

Universität Salzburg  
Naturwissenschaftliche Fakultät  
Institut für Geographie und angewandte GeoInformatik

---

**Die Integration von Geodaten in  
Data-Warehouse-Systeme**  
am Beispiel von endkundenrelevanten Daten in  
Energie-Versorgungsunternehmen

Abschlussarbeit des Universitätslehrgangs „UNIGIS professionell“

Vorgelegt bei Prof. Dr. Josef Strobl

Von  
Dipl.-Volksw. Jochen Pogrzeba  
Gottesauer Str. 18  
D-76131 Karlsruhe  
Up10196

24. Juli 2002

## **Abstract**

Die vorliegende Arbeit soll aufzeigen, wie sich Geodaten in ein Data-Warehouse-Konzept integrieren lassen.

Ein aktiv agierendes Unternehmen benötigt vielfältige Informationen über sein Marktumfeld, um Entscheidungen zu treffen und seine gesteckten Ziele erreichen zu können. Das Data-Warehouse-Konzept bietet dazu eine Grundlage, unternehmensrelevante Informationen in strukturierter Form, speziell für analytische und strategische Zwecke, zu speichern. Dazu gehören heute verstärkt Informationen, die einen räumlichen Bezug beinhalten. Allerdings ist der Data-Warehouse-Ansatz für den Bereich der geographischen Daten noch nicht allzu weit fortgeschritten. So findet man in der Praxis selten Data-Warehouse-Lösungen in denen ganz explizit Geodaten integriert sind. Auch die Anbieter geographischer Daten bieten diese zumeist als „Stand-alone-Lösung“ an. Die Integration Ihrer Daten in den unternehmensweiten Datenbestand spielt in Ihren Beratungsleistungen eine eher untergeordnete Rolle.

Eine Besonderheit an geographischen Daten, im Gegensatz zu anderen Unternehmensdaten, liegt darin, dass Geodaten nicht nur aus alpha-numerischen Sachdaten bestehen, sondern auch eine graphische Komponente beinhalten. Diese Komponente gilt es in die Struktur eines Data-Warehouses zu integrieren.

Diese Arbeit will daher Möglichkeiten skizzieren, wie man geographische Daten mit einem eher alpha-numerisch-geprägten Datenumfeld verknüpfen könnte. Nach der Darstellung möglicher Alternativen soll eine dieser Alternativen näher betrachtet werden. Unter diesem „rationalen Geo-Data-Mart“ soll eine praktikable Data-Warehouse-Komponente verstanden werden, die für die Arbeit mit Geodaten geeignet ist. Hier sollen typische Data-Warehousing-Prozesse vor dem Hintergrund der Integration von Geoinformation behandelt werden.

Als Praxisbezug soll sich die das Management von endkundenorientierten Daten von Energie-Versorgungsunternehmen mit Geomarketing-Anwendungen durch die Arbeit ziehen.

## Gliederungsübersicht

1. Einführung	1
1.1 Problemstellung	1
1.2 Zielsetzung und Vorgehensweise	1
2. Data-Warehousing, OLAP und Data-Mining	3
2.1 Data-Warehousing	3
2.1.1 Funktionen	3
2.1.2 Architektur	4
2.1.3 Data-Marts	6
2.2 OLAP (Online Analytical Processing)	7
2.3 Data-Mining	10
3. Die spezielle Situation von Energie-Versorgungsunternehmen	12
3.1 Informationsmanagement in Energie-Versorgungsunternehmen	12
3.2 GeoMarketing als Entscheidungsgrundlage in Energie-Versorgungsunternehmen	14
3.3 Geodaten in Energie-Versorgungsunternehmen	16
4. Geodaten in Data-Warehouse-Systemen	19
4.1 Integrationskonzepte	19
4.2 Monolithische GIS-Arbeitsplätze	23
4.3 Georelationale Speicherung	23
4.4 Geocodierung	26
4.5 Georationale Speicherung	29
4.6 (Geo-)Datenmodellierung	32

5. Geo-Data-Warehousing	35
5.1 Der georationale Data-Mart	35
5.2 Prozesse	37
5.2.1 Analyse und Extraktion der Informationen	37
5.2.2 Transformation	38
5.2.3 Integration bzw. Laden in das Data-Warehouse	38
5.3 Meta-Daten	39
5.4 Data-Mining mit Geodaten	42
5.5 Das (Geo-)Data-Warehouse als Bestandteil eines Marketing-Systems	45
6. Fallbeispiel: Penetrationsanalyse für den Strom-Endkundenmarkt	48
6.1 Aufgabenstellung	48
6.2 Data-Warehouse-Modell	49
6.3 Ladeprozesse	51
6.4 OLAP-Prozesse	54
6.5 GIS-Integration	58
6.5.1 Einrichten der Datenquelle	59
6.5.2 Verknüpfung der Attributdaten mit den Geometrien	59

Anhang: Literaturverzeichnis

---

## Abbildungsverzeichnis

Abb. 1: Data-Warehouse-Referenz-Architektur	5
Abb. 2: Tabelle (Beispiel)	8
Abb. 3: Tabelle in Matrix überführt (Beispiel)	8
Abb. 4: Data Cube (Beispiel)	9
Abb. 5: Prozesskette eines Energie-Versorgungsunternehmens	13
Abb. 6: Beispiel einer Trassendokumentation mit SICAD	15
Abb. 7: Ablauf der Datenbankintegration	19
Abb. 8: Integrationsvarianten für Geodaten	22
Abb. 9: Objektrelationales Modell (Beispiel)	24
Abb. 10: GIS-DBMS-Architektur mit ArcSDE	25
Abb. 11: Geocoding	27
Abb. 12: Geocodierung im Batch-Betrieb	28
Abb. 13: Geoschlüssel	31
Abb. 14: Struktur eines logischen Datenmodells	33
Abb. 15: Relationales Datenmodell	34
Abb. 16: Geographischer Data-Mart	36
Abb. 17: Metadaten in der Datenhierarchie	39
Abb. 18: Metadatenelemente nach DIN ENV 12657	41
Abb. 19: Der geographische Datenquader	44
Abb. 20: OLAP-Funktionalität mit Karten	45
Abb. 21: Data-Warehouse-Modell für einen „geographischen Data-Mart“ (Auszug)	50
Abb. 22: Tabellenstruktur des operativen Systems „Anschlussverwaltung“ (Auszug)	51
Abb. 23: Tabellenstruktur der externen Demographiedaten (Auszug)	52
Abb. 24: Tabellenstruktur der Pärchentabelle „Mikrogebiet vs. Strassen“ (Auszug)	53
Abb. 25: Auftragserstellungs-Assistent (SQL-Server 2000)	54
Abb. 26: Ergebnistabelle der Sicht olap_anchluesse_gkz	55
Abb. 27: Ergebnistabelle der Sicht olap_demogr_gkz	56
Abb. 28: Ergebnistabelle der Sicht olap_anchluesse_mikro	57
Abb. 29: Ergebnistabelle der Sicht olap_demogr_mikro	57
Abb. 30: Ergebnistabelle der Sicht olap_penetration_mikro	58
Abb. 31: Einrichten der ODBC-Datenquelle (MapInfo professional 5.5)	59
Abb. 32: Verknüpfung der Sachdaten-Relation mit der Geometrie-Relation	60
Abb. 33: Beispiel für Penetrationsanalyse für GKZ-Gebiete	61

Abb. 34: Beispiel für Penetrationsanalyse für Mikro-Gebiete	62
Abb. 35: Beispiel für Penetrationsanalyse für Mikro-Gebiete (Auszug)	63

*“Ten years ago I could have told you how Doritos were selling west of the Mississippi. Today... I can tell you how well they are selling in California, in Orange County, in the town of Irvine, in the local Von’s Supermarket...”*

*D.W. Calloway, CEO von PepsiCo<sup>1</sup>*

## **1 Einführung**

### **1.1 Problemstellung**

Neben den klassischen Produktionsfaktoren rückt der Faktor „Information“ immer mehr in das Bewusstsein der Unternehmen. Ein aktiv agierendes Unternehmen benötigt vielfältige Informationen über sein Marktumfeld, um Entscheidungen zu treffen und seine gesteckten Ziele erreichen zu können. Dazu gehören heute verstärkt Informationen, die einen räumlichen Bezug beinhalten. Gerade im Marketing ist diese räumliche d.h. geographische Komponente einer Information gegeben, denn ein Markt definiert sich nicht nur durch seine sachlichen, sondern auch durch seine regionalen Eigenschaften. Berücksichtigt man dies, ist es möglich, Marktgegebenheiten besser zu verstehen und somit eine effizientere Marktbearbeitung zu bewerkstelligen.

Um aus Informationen Wissen zu generieren, ist es unabdingbar, diese als „Daten“ zu begreifen und zu speichern. Ein relativ junger Ansatz dazu ist das Data-Warehouse-Konzept, welches speziell für analytische Zwecke geschaffen wurde. Allerdings ist dieser Ansatz für den Bereich der geographischen Daten noch nicht allzu weit fortgeschritten. So findet man in der Praxis selten Data-Warehouse-Lösungen in denen ganz explizit Geodaten integriert sind. Auch die Anbieter geographischer Daten bieten diese zumeist als „Stand-alone-Lösung“ an. Die Integration Ihrer Daten in den unternehmensweiten Datenbestand spielt in Ihren Beratungsleistungen eine eher untergeordnete Rolle.

### **1.2 Zielsetzung und Vorgehensweise**

Ziel der vorliegenden Arbeit ist es, aufzuzeigen wie sich Geodaten in ein Data-Warehouse-Konzept integrieren lassen. Im Gegensatz zu anderen Unternehmensdaten ist dabei dem Umstand Rechnung zu tragen, dass Geodaten nicht nur aus alpha-numerischen Sachdaten bestehen, sondern auch eine graphische Komponente beinhalten.

Ziel der Arbeit ist es nicht, eine umfassende praktische Projektbeschreibung vorzulegen. Es soll vielmehr eine theoretische Grundlage bieten, was bei der Integration von Geodaten in

---

<sup>1</sup> Zitiert in: Lusti (1999), S. 113

ein Data-Warehouse-System zu beachten ist, und wie diese Integration grundsätzlich aussehen könnte. Sie soll dazu beitragen, Ideen zu generieren, wie man diese Thematik angehen könnte, bevor man sich an die konkrete Umsetzung mittels Softwarelösungen macht. Sie richtet sich daher hauptsächlich an Data-Warehouse-Verantwortliche, die keinen Bezug zu geoinformatischen Fragestellungen besitzen, oder auch an Leser, die mit geographische Datenbeständen bisher nur als „Stand-alone-Daten“, also außerhalb eines Data-Warehouses, gearbeitet haben.

Kapitel 2 gibt einführend einen Überblick über das theoretische Grundgerüst und die Basisfunktionen des Data-Warehouse-Konzeptes. Um ein Data-Warehouse nicht nur vom Aspekt der Datenhaltung zu beschreiben, sondern auch von der Transaktionsorientierung, werden auch damit zusammenhängende Begriffe wie OLAP oder Data-Mining umrissen.

In Kapitel 3 wird der Rahmen dessen beschrieben, was sich als Praxisbezug durch die Arbeit ziehen soll: Das Management von endkundenorientierten Daten von Energie-Versorgungsunternehmen mit Geomarketing-Anwendungen. Hierbei soll der veränderte Energiemarkt und seine Auswirkungen auf die Informationsbedürfnisse und Geomarketing-Notwendigkeiten dargestellt werden. Es soll dann der Bogen geschlagen werden zur Rolle von Geodaten in diesen Anwendungen.

Kapitel 4 geht dann auf die eigentliche Integration von Geodaten in bestehende Data-Warehouses ein. Dabei soll auf mögliche praktische Varianten eingegangen werden, wie Geodaten mit dem Unternehmensdatenbestand verknüpft werden können.

In Kapitel 5 wird beschrieben, wie ein sogenannter „geographischer Datamart“ innerhalb der Data-Warehouse-Konzeption aufgestellt werden kann. Darunter kann man eine Data-Warehouse-Komponente verstehen, die für die Arbeit mit Geodaten und Geomarketing-Anwendungen geeignet ist. Hier sollen typische Data-Warehousing-Prozesse, wie ETL oder Data-Mining, vor dem Hintergrund der Integration von Geoinformation beschrieben werden.

Kapitel 6 schließt die Arbeit mit einem Fallbeispiel. Anhand einer Penetrationsanalyse soll ein kurzes Beispiel zeigen, wie ein typischer Data-Warehouse-Prozess innerhalb dieses „geographischer Datamart“ ablaufen könnte.

## 2 Data-Warehousing, OLAP und Data-Mining

### 2.1 Data-Warehousing

#### 2.1.1 Funktionen

Die Grundlage jeder Information sind Daten. Daten werden aber erst dann zu einer Information, wenn Sie dazu beitragen, Erkenntnisse zu generieren, die den Empfänger dazu befähigen, eine gestellte Aufgabe zu lösen. In den meisten Fällen wird die Datenquelle nicht mit dem Datenempfänger übereinstimmen, so dass Daten eine Metamorphose durchlaufen müssen, um dann für den Empfänger eine Information darzustellen. Dieser Prozess der Metamorphose ist sehr oft mehrstufig, meistens iterativ und nicht immer in Prozessen abzubilden.

