

# **Semantische Anreicherung von strukturierten Daten und Prozessen in analytischen Informationssystemen am Beispiel von MUSTANG**

Matthias Mertens<sup>1</sup>, Yvette Teiken<sup>1</sup>, Jürgen Appelrath<sup>2</sup>

<sup>1</sup> OFFIS - Institut für Informatik Oldenburg

<sup>2</sup> Universität Oldenburg, Praktische Informatik, Abteilung Informationssysteme

## **1 Einleitung**

Analytische Informationssysteme ermöglichen es multidimensionale Ad-Hoc-Analysen mittels On-Line Analytical Processing (OLAP) Operationen auf quantitativen Daten eines Data-Warehouses durchzuführen. Hierbei werden die Auswertungen oftmals in Form von Berichten und zugehörigen Berichts-elementen – Tabellen, Diagrammen, Karten, etc. – visualisiert, gespeichert und gegebenenfalls mit zusätzlichen unstrukturierten Informationen zur besseren Verständlichkeit beschrieben. Diese Berichte können im Sinne einer Informationslogistik dazu dienen, betriebliche Fach- und Führungskräfte zeitnah mit allen relevanten Informationen über interne und externe Sachverhalte und deren Zusammenhängen zu versorgen und sie somit bei der Entscheidungsfindung zu unterstützen (Gluchowski, 2006; Winter, 2008). Eine wesentliche Rolle sollte jedoch auch die Dokumentation von Prozessen und Ergebnissen in analytischen Informationssystemen einnehmen. Dies umfasst sowohl die zugrunde liegenden Fragestellungen als auch die Analyse-durchführung und die Ergebnissicherung. So stellen sich die Fragen: Welche Analysten haben Ad-Hoc Analysen zu spezifischen Fragestellungen durchgeführt? Welche Verfahren, Parameter und OLAP-Operationen wurden im Rahmen dieser Analysen in den einzelnen Analyseschritten auf den quantitativen Daten durchgeführt? Welche Erkenntnisse und Zusammenhänge wurden in den quantitativen Daten gewonnen und welche Fragestellungen konnten durch die Analyse be- bzw. widerlegt werden? Des Weiteren könnte auch von Interesse sein, welche Entscheidungen auf Basis eines Analyseberichts getroffen wurden und wie diese sich in weiteren quantitativen Daten des DWH widerspiegeln.

Die Beantwortung solcher und ähnlicher Fragen spielt insbesondere abseits des Standard Reportings bei Untersuchung komplizierter Fragestellungen eine große Rolle. Die Dokumentationsmöglichkeiten bestehender Ad-hoc Reporting Werkzeuge, die allenfalls in Form unstrukturierter Zusatzinformationen erfolgen (Willenborg), stoßen hierbei schnell

an ihre Grenzen. Sie bieten weder Speicher- und Verwaltungs-, noch Such- und Navigationsmöglichkeiten für die beschriebenen Informationen. Daher wäre es wünschenswert, in einem analytischen Informationssystem über eine Metaebene zu verfügen, die es erlaubt, Informationen zu Fragestellungen, zu Analyseprozessen sowie zu Ergebnissen und deren Beziehungen in Form von strukturierten maschinenlesbaren Metadaten festzuhalten. In diesem Zusammenhang wird in diesem Forschungsansatz auch von semantischen Annotationen auf Basis von Semantic Web Technologien gesprochen, die zur Anreicherung von Prozessen und insbesondere quantitativen Daten verwendet werden. Die hierfür zugrunde liegende Wissensbasis wird durch verschiedene Ontologien umgesetzt.

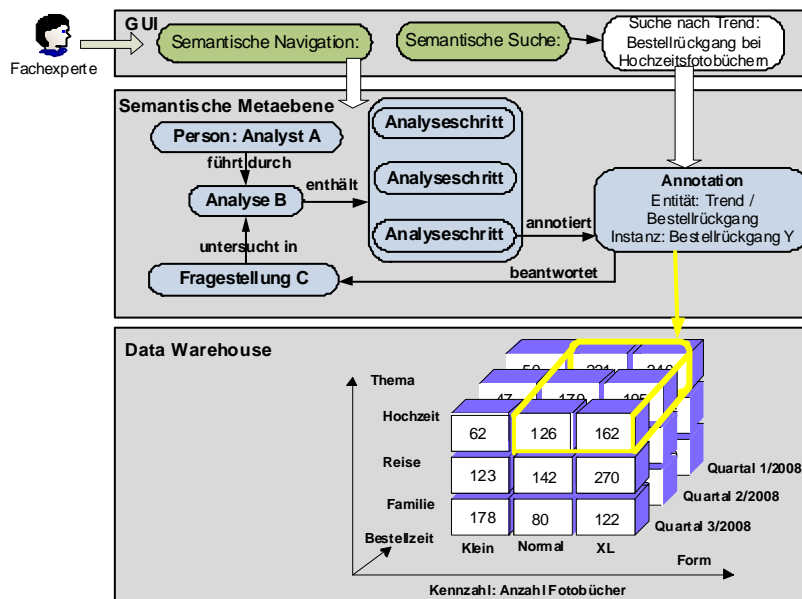


Abbildung 1: Beispiel semantischer Annotationen

Die Vorteile dieses Ansatzes bestehen darin, auf Basis strukturierter semantischer Metadaten auf den quantitativen Daten und Prozessen neue Funktionalitäten in einem analytischen Informationssystem anbieten zu können. Wie das Beispiel in Abb. 1 zeigt, können mit einer semantischen Suche annotierte quantitative Daten des Data-Warehouse mit einer spezifischen Semantik, wie z.B. Trend, Ereignis, Anomalie, Effekt, Muster, etc. gefunden werden. Aber auch die Suche nach komplexeren Annotationen, z.B. zu Analysen, die in einer Beziehung zu einer Fragestellung und zu einer Personen stehen, ist möglich. Des Weiteren ermöglicht dieser Ansatz eine Visualisierung von bereits bestehenden Ergebnissen und hilft so Analysten und Entscheidungsträgern relevante Daten samt ihrer Semantik in Berichtselementen zu erkennen. Darüber hinaus kann, eine semantische Navigation entlang der modellierten „benannten“ Beziehungen in den Metadaten erfolgen.

