgewählt wurden. Eine genauere Erläuterung der einzelnen Packages findet sich in (OMG 2001a).

- Das Package *Record* liefert die grundlegenden Klassen *RecordDef* und *Field* zur Modellierung von Recordbehältern und Recordfeldern.
- Das Package *BusinessInformation* erlaubt über die Klassen *Description* und *ResponsibleParty* die Modellierung von Beschreibungsattributen sowie Verantwortlichkeiten.



**Abb. 4:** Beispiel für die Abbildung verschiedener MDR-Attribute auf das gleiche CWM-Konstrukt

- Mithilfe des Packages *SoftwareDeployment* können Metadaten zu eingesetzten Informationssystemen (Klasse *SoftwareSystem*) und deren Systemkomponenten (Klasse *Component*) modelliert werden.
- Das Package *Core* stellt als Kernbestandteil des CWM zudem die Superklassen sowie grundlegende Assoziationen zur Verfügung, die im Rahmen der Wiederverwendung in anderen Packages benötigt werden.

Die im CWM verwendeten Begriffe unterscheiden sich zum Grossteil von den Begriffen des MDR, sodass eine Begriffsklärung erfolgen muss. Für die Erreichung einer möglichst hohen Standardkonformität erscheint es sinnvoll, die CWM-Begriffe zu übernehmen und sie zusammen mit den MDR-Begriffen und entsprechenden Synonymbeziehungen in das Fachbegriffssystem des Metadatensystems aufzunehmen (s. auch Kap. 4.1). Auf diese Weise sind sowohl die ursprünglichen Bezeichnungen des MDR als auch die des CWM verfügbar, und ihr Bezug ist klar definiert. Neben den 50 genannten MDR-Attributen konnten die übrigen 66 untersuchten Attribute nicht durch bestehende CWM-Konstrukte rekonstruiert werden. Hieraus ergibt sich ein Bedarf zur unternehmensspezifischen Erweiterung des Referenzmodells von recht grossem Umfang.

### Modellmodifikation

Für die Erweiterung des CWM stehen mit Stereotypen, Stereotypen in Verbindung mit Tagged Values sowie Erweiterungsklassen drei Mechanismen zur Verfügung, von denen der zweckmässigste auszuwählen ist (vgl. Poole et al. 2002, S. 207-219;

Melchert 2003, S. 102). Im Folgenden sollen die drei Mechanismen hinsichtlich ihrer Eignung im konkreten Anwendungsfall des UBS DWP kurz diskutiert werden:

1. Die Erweiterung auf Basis von Stereotypen allein redefiniert lediglich die bestehenden CWM-Klassen, um diesen eine alternative Bedeutungen zuzuweisen. Für die benötigte Neuanlage von Attributen reicht dieser Mechanismus daher nicht aus.
2. Die Verwendung von Stereotypen i. V. m. sog. Tagged Values erlaubt es, die zusätzlich benötigten Attribute mithilfe des bereits in das CWM integrierten Tagged-Value-Konstrukts CWM-konform abzubilden, ohne das Referenzmodell selbst erweitern zu müssen. Jedes MDR-Attribut der Form (<Attributname>, <Attributwert>) könnte dann als eine Instanz der CWM-Klasse *TaggedValue* mit den Attributwerten (tag = <Attributname>, value = <Attributwert>) konstruiert werden. Durch eine Assoziation ist jede Instanz der Tagged-Value-Klasse einem spezifischen CWM-Objekt zugeordnet. Darüber hinaus können TaggedValue-Instanzen über Stereotypen zusammengefasst werden, was die Übersichtlichkeit erhöht. Insgesamt ist es damit möglich, durch die Nutzung von Stereotypen und Tagged Values alle MDR-Attribute CWM-konform zu modellieren, ohne das Referenzmodell tatsächlich erweitern zu müssen. Dieser Vorteil geht jedoch mit dem Nachteil einer geringen Flexibilität einher: Das CWM erlaubt für TaggedValue-Instanzen lediglich Zeichenketten als Attributwerte. Werden andere Datentypen benötigt, so wäre eine Modifikation des Referenzmodells erforderlich. Weiterhin können auf Basis des TaggedValue-Ansatzes keine zusätzlichen Beziehungen zwischen einzelnen Tagged Values modelliert werden.
3. Mithilfe eigener Erweiterungsklassen zum CWM lassen sich nahezu beliebige Anpassungen des Referenzmodells vornehmen. Durch die Nutzung des für das CWM vorgeschlagenen Vererbungsprinzips können zusätzliche Attribute als Teil neu erstellter Subklassen modelliert werden. Dies hat den Vorteil, dass die vererbten Eigenschaften der Superklasse weiterhin CWM-konform sind und die neu hinzugefügten Attribute möglichst eng in das Modell integriert werden.