Im Allgemeinen müssen Daten gesammelt, gespeichert, aufbereitet und analysiert werden um daraus Informationen abzuleiten. Die Qualität dieser Schritte bestimmt schließlich auch, ob aus den Informationen letztendlich verwertbares Wissen wird.

Im alltäglichen Umgang mit Daten kommt besonders dem Aspekt der Speicherung eine bedeutende Rolle zu: Sie schafft die Grundlage für die weitere Verwertbarkeit.

Daten, die miteinander thematisch und konzeptionell in Verbindung stehen, lassen sich in einer Datenbank sammeln, welche wiederum Teil eines computergestützten Informationssystems darstellt.<sup>2</sup>

Datenbanken lassen sich nach vielfältigen Kriterien definieren und differenzieren. Je nachdem, inwieweit die dort gespeicherten Daten in den operativen bzw. strategischen Prozessen des Unternehmens eingebunden sind, kann man in produktiven- und analytischen Datenbanken unterscheiden.<sup>3</sup> Eine **Produktionsdatenbank** unterstützt das operative Geschäft, indem sie Transaktionen, „...laufend, vollständig und redundanzarm fortschreibt.“<sup>4</sup>

Die Datenstrukturen und -modelle dieser Datenbanken sind für laufende Unternehmensprozesse optimiert, so dass diese Datenbanken eher unübersichtlich und wenig benutzerfreundlich sind. Auswertungen zur Informationsgenerierung sind mit diesen Datenbanken kaum effizient realisierbar.

Aus diesem Grunde wurden Anfang der 90er-Jahre vermehrt **analytische Datenbanken** eingesetzt, die diese Mängel beseitigen sollten. Für diese Datenbanken wurde der Begriff des **Data-Warehouses** geprägt, der auf den Unternehmensberater W.H. Inmon zurückgeht:

---

<sup>2</sup> Vgl. Bill / Zehner (2001), S. 54

<sup>3</sup> Vgl. Lusti (1999), S. 124

<sup>4</sup> Lusti (1999), S. 123

*„A Data-Warehouse is a subject-oriented, integrated, time-variant and nonvolatile collection of data in support of management decision support process”<sup>5</sup>*

Ein Data-Warehouse stellt somit eine Datenbasis dar, die aus mehreren operativen Datenbanken bereitgestellt wird, von diesen aber physisch getrennt betrieben wird. Ein Data-Warehouse ist von seiner Struktur her nicht an praktikable Datenmodelle gebunden, sondern eher am betriebswirtschaftlichen Umfeld des Unternehmens ausgerichtet.<sup>6</sup> Dies kann sich z.B. darin äußern, dass eine Datenbank, die in einem Data-Warehouse vorliegt, nicht zwangsläufig normalisiert sein muss. Auch kann das Postulat der Redundanzfreiheit aufgehoben sein, da auch abgeleitete bzw. schon aufbereitete Daten enthalten sein können. Das Data-Warehouse-Konzept beinhaltet jedoch nicht nur das Management von Daten und deren Speicherung, sondern auch Algorithmen und Methoden zu deren Analyse. Somit wird deutlich, dass die Hauptaufgabe eines Data-Warehouse-Konzeptes, im Gegensatz zu operativen Datenbanken, in der unternehmensweiten und fachübergreifenden Bereitstellungen von Analysemöglichkeiten des Unternehmens-Datenbestand besteht.

### **2.1.2 Architektur**

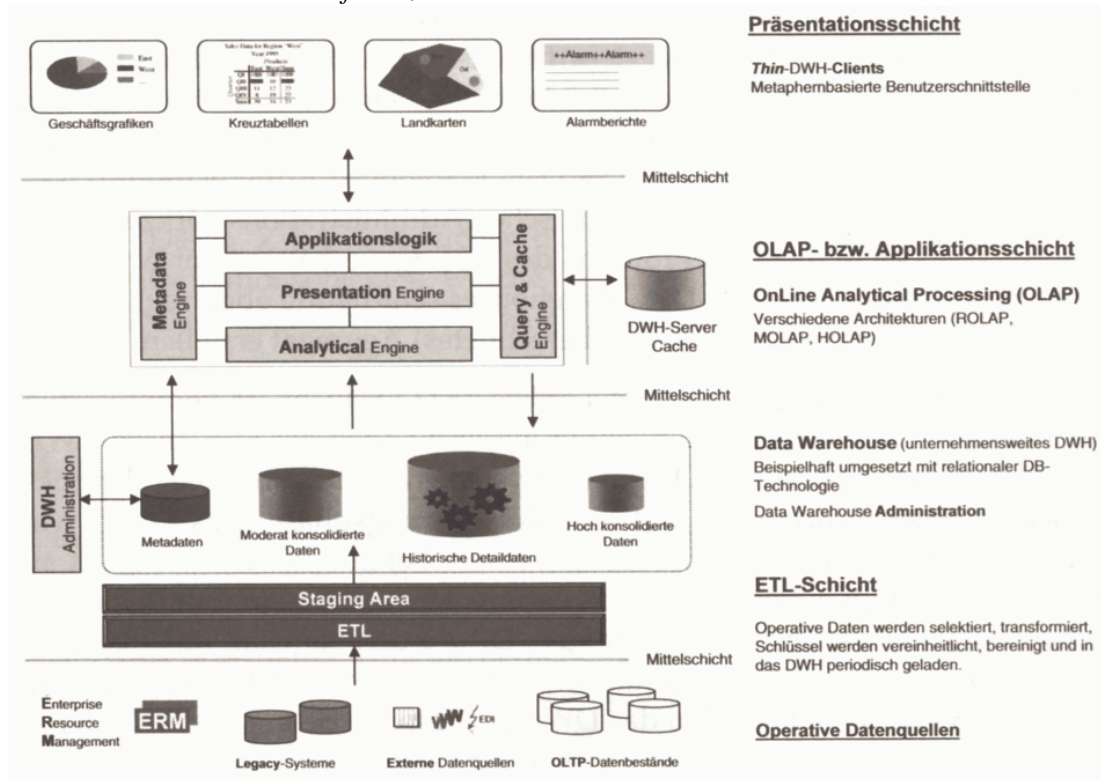
Die technische Architektur eines Data-Warehouse hängt stark von dessen Komplexität ab. So können mehrere Server-Anwendungen integriert sein, die die einzelnen Komponenten wie Datenbank oder Analyse-Anwendungen verwalten und bereitstellen. Soll auf ein Data-Warehouse über Intra-/Internet zugegriffen werden, stellt sich noch die Web-Komponente dazu. Nachfolgend soll auf die grundlegende logische Architektur eingegangen werden, die einem Data-Warehouse zugrunde liegt:

---

<sup>5</sup> Zitiert in: Kurz (1999), S. 49

<sup>6</sup> Vgl.: Kurz (1999), S. 49

Abb. 1: Data-Warehouse-Referenz-Architektur<sup>7</sup>



Grundlage eines Data-Warehouse bilden wie in Abb. 1 dargelegt **operative Datenquellen**, die aus diversen Unternehmensdatenbanken zusammengeführt werden können. Dies können Daten aus CRM<sup>8</sup>, Warenwirtschafts oder ähnlichen Systemen sein, sowie auch externe Daten wie mikrogeographische Haushaltsdaten.

Darüber befindet sich die **ETL<sup>9</sup>-Schicht**, welche für den Transfer und Aufbereitung der Daten sorgt. Die Daten werden nach einem vorher definierten Prozess extrahiert, transformiert und über die Staging Area<sup>10</sup> in die **Data-Warehouse-Schicht** geladen, welche das Herzstück eines Data-Warehouses darstellt. Da in diesem Stadium die Daten bereits an Datenmodelle oder Entscheidungsregeln geknüpft werden, spricht man auch von DSP-Daten („Decision Support Processing“)<sup>11</sup>

Diese Komponente ist für die eigentliche Datenhaltung zuständig und wird durch ein DBMS<sup>12</sup> realisiert. Somit ist hier auch das Data-Dictionary implementiert, welche die Meta-Informationen<sup>13</sup> des Data-Warehouse bereitstellt.

<sup>7</sup> Kurz (1999), S. 187

<sup>8</sup> Customer Relationship Management

<sup>9</sup> Extracting, Transformation and Loading

<sup>10</sup> Die Staging Area stellt den Arbeitsbereich dieser Schicht dar. Hier können Datenextrakte aufgebaut werden, die nach festgelegten Übernahmekriterien und -zeitpunkten weiterverarbeitet werden.

<sup>11</sup> Vgl.: Hummeltenberg (1998), S. 55

<sup>12</sup> Datenbank-Management-System

<sup>13</sup> Meta-Informationen (oder Meta-Daten) sind Daten über Daten, d.h. sie beschreiben den Inhalt und die Struktur der Daten und dienen somit zu deren Verwaltung; vgl.: Bill/Zehner (2001) S. 174

Die übergeordnete Schicht, die sogenannte **Applikationsschicht**, stellt die Schnittstelle zwischen den Informationsbedürfnissen der Benutzer und den Daten dar. Diese Schicht wird auch als **OLAP<sup>14</sup>-Schicht** bezeichnet. Hier ist somit die Logik implementiert, wie aus den integrierten Daten mittels Analyse-Algorithmen Erkenntnisse zu gewinnen sind. Die Benutzer-Abfragen werden hier mittels Abfragesprachen (z.B. SQL<sup>15</sup>) an die Data-Warehouse-Schicht weitergeleitet, welche die Ergebnisse über die Applikationsschicht an die Benutzerschnittstelle weiterleitet.

Diese Benutzerschnittstelle ist in der **Präsentations- oder Visualisierungs-Schicht** realisiert. Diese ist client-basiert, d.h. die Visualisierungs-Anwendungen laufen auf den Rechnern der Endanwender, und sollten sich somit an deren Bedürfnisse und Kenntnisse anpassen lassen. Diese Anwendungen sind im Allgemeinen klassische Reporting-Tools, die Berichte generieren und damit Daten in Kreuztabellen oder Präsentationsgrafiken visualisieren können. Aber auch andere entscheidungsunterstützende Systeme, wie Statistik-Programme oder Geo-Informationssysteme können in dieser Schicht implementiert sein. Der Begriff Präsentation sollte aber nicht vergessen lassen, dass die Benutzerschnittstellen nicht nur zum Auslesen von Daten dienen, sondern auch die Benutzeranfragen (nach vorher definierten Prozessen) an die Applikationsschicht weiterleiten.

### 2.1.3 Data-Marts

Ein Data-Warehouse dient als analytische Informationsbasis für ein ganzes Unternehmen, oder zumindest für einen größeren Teilbereich. Dies hat den Vorteil, dass sowohl Administrationstätigkeiten als auch die Datenintegration zentral erfolgen können. Nachteil ist jedoch, dass individuellen Bedürfnissen bei den Data-Warehouse-Nutzern eher schlecht entsprochen werden kann. So wird beispielsweise der Vertrieb andere Analysen durchführen als der technische Betrieb. Um diesem Umstand gerecht zu werden, wird man in der Praxis aus einem großen unternehmensweiten Data-Warehouse einzelne „kleine“ Data-Warehouses ableiten, die sogenannten **Data-Marts**. Diese sind den speziellen Bedürfnissen bestimmter Einheiten angepasst, indem sie mit bestimmten Analysemethoden und/oder einem definierten Datenextrakt ausgestattet sind. Diese Einheiten können Abteilungen, Projekte oder Prozesse sein. Man spricht dabei von „Subjektbereichen“<sup>16</sup>.

Data-Marts stellen somit einen Ausschnitt aus dem übergeordneten unternehmensweiten Data-Warehouse dar. Neben diesen sog. *abgeleiteten* oder *abhängigen* Data-Marts kann man auch *proprietäre* Data-Marts unterscheiden. Diese sind eigenständige Lösungen, welche nicht aus einem übergeordneten zentralen Data-Warehouse gespeist werden, sondern für die Benutzergruppen einzeln aufgesetzt sind. Meist sind dies Lösungen, die historisch

---

<sup>14</sup> Online Analytical Processing, siehe Kapitel 2.2

<sup>15</sup> *Structured Query Language*, eine standardisierte Abfragesprache für DBMS

<sup>16</sup> Vgl.: Degen (1998), S. 100

gewachsen sind, da sie den Erfordernissen einer einzelnen Abteilung entsprechen. Jedoch führt dieses Data-Mart-Konzept zu einer Vielzahl von Insellösungen, die der ursprünglichen Idee eines Data-Warehouses nicht mehr gerecht werden können. Daher kann dies nicht als wirtschaftlich sinnvolle Lösung angesehen werden<sup>17</sup>.

## 2.2 OLAP (Online Analytical Processing)

Der Begriff OLAP<sup>18</sup> ist eng mit dem Begriff des Data-Warehousing verbunden. OLAP-Systeme haben zum Ziel, für Entscheider und Entscheidungsvorbereiter auf möglichst effiziente und direkte Weise („online“) den Zugang zu relevanten Informationen zu sichern. OLAP stellt also eine Software-Technologie dar, welche Funktionen und Methoden zur Analyse der aufbereiteten Daten des Data-Warehouses liefert. Dabei lassen sich auch Informationen ableiten, wenn sie sich nicht direkt aus der vorgegebenen Datenbankstruktur ergeben.