Ein Beispiel hierfür ist die Navigation von einem Analysten A zu einer Analyse B über den Pfad „führt durch“.

## 2 Verwandte Arbeiten

Schwerpunkt dieser Arbeit ist die Erfassung und Nutzung von Metadaten zu Analyseprozessen und deren Ergebnissen im Business Intelligence (BI) Bereich auf Basis von Semantic Web Technologien. Im Folgenden werden verschiedenen Forschungsbereiche und -arbeiten beschrieben, die sich zum Einen ebenfalls mit der Erfassung und Nutzung solcher Metadaten beschäftigen (Baars, 2006). Zum Anderen werden aber auch Forschungsbereiche aufgezeigt, die nicht unmittelbar die gleiche Forschungsfrage, wie der hier vorgestellte Ansatz verfolgen, jedoch die Potentiale der Verknüpfung des Semantic Webs mit dem BI-Bereich verdeutlichen. Dazu gehören die Bereiche: Verknüpfung von strukturierten und unstrukturierten Daten, Verknüpfung verschiedener BI-Systeme auf Basis einer semantischen Mittelschicht und die Überwindung bestehender Begrenzungen aktueller Data-Warehouse-Systeme.

Erste Synergiepotentiale zwischen dem Semantic Web und der Business Intelligence konnten bereits 2003 von Priebe mit der noch heute relevanten Forschungsrichtung, der Integration und Analyse von strukturierten und unstrukturierten Daten identifiziert werden (Priebe, 2003). Konkret ging es in den Arbeiten um die Verwendung von Ontologien für den Aufbau eines Enterprise Knowledge Portals, das sowohl OLAP-, als auch Information Retrieval- Funktionalitäten in quantitativen Data-Warehouse-Daten und zugehörigen qualitativen Dokument-Daten ermöglicht. Ein ähnlicher Themenschwerpunkt wurde auch in (Haak, 2007) verfolgt. Hier stand die Frage im Fokus, inwieweit sich strukturierte Data-Warehouse-Daten und unstrukturierte Content-Management-Daten auf Basis ihrer Semantik integrieren lassen. Anlog beschrieb Ludwig, dass Wissenrepräsentationssprachen wie RDF unstrukturierte als auch strukturierte Daten (semi-) strukturieren können und so eine Klammer zwischen Dokument und Tabelle, zwischen CMS und DWH, gebildet werden kann (Ludwig, 2005).

Soll ein Corporate Knowledge Center (Klesse, 2003) zur ganzheitlichen Entscheidungsunterstützung des Managements aufgebaut werden, so muss bei der Schaffung einer Integrationsinfrastruktur für die BI-Applikationen eine semantische Verknüpfung strukturierter Daten auf logischer Ebene erfolgen. Für die Umsetzung eines solchen Systems kann eine ontologiebasierte Architektur verwendet werden. In (Spahn, 2008) wird ein ähnlicher Ansatz verfolgt, wobei mit Hilfe einer „business level Ontology“ eine semantischen Mittelschicht gebildet wird, die Endnutzern die Möglichkeit gibt, in einem einheitlichen Vokabular verschiedene BI-Systeme (CRM, ERP, ...) konsistent anzufragen.

Einen besonderen Stellenwert nimmt das Semantic Web in der Überwindung bestehender Begrenzungen von aktuellen Data-Warehouse-Systemen ein, die laut Inmon meist auf die Vernachlässigung von Metadaten zurückzuführen sind (Inmon, 2008). So werden in verschiedenen Forschungsansätzen Business-Ontologien aufgebaut, welche „Business Semantics“ modellieren und diese für das BI-System zur Verfügung stellen (O’Neil, 2007). Sie ermöglichen es, den semantischen Kontext - Annahmen, Definitionen, Business Regeln, Terminologie, Hintergrundinformationen - eines Business Objekts zu berücksichtigen und für neue Funktionen zu nutzen. Xie verwendet diese Technik, um schnell und flexibel die Analysen auf sich dynamisch ändernde Anforderungen anpassen zu können (Xie, 2007). Mit den Cerebra Tools sollen so flexible und erweiterbare Datenrepräsentationen sowie komplexe und feingranulare Business Rules modelliert und berücksichtigt werden (Cerebra, 2004). In Diamantini wird neben einer Domänenontologie für die Business Semantics eine mathematische Ontologie einbezogen, die es erlaubt, die mathematischen Formeln auf den Kennzahlen samt der Semantik von Operatoren und Operanden zu beschreiben, um so neue OLAP Operationen anbieten zu können (Diamantini, 2008). Sell beschreibt mit ihrem Semantic Business Intelligence Framework ebenfalls eine vielversprechende Möglichkeit Business Intelligence mit weiterführenden Semantiken zu verknüpfen. Verbesserte Such- und Navigationsmöglichkeiten in den zu untersuchenden Daten, die Filterung multidimensionaler Daten anhand ihrer Semantik, die Berücksichtigung von Business Rules oder das Vorschlagen weiterführender Analysemöglichkeiten sind nur einige Funktionen einer verbesserten explorativen Analyse. Das automatische Query Rewriting zur semantischen Anreicherung der DWH-Anfrage-Ergebnisse bildet einen Schwerpunkt in diesem Arbeiten (Sell, 2005/2008).

Im Bereich der Dokumentation von Analyseprozessen und -ergebnissen ist die Arbeit von Baars zur Distribution von Business-Intelligence-Wissen zu nennen. Die Arbeit zielt auf die kontrollierte und einfache organisationsweite Verbreitung und Weiterverwendung von Berichten und Analyseansätzen (Templates) durch Einstellen der BI-Inhalte in Wissensmanagementsysteme, jedoch ohne auf eine konkrete technische Umsetzung einzugehen (Baars, 2006).