Neben der Entscheidung für einen der Erweiterungsmechanismen ist auch die Kombination der Mechanismen denkbar. Jede der diskutierten Erweiterungen bringt allerdings den Nachteil mit sich, dass die neu hinzugefügten, unternehmensspezifischen Modellelemente nicht Teil des CWM-Standards und somit auch nur einem eingeschränkten Nutzerkreis bekannt sind. Bei einem Austausch von Metadaten, die auf solchen zusätzlichen Modellelementen basieren, ist daher zunächst in jedem Fall eine Verständigung der austauschenden Parteien über die vorgenommenen Erweiterungen notwendig. Der sich durch die Verwendung von Tagged Values ergebende Vorteil beschränkt sich daher auf den im Vergleich zur Bildung von Erweiterungsklassen geringeren Aufwand zur Modellerweiterung. Hinsichtlich der Standardkonformität sind die Unterschiede zwischen Mechanismus 2 und 3 jedoch eher unwesentlich. Im Kontext des konkreten Anwendungsfalls entschied man sich für die Re-

ferenzmodellmodifikation mithilfe von Erweiterungsklassen, da dies eine grössere Flexibilität in der Modellierung erlaubt und zudem dem Modularitätsgedanken der CWM-Packages und somit der Übersichtlichkeit besser Rechnung trägt als eine Lösung über Stereotypen und Tagged Values. Darüber hinaus besteht zusätzlich die Möglichkeit, die konstruierten Erweiterungsklassen inhaltlich zu Erweiterungs-Packages zusammenzufassen und diese als CWM Extension Packages anzubieten. Dieses Vorgehen wird auch von der OMG begrüsst, die einige bereits fertig gestellte Erweiterungs-Packages in einem Package *CWM Extensions* (CWMX) zusammenfasst und in Ergänzung des CWM-Standards anbietet (vgl. OMG 2001b).

Zur Komplettierung des Testszenarios wurde schliesslich die Modellierung von UBS-spezifischen Erweiterungsklassen durchgeführt. Um eine möglichst hohe Konformität der neu zu konstruierenden Elemente mit den bestehenden CWM-Bestandteilen sicherzustellen, machte man hierbei intensiven Gebrauch vom Vererbungsprinzip des CWM. So konnten die meisten zusätzlich benötigten Attribute als Bestandteile neu angelegter Subklassen definiert werden, die ihre Eigenschaften von den bereits zur Verwendung ausgewählten CWM-Klassen *SoftwareSystem*, *Component*, *RecordDef* und *Field* erben. Da es sich bei den Attributen meist lediglich um spezifische Eigenschaften der entsprechenden Klasse handelt, wurden kaum Beziehungen benötigt, die nicht bereits durch vererbte Assoziation zwischen CWM-Superklassen zur Verfügung standen.

Die konstruierten Erweiterungsklassen liessen sich zudem inhaltlich vier unterschiedlichen Typen von Legacy-Systemen zuordnen und wurden – dem Leitgedanken einer möglichst starken Modularität des Referenzmodells folgend – in vier neue CWM Extension Packages aufgeteilt. Da es sich bei den Erweiterungen ausschliesslich um Datenstrukturdefinitionen handelt, können alle vier Packages der Ebene *Resource* des CWM-Extension-Packages zugeordnet werden.