Obwohl OLAP-Systeme konzeptuell unabhängig von Data-Warehouse-Konzepten gesehen werden können, werden sie in der Praxis als Bestandteil der Data-Warehouse-Philosophie akzeptiert.<sup>19</sup> Zentrale Eigenschaft eines OLAP-Systems besteht in der Multidimensionalität der zu auswertenden Variablen.<sup>20</sup> Daher wurde der Begriff des „**OLAP-Würfels**“ oder „Data Cube“ geprägt, welcher eine konkrete analytische Sicht auf die Daten ermöglicht.

Zur Erläuterung ein Beispiel:

---

<sup>17</sup> Vgl.: Kurz (1999), S. 599

<sup>18</sup> Der Begriff OLAP geht auf *E.F. Codd* zurück, der auch das Gedankengerüst für relationale Datenbanken lieferte.

<sup>19</sup> Vgl.: Alpar, Niedereichholz (2000), S. 14

<sup>20</sup> Vgl.: Chamoni, Gluchowski (1998), S. 402

Abb. 2: Tabelle (Beispiel)

Region	Produkt	Kundenzahl
Saarland	Allg. Tarif	4.100
Bayern	Umwelttarif	1.200
Bayern	Solartarif	3.100
Niedersachsen	Allg. Tarif	2.800
Niedersachsen	Umwelttarif	1.600
Niedersachsen	Solartarif	1.200
Thüringen	Umwelttarif	1.000
Thüringen	Solartarif	1.000
Saarland	Solartarif	3.500

Diese Tabellendarstellung ist geeignet für eine Speicherung der Daten, doch für Analysezwecke eher unübersichtlich. So müsste für eine Ermittlung der Gesamtkunden des Produktes „Solartarif“ eine Auswahlabfrage über die komplette Tabelle durchgeführt werden. Besser wäre eine Darstellung in Matrixform:

Abb. 3: Tabelle in Matrix überführt (Beispiel)

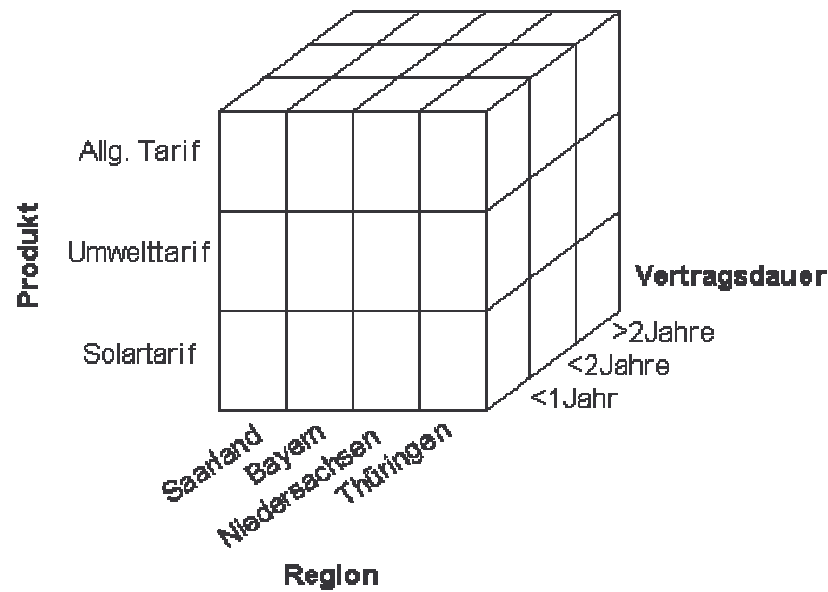
		Region			
		Saarland	Bayern	Niedersachsen	Thüringen
Produkt	Allg. Tarif	4.100	0	2.800	0
	Umwelttarif	0	1.200	1.600	1.000
	Solartarif	3.500	3.100	1.200	1.000

Hier besteht der Vorteil darin, dass die Abfrage nach den Gesamtkunden des Produktes „Solartarif“ ein einfaches Aufsummieren der Zeile „Solartarif“ darstellt. Die Schlüsselworte „Region“ und „Produkt“ stellen die **Dimensionen** dar, die Namen der Ausprägungen („Saarland“ usw. ) die **Positionen**.<sup>21</sup>

Fügt man zu dieser zwei-dimensionalen Darstellung noch eine weitere Dimension hinzu, bspw. die Vertragsdauer, erhält man einen Würfel:

<sup>21</sup> Vgl.: Breitner u.a. (1996), S.110

Abb. 4: Data Cube (Beispiel)



Der Begriff „**OLAP-Würfel**“ ist allerdings irreführend, da auch mehr als drei Dimensionen integriert sein können. Somit kann man im geometrischen Sinne nicht mehr von einem Würfel sprechen, der Begriff ist aber mittlerweile Synonym für Analysemöglichkeiten mit OLAP geworden.

Um die Daten so anzuzeigen, wie es für die individuellen Auswertebedürfnisse ausreichend ist, existieren mehrere Funktionen, die auf OLAP-Würfel angewandt werden können<sup>22</sup>:

- **Data Slicing** (bzw. -Rotation oder -Pivoting):  
Um je nach Auswertung verschiedene Daten des OLAP-Würfels abzufragen, rotiert man - bildlich gesprochen - den OLAP-Würfel. Dadurch werden unterschiedliche Sichten generiert.
- **Data Dicing** (bzw. -Ranging):  
Mit dieser Technik werden Daten gefiltert, d.h. nur ein Ausschnitt des OLAP-Würfels wird zur Verfügung gestellt.
- **Roll-up** (bzw. Drill-up):  
Hiermit wird eine Aufwärts-Bewegung innerhalb eines Hierarchiebaumes bezeichnet. Ein Hierarchiebaum könnten z.B. die Gebietskörperschaften darstellen. Aggregiert man Gemeindedaten auf Bundesländerebene bezeichnet man dies als Roll-up. Der Gegensatz dazu bildet das
- **Drill-down**:  
Dies kennzeichnet eine Abwärts-Bewegung innerhalb eines Hierarchiebaumes. Großräumige Daten werden auf kleinere Gebietseinheiten disaggregiert.

<sup>22</sup> Vgl.: Breitner u.a. (1996), S.112

Diese Funktionen sind aufgrund Ihrer Komplexität nur bedingt mit einer relationalen Datenbank abzubilden. So müssen noch zusätzliche Softwarekomponenten zu dem DBMS hinzugefügt werden, um Mehrdimensionalität in den Daten abzubilden. In diesem Falle spricht man dann von einem **relationalem OLAP** (ROLAP).

Wird die OLAP-Funktionalität direkt auf dem Client-PC ausgeführt, spricht man von einem **Desktop-OLAP** (DOLAP) .

### 2.3 Data-Mining

Unter **“Data-Mining“** versteht man umgangssprachlich das „Schürfen“ in den Daten zum Zwecke der Wissensbildung. Diese Daten werden idealerweise durch ein Data-Warehouse bereitgestellt, da hier die Daten aufbereitet und zentral vorliegen. So kann man Data-Mining als ein Baustein des Data-Warehouse-Konzeptes ansehen.<sup>23</sup>

Im Gegensatz zum OLAP-Ansatz, welcher primär auf die Untersuchung von definierten Gegebenheiten abzielt, versucht man mit Data-Mining-Methoden *Muster und Zusammenhänge* in den Daten zu finden. Data-Mining kann daher definiert werden als „... die Anwendung spezifischer Algorithmen zur Extraktion von Mustern aus Daten“.<sup>24</sup> Diese Algorithmen werden aus diversen Disziplinen entlehnt, so z.B. aus der Statistik, der Datenbankforschung, der theoretischen Informatik oder der Expertensystemforschung. Im Gegensatz zur schließenden Statistik, welche formulierte Hypothesen untersucht, läuft das Data-Mining hypothesenfrei ab, d.h. die Erkennung von Auffälligkeiten und die Ableitung von Aussagen steht im Vordergrund.<sup>25</sup>

Ein Data-Mining-System beinhaltet mehrere Verfahren und Techniken. Diese kann man in zu lösende „*Aufgaben*“ und in „*Methoden*“ diese zu lösen, unterscheiden<sup>26</sup>. Zu den „*Aufgaben*“ oder „*Anwendungsklassen*“<sup>27</sup>, die ein Data-Mining-System bearbeiten kann, zählen hauptsächlich:

- **Klassifikation:**  
Objekte werden nach bestimmten Regeln in vorher definierten Klassen eingeordnet, die diese Objekte dann beschreiben helfen. Eine Methode hierfür wäre beispielsweise die Diskriminanzanalyse.

---

<sup>23</sup> Theoretisch ist eine Data-Mining-Aktivität auch ohne ein Data-Warehouse denkbar, doch wird man sich in der Praxis schwer tun, Data-Mining direkt mit operativen Daten, also mit nicht-aufbereiteten Daten, durchzuführen.

<sup>24</sup> Fayyad, U. zitiert in Alpar, Niedereichholz (2000), S. 3

<sup>25</sup> Dies betrifft die Definition für Data-Mining **i.e.S.**. Für eine Data-Mining-Definition **i.w.S.** werden auch klassische Methoden der schließenden Statistik und OLAP-Methoden hinzugezählt.

<sup>26</sup> Vgl.: Alpar, Niedereichholz (2000)

<sup>27</sup> Vgl.: Lusti (1999), S. 250

- **Segmentation (Clustering-Systeme):**  
Objekte werden nach bestimmten Regeln in Gruppen zusammengefasst, welche dadurch eine Eigenschaft erhalten. Eine Methode hierfür wäre beispielsweise die Clusteranalyse.
- **Abhängigkeitsanalyse:**  
Hier werden Beziehungen zwischen untersuchenden Objekten untersucht und daraus Regeln abgeleitet, mit denen dann in Form einer „If...then“-Regelsuche gearbeitet werden kann.
- **Abweichungsanalyse:**  
Mit diesen Analyseverfahren werden Objekte (z.B. betriebswirtschaftliche Kennzahlen) identifiziert, die den Regelmäßigkeiten, welche mittels der Abhängigkeitsanalyse gefunden wurden, nicht gehorchen.
- **Prognose:**  
Merkmalswerte in einer unbekanntem Zukunft werden auf der Basis von bekannten Merkmalswerten der Vergangenheit vorhergesagt.
- **Entscheidungsbaumverfahren:**  
Diese Verfahren bilden Cluster aus Daten. Als „Nebenprodukt“ werden die Entscheidungen, welche zur Clusterbildung geführt haben, als Regeln definiert. Da das System dadurch lernt („Maschinelles Lernen“), können diese dann auf ähnliche Daten angewandt werden.
- **Künstliche neuronale Netze:**  
Nachbildung der Vernetzung von Nervenzellen in Form von Software. Ziel ist das Lernen aus Fehlern und die Anwendung des Gelernten auf neue Daten.
- **Sequentielle Muster:**  
Es sollen Muster und Gesetzmäßigkeiten gefunden werden, die alleine aus der Anordnung der Daten resultieren. Ein Beispiel ist die klassische Zeitreihenanalyse zur Exploration von Trends oder Periodizitäten.
- **Text-Mining:**  
Hierunter versteht man die Suche nach Mustern in Texten, z.B. „Webcrawler“, also Tools zur systematischen Suche nach Schlüsselwörtern im Internet
- **Graphische Systeme:**  
Hierunter fallen Systeme, die Datenmuster visuell erkennen lassen, bspw. komplexere Graphiktools, oder auch das Kartographie-Modul eines Geo-Informationssystems.

### 3 Die spezielle Situation von Energie-Versorgungsunternehmen

#### 3.1 Informationsmanagement in Energie-Versorgungsunternehmen

Seit der EU-Richtlinie zur Liberalisierung der europäischen Energiemärkte aus dem Jahr 1996, welche die Mitglieder der Europäischen Union zur schrittweisen Öffnung ihrer nationalen Strommärkte verpflichtet, haben sich die Rahmenbedingungen in den nationalen Energiemärkten grundlegend verändert.

So gab es vor dem Inkrafttreten des neuen Energierechts im April 1998 für jedes Unternehmen der öffentlichen Stromversorgung genau abgegrenzte Versorgungsgebiete ohne jegliche Wahlmöglichkeiten für den Energiebezug der Verbraucher.

Die Neuregelung in Deutschland begnügte sich nicht mit einer abgestuften Öffnung des Marktes, sondern entließ die Stromwirtschaft von einem Tag auf den anderen in den vollen Wettbewerb. In Österreich wurde zweistufig vorgegangen: Im Februar 1999 wurde der Markt für Industriekunden geöffnet und im Oktober 2001 der Markt für die restlichen Kunden, wie Privathaushalte oder mittelständische Betriebe.

Geschützte Versorgungsgebiete gibt es seitdem nicht mehr. Technisch bleiben die Verbraucher zwar weiterhin dem Netz des angestammten Versorgers verbunden, die Rechnung können sie sich aber vom jeweils günstigsten Anbieter ausstellen lassen. Über die alten, aufgehobenen Grenzen der Versorgungsgebiete hinweg vermischen sich deshalb die Kundenstämme.