Im eigenen Ansatz wird ebenfalls die Erfassung von maschinenlesbaren Metadaten zur Durchführung des Analyseprozesses sowie zur Ergebnissicherung angestrebt. Allerdings wird der Fokus auf die Semantik der Entitäten sowie deren Beziehungen zueinander gelegt, um darauf aufbauend semantische Such- und Navigationsfunktionen im analytischen Informationssystem zu realisieren, wie sie in den Arbeiten von (Sell, 2005/2008) auf den zu untersuchenden quantitativen Daten möglich sind. Hierbei verbleiben die semantischen Metadaten jedoch im Gegensatz zu (Baars, 2006) im analytischen Informationssystem und werden nicht in Form von Templates für neue ähnliche Analysen an ein Wis-

sensmanagementsystem weitergegeben. Die Verknüpfung von strukturierten und unstrukturierten Daten sowie die Verknüpfung zu anderen BI-Systemen werden in diesem Ansatz nicht angestrebt. Auch spielt die Modellierung und Berücksichtigung der Business Semantics der zu untersuchenden Objekte in diesem Ansatz zunächst keine Rolle.

### 3 Konzept

Ziel dieses Forschungsansatzes ist die Erfassung und Verwaltung maschinenverständlicher Metadaten zu Analyseprozessen und Ergebnissen und deren Nutzung für semantische Such- und Navigationsmöglichkeiten in einem analytischen Informationssystem. Technologisch bilden Ontologien auf Basis der standardisierten Web-Ontology-Language (OWL) die benötigte Metaebene für die strukturierte Dokumentation. Sie definieren sowohl die Entitäten und ihre Semantik als auch ihre hierarchischen und semantischen Beziehungen zueinander.

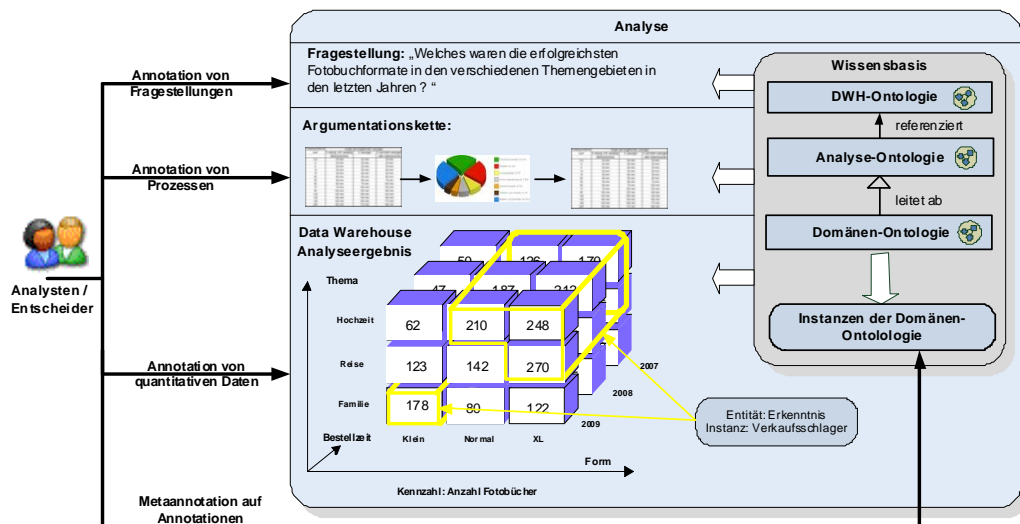


Abbildung 2: Semantische Annotation von Daten und Prozessen

In Abbildung 2 werden die Ontologien unter dem Begriff Wissensbasis zusammengefasst, wobei sie als Grundlage für die semantischen Annotationen dienen. Diese repräsentieren konkrete Instanzen der verschiedenen Entitäten und stellen diese miteinander in Beziehung. So kann z.B. in Abb. 2 eine spezifische Fragestellung erfasst und mit einem konkreten Analyseprozess in eine „untersucht“ Beziehung verknüpft werden. Auch ist es möglich, verschiedene Analyseelemente als Bestandteil eines Analyseprozesses zu kennzeichnen und diese in einer Vorgänger-Nachfolger Beziehung zeitlich anzuordnen. Darüber hinaus können Analyseergebnisse mit einer spezifischen Semantik wie „Verkaufsschlager“ auf quantitativen Daten annotiert werden, was in Abb. 2 durch die gelbe Um-

randung auf dem Datenwürfel visualisiert wird. Metadatenannotationen ermöglichen es Aussagen über Annotationen zu treffen und so z.B. eine Ursache / Wirkungsbeziehung zwischen zwei Annotationen – z.B. Marketingaktion und Verkaufsschlager – zu beschreiben. Im Folgenden werden der Aufbau der jeweiligen Ontologien, deren Beziehungen zueinander sowie ihre Entstehung näher erläutert.

### **3.1 DWH-Ontologie**

Um eine Schnittstelle zwischen den Realweltentitäten des Data-Warehouses und den Analyseprozess- und Ergebnisentitäten herstellen zu können, wird eine Data-Warehouse-Ontologie benötigt, die ein Mapping auf die Entitäten des DWH-Schemas, wie Kennzahlen, Dimensionen, Dimensionshierarchien, etc. beschreibt. Hartmann stellt in diesem Zusammenhang eine OWL-Ontologie zur semantisch reichhaltigen Beschreibung von DWH-Metadaten zur Verfügung, die auf dem Semantic-Data-Warehouse-Modell (SDWM) basiert (Hartmann, 2008; Böhnlein, 2001). Ergänzt um weitere Konzepte aus dem am OFFIS entwickelten Metamodell MADEIRA (Wietek, 2000) und reduziert um nicht benötigte Entitäten ergibt sich das in Abb. 3 visualisierte Schema.