## 4.2 Konzeptionelle Modellierung fachlicher Metadaten auf Basis des CWM

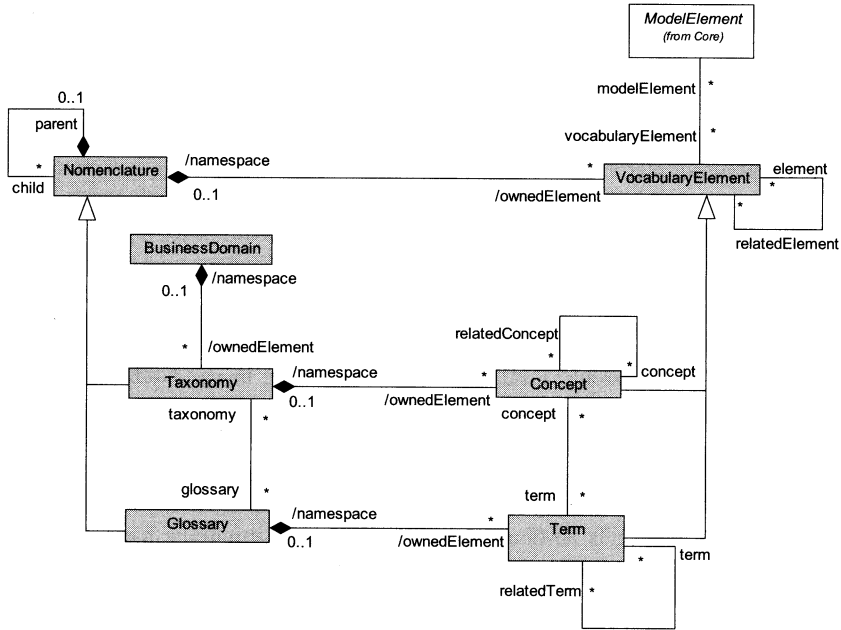
Über die Vielzahl technischer Metadaten hinaus enthält das UBS MDR im Bereich der fachlichen Metadaten vor allem Fachbegriffe und Abkürzungen, welche nicht nur aus unterschiedlichen Systemen stammen, sondern sich auch auf sehr unterschiedliche Anwendungsbereiche beziehen. So sind sowohl betriebswirtschaftliche Begriffe unterschiedlicher Fachbereiche als auch technische Terme aus der Informationsverarbeitung vorhanden. In der ersten Ausbaustufe des Systems gelang es dem Metadatenteam, eine Reihe von Beziehungen zwischen betriebswirtschaftlichen und technischen Begriffen zu identifizieren, die die Grundlage für die anschliessend implementierte Navigationsfunktionalität bildete. Das vorliegende System lieferte somit zwar schon eine recht gute Verbindung zwischen Metadaten der betriebswirtschaftlichen und der technischen Ebene, konnte allerdings einige Funktionalitäten noch nicht zufrieden stellend abdecken. So existierten kaum Möglichkeiten, Synonym- oder Homonymbeziehungen zwischen Termen zu verwalten, um bisher unge-

klärte Redundanzen aufzuheben, die sich durch die Zusammenführung sich teilweise überschneidender Glossare und Begriffssammlungen ergeben hatten. Des Weiteren bestanden lediglich Möglichkeiten, Beziehungen zwischen einander entsprechenden Begriffen verschiedener Entwicklungsebenen zu verwalten. So war es bspw. möglich, einen Fachbegriff der konzeptionellen Ebene einem Element des logischen Datenmodells zuzuordnen. Begriffe auf derselben Ebene konnten hingegen nicht zueinander in Beziehung gesetzt werden. Auch hierarchische Beziehungen zwischen Fachbegriffen wie sie z. B. für Kennzahlssysteme üblich sind, (vgl. Holten, Dreiling, Schmid 2003, S. 429f.) konnte das System somit nicht abbilden. Darüber hinaus war es nicht möglich, den Geltungsbereich einzelner Begriffe festzulegen.