Es muss nun um jeden Kunden regelrecht geworben werden. Die Liberalisierung schuf eine völlig neue Situation, die alte Strukturen sterben und neue entstehen lässt. So verschwindet das Energie-Versorgungsunternehmen alten Typs, das häufig Stromerzeugung, Netzbetrieb und Verkauf unter einem Dach vereinte. Das Kartellgesetz schreibt vor, dass die Bereiche Erzeugung, Übertragung, Verteilung und Vertrieb, sowie Aktivitäten außerhalb des Elektrizitätsbereichs, organisatorisch und buchhalterisch strikt voneinander zu trennen sind. Diese Differenzierung, das sogenannte „Unbundling“, stellt die Unternehmen vor neue Aufgaben im Prozessablauf der einzelnen Konzernbereiche. Diese Konzernbereiche sind daher als selbstständige Bereiche organisiert, welche die komplette Prozesskette von der Stromerzeugung bis zur Kundenabrechnung betreuen.

Diese Prozesskette stellt sich am Beispiel der EnBW AG<sup>28</sup> wie folgt dar:

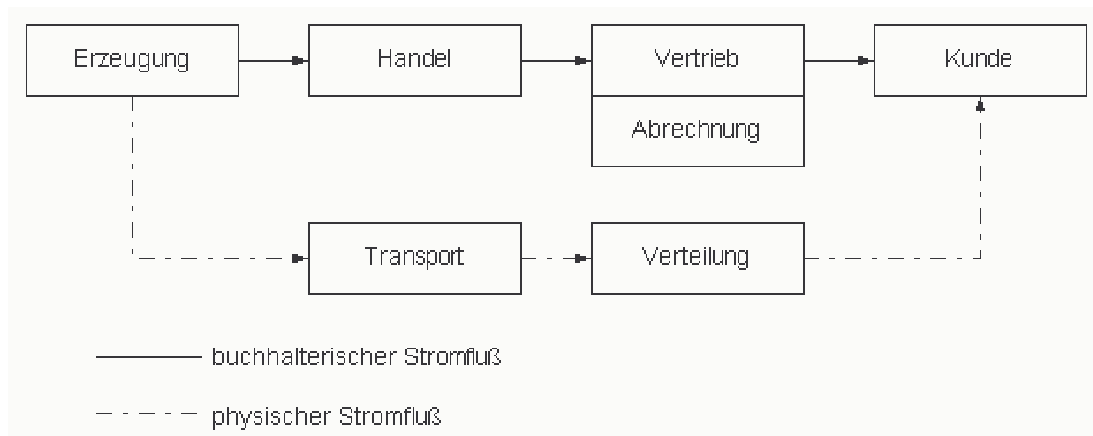
- Stromerzeugung (EnBW Kraftwerke AG)
- Stromhandel (EnBW Ges. für Stromhandel mbH)
- Stromvertrieb (EnBW Vertriebsgesellschaft AG)

---

<sup>28</sup> Die Energie Baden-Württemberg AG mit Sitz in Karlsruhe ist das drittgrößte deutsche Energie-Versorgungsunternehmen mit einem Gesamtumsatz von ca. 7,86 Mrd. Euro und 44.500 Mitarbeiter im Jahr 2001

- Stromtransport (EnBW Transportnetze AG)
- Stromverteilung (EnBW Regional AG)
- Stromabrechnung (EnBW Kundenservice GmbH)

Abb. 5: Prozesskette eines Energie-Versorgungsunternehmens



Der Strom wird **erzeugt** und muss nun dem liberalisierten Stromgroßmarkt zur Verfügung gestellt werden.<sup>29</sup> Dort wird der Strom an diversen europäischen Strombörsen<sup>30</sup> **gehandelt**, d.h. entweder verkauft oder zusätzlich Strom hinzugekauft. Der **Vertrieb** übernimmt diesen Strom, der dann per Überlandleitung an seinen Bestimmungsort **transportiert**<sup>31</sup>, und schließlich an den Endkunden bzw. die Stadtwerke **verteilt** wird. In einem letzten Schritt wird der gelieferte Strom **abgerechnet**.

Durch diese Prozesskette, bei der mehrere unabhängige Gesellschaften involviert sind, ergibt sich eine Fülle von neuen Anforderungen an die Unternehmensstrategien, v.a. im Bereich der IT-Landschaft. So sind die vielfältigen Prozesse zwischen den Bausteinen der Wertschöpfungskette abzubilden und der Austausch von relevanten Daten sicherzustellen. Dazu gehören sowohl Daten zu technisch-operativen Prozessen wie bereitzustellende Energiemengen und Strom-Lastverläufe<sup>32</sup>, oder Daten im Management- und Controlling-Bereich. Diese Daten dienen dabei nicht nur zur effizienten Abwicklung des Energietransfers, sondern diese Datenströme bilden auch die Grundlage für strategische Entscheidungen.

Aus diesem Grunde bietet sich für ein Energie-Versorgungsunternehmen eine Data-Warehouse-Lösung an, in welches aufbereitete Daten aus dem operativen Betrieb fließen,

<sup>29</sup> Dies geschieht vor dem kartellrechtlichen Hintergrund, dass sich anderenfalls eingessene EVU's durch „interne Subvention“ eine marktherrschende Stellung schaffen könnten, indem sie zu Preisen anbieten, zu denen neue Wettbewerber nicht konkurrieren können.

<sup>30</sup> z.B. die Deutsche Strombörse in Leipzig oder die Strombörse Austria-Adria in Graz.

<sup>31</sup> Unter „Transport“ ist aber eher ein „den Strom im Stromnetz verfügbar machen“ zu verstehen.

<sup>32</sup> Unter einem Lastverlauf versteht man die verbrauchte Menge an elektrischer Energie gemessen in kWh zu jedem Zeitpunkt

um für Analysen zur Verfügung zu stehen. Ein Data-Warehouse-System für Energie-Versorgungsunternehmen wird sich in seiner Grundarchitektur nicht wesentlich von einem „klassischen“ Data-Warehouse unterscheiden. Jedoch werden die „äußeren“ Schichten<sup>33</sup> des Data-Warehouses extrem vielschichtig sein. So werden die Daten aus den unterschiedlichsten Quellsystemen stammen, die in einen einheitlichen Kontext gebracht werden müssen. Auch werden die Analysebedürfnisse der jeweiligen Adressaten stark differieren. Dies lässt schon eine erste Aussage zu, dass ein Data-Warehouse-System eines Energie-Versorgungsunternehmens idealerweise ein System von verschiedenen Data-Marts sein wird.

Die Trennung der Unternehmensbereiche im Zuge des Unbundlings bedeutet auch im Daten-Kontext eine starke Einschränkung. So darf bspw. der Netzbetreiber nur Daten an die Vertriebsgesellschaft weitergeben, die er auch allen anderen Marktteilnehmern zur Verfügung stellt. Da dies natürlich nur bedingt erwünscht ist, wird der Datenaustausch, der über das reine Energiemanagement hinausgeht, dadurch ziemlich erschwert. Dies bedeutet, dass ein Data-Warehouse diesem Umstand Rechnung leisten muss.

Daten, welche in einem Data-Warehouse eines Energie-Versorgungsunternehmens zu berücksichtigen sind, sind vielseitiger Natur, im Rahmen dieser Arbeit sollen endkundenrelevante Daten betrachtet werden.

### **3.2 Geomarketing als Entscheidungsgrundlage in Energie-Versorgungsunternehmen**

Die klassischen Einsatzfelder von Geo-Informationssystemen in einem Energie-Versorgungsunternehmen waren vor der Liberalisierung auf den technischen Bereich konzentriert. So wurde und wird GIS v.a. für die Netz- und Leitungsdokumentation eingesetzt. Hier wurden sehr anspruchsvolle und komplexe GIS-Lösungen geschaffen, die ein effizientes Netzmanagement ermöglichen. Dazu gehören beispielsweise Systeme zur:

- Netzbetriebsführung
- Netzplanung
- Anlagebewertung
- Störfallmanagement
- Durchleitungsmanagement

---

<sup>33</sup> Hiermit sind die Quellsysteme und die Präsentations-Schicht gemeint.

Abb. 6: Beispiel einer Trassendokumentation mit SICAD



Dieses Know-How im GIS-Bereich wird zunehmend auch für marketingrelevante Fragestellungen genutzt. So stehen beim Wettbewerb um den einzelnen Kunden die Energie-Versorgungsunternehmen nun vor völlig neuen Herausforderungen, v.a. was den Einsatz der Marketing-Instrumente angeht. So ist das Umwerben der Stromkunden mit klassischen Massenmedien, sei es TV-Werbung, Plakatwerbung oder Sport-Sponsoring, eine Aufgabe in die bisher viel Geld geflossen ist, aber bisher kaum nennenswerten Erfolg gebracht hat<sup>34</sup>. So werden im Energiesektor Geomarketing-Themen zu einer effizienteren Marktbearbeitung herangezogen. Unter **Geomarketing** versteht man Techniken, die Marketing- und Vertriebsaktivitäten auslösen, steuern, begleiten und kontrollieren, und dabei räumliche Phänomene berücksichtigen<sup>35</sup>. Dies insbesondere in den Bereichen Distributionspolitik, Kommunikationspolitik und Marketingforschung. Räumliche Fragestellungen ergeben sich bei einem Energie-Versorgungsunternehmen schon alleine dadurch, dass das Gut von einem Ort (dem Einspeisepunkt) zu einem anderen Ort (dem Kunden) gebracht werden muss. Welchen „Weg“ der Strom dabei geht, kann sich täglich ändern und hängt von vielen Faktoren ab<sup>36</sup>, ist also nicht vorneherein festgelegt. Dies hat einen direkten Einfluss auf die tägliche Vertriebs- und Marketing-Praxis und wirft z.B. folgende Fragestellungen auf:<sup>37</sup>

- Identifizierung der Gebiete mit hoher Kaufkraft
- Ermittlung des Kundenpotentials
- Identifizierung der Wirtschaftsstruktur der Gebiete (Haushalte vs. Industrie)
- Identifizierung der Nähe zu den Einspeisepunkten
- Vertriebs- und Kundendienststeuerung
- Lokalisierung von Versorgungslücken
- Optimierung von Werbemaßnahmen

<sup>34</sup> So haben schätzungsweise seit der Liberalisierung weniger als drei Prozent der deutschen Haushalte den Energielieferanten gewechselt, vgl. Allgayer (2002)

<sup>35</sup> Vgl. Czeranka (2000 I), S. 1, sowie Keller-Griesbach (2000), S. 95

<sup>36</sup> Beispielsweise Kraftwerkverfügbarkeit, Lastgang des Verbrauchers, Durchleitungsentgelte u.ä.

<sup>37</sup> Vgl. Keller-Griesbach (2000), S. 95

Bei Energie-Versorgungsunternehmen werden sich klassische Geomarketing-Fragestellungen sehr stark mit Fragestellungen des **Mikro-Marketings** überlappen. Unter Mikro-Marketing versteht man Marketing-Maßnahmen, welche sich individuell auf Kunden- bzw. Kundengruppen beziehen dabei und stark raum- und adreßfokussiert sind.<sup>38</sup> Will man die Interaktivität der Maßnahmen und die individuelle Ansprache herausstellen, spricht man auch vom **Direkt-Marketing**. Eine Raum- oder Adressfokussierung von Marketing-Maßnahmen hängt sehr stark von einer gut ausgebauten Datenbasis ab. Das Speichern und Aufbereiten von dieser Kunden- und Adressdaten in einer Datenbank zu Marketing-Zwecken wird auch mit dem Begriff **Database-Marketing** belegt.

### 3.3 Geodaten in Energie-Versorgungsunternehmen

Aus diesen Tätigkeitsbildern folgert sich ein großer Bedarf an Daten, um diese Aufgaben effizient zu bewältigen. Im Geomarketing-Zusammenhang kommt dabei besonders den **Geodaten** eine wichtige Bedeutung zu. Als Geodaten werden dabei i.A. Daten bezeichnet, die räumliche Objekte und deren Beziehungen untereinander beschreiben.

Die DIN-Norm ENV<sup>39</sup> 12009 definiert Geodaten als eine datenverarbeitungstechnische Form von „...Informationen zu Phänomenen, die direkt oder indirekt in Bezug zu einem Ort auf der Erde stehen“<sup>40</sup>. Zweck von Geodaten ist also, diesen Objekte eine Position im Raum zuzuordnen.<sup>41</sup>

Geodaten besitzen zwei Komponenten: Zum einen die graphische Ausprägung, bspw. als Punkt auf einer Landkarte; zum anderen die Attribute, die diesem Punkt in alphanumerischer Form zugeordnet sind. Während man ersteres als die **Geometrie** (Geobasisdaten) eines Geodatensatzes bezeichnet, stellen zweite die **Attribute** (Geosachdaten) dar.