Die dargestellten Entitäten und Beziehungen ermöglichen es Metadaten zu den Datenstrukturen des Data-Warehouses sowie deren Analysemöglichkeiten festzuhalten und diese mit weiteren Metadaten des Analyseprozesses in Verbindung zu setzen. Die in Abb.3. visualisierten Entitäten und Beziehungen reichen für das hier verfolgte Ziel aus, jedoch wäre es durch die Erweiterung der DWH-Ontologie theoretisch möglich eine ähnliche Business Ontologie, wie in (Sell ,2005/2008) oder (Xie, 2007) aufzubauen. Hartmann sieht bereits weitere Strukturen zu den Bereichen Terminologie, Organisations- und Systembezug, Datenqualität, Datentransformationen und Historie vor. Daneben wäre es aber auch möglich Konzepte zu modellieren, die spezifische Aussagen über die zu analysierenden quantitativen Daten enthalten. Dazu zählen insbesondere semantische Beziehungen, die in der Realwelt existieren, von bestehenden hierarchischen DWH-Strukturen aber nicht erfasst werden können. So können spezifische Business Semantics (O’Neil, 2007) der im DWH untersuchten Objekte explizit festgehalten werden.

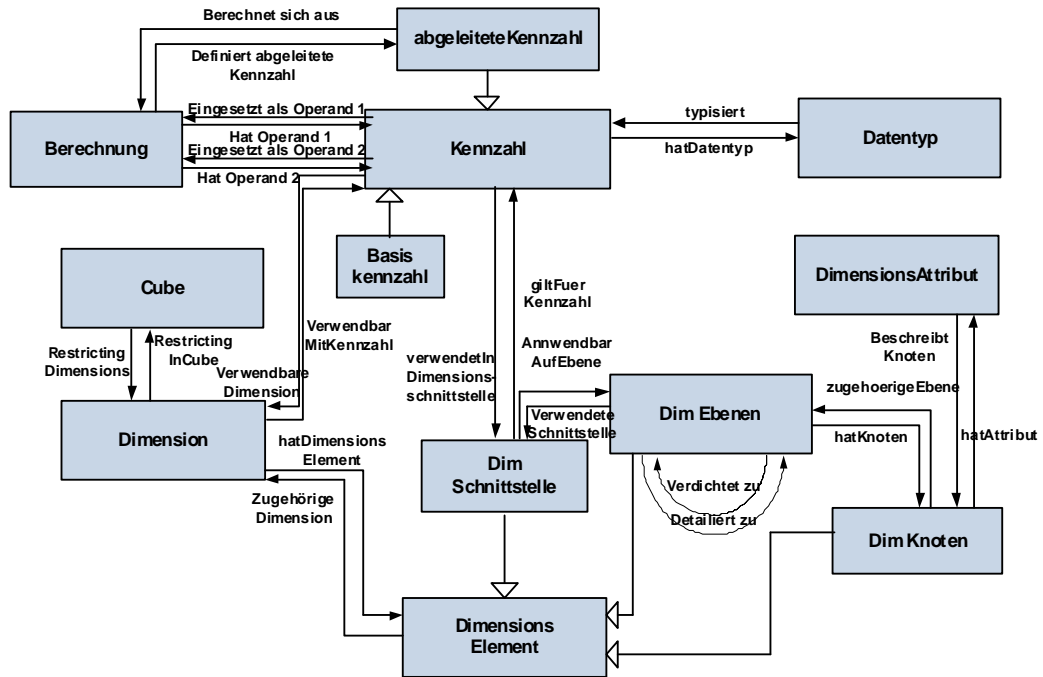


Abbildung 3: Entitäten und Beziehungen der DWH - Ontologie

### 3.2 Analyse- und Domänen-Ontologie

Die Analyse-Ontologie gliedert sich in zwei Bereiche, wobei der erste die allgemeine Definition von Entitäten und Relationen aus dem Umfeld analytischer Informationssysteme fokussiert. Er erlaubt es, Konzepte wie Personen, Analysen, Fragestellungen, Berichtselemente sowie deren semantische Beziehungen zu beschreiben (s. Abb. 4). Darauf basierend können semantische Annotationen zur Dokumentation von Fragestellungen und Analyseprozessen erfolgen. Der zweite Bereich dient der Ergebnissicherung. Er definiert zum Einen Entitäten, die von „direkte Annotation“ ableiten und es dem Analysten erlauben, erkannte Ergebnisse in den quantitativen Daten festzuhalten. Zum Anderen definiert der zweite Bereich aber auch Metadatenannotationen, um Aussagen über die Annotationen zu treffen. Sowohl direkte Annotationen als auch Metadatenannotationen verfügen über eine Vererbungshierarchie, um spezifischere Annotationstypen, wie Effekte, Anomalien, Trends, etc. beziehungsweise Zusammenhänge, Sachverhalte, etc. zu beschreiben. Zu beachten ist, dass diese Entitäten zunächst domänenunabhängig modelliert sind, um die Ontologie in verschiedenen Analyseszenarien einsetzen zu können. Da jedoch in verschiedenen Domänen unterschiedliche Ergebnisse annotiert werden können, werden Domänenontologien definiert, welche die Entitäten der Analyse-Ontologie um domänenspezifische Konzepte erweitern. So ist es möglich in der Domäne epidemiologisches Krebsregister eine Entität „Erhöhung der altersspezifischen Inzidenz“ zu modellieren die

ebenso wie die Entität „Bestelleinbruch“ in der Domäne Fotobuch von der Entität „Trend“ ableitet.