Grundlage für die gewünschte Funktionalitätserweiterung in der Fachbegriffsverwaltung ist die Erstellung eines konzeptionellen Fachbegriffsmodells, das die benötigten Beziehungen zwischen Fachbegriffen abzubilden vermag. Im Rahmen der Evaluation des CWM bot es sich an, die für den Bereich Fachbegriffsmodellierung verfügbaren CWM-Konstrukte auf ihre Eignung zur Erstellung eines solchen mächtigeren Fachbegriffsmodells zu überprüfen. Wesentliche Grundlage hierfür bildete das CWM-Package *BusinessNomenclature*, das speziell der Definition von Fachbegriffsmodellen dient und entsprechende Modellelemente zur Verfügung stellt (vgl. OMG 2001, S. 14-1 – 14-17). Abbildung 5 zeigt die wesentlichen Elemente und Beziehungen des Packages, die sich grösstenteils aus zwei Grundkonstrukten ableiten lassen.

Die Klasse *VocabularyElement* als das eine Grundkonstrukt bezeichnet einzelne Elemente (im Folgenden Vokabeln) eines Fachbegriffssystems und wird in die Klassen *Term* und *Concept* spezialisiert. Ein Konzept (*Concept*) bezeichnet einen abstrakten Fachbegriff, der mithilfe von Termen dargestellt werden kann und soll einen unternehmensweit einheitlich verstandenen Begriff repräsentieren. Ein Term hingegen stellt einen i. d. R. domänenspezifischen Fachausdruck dar, der die Bedeutung eines oder mehrerer Konzepte wiedergibt. Über eine n:m-Assoziation werden Terme und Konzepte zueinander in Beziehung gesetzt. Eine weitere von der Superklasse *VocabularyElement* geerbte Assoziation ermöglicht, dass Konzepte Beziehungen zu anderen Konzepten, und Terme mit unspezifischen, Synonym- und Spezialisierungs-/Generalisierungsbeziehungen sogar drei Arten von Beziehungen zu anderen Termen unterhalten können. Jede Vokabel (*VocabularyElement*) muss durch einen Definitionstext sowie die textuelle Angabe von Beispielinstanzen und Nutzungskonventionen näher beschrieben werden.

Die Klasse *Nomenclature* als das zweite Grundkonstrukt des Packages *BusinessNomenclature* definiert Behälter für eine oder mehrere Vokabeln. Ihre Subklasse *Taxonomy* dient dabei der Zusammenfassung mehrerer Konzepte, welche einen Begriffsraum in Bezug zu einer bestimmten Geschäftsdomäne (*BusinessDomain*) aufspannen. Die Subklasse *Glossary* dient entsprechend der Zusammenfassung mehrerer Terme zu Glossaren. Über eine von der Superklasse *Nomenclature* geerbte



**Abb. 5:** Übersicht über das CWM-Package BusinessNomenclature (vgl. Poole 2002, S. 128; OMG 2001a, S. 14-3f.)

Komposition<sup>3</sup> können sowohl Glossare als auch Taxonomien Teile-Ganzes-Beziehungen zu anderen Glossaren bzw. Taxonomien unterhalten.

Mithilfe einer n:m-Assoziation zwischen der Klasse VocabularyElement und der Klasse ModelElement, die die Superklasse nahezu aller CWM-Bestandteile darstellt, können die Fachbegriffsmetadaten mit anderen Metadaten in Verbindung gebracht werden.

Die genannten Konstrukte bieten eine Reihe von Möglichkeiten zur weitgehenden Systematisierung von Fachbegriffen, die im Rahmen des bisher verfügbaren UBS Metadaten Repository nicht bestanden. Der Nutzen des BusinessNomenclature-Packages als Referenzmodell für das UBS Metadaten Repository lässt sich an folgenden Punkten nachvollziehen.