- **Geobasisdaten** (Geodaten „im engeren Sinne“):

Hierunter versteht man die reine **Geometrie**, also die Beschreibung der Lage und Form des Objektes im Raum. Diese Daten kann man in einem GIS durch graphische Elemente abbilden. Hierbei kann man unterscheiden zwischen Vektor- und Rasterdaten. Ersteres verwendet man i.A. zur Darstellung linearer Elemente wie Punkte, Linien und Polygone, während man unter zweitem hauptsächlich georeferenzierte Bilddateien versteht, wie topographische Karten oder Orthophotos. Diese Daten können also **direkt** einer Position im Raum zugeordnet werden.

---

<sup>38</sup> Vgl.: Nitsche (1998), S. 14

<sup>39</sup> Europäische Vornorm

<sup>40</sup> zitiert in: Wegner (2000 I), S. 18

<sup>41</sup> Vgl. Bill / Zehner (2001), S. 106

- **Geosachdaten** (Geodaten „im weiteren Sinne“):

Hierunter versteht man die **Sach- oder Attributdaten**, die einem geometrischen Objekt zugeordnet werden können und dessen Eigenschaften beschreiben. Diese Attribute beschreiben dabei den räumlichen Bezug des Objektes durch eine entsprechende Vercodung, die ebenfalls in den Attributen hinterlegt ist, z.B. als Adresse, Koordinaten, ZIP-Codes u.ä. Diesen Attributen kann somit **indirekt** über Geocodierung oder der Verknüpfung zu Geobasisdaten eine Position im Raum zugeordnet werden.

Am Beispiel einer Gemeindedatenbank in einem GIS, würden die graphischen Gemeindegrenzen die *Geometrie* darstellen, während die zugeordneten Daten wie bspw. die Gemeindekennziffer die *Attribute* repräsentieren.

Geodaten können sowohl käuflich erworben werden, als auch durch Evaluierung und Anreicherung interner Daten gewonnen werden. So schätzt man, dass mehr als 50 % aller Unternehmensdaten einen räumlichen Bezug haben, und somit potentiell für Geomarketing-Analysen zur Verfügung stehen.<sup>42</sup> Im Gegensatz zu anderen Branchen die Geodaten einsetzen, fallen bei einem Energie-Versorgungsunternehmen Geodaten durch den Betriebsprozess an, so z.B. durch die Leitungsdokumentation. Diese Daten werden i.a. durch Digitalisierung oder GPS-Messungen gewonnen. Die Nutzung dieser Daten für Marketing- oder vertriebsrelevante Analysen wird aber durch das „Unbundling“ sehr erschwert, da die Datenweitergabe netztechnischer Daten an die Vertriebsseinheit praktisch nicht möglich ist.<sup>43</sup> Im Vertriebs- und Marketingbereich gilt ähnliches wie bei anderen Unternehmen, die sich mit ihren Produkten direkt an den Endverbraucher wenden. Hier lassen sich in erster Linie die Kundendaten nennen, die entsprechend attributisiert sind, oder auch räumlich gegliederte Absatz- und Umsatzzahlen.

Bei den extern zugekauften Geodaten sind hauptsächlich zu nennen:<sup>44</sup>

- Geometriedaten für administrative Gebietseinheiten
- Kaufkraft- oder Konsumdaten
- Daten über soziographische und sozioökonomische Merkmale<sup>45</sup>
- (alphanumerische) Adressdaten
- Topographische Karten im Rasterformat für Kartographiezwecke

Der Geodaten-Bedarf eines Energie-Versorgungsunternehmen hängt sehr stark von der Aufgabenstellung und der Analysebedürfnisse ab. Dies gilt besonders für externe Geodaten,

---

<sup>42</sup> Vgl. Czeranka (2000 I), S. 4

<sup>43</sup> siehe Kapitel 3.1

<sup>44</sup> Vgl.: Nitsche (1998), S. 47

<sup>45</sup> Bspw. Lifestyle-Daten, über Lebensstile von Konsumenten. So kann man davon ausgehen, dass die Entscheidung den Stromversorger zu wechseln, nicht nur eine Entscheidung des Preises ist, sondern eine Frage der persönlichen Einstellung und der eigenen sozialen Rolle.

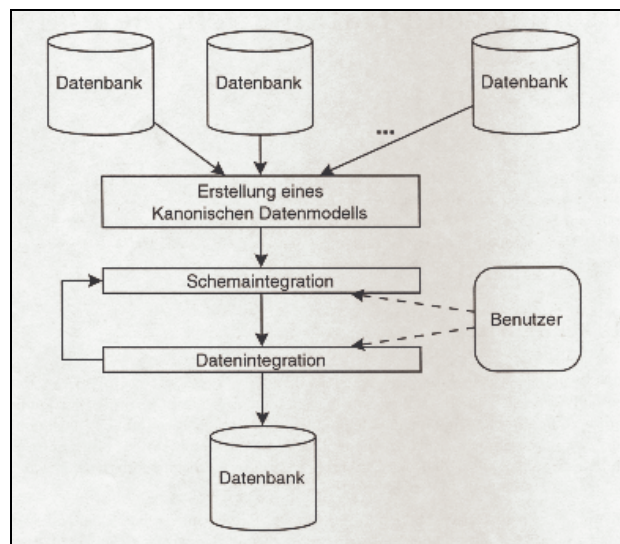
deren Beschaffung zumeist aufgrund der hohen Kosten eine hohe Investition darstellt. Die effiziente Nutzung von Geodaten ist aufgrund der fachlichen Anforderungen an den Nutzer nicht einfach. Daher ist es umso wichtiger, diese in einem Data-Warehouse in leicht zu handhabender, aber aussagekräftiger Form zur Verfügung zu stellen.

## 4 Geodaten in Data-Warehouse-Systemen

### 4.1 Integrationskonzepte

„Integration“ bedeutet „...das Kombinieren von Daten ... unterschiedlicher Typen und aus ... unterschiedlichen Quellen und Systemen...“<sup>46</sup>. Diese Definition zeigt, dass eine Datenintegration von existierenden operativen Daten in ein Data-Warehouse, auch immer eine Integration von Daten aus verteilten und unabhängigen Datenbanken bedeutet. Diese werden in den seltensten Fällen in der gleichen Struktur vorliegen. Um diese Datenbanken in eine Datenbank zu integrieren, geht man von einem dreistufigen Integrationsprozess aus:

Abb. 7: Ablauf der Datenbankintegration<sup>47</sup>



Im ersten Schritt werden die Daten zur Vereinheitlichung in ein grobes Datenmodell überführt, welches alle Attribute aufnehmen kann, und als Grundlage zur weiteren Integration dient. Dieses Datenmodell wird das *kanonische Datenmodell* genannt. In einem zweiten Schritt werden die einzelnen Datenbank-Schematas in ein neues globales Schema überführt. Hierbei sind insbesondere die ursprünglichen Relationen der einzelnen Entitäten zu berücksichtigen. In einem dritten Schritt werden dann die Datensätze aus den unterschiedlichen Quellen in der neuen Datenbank zusammengeführt. Hierbei ist auf Datenverfälschung durch Verschmelzung zweier verschiedener Datensätze, bzw. der Nicht-Zusammenführung zwei gleicher Datensätze, zu achten.

Dies legt nahe, dass eine Integration von Daten in ein bestehendes Data-Warehouse sich nicht in einem Gedanken- und Arbeitsschritt bewerkstelligen lässt, sondern ein geplantes und

<sup>46</sup> Bill/Zehner (2001), S. 139

<sup>47</sup> Breitner u.a. (1996), S. 46

zielorientiertes Vorgehen erfordert. Idealerweise sollte dieses Vorhaben als offizielles Projekt<sup>48</sup> etabliert werden, für welches Ressourcen abgestellt werden sollten. Als Arbeitspakete eines entsprechenden Projektes mit geographischem Hintergrund kann man u.a. folgende Aufgaben definieren:<sup>49</sup>

- Evaluierung der bestehenden Datenbasis
- Evaluierung der geographischen Fragestellungen und des Analysebedarfs
- Daraus abgeleitet den Bedarf an Geodaten
- Erhebung der internen und externen Datenquellen
- Sichtung des bestehenden Datenmodells
- Evaluierung der Entitäten, für welche eine geographische Referenzierung sinnvoll ist
- Einpassung der neuen geographischen Entitäten
- Verknüpfung mit den bestehenden Entitäten
- Aktualisierung der Meta-Information
- Aufsetzen von Prozessen für die Aktualisierung der dynamischen Daten

Ein üblicher Ansatz zur Integration von Geodaten in unternehmerische Entscheidungsprozesse wird häufig in der Visualisierung gesehen<sup>50</sup>. Auch wenn dies als ein Aufgabenschwerpunkt jedes geographischen Datenbestandes gesehen werden kann, geht eine Integration in ein Data-Warehouse aber wesentliche Schritte weiter. Die Unternehmensdaten sollen nicht nur „geographisch bearbeitet“, sondern erweitert und angereichert werden. Darüber hinaus soll das Geodaten-System kein proprietäres System bilden, welches nur für wenige Anwendungen genutzt werden kann und in der IT-Landschaft isoliert betrieben wird. Geodaten sollten zu Business Mapping-Zwecken in den Unternehmensdatenbestand integriert werden, im operativen Betrieb, ganz besonders aber im strategischen Bereich. Daher bietet sich eine Integration in ein unternehmensweites Data-Warehouse-System an. Dies ist eine Grundvoraussetzung dafür, dass effiziente Geomarketing-Prozesse etabliert werden können, die durch Ihre Planungs-, Analyse und Reporting-Funktionen die strategischen Entscheidungen unterstützen.

Dies hat zur Folge, dass Geodaten mit den „Nicht-Geo“-Daten eines Unternehmens effizient vernetzt werden müssen. Eine große Herausforderung besteht dabei darin, die Geodaten mit ihrer graphischen Komponente, also einer Objektorientierung, mit den übrigen, eher attributorientierten Unternehmensdaten in Einklang zu bringen. Dazu müssen Methoden und

---

<sup>48</sup> Nach DIN-Norm 69901 ist ein Projekt ein Vorhaben „...das im Wesentlichen durch die Einmaligkeit der Bedingungen in Ihrer Gesamtheit gekennzeichnet ist, wie z.B. Zielvorgabe, zeitliche, finanzielle, personelle und andere Begrenzungen, Abgrenzung gegenüber anderen Vorhaben, projektspezifische Organisation.“

<sup>49</sup> Vgl.: Zincke (2000), S. 61

<sup>50</sup> Vgl.: Zincke (2000), S. 59

Schnittstellen definiert werden, die der Relevanz der Geodaten für die betrieblichen Zwecke gerecht werden.

Prinzipiell lassen sich –grob formuliert- vier Grundkonzepte zur Verknüpfung von Geodaten mit den attributisierten Daten skizzieren.

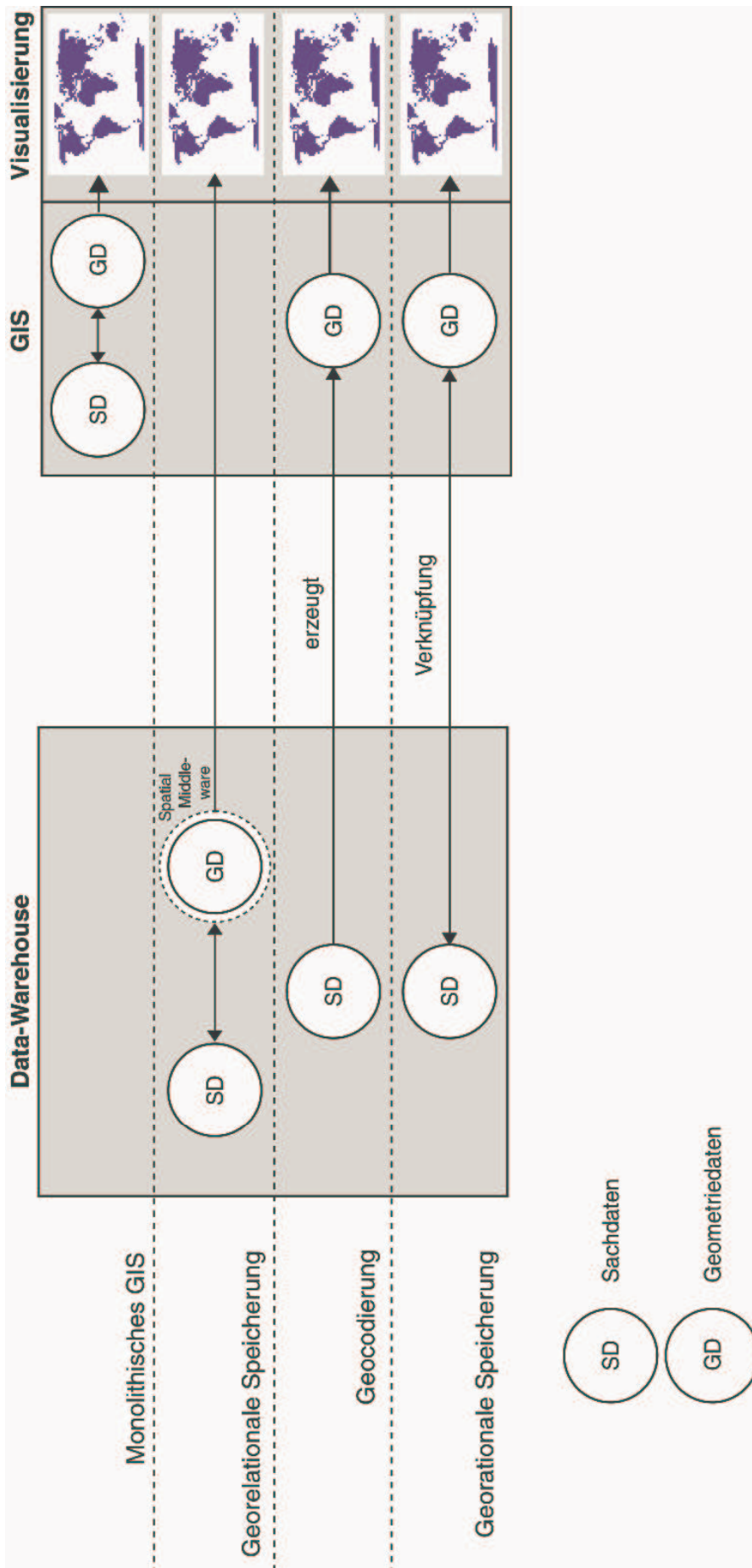
- **Monolithisches GIS**<sup>51</sup>:  
Speicherung der Geokomponente (Geometrien, Topographie) sowie der zugehörigen Sachdaten in einem GIS.
- **Georelationale Speicherung**:  
Speicherung der Attribute und der Geometrie in einem Data-Warehouse, welches mit einer Technologie ausgestattet ist, die dessen nativen Datentypen erweitert, um Geobjekte abzulegen.
- **Geocodierung**:  
Speicherung der Sachdaten in einem Data-Warehouse, bei Bedarf „On-the-Fly“-Erzeugung der Geokomponente aus diesen Sachdaten, bspw. durch gespeicherte X,Y-Koordinaten oder Adressdaten
- **Georationale Speicherung**:  
Speicherung der Geokomponente in einem GIS; Speicherung der zugehörigen Sachdaten in einem Data-Warehouse; Verknüpfung über entsprechende Software-Schnittstellen