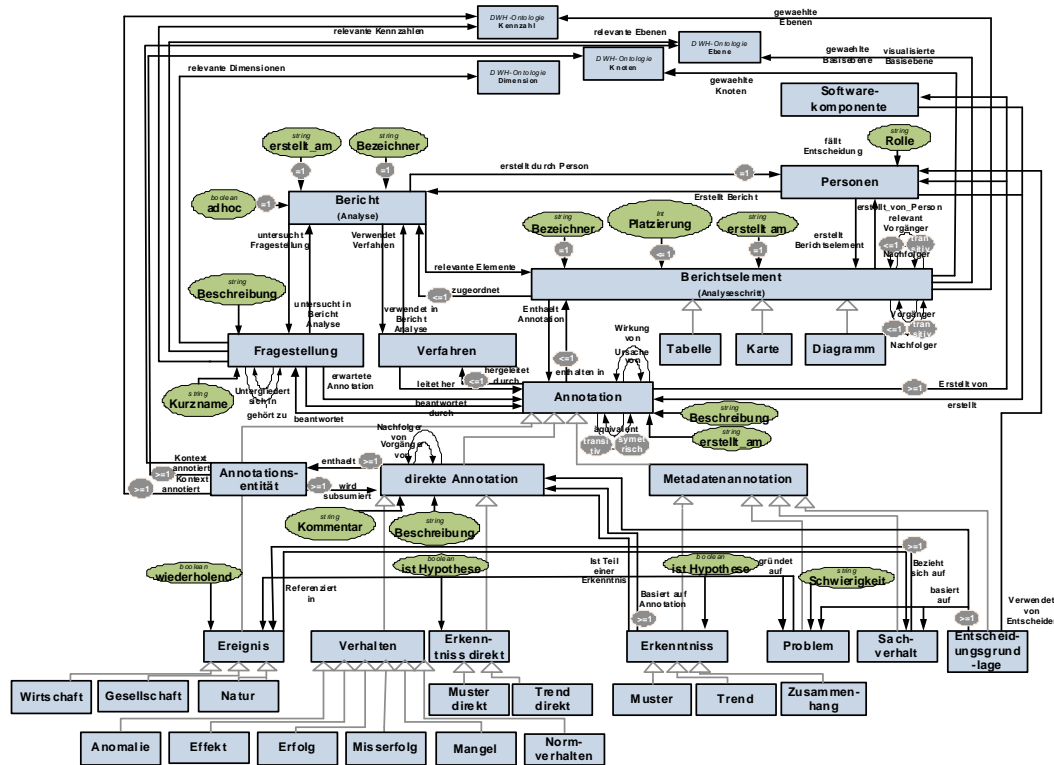


Abbildung 4: Entitäten und Beziehungen der Analyse-Ontologie

Eine Verknüpfung der Analyse-Ontologie und der DWH-Ontologie wird über die semantischen Beziehungen mehrerer Entitäten zueinander hergestellt. Berichtselemente speichern beispielsweise die Informationen, welche Kennzahlen in welchen Dimensionsebenen und Knoten dargestellt werden und welche Dimensionsebenen und Knoten im Kontext gewählt sein müssen. Auch spezifizieren Fragestellungen, die Kennzahlen, Dimensionen, Ebenen und Knoten, die in der DWH-Ontologie instanziiert und im Rahmen einer Untersuchung von Interesse sind. Um semantische Annotationen auf den quantitativen Daten auszuführen, die z.B. einen spezifischen Trend beschreiben, müssen ebenfalls Beziehungen zu der DWH-Ontologie bestehen. Hierbei ist zu beachten, dass eine direkte Annotation sich nicht auf die Faktdaten selbst, sondern auf die DWH-Metadaten (Dimensionen, Ebenen, Knoten), welche die quantitativen Kennzahlausprägungen aufspannen, beziehen. Über das Konzept der Annotationsentitäten ist es möglich, einzelne Kennzahlenbereiche (s. Abb.2) zu wählen. Hierbei werden mehrere Annotationsentitäten in einer direkten Annotation subsumiert.



Da die hier gewählte OWL-DL-Ontologie auf Beschreibungslogik basiert, ist es möglich, Restriktionen auf den Beziehungen der Entitäten zu definieren und diese darüber hinaus als transitiv, symmetrisch und funktional zu markieren. So kann z.B. spezifiziert werden, dass eine Annotation von mindestens einer Person oder Softwarekomponente erstellt werden muss. Dies erlaubt eine Konsistenzprüfung der modellierten Ontologieinstanzen.

### 3.3 Semantische Anwendungen

Auf Basis der vorgestellten Ontologien lässt sich eine semantische Metaebene in das analytische Informationssystem einbinden, die es ermöglicht, Analyseprozesse und Ergebnisse strukturiert und maschinenverständlich zu dokumentieren. Dies ermöglicht nicht nur die Visualisierung bestehender Ergebnisse samt ihrer Semantik während der Exploration, sondern erlaubt auch mit Hilfe einer semantischen Suche komplexe Anfragen an die Wissensbasis zu senden. Hierbei kann nach allen Entitäten der Analyse-Ontologie insbesondere unter Berücksichtigung der ein- und ausgehenden semantischen Beziehungen gesucht werden. Technisch wird eine Umsetzung auf Basis der „SPARQL Query Language for RDF“ erfolgen. Die semantischen Beziehungen zwischen den aufgezeichneten Metadaten erlauben eine explorative Navigation entlang benannter Pfade.

## 4 Implementierung in der MUSTANG Plattform

OFFIS beschäftigt sich seit vielen Jahren mit epidemiologischen Fragestellungen. Ursprung dieser Aktivitäten war die Bereitstellung eines analytischen Informationssystems für das Epidemiologische *Krebsregister* Niedersachsen (EKN). Über die Zeit hat sich dieses Werkzeug als sehr mächtig erwiesen, so dass es neben dem Public Health Bereich vielfältig auch in anderen Projekten beispielsweise als Management Unterstützungssystem zum Einsatz gebracht worden ist. Da diese Analyse-Infrastruktur ursprünglich als Expertensystem für die Epidemiologie konzipiert worden war, stellte sich bei Einsätzen in anderen Domänen die intuitive Bedienbarkeit nicht immer ein. Auch erwies sich die ursprüngliche Systemarchitektur als unflexibel, so dass Anpassungen an neue Bereiche und Anforderungen sich als schwierig erwiesen. Aus diesem Grund wurde mit der Neuentwicklung der multidimensionalen, geografischen statistischen Datenanalyse (MUSTANG) Plattform begonnen. Hierbei spielte der Aspekt der Serviceorientierung eine wesentliche Rolle. Im Folgenden werden kurz die drei Kerngebiete von MUSTANG aufgelistet.

- **Multidimensional:** Daten, die mittels der MUSTANG Plattform zur Analysen verwendet werden sind multidimensional aufbereitet.