- Um dem Mangel an Übersicht zu begegnen, der durch die Zusammenführung der Fachbegriffe unterschiedlichster Quellen in eine teilweise redundante Begriffssammlung entstanden ist, bietet es sich an, das bisher vorliegende Gesamtglossar als Komposition mehrere einzelner Subglossare zu modellieren.

<sup>3</sup> Die Komposition zwischen den Klassen Nomenclature und VocabularyElement ist ebenso wenig expliziter Modellbestandteil wie die Kompositionen zwischen ihren Subklassen, da es sich jeweils um die vererbte Assoziation *ElementOwnership* handelt, die bereits im Core-Paket zwischen den Klassen ModelElement und Package definiert ist.



Realisiert man ein Glossar pro Verantwortungsbereich, ergeben sich hieraus gleich mehrere Vorteile: Erstens lässt sich so auf einfache Weise zu jedem Fachbegriff der entsprechende Geltungsbereich ermitteln. Zweitens führen gleich lautende Terme aus verschiedenen Quellen nicht zu einer unerklärten Redundanz sondern können stattdessen besser in ihrem spezifischen Kontext eingeordnet und zueinander in Beziehung gesetzt werden. Drittens ermöglicht es das CWM, jedem Glossar verantwortliche organisatorische Stellen zuzuordnen, die bspw. für die Glossarpflege zuständig sind. Dies kann als entscheidende Voraussetzung für eine dauerhafte organisatorische Etablierung des Metadatenmanagements im Allgemeinen und des Terminologiemanagements im Speziellen gesehen werden (vgl. Auth 2003, S. 274f.). Über diese Vorteile hinaus ermöglicht das CWM durch die Modellierung von Glossarhierarchien die Herstellung einer Gesamtsicht auf alle bereichsspezifischen Glossare ohne deren physische Integration. Auf diese Weise lässt sich auch das bisher realisierte Gesamtglossar rekonstruieren.

- Über ein spezielles Attribut *language* kann zudem die Sprache spezifiziert werden, in der ein Glossar verfasst ist. Gerade in einem multinationalen Unternehmen wie der UBS AG kann dieses Konstrukt helfen, mit der Vielsprachigkeit umzugehen und die Kommunikation zu vereinfachen. Um das Konzept anwenden zu können, muss allerdings jeder Term für jede verwendete Sprache separat definiert werden.
- Die Glossare können in einer nächsten Modellierungsphase um Taxonomien ergänzt werden, indem die bereichsspezifisch definierten Terme mithilfe bereichsübergreifender Konzepte beschrieben und mit diesen Konzepten in Beziehung gesetzt werden. Die Zuordnung der Terme zu anderen Metadaten (z. B. Datenstrukturdefinitionen) wie sie im bisherigen MDR bereits vorhanden waren, können über die Beziehung zwischen den beiden CWM-Klassen *VocabularyElement* und *ModelElement* problemlos in ein CWM-basiertes Modell übernommen werden.

Insgesamt lässt sich daher festhalten, dass die Metamodellkonstrukte des CWM ausreichend erscheinen, um mit Ihnen eine Fachbegriffsmodellierung im UBS DWP durchzuführen. Anzumerken bleibt hierbei, dass es sich bei den CWM-Konstrukten lediglich um das Fachbegriffs-Metamodell handelt und das eigentliche Fachbegriffsmodell erst noch konstruiert werden muss. Die bisher vorliegenden Informationen über Fachbegriffe und ihre Beziehungen lassen sich dabei automatisch in das neue Format überführen, die Konsolidierung der Terme und Beziehungen sowie die Erstellung von Taxonomien bedürfen allerdings eines manuellen Modellierungsprozesses, welcher kontinuierlich durchlaufen wird. Ansätze zur Gestaltung und Etablierung entsprechender organisatorischer Prozesse finden sich in (Auth 2003, S. 196ff.; S. 220-237; S. 268).

### 4.3 Beurteilung der Eignung des CWM als Referenzmodell für ein konzeptionelles DWP-Metamodell

Die Anwendung des CWM in den geschilderten Testszenarien hat zu recht unterschiedlichen Ergebnissen geführt.