Abb. 8 stellt die vier beschriebenen Varianten dar, auf die in den nachfolgenden Kapiteln noch näher eingegangen wird.

---

<sup>51</sup> Monolithisch = aus einer Einheit bestehend

Abb. 8: Integrationsvarianten für Geodaten



## 4.2 Monolithische GIS-Arbeitsplätze

Auch in Zeiten von Data-Warehouse-Systemen und unternehmensweiten Rechnernetzen sind GIS-Arbeitsplätze oftmals als „Stand-alone“-Lösung anzutreffen<sup>52</sup>. Dies bedeutet, dass ein einzelner Arbeitsplatz mit GIS-Software und Geodaten ausgerüstet ist, welcher von einer geschulten Fachkraft bedient wird. Dieser PC ist zwar an das Unternehmensnetzwerk angebunden, jedoch bezieht er keine oder nicht regelmäßig Daten aus diesem. Die Sachdaten werden zusammen mit den Geodaten in einem System gehalten, bspw. in ArcView-Shapefiles oder ArcInfo-Coverages. Vorteil dieser Lösung wäre eine relativ einfache Handhabung der Geodaten, klarer Nachteil allerdings die Abschottung dieser Daten von Nicht-GIS-Anwendungen.

Diese Variante wird häufig bei Unternehmen angetroffen, in denen das Geo-Marketing nicht konsequent in organisatorische Abläufe eingebunden ist. Da man hierbei nicht mal ansatzweise von einer Data-Warehouse-Lösung sprechen kann, soll diese Variante keine weitere Beachtung finden, obwohl diese noch sehr oft in der Praxis anzutreffen ist.

## 4.3 Georelationale Speicherung

Während traditionelle Geometrien in einem GIS gespeichert werden, sind seit einiger Zeit Technologien in den Vordergrund gerückt, die es ermöglichen nicht nur die Attributdaten in einem industriellen Standard-DBMS abzulegen, sondern auch die Geometrie. Diese **DBMS-basierte** Vorgehensweise hat im Gegensatz zu o.g. file-basierten Lösungen den Vorteil, dass die Vorteile eines DBMS gewährleistet bleiben:

- verteilte Nutzung
- Multi-User-Betrieb
- hohes Datenaufkommen
- hohe Performance
- sichere Datenverwaltung
- zentraler Zugang

Dazu ist es nötig die nativen Datentypen des DBMS um Datentypen zu erweitern, die es ermöglichen, graphische Geobjekte abzulegen. Bei den Produkten die diese Erweiterung bewerkstelligen, gibt es sowohl Datenbank-integrierte Produkte wie Oracle SDO, als auch Middleware-Komponenten wie ESRI ArcSDE. Ziel dieser Produkte ist es, Geometrie-Daten in der selben Weise wie alpha-numerische Daten in einer DBMS-basierten Datenbank

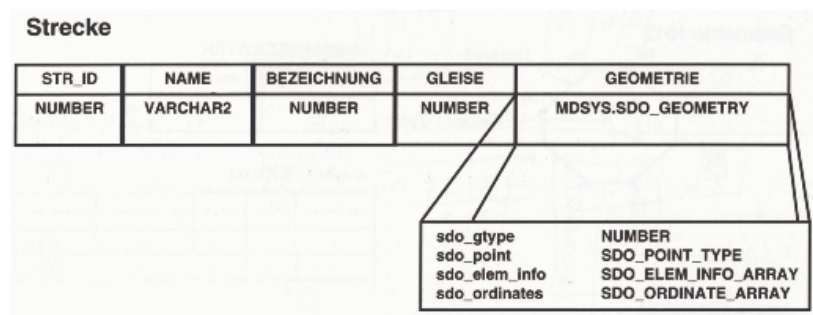
---

<sup>52</sup> Vgl.: Tappert (1997), S. 155

abzulegen und zu verwalten. Dazu ist es nötig, dass diese Produkte Datenmodelle für räumliche Daten bereitstellen.

**Oracle SDO**<sup>53</sup> ist ein Aufsatz, welcher eine Oracle-DBMS um räumliche Datenmodelle, Administrationswerkzeuge, räumliche Induzierungsmechanismen und relevante Abfragemethoden erweitert. Diese wird somit direkt in ein bestehendes Oracle-DBMS integriert. Bei den Datenmodellen unterstützt Oracle SDO sowohl das relationale als auch das objektrelationale Datenmodell. Diese unterscheiden sich dadurch, wie die Geometrie mit den Attributdaten verbunden ist. Im **relationalen Modell** sind die Geometrien in sogenannten Layern organisiert. Für ein Layer werden in diesem Datenmodell drei Tabellen angelegt<sup>54</sup>, welche Metainformationen und Lagepunkte der Geometrien enthalten. Im Gegensatz dazu sind im **objektrelationalen Modell** die Geometriewerte jedem einzelnen Datensatz der Attributtabelle zugeordnet. Jede Geometrie bildet also ein Objekt des zugehörigen Datensatzes in dieser Attributtabelle, und ist auch in dieser gespeichert. So ist in der Attributtabelle ein Feld enthalten, welches einen speziellen SDO-spezifischen Datentyp enthält.<sup>55</sup>

Abb. 9: Objektrelationales Modell (Beispiel)<sup>56</sup>



**ArcSDE**<sup>57</sup> der Firma ESRI verfolgt den selben Zweck wie SDO, nämlich die Einbindung von Geometrien in ein DBMS. ArcSDE wird dagegen nicht in ein DBMS integriert sondern fungiert als Middleware-Lösung, die als GIS-Gateway ein frei wählbares<sup>58</sup> DBMS mit GIS-Software der „ESRI-Welt“ verbindet. ArcSDE benutzt die nativen Datentypen und Operatoren des DBMS um die entsprechenden Geometrien darin ablegen zu können. Man kann dann von einer **Geo-Datenbank** sprechen.<sup>59</sup>

Für die Anbindung der Daten an GIS-Applikationen mittels ArcSDE, bietet dieses die Möglichkeit entweder eine **2-tier** oder eine **3-tier** Architektur einzusetzen.

<sup>53</sup> *Spatial Data Option*

<sup>54</sup> *<layer>\_SDOLAYER* und *<layer>\_SDODIM* enthalten Metainformationen zum Layer, während *<layer>\_SDOGEOM* die Werte der Geometrien abgelegt sind, vgl.: Fleischmann, Kaiser (2000), S. 162

<sup>55</sup> *SDO\_GEOMETRY*, vgl.: Fleischmann, Kaiser (2000), S. 164

<sup>56</sup> Fleischmann, Kaiser (2000), S. 164

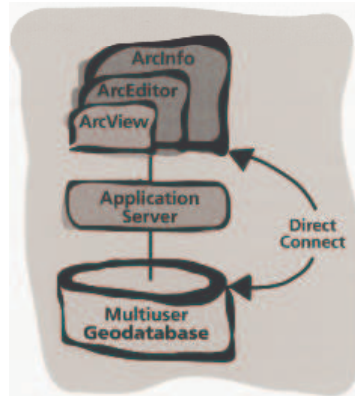
<sup>57</sup> *Spatial Data Engine*

<sup>58</sup> ArcSDE kann mit Oracle, IBM DB2, MS-SQL-Server und Informix betrieben werden

<sup>59</sup> Die Firma ESRI spricht hierbei von einer **Enterprise Geo-Database**

Bei der zweiten Lösung wird ein **Application Server** auf der Hardware aufgesetzt auf der sich die Geo-Datenbank befindet. Dieser bündelt die Anfragen der GIS-Applikationen und leitet diese an die Geo-Datenbank weiter. Der Application Server berechnet anschließend das Ergebnis und leitet dieses an die GIS-Applikationen zurück.

Abb. 10: GIS-DBMS-Architektur mit ArcSDE<sup>60</sup>



Diese hier gezeigte Variante stellt sicherlich die professionellste und eleganteste Möglichkeit dar, geographische Informationen zu verwalten. Jedoch ist diese Methode nur effizient, wenn es sich einerseits um sehr große (geographische) Datenmengen, als auch um „reine“ Geo-Datenbanken handelt.

Diese Technologien verfolgen ein ganzheitliches Konzept im Umgang mit geographischer Information, d.h. eine Datenbank in einem DBMS wird durch die Erweiterungen um diese Technologien gänzlich neu aufgestellt werden. So sind diese nicht dazu konzeptioniert, einfach eine Kundendatenbank mit „geographischen Datenbank-Features“ aufzuwerten, sondern sie wandeln eine Datenbank in eine Datenbank um, welche nun den Hauptzweck hat, reine geographische Daten, wie Geometrien oder Topologien, zu verwalten. Eine Geo-Datenbank mit georelationaler Speicherung stellt daher immer eine „Insellösung“ innerhalb des Gesamt-Datenbestands des Unternehmens dar.

Idealerweise kommt diese Technologie also dort zur Anwendung, wo sehr große geographische Datenmengen anfallen, die hauptsächlich operativ eingesetzt werden, so z.B. bei der Fernerkundung oder dem Vermessungswesen. Im Falle endkundenrelevanter Datenbestände für Geomarketing-Anwendungen bedarf es daher einer Analyse der Datenmengen, -typen und -ströme innerhalb des Unternehmens, um die Praktikabilität dieser Technologie zu erfassen. In einer Umgebung in der Kundendaten in erster Linie als alphanumerische Daten aufgefasst werden, und erst in zweiter Linie als Träger von Geo-Information, lässt sich der in Kap. 4.5 beschriebene file-basierte Ansatz besser in die Data-Warehouse-Struktur des Unternehmens einpassen.

<sup>60</sup> ESRI-Broschüre „ArcSDE“

#### 4.4 Geocodierung

Bei einfachen Geometrien kann es auch sinnvoll sein, diese nicht explizit in einem GIS zu halten, sondern diese bei Bedarf zu erzeugen. Handelt es sich beispielsweise um Daten, welche sich geographisch durch einen Punkt repräsentieren lassen, wäre es u.U. zweckmäßig, diese Informationen als X,Y-Koordinate in einem DBMS zu hinterlegen, und die Geometrie bei Anforderung in einem GIS zu erzeugen. Diese Geodaten wären somit dadurch gekennzeichnet, dass Ihnen keine Geometrie zugeordnet ist, sie aber implizit Geoinformation enthalten. Ähnlich verhält es sich auch bei Attributen wie Adressinformationen, die einen räumlichen Bezug aufweisen und mittels **Geocodierung** zu geographischen Objekten zugeordnet werden. Die Geo-Komponente der Geodaten wäre somit in den Sachdaten integriert und kann aus diesen generiert werden. Diese Daten lassen sich gänzlich in einem DBMS halten, und stehen somit auch allen „nicht-geographischen“ Datenbank-Methoden zur Verfügung. Sind diese Daten dann mit einem GIS verbunden, können diese im operativen Betrieb dann durch GIS-Funktionen in Geometrie-Informationen umgewandelt werden.

Zweckmäßig ist diese Vorgehensweise dann, wenn die erzeugten Geometrien nicht allzu komplex sind. So sind Punkte<sup>61</sup> oder einfache Linien ohne getrennte Abschnitte<sup>62</sup> durch Punkt-Koordinaten repräsentierbar, welche als numerische Werte gespeichert werden können. Die Generierung einer Geometrie aus diesen Werten durch die GIS-Komponente ist dann eine relativ einfache Funktion.