- **Erweiterte Statistik:** Für die Analysen können vielfältige erweiterte statistische Verfahren angewendet werden. Diese Eigenschaft hat ihren Ursprung in der Epidemiologie, da hier komplexe Verfahren realisiert werden müssen.
- **Geographisch:** Daten können einen Geographiebezug besitzen, der dann zur räumlichen statistischen Analyse herangezogen wird. Hierbei unterstützen wir sowohl kleinräumige wie Flächenanalysen.

#### 4.1 Beschreibung der MUSTANG Plattform

Die MUSTANG Plattform basiert auf einer rekonfigurierbaren serviceorientierten Architektur, wobei die Services zustandslos sind. Daten werden über sogenannte Datentransfer Objekte ausgetauscht. Diese Objekte beschreiben zusammen den Zustand des Systems. Die Services lassen sich in einem gewissen Rahmen zu Anwendungen komponieren. Services werden abstrakt als Interfaces konzipiert und werden dann durch ggf. verschiedene Realisierungen konkretisiert. Sowohl bei den Services wie auch bei den Datentransfer Objekten muss eine größtmögliche Stabilität gewährleistet werden, da Änderungen an diesen Definitionen große Umstrukturierungen mit sich ziehen.

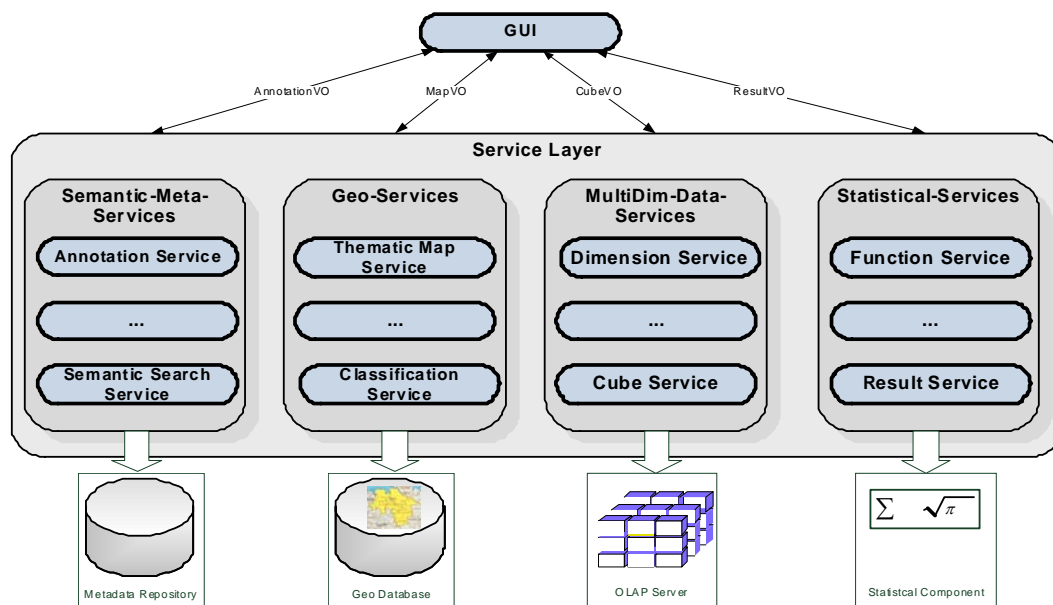


Abbildung 5: Erweiterte MUSTANG Architektur

In Abbildung 5 ist die Architektur von MUSTANG mit ihren verschiedenen Schichten zu sehen. Auf der untersten Schicht befinden sich die Standardkomponenten, wie Geo-Datenbank, OLAP Server und Statistik-Komponente. Diese werden in einer darüber liegenden Service-Schicht in eine einheitliche Sicht integriert, so dass diese Fremdkompo-

nen transparent zugreifbar sind. Die oberste Schicht bildet schließlich die graphische Benutzungsschnittstelle (GUI).

Die Service-Schicht unterteilt sich in vier voneinander abgegrenzte Anwendungsblöcke mit unterschiedlichem Funktionsumfang. Die Realisierung dieser Blöcke erfolgt durch konkrete Services.

- **Geo-Services:** Hier werden Funktionen zum Umgang mit geografischen Daten umgesetzt, aktuell der Classification-Service zur Anfrage von Geo-Objekten aus einer Geodatenbank, sowie der Thematic-Map-Service zum Erzeugen thematischer Karten. Zurzeit wird PostGIS als Standard-Geodatenbank integriert.
- **Multidimensional-Data-Services:** Diese Services kapseln den Zugriff auf den OLAP-Server. Der Dimension-Service ist für die Abfragen von Dimensionen und deren Elementen zuständig, der Cube-Service für das Abfragen von Cubes innerhalb des OLAP-Servers. Der eigentliche Zugriff auf den OLAP-Server erfolgt mittels der Abfragesprache XMLA. Dies ermöglicht es, auf einfache Weise andere OLAP-Server an die Plattform anzubinden. Unterstützte Standard-OLAP-Server sind zurzeit Microsoft Analysis Services und Mondrian.
- **Statistical-Services:** Hier werden die von MUSTANG unterstützten statistischen Verfahren, für komplexe Berechnungsmethoden auf OLAP-Cubes, realisiert. Hierbei ist, auch über die Grenzen einzelner OLAP-Cubes hinweg, eine große Vielzahl unterschiedlicher statistischer Verfahren und Berechnungsmöglichkeiten realisierbar. Der Function-Service stellt dabei die im System integrierten statistischen Verfahren bereit. Der Result-Service dient zum Abruf der um statistische Daten angereicherten OLAP-Cubes bzw. Kennzahlen. Die Berechnungen werden nicht innerhalb der Plattform durchgeführt, sondern durch Einbindung der Standard-Statistik-Komponente RProject.
- **Semantic-Meta-Services:** Realisiert die in diesem Artikel vorgestellte Verfahren zur strukturierten Erfassung semantischer Metadaten im Analyseprozess. Die Einbettung in die MUSTANG Plattform wird im folgenden Abschnitt erläutert.