Im Bereich der betrachteten technischen Metadaten des UBS DWP musste das Metadatenteam erkennen, dass die unternehmensspezifischen Konstrukte sich mithilfe des CWM nur teilweise abbilden lassen. Der Grund hierfür liegt in einer geringeren Definitionstiefe des Referenzmodells, welches eher breit im Sinne der Unterstützung möglichst vieler verschiedener Metadatenkategorien ausgelegt ist. In den einzelnen Kategorien werden i. d. R. nur die allgemein benötigten Elemente und Beziehungen definiert. Im Bereich der technischen Metadaten ist daher damit zu rechnen, dass Spezifika bestimmter Softwarelösungen – wie sie bspw. bei der UBS existieren – nicht Bestandteil des Referenzmodells sind. Dennoch scheint es auch in diesem Bereich aus drei Gründen durchaus sinnvoll zu sein, die Konstruktion eines konzeptionellen Modells auf dem CWM basieren zu lassen. Zum Ersten kann auf diese Weise zumindest ein Teil der vorhandenen technischen Metadaten standardkonform abgebildet werden, wodurch sich die Möglichkeiten zum Metadatenaustausch verbessern. Zum Zweiten ermöglichen die CWM-Erweiterungsmechanismen bei richtig verstandener Anwendung der zugrunde liegenden Modellierungsprinzipien die flexible Anpassung des Referenzmodells an unternehmensspezifische Anforderungen bei gleichzeitiger Wahrung eines engen Bezugs zu den Standardelementen (vgl. Poole et al. 2002, S. 223 ff.) Ein auf diese Weise konstruiertes Modell zeichnet sich sowohl durch eine möglichst hohe Standardkonformität als auch durch eine sehr durchgängige Struktur und einen hohen Integrationsgrad der Modellbestandteile aus. Zum Dritten bietet sich durch die von der OMG propagierten CWM Extension Packages die Möglichkeit, eigene Modellteile zu verbreiten bzw. Modellteile anderer CWM-Nutzer wieder zu verwenden.

Im Hinblick auf die fachlichen Metadaten konnte festgestellt werden, dass das CWM durchaus geeignet ist, die gewünschten Konstrukte für ein umfassendes Fachbegriffsmodell bereitzustellen. Im Gegensatz zum technischen Szenario, in dem die Metadaten bereits zu grossen Teilen vorliegen, ist die Fachbegriffssammlung des UBS DWP jedoch unvollständig und muss um eine Reihe von Informationen insb. zu Beziehungen zwischen Fachbegriffen ergänzt werden. Das CWM liefert als Metamodell lediglich die Sprache zur Formulierung eines Fachbegriffsmodells, das Modell selbst muss allerdings durch den CWM-Anwender, in diesem Fall das Metadatenteam der UBS, erst erstellt werden. Und da es hierbei vor allem auf die semantische Interpretation der vorliegenden Metadaten ankommt, wird sich die Modellierung nur in geringem Masse automatisieren lassen. Ein konsistentes und aktuelles Fachbegriffsmodell wird sich zudem nur auf Basis klar definierter organisatorischer Prozesse zu Aufbau und Pflege der Fachbegriffe realisieren lassen.