Komplexer wird diese Vorgehensweise, wenn die Zuordnung der Geometrie nicht direkt über Raumkoordinaten, sondern **indirekt** über räumliche Informationen wie Straßennamen oder Gebietseinheiten erfolgen soll.

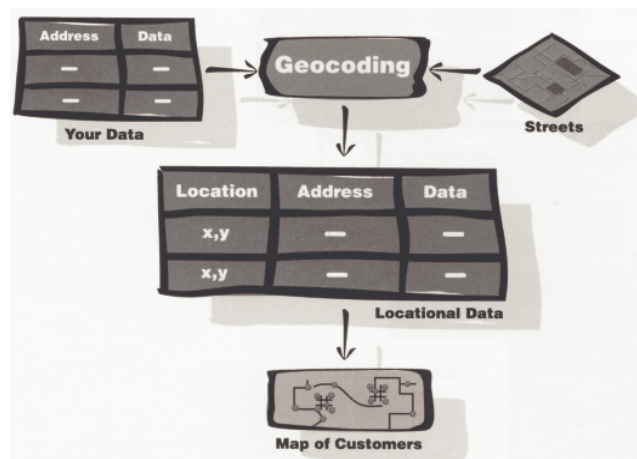
Bei dieser Methode, der **Geocodierung**, wird ein Datensatz über ein systematisches Vorgehen geographischen Objekten zugeordnet, und somit seine Lage im Raum bestimmt.

---

<sup>61</sup> repräsentiert durch ein Koordinatenpaar

<sup>62</sup> repräsentiert durch zwei Koordinatenpaare

Abb. 11: Geocoding<sup>63</sup>



Dieses Vorgehen kann man als Datenbankabgleich auffassen, bei dem über einen Schlüssel (PLZ oder Adresse) zwei Datenbanken miteinander in Beziehung gesetzt werden. Um dies zu realisieren, benötigt man eine spezielle Software, den sogenannten **Geocoder**. Dieser hat die Anforderungen die ihm „...zugeführten Adressen so richtig, schnell und vollständig wie möglich zu geocodieren...“<sup>64</sup>. Um dies zu gewährleisten muss ein Geocoder mit der Komplexität von Adress-Strings umgehen können. So wird sich die Qualität von Geocoder-Software v.a. daran messen lassen wie sie mit Zuordnungsproblemen wie Schreibfehlern, Abkürzungen, Sonderzeichen oder Unvollständigkeiten umgeht. So wird sich eine genaue 100%-ige Geocodierung bei einem großen, komplexen Adressbestand kaum realisieren lassen. So sollten nicht nur ganze Strings verglichen werden können, sondern auch Wortteile, Abkürzungen und Zeichen sollten interpretiert werden können. Dies erfordert intelligente Algorithmen, wie ein phonetischer Match, der gleichklingende Worte erkennt<sup>65</sup> oder ein „Fuzzy-Match“, der auch klassische Tippfehler wie Buchstabendreher oder -auslasser erkennt. Diese näherungsweisen Treffer sollten nach der Genauigkeit skaliert werden können<sup>66</sup>, was die Bewertung und die Nachbearbeitung erleichtert.

Eine Geocodierung ist idealerweise eine einmalige Tätigkeit, in der ein gegebener Adressbestand einmal bearbeitet wird. Die Integration von Geocodierungs-Algorithmen, die ein Adressbestand bei jedem Bedarf neu geocodieren, um in geographisch zu bearbeiten, macht wenig Sinn.

Die Frage nach der Aufsetzung eines Geocodierungs-Prozesses in einem Data-Warehouse stellt sich besonders dort, wo besonders viel neue Adressen regelmäßig hinzugefügt oder geändert werden müssen. Energie-Versorgungsunternehmen, die mit Massendaten arbeiten,

<sup>63</sup> ESRI-Broschüre „ArcSDE“

<sup>64</sup> Nitsche (1998), S. 78

<sup>65</sup> Auch „Franz-Meier-Weg“ bringt eine Übereinstimmung mit „Franz-Meyer-Weg“

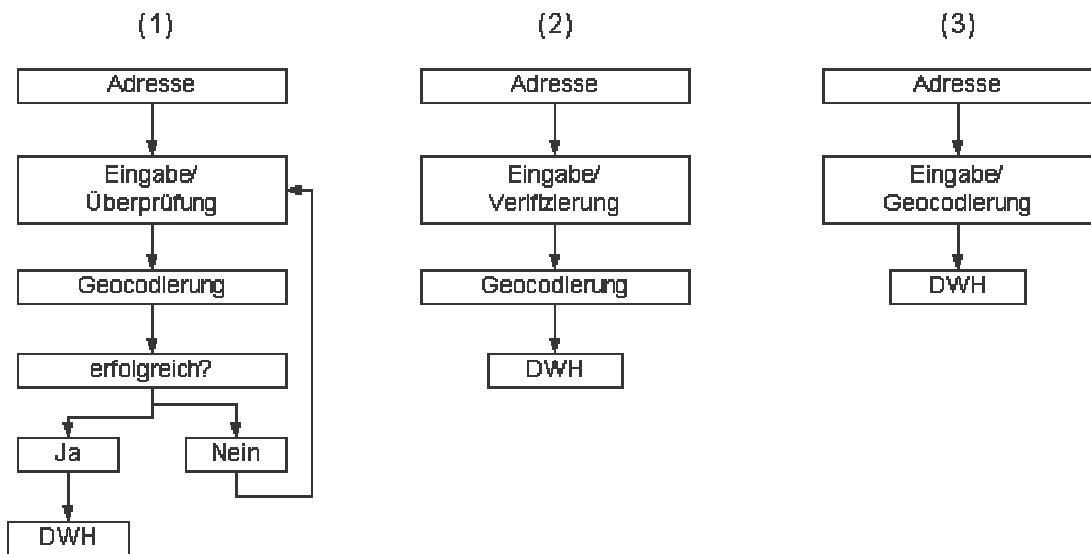
<sup>66</sup> 100% → genauer Treffer; 0% → keine Übereinstimmung

stehen oft vor dem Problem, dass täglich viele hundert Kundenadressen hinzugefügt oder geändert werden müssen.

Sollten diese nur geocodiert im Data-Warehouse zur Verfügung stehen, muss die Geocodierung vor der Aufnahme in die Staging-Area durchgeführt werden. Denkbar wäre ein nächtlicher Batch-Lauf, der die Geocodierung automatisiert durchführt und die Nicht-Treffer kennzeichnet, so dass diese am nächsten Arbeitstag bearbeitet werden können. (Beispiel (1) in Abb. 12)

Eine weitere Variante wäre eine „On-the-fly-Geocodierung“ direkt bei der Eingabe der Adressdaten durch das Call-Center-Personal. Hierbei würde man dem Datenerfasser die Möglichkeit nehmen, die Adresse frei zu vergeben. Er muss diese aus einer Auflistung der möglichen Adressen auswählen, was Fehler wie Buchstabendreher oder falsche Abkürzungen verhindert. Diese Adressen könnten dann entweder im nächtlichen Batch-Lauf (2) mit einer 100%-igen Trefferquote geocodiert werden, oder direkt nach der Eingabe. (3) Diese Variante würde aber die Verfügbarkeit eines Geocodierungsalgorithmus in der Datenerfassungssoftware erfordern.

Abb. 12: Geocodierung im Batch-Betrieb



Prinzipiell sollten Adressdaten bereits geocodiert in einem Data-Warehouse vorliegen, um zeitnah geographische Analysen durchzuführen. Eine Lagerung von ungecodierten Daten im Data-Warehouse, deren Geo-Komponente bei Bedarf generiert wird, lohnt sich nur bei einem überschaubaren und unregelmäßig benötigten Datenbestand, bei einer geringen Komplexität der Geometrie (z.B. Punkte oder Linien mit zwei Knoten), oder bei einer geringen Komplexität des Geoschlüssels (z.B. eindeutige Postleitzahlen oder Gemeindekennziffern).

#### 4.5 Georationale Speicherung

Neben den beiden Extremen (Speicherung der Geo- und Sachdaten in einem GIS / in einem DBMS) könnte sich deren Kompromiss als effiziente Lösung erweisen. In diesem hybriden System wird die Geokomponente in einem GIS gespeichert, und die Sachdaten in einem DBMS, wobei beide Komponenten via **Datenbank-Verknüpfungen** verbunden werden. Auf diese Weise werden die Nachteile der beiden Extrem Lösungen abgeschwächt, ohne dass die Vorteile gänzlich aufgegeben werden müssten. Die Sachdaten ständen im Einklang mit den gesamten Unternehmensdaten, und somit dem ganzen Unternehmen zur Verfügung. Die Geokomponente wird in einem GIS gespeichert und sichert somit den schnellen Zugriff durch ein GIS.

Die Herausforderung bei dieser Variante besteht im Zusammenfügen dieser beiden Komponenten, sowohl in Fragen der Datenmodellierung als auch in Fragen der Software-Technologie.

In Fragen der Datenmodellierung muss man von zwei Tabellen ausgehen, welche miteinander in Beziehung gebracht werden müssen. Eine Tabelle ist in einem GIS gespeichert und stellt die **Geokomponente** dar, da sie als Speicher der Geometrie dient. Die andere Tabelle stammt aus einem Data-Warehouse und stellt die **Sachkomponente** dar, da sie die reinen alpha-numerischen Daten enthält. Die Geokomponente wird physisch durch ein GIS-kompatibles File<sup>67</sup> repräsentiert, so dass man hier auch von einem **File-basierten** Datenmodell sprechen kann.

Um diese zwei Komponenten miteinander zu verbinden, muss in beiden Komponenten ein Verbindungselement in Form eines *Schlüsselfeldes* existieren, welches in einem ersten Schritt erlaubt die jeweiligen Datensätze eindeutig zu identifizieren und in einem zweiten Schritt zueinander in Beziehung zu setzen.

Um dies optimal zu gewährleisten, muss dieses Schlüsselfeld in beiden zu verbindenden Tabellen *eindeutig* sein, d.h. er muss jeden Datensatz einer Tabelle eindeutig identifizieren. Diese Schlüssel müssen nicht zwangsläufig als Primärschlüssel der Tabelle angelegt sein, es erleichtert jedoch die klare Zuordenbarkeit. Dieser Schlüssel kann auch als zusammengesetzter Schlüssel realisiert sein, d.h. er kann sich aus mehreren Attributen bilden.

In der Tabelle der **Geokomponente** sind im effizientesten Fall zwei Attribute enthalten: Ein Objekt-Attribut<sup>68</sup> welches die Geometrie enthält und ein ID-Feld, welches dieser Geometrie einen eindeutigen Identifier zuweist. In dieser Tabelle stellt das Geometrie-Attribut per definitione den Primärschlüssel dar, da jeder Datensatz durch seine graphische Ausprägung

---

<sup>67</sup> Shape-Files, MapInfo-Tabs u.ä.

<sup>68</sup> Bspw. das Feld *Shape* in der Tabellenansicht eines ArcView-Shape-Files

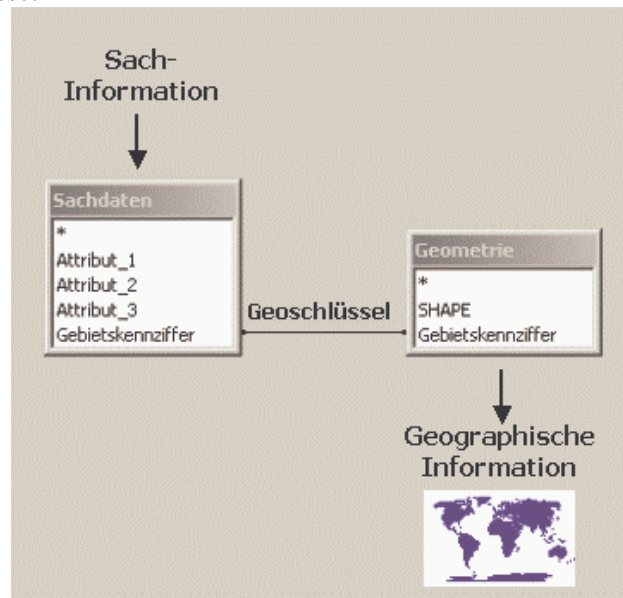
eindeutig identifiziert wird. Im ID-Feld wird ein Wert gespeichert, welcher der Geometrie semantisch zugeordnet ist, objektiv und eindeutig. Würde man z.B. in einer solchen Tabelle alle Gemeindegrenzen der Bundesrepublik Deutschland als Polygone halten, würde sich hierbei als ID die achtstellige Gemeindekennziffer des statistischen Bundesamtes anbieten. Diesen Identifier kann man als „**Geo-Schlüssel**“ bezeichnen, da seine Ausgestaltung im unmittelbaren Zusammenhang mit der Gebietseinheit steht, die er repräsentiert. Mit diesem Geo-Schlüssel wäre es somit möglich, an jeder Stelle des Data-Warehouse raumrelevante Daten mit einer Geometrie zu verknüpfen, sofern an dieser Stelle der Geo-Schlüssel als Attribut abgelegt ist.