Der Service-Layer ist flexibel gehalten, so dass auch weitere Standardkomponenten, wie beispielsweise weitere Geo- Datenbanken oder andere OLAP-Server, integriert werden können.

## 4.2 Einbettung des Forschungsansatzes in die MUSTANG Plattform

Mit dem hier vorgestellten Forschungsansatz wird ein neuer Anwendungsblock in den Servicelayer eingeführt, der mit neuen Funktionalitäten bestehende Probleme analytischer Informationssysteme im Allgemeinen und MUSTANG im Speziellen im Bereich der Dokumentation relevanter Zusatzinformationen zu Fragestellungen, Analyseprozessen und Ergebnissen sowie deren Semantiken adressiert. Durch die Erfassung solcher Informationen in Form von strukturierten Metadaten, die als Instanzen eines DWH-, BI- und Domänenmodells in einen Kontext eingebettet werden, ergeben sich neue Visualisierungs-, Such- und Navigationsmöglichkeiten, die insbesondere eine bessere Nachvollziehbarkeit von Analysewegen zur Ergebnisfindung, den Ergebnissen selbst und darauf basierenden Entscheidungen ermöglichen.

Die Semantic-Meta-Services:

- **OntologyManagementService:** Dieser Service dient der Verwaltung der in Kapitel 3. vorgestellten Ontologien, wobei er das Erzeugen und Ändern sowie Laden und Speichern von Ontologiemodellen, sowie konkreten Entitäten und Instanzen kapselt. So kann im Hintergrund ein beliebiges Ontologie-Verwaltungs-System eingesetzt werden, ohne das weitere Services angepasst werden müssen. Eine Umsetzung erfolgt durch die Jena2 Ontology API.
- **AnalyseManagementService:** Dieser Service dient unter anderem dem Erfassen strukturierter Metadaten zum Einloggen von spezifischen Nutzern, zum Erstellen von Analysen und dem Erstellen von Fragestellungen.
- **DWHOntologyService:** Dieser Service verwaltet die Data-Warehouse-Ontologie. Durch ihn werden alle relevanten Instanzen des DWH-Schema in die Ontologie geladen, verwaltet und wieder auf das DWH-Schema abgebildet.
- **AnnotationService:** Dieser Service erlaubt das Erstellen von direkten Annotationen auf den quantitativen Daten und das Erstellen von Metaannotationen auf diesen.
- **LoggingService:** Mit diesem Service werden die strukturierten Daten rund um die Analysedurchführung strukturiert festgehalten. Dies umfasst unter anderem die gewählten Parameter, die durchgeführten OLAP Operationen sowie die gewählten Visualisierungen der Daten.

- **VisualisationService:** Dieser Service realisiert die Visualisierung bestehender Ergebnisse in Form von Annotationen auf quantitativen Daten während der Exploration im analytischen Informationssystem.
- **SearchService:** Dieser Service realisiert die Suche nach Metadaten mit einer spezifischen Semantik. Intern werden Fragen in der SPARQL Query Language for RDF formuliert und an das Ontologie-Verwaltungs-System geschickt. Die zurückgesendeten Antworten werden verarbeitet und für den Systembenutzer aufbereitet.
- **NavigationService:** Dieser Service erlaubt die Durchführung von Navigationen entlang semantischer Pfade, wie sie in den zugrunde liegenden Analyse- und Domänen-Ontologien modelliert sind.

Alle Services können intern und mit der GUI über Datentransferobjekte kommunizieren. Die hier aufgelisteten Services sind dabei nicht isoliert zu betrachten sondern stehen ebenfalls mit den MultiDim- Data-Services in Verbindung. Analog zu den Services werden entsprechende GUI Elemente realisiert, welche die im Forschungsansatz beschriebenen Funktionalitäten umsetzen.

## 5 Fazit und Ausblick

Ziel des vorgestellten Forschungsansatzes ist es, strukturierte Dokumentationen von Analysefragestellungen, Analyseprozessen und erkannten Erkenntnissen und Zusammenhängen auf Basis semantischer Metadaten in analytischen Informationssystemen umsetzen zu können. Insbesondere abseits des Standardreportings kann mit diesem Ansatz die Nachvollziehbarkeit komplexer Analysen, deren Ergebnissen und darauf basierender Entscheidungen verbessert werden, indem semantische Such- und Navigationsfunktionalitäten in dem analytischen Informationssystem angeboten werden. Es wurde beschrieben, dass mit Hilfe von Semantic Web Technologien eine maschinenlesbare und -verständliche Metaebene aufgebaut werden kann, die es erlaubt, die semantischen Annotation zu erfassen, zu speichern und zu verwalten und schließlich in Form neuer Funktionalitäten dem Analysen oder auch der Fach- oder Führungskraft wieder zur Verfügung zu stellen. In diesem Zusammenhang wurden die Data-Warehouse- und die Analyse-Ontologie vorgestellt, die den Kern der Metadatenverwaltung bilden. Erstere bildet Entitäten und deren Zusammenhänge des DWH ab, um diese in anderen Ontologien zugreifen zu können und eignet sich darüber hinaus zur Modellierung sogenannter Business Semantics. Die Analyse-Ontologie hingegen modelliert Entitäten und deren hierarchische und semantische Beziehungen, um Fragestellungen und Prozesse strukturiert festhalten zu können. Des Weiteren definiert sie domänenunabhängige Entitäten, um erkannte Analyseergebnisse in quantita-

tiven DWH-Daten dokumentieren zu können. Spezifischere Domänenontologien leiten sich von der Analyse-Ontologie ab, um konkrete Annotationsmöglichkeiten für die jeweilige zu untersuchende Domäne zu modellieren.

Eine Umsetzung der semantischen Metaebene zur Speicherung und Verwaltung erfasster Metadaten sowie darauf aufbauende Such-, Navigations- und Visualisierungsmöglichkeiten werden in der am OFFIS entwickelten MUSTANG-Plattform prototypisch umgesetzt. Eine Implementierung von entsprechenden Services und GUI-Elementen wird derzeit realisiert.