## 5 Zusammenfassung und Ausblick

Der Artikel zeigt den Weg der UBS AG zur Metadatenintegration im Data Warehousing. Hierbei ist es bereits gelungen, die Bestände an technischen und fachlichen Metadaten aus verschiedenen Quellen im Umfeld des UBS Data Warehouse Programms zusammenzuführen und über ein intranetbasiertes Metadaten Reporting einer breiten Nutzergruppe zur Verfügung zu stellen. Die bereits realisierte Verknüpfung von fachlichen und technischen Metadaten ermöglicht schon in vielen Fällen eine durchgängige Navigation von Fachbegriffsdefinitionen zu korrespondierenden Datenbankfeldern und von liefernden Quellsystemen zu Zieltabellen im Data Warehouse. Mithilfe der durch das System hergestellten Übersicht offenbarten sich jedoch auch bisher nicht gekannte Probleme wie unerklärbare Redundanzen und die Schwierigkeit der Identifikation von Beziehungen zwischen Metadaten. Die angestrebte Verbesserung des Systems sieht ein integriertes Metadaten Repository vor, welchem ein einheitliches konzeptionelles Modell des gesamten Metadatenbestandes zugrunde liegt. Anhand zweier Beispielszenarien wird gezeigt, dass der Metadatenstandard Common Object Metamodel der OMG geeignet erscheint, um als Referenzmodell zur Konstruktion eines solchen konzeptionellen Modells herangezogen zu werden.

Die nächsten Schritte zur Realisierung eines integrierten Metadaten Repository wären die Erstellung des konzeptionellen Metadatenmodells sowie dessen anschließende Implementierung. Aufgrund der bereits im Einsatz befindlichen Software sowie des im Unternehmen vorhandenen Know-hows erscheint die Realisierung in einer relationalen Datenbank als sinnvoll. Eine Ausgangsbasis für die Entwicklung könnte das vom CWM-Entwicklerteam propagierte Datenbankschema Metastore sein, welches das CWM-Modell auf eine relationale Datenbank abbildet und mit Datenzugriffsfunktionen anreichert (vgl. Tolbert 2002).

Mit der Implementierung des integrierten Metadaten Repository würde eine wesentliche Basis für ein umfassendes Metadatenmanagement geschaffen, dass neben der integrierten Verwaltung auch die Pflege sowie die Nutzung des Metadaten umfasst und organisatorisch in das Data Warehousing integriert.

## Literatur

- Auth, G.: Prozessorientierte Organisation des Metadatenmanagements für Data-Warehouse-Systeme. Dissertation Universität St. Gallen, Bamberg 2003.
- Bauer A., Günzel. H.: Data Warehouse Systeme: Architektur, Entwicklung, Anwendung. dpunkt.verlag, Heidelberg 2001.
- Devlin, B.: Data Warehouse: from Architecture to Implementation. Addison Wesley, Reading et al. 1997.

- Holten, R.; Dreiling, A.; Schmid, B.: Management Report Engineering - A Swiss Re Business Case. In: von Maur, E.; Winter, R. (Hrsg.): Vom Data Warehouse zum Corporate Knowledge Center. Proceedings der Data Warehousing 2002. Physica, Heidelberg 2002, S. 421-437.
- Do, H.H.; Rahm, E.: On Metadata Interoperability in Data Warehouses. Technical Report Nr. 01(2000), Institut für Informatik, Universität Leipzig, Leipzig 2000.
- Melchert, F.: Das Common Warehouse Metamodel als Standard Metadaten im Data Warehousing. In: von Maur, E.; Winter, R. (Hrsg.): Data Warehouse Management – Das St. Galler Konzept zur ganzheitlichen Gestaltung der Informationslogistik. Springer, Berlin et al. 2003.
- Meyer, M.: Organisatorische Gestaltung des unternehmensweiten Data Warehousing - Konzeption der Rollen, Verantwortlichkeiten und Prozesse am Beispiel einer Schweizer Universalbank. Bamberg 2000.
- OMG: Common Warehouse Metamodel (CWM) Specification. Volume 1. Version 1.0. o. O. 2001a.
- OMG: Common Warehouse Metamodel (CWM) Specification. Volume 2 - Extensions. Version 1.0. o. O., 2001b.
- Poole, J.; Chang, D.; Tolbert, D.; Mellor, D.: Common warehouse metamodel: An Introduction to the Standard for Data Warehouse Integration. John Wiley & Son, Inc., New York et al. 2002.
- Schütte, R.: Grundsätze ordnungsmässiger Referenzmodellierung. Gabler, Wiesbaden 1998.
- Tolbert, D.: CWM Metastore. Unisys, o. O. 2002.