Weiterführende Arbeiten werden sich mit der intelligenten Nutzung erhobener Analyse- und Ergebnismetadaten, insbesondere in Kombination mit modellierten Business Semantics der zu untersuchenden quantitativen Daten, beschäftigen. Im Fokus steht zum Einen die Bildung von Klassifikatoren auf Basis bestehender Analyseergebnisse zur semiautomatischen Exploration und Annotation von neuen Analyseergebnissen während des Analyseprozesses. Zum Anderen wird eine zielgerichtete und weiterführende Analyseunterstützung für Ad-Hoc Analysen im analytischen Informationssystem angestrebt, von der insbesondere Fach- und Führungskräfte, aber auch Analysten profitieren sollen.

## 6 Literatur

Baars, H. (2006). Distribution von Business-Intelligence-Wissen: Diskussion eines Ansatzes zur Nutzung von Wissensmanagement-Systemen für die Verbreitung von Analyseergebnissen und Analysetemplates. In P. Chamoni, & P. Gluchowski, *Analytische Informationssysteme* (pp.409-424). Berlin, Heidelberg: Springer.

Böhnlein, M. & vom Ende, A.U. (2001). Semantisches Data Warehouse-Modell (SDWM): Ein konzeptuelles Modell für die Erstellung multidimensionaler Datenstrukturen. Retrieved February 8, 2009 from [http://www.ceushb.de/forschung/downloads/SDWM\\_Rundbrief.pdf](http://www.ceushb.de/forschung/downloads/SDWM_Rundbrief.pdf)

Diamantini, C. & Potena, D. (2008). Semantic Enrichment of Strategic Datacubes. In I.Song & A. Abelló (Eds.), *DOLAP 2008, ACM 11th International Workshop on Data Warehousing and OLAP*, ACM

Gluchowski, P. (2006). Techniken und Werkzeuge zum Aufbau betrieblicher Berichtssysteme In P.Chamoni, & P.Gluchowski, *Analytische Informationssysteme* (pp.207-226). Berlin, Heidelberg: Springer.

Haak, L. (2007). *Semantische Integration von Data Warehousing und Wissensmanagement*. Berlin: dissertation.de – Verlag im Internet GmbH.

Hartmann, S. (2008). *Überwindung semantischer Heterogenität bei multiplen Data-Warehouse-Systemen*, Bamberg: University of Bamberg Press.

Inmon, W.H., Strauss, D. & Neushloss, G. (2008). *DW 2.0: The Architecture for the Next Generation of Data Warehousing* (Morgan Kaufman Series in Data Management Systems), Burlington: Elsevier Science.

Klesse, M., Melchert, F. & von Maur, E. (2003). Corporate Knowledge Center als Grundlage integrierter Entscheidungsunterstützung In U. Reimer, A. Abecker, S. Staab & G. Stumme (Eds.), WM 2003: Professionelles Wissensmanagement - Erfahrungen und Visionen. GI.

Ludwig, L. (2005). Business Intelligence und das Semantic Web: ein Traumpaar. Retrieved May 15, 2009 from <http://www.competence-site.de/business-intelligence-hr/Business-Intelligence-und-das-Semantic-Web-39683>

Network Inference Inc. (2004). Ontology And Data Warehousing: How Data Warehousing Investments can deliver continous ROI by augmenting them with ontologies and inference capabilities Retrieved March 22, 2009 from [http://me.jtpollock.us/pubs/2005.10-Whitepaper\\_DataWarehouse.pdf](http://me.jtpollock.us/pubs/2005.10-Whitepaper_DataWarehouse.pdf)

O'Neil, B. (2007). Semantics and Business In The Data Administration Newsletter. Retrieved September 9, 2009 from <http://www.tdan.com/view-articles/4934>

Priebe, T. & Pernul, G. (2003). Ontology-based Integration of OLAP and Information Retrieval. In V. Marik, W. Retschitzegger & O. Stepankova (Eds.), 14th International Workshop on Database and Expert Systems Applications (pp. 610-614). Heidelberg: Springer.

Sell, D., Caroso da Silva, D., Beppler, F.B., Napoli, M., Ghisi, F.B., Pacheco, R.C.S. & Todesco, J.L. (2008). SBI: a semantic framework to support business intelligence. In A. Duke, M. Hepp, K. Bontcheva & M.B. Vilain (Eds.), Proceedings of the First International Workshop on Ontology-supported Business Intelligence. ACM.

Sell, D., Cabral, L., Motta, E., Domingue, J., Hakimpour, F. & Pacheco, R. (2005). A Semantic Web based Architecture for Analytical Tools In 7th IEEE International Conference on E-Commerce Technology (CEC 2005)

Spahn, M., Kleb, J., Grimm, S. & Scheidl, S. (2008). Supporting Business Intelligence by Providing Ontology-Based End-User Information Self-Service intelligence. In A. Duke, M. Hepp, K. Bontcheva & M.B. Vilain (Eds.), Proceedings of the First International Workshop on Ontology-supported Business Intelligence. ACM.

Wietek, F. (2000). Intelligente Analyse multidimensionaler Daten in einer visuellen Programmierumgebung und deren Anwendung in der Krebsepidemiologie. Universität Oldenburg.

Willenborg, K., Berichtswesen - Vereinigung von Gegensätzen. Retrieved July 10, 2009 from <http://www.sapdesignguild.org/editions/edition2/willenborg/index.html>

Winter, R., Schmaltz, M., Dinter, B. & Bucher, T. (2008). Das St. Galler Konzept der Informationslogistik. In B. Dinter & R. Winter, Integrierte Informationslogistik. (pp.1-16). Berlin, Heidelberg: Springer.

Xie, G., Yang, Y., Liu, S., Qiu, Z., Pan, Y. & Zhou, X. (2007). EIAW: Towards a Business-Friendly Data Warehouse Using Semantic Web Technologies In K. Aberer, K. Choi, N. Fridman Noy, D. Allemang, K. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, P. Cudré-Mauroux (Eds.), The Semantic Web, 6th International Semantic Web Conference. Berlin, Heidelberg: Springer.