Um diese Funktionalität sicherzustellen sind an die Ausgestaltung des Geo-Schlüssels Bedingungen geknüpft. Er sollte:

- eindeutig sein, d.h. eine genau definierte Raumeinheit definieren.
- unternehmensweit gleich definiert sein.
- möglichst vollständig sein, d.h. auch übergeordnete Gebietseinheiten enthalten. Ein Beispiel wäre die deutsche Gemeindekennziffer, in welcher auch das Bundesland und der Regierungsbezirk, in dem die Gemeinde liegt, kodiert ist.
- objektiv sein, d.h. er sollte sich idealerweise an amtlichen Gebietszuordnungen wie Postleitzahlen oder sonstigen Kennziffern orientieren.

Auf Seiten der **Sachkomponente** handelt es sich um reine alpha-numerische Daten, welche aus dem Unternehmens-Datenbestand stammen und dafür aufbereitet wurden, um mit dem Geodaten verbunden zu werden. Hier muss ein Attribut definiert sein, welches mit dem Attribut „ID-Feld“ in der Geokomponente korrespondiert. In diesem Feld wird der Geo-Schlüssel hinterlegt, der in der Tabelle der Geokomponente seine Entsprechung hat. Über dieses Feld werden die beiden Tabellen verbunden, um die Sachdaten der Geometrie zuzuordnen und sie somit als „Geodaten“ nutzbar zu machen. Bei der Verknüpfung gelten daher die Regeln, wie sie im **relationalen Modell** definiert ist. Die Idee hierbei besteht darin, dass einzelne Tabellen („Entitäten“), miteinander in Beziehung gesetzt werden. In unserem Anwendungsfall werden die Entitäten „Geometrie“ und „Sachdaten“ durch den Geoschlüssel miteinander verbunden, um so aus zwei Fragmenten eine Einheit zu bilden.

Abb. 13: Geoschlüssel



Die software-technische Umsetzung dieser Datenbankverknüpfung lässt sich über die Datenbank-Beschreibungssprache SQL (*Structured Query Language*) verwirklichen. Diese standardisierte Sprache beinhaltet das Instrumentarium zur Benutzung von DBMS und ist insbesondere auf das relationale Modell ausgelegt. So lassen sich Abfragen oder Verknüpfungen von Entitäten über die Data Definition Language-Komponente von SQL verwirklichen.

Auf dieser Technologie bauen daher auch Standard-Datenbankschnittstellen wie **ODBC** (Open Database Connectivity) von Microsoft auf. Diese Schnittstelle verbindet eine Datenbank mit einer Anwendung, in unserem Fall das DBMS mit dem GIS.<sup>69</sup> Die ODBC-Technologie hat den Vorteil, dass nicht jede Anwendung auf jedes Datenbank-System angepasst werden muss um Daten auszutauschen. Es genügt, wenn für das Datenbank-System ein ODBC-Treiber installiert ist, und die Anwendung ODBC-fähig ist. Dann stellt die Anwendung über die Funktionsbibliothek des ODBC-Treibers die Verbindung mit der entsprechenden Datenbank her, ohne dass die Anwendung die spezifischen Eigenheiten der Datenquelle kennen muss. ODBC ist nach einem „Baukastenprinzip“ aufgebaut. Die Komponente, die auf der Seite der Anwendung sitzt, nimmt über den ODBC-Standard die Anfragen entgegen, verarbeitet diese und gibt sie auf SQL-Basis an die Datenquelle weiter. Der ODBC-Standard wird mittlerweile durch den **Universal Data Access** von Microsoft abgelöst. Dieser basiert auf der **OLE DB**-Technologie welche gegenüber ODBC einige Vorteile aufweist. So kann hier alles als Datenquelle definiert werden, was Daten enthält; es ist also nicht an ein DBMS gebunden. Außerdem wird die Abhängigkeit von SQL als Abfragesprache aufgehoben. Der OLE DB-Standard unterstützt ADO (Active Data Objects), eine flexible objektorientierte Datenzugriffsmethode für Datenbankentwickler.

<sup>69</sup> Vgl.: Monadjemi (1999), S. 60

Da im Client/Server-Betrieb eine Datenbank meistens auf einem entfernten Rechner läuft („Remote-Datenbank“), und somit ein Zugriff nicht einfach über den Netzwerkpfad erfolgen kann, erscheint diese Datenquelle in der Anwendung nur als Namen, d.h. im Falle von ODBC als *Data Source Name*, und im Falle von OLE DB als *Universal Datalink*. Darüber kann die Datenquelle angesprochen werden, ohne dass die Anwendung den genauen Ort der Datenbank kennen muss.

#### 4.6 (Geo-)Datenmodellierung

Unter einem Datenmodell versteht man „künstlich geschaffenes abstraktes Abbild eines Ausschnittes aus der Wirklichkeit mit dem Ziel, bestimmte Gegebenheiten ... abbilden zu können“<sup>70</sup>. Ein Datenmodell erklärt und visualisiert die Struktur einer Datenbank, indem die einzelnen Entitäten und ihre Relationen erarbeitet und dargestellt werden.

Ein Datenmodell bekommt dadurch eine konzeptuelle Komponente, indem es dazu dient, die integrierenden Daten vorzustrukturieren und ihre Stellung im Zusammenspiel mit anderen Daten im Vorfeld zu klären. Ein Ziel jedes Datenmodells sollte sein, den Erklärungswert und die Zweckorientierung der Daten sicherzustellen bzw. zu optimieren.

Während bei operativen Datenbanken das Augenmerk auf Eigenschaften wie Redundanzfreiheit oder eine hohe Transaktionsdichte gelegt wird, muss die Datenstruktur eines Data-Warehouses in erster Linie verständlich für den Benutzer sein. Diesem Ziel der strategischen Entscheidungsunterstützung müssen sich die technischen Gegebenheiten und das Datenmodell unterordnen. Dies hat zur Folge, dass in der Praxis auf Redundanzfreiheit oder Normalisierung verzichtet wird, wenn dadurch eine bessere Handhabbarkeit gewährleistet wird. Eine vollständige Normalisierung würde zu einer Vielzahl von einzelnen Relationen führen, was die Analyse extrem verkompliziert. So kann es sein, dass exakt die selben Daten mehrfach vorliegen, bspw. in verschiedenen Zeitaufösungen.

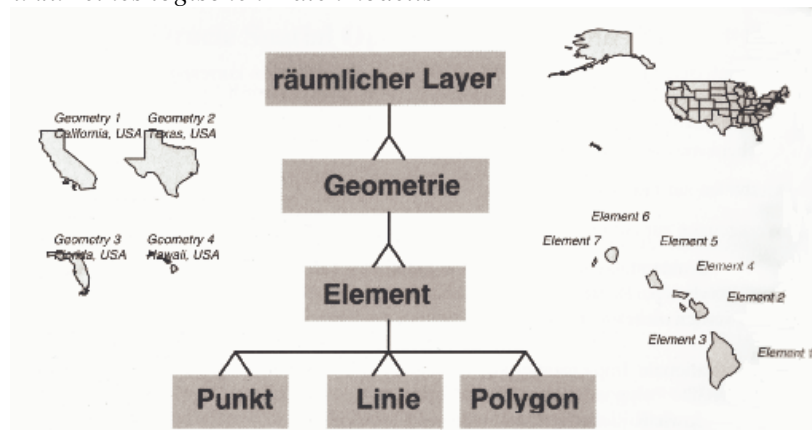
Erstellt man ein Datenmodell in dem Geodaten eine Rolle spielen, ist es bedeutend, inwieweit die Speicherung der Geometrie, also der graphischen Ausprägung, in das Data-Warehouse integriert ist. Im „Normalfall“ wird sich der Datenmodellierer intensiv Gedanken über die Modellierung der alpha-numerischen Sachdaten und deren Zusammenspiel mit der Geometrie-Komponente machen, während die Technologie zur reinen Speicherung der Geometrie durch ein GIS-Produkt abgedeckt wird. Er muss sich keine Gedanken über die Modellierung von geographischen Objekten mittels geographischen Primitiven wie Punkte, Linien oder Polygone machen. Diese Aufgaben werden von einer GIS-Komponente übernommen, welche wie eine „Black-Box“ im Data-Warehouse steht. Der Datenmodellierer

---

<sup>70</sup> Bill/Zehner (2001), S. 58

muss aber wissen, welche geometrischen Ausprägungen einen einzelnen Datensatz darstellen, um ihn einem korrespondierenden Datensatz in den Sachdaten zuordnen zu können. So muss zwar jedem Datensatz eine einzige Geometrie zugeordnet werden, diese muss aber nicht zwingend nur aus einem einzigem geometrischen Element bestehen. Diese Zusammenhänge beschreibt das **logische** Datenmodell von Vektordaten:

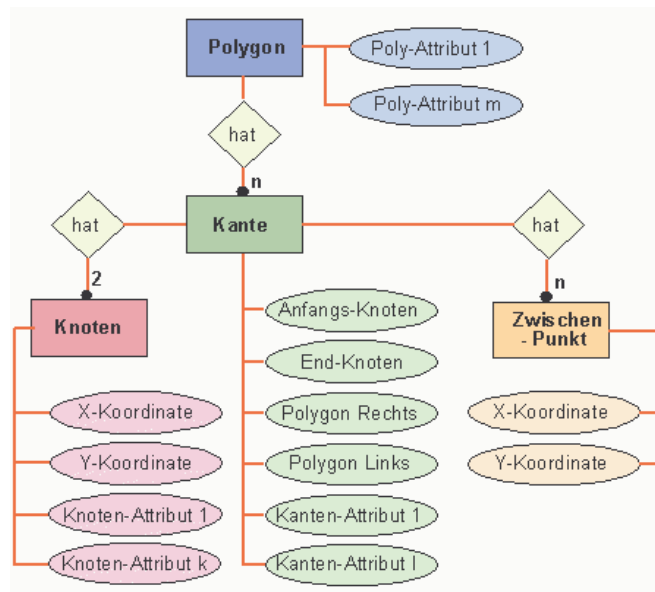
Abb. 14: Struktur eines logischen Datenmodells<sup>71</sup>



So stellen bspw. bei einer Datenbank mit zugeordneten Staatsgrenzen diese die Geometrie eines jeden Datensatzes dar, welche durch Attribute wie Staatsnamen oder ID beschrieben werden. Besteht eine Geometrie aus mehreren nicht-miteinander verbundenen Teilen, spricht man von **Elementen**. Dies wäre z.B. bei Exklaven der Fall, oder bei Staaten die über einzelne Inseln verstreut sind. Ein Element setzt sich wiederum aus geometrischen **Geo-Objekten** oder **Features** zusammen. Dies liegen zumeist in der nicht-komplexen Variante Punkte, Linien, und Polygone vor. Diese werden wiederum aus **Koordinaten-Tupel** aufgebaut. Ein **phisches** oder **relationales** Datenmodell beschreibt, wie diese geometrischen Primitiven in Tabellen gespeichert werden:

<sup>71</sup> Fleischmann, Kaiser (2000), S. 162

Abb. 15: Relationales Datenmodell<sup>72</sup>



<sup>72</sup> Streit (2001), Kap. 9.3

## 5 Geo-Data-Warehousing

Unter einem „Geo-Data-Warehouse“ soll nachfolgend ein Data-Warehouse verstanden werden, welches durch Integration von Geodaten dem Benutzer ermöglicht, räumliche Analysen mittels einem GIS auf den Sachdatenbestand des Data-Warehouse anzuwenden. Dies gilt auch unabhängig davon, welcher der in Kapitel 4 behandelten Integrationsvarianten angewandt wird.

### 5.1 Der georationale Data-Mart

Gleich welche der oben genannten Varianten zur Anbindung der Geodaten benutzt wird, bleibt das Grundproblem das Selbe: Es müssen Massendaten (die Sachdaten) mit eher speziellen Daten (die Geodaten) in Verbindung gebracht werden. Während erstere dem ganzen Unternehmen zur Verfügung stehen sollten, besitzen zweite einen eher eingeschränkten Nutzerkreis. So werden bspw. Kundendaten eines Energie-Versorgungsunternehmens in nahezu allen Bereichen und Gliedern der Wertschöpfungskette benötigt: Im Vertrieb, im Kundenservice, in der Kundenabrechnung, in der Messzählerkontrolle usw. Die Verbindung dieser Daten mit Geodaten wird aber eher nur den kleinen Kreis von Mitarbeitern von Nutzen sein, die sich mit Geo-Marketing-Fragestellungen, wie Mikro-Marketing beschäftigen.<sup>73</sup> Daher bietet es sich an, diese Geodaten nur diesem Nutzerkreis zur Verfügung zu stellen.

Im Data-Warehouse-Kontext bietet sich somit der Aufbau eines „**geographischen Data-Marts**“ an. D.h. eine Teilmenge des unternehmensweiten Data-Warehouse wird aufgesetzt, der speziell dafür konfiguriert wird, Daten geographisch aufzuarbeiten und mit einem GIS zu bearbeiten.

Dieser „Geo-Data-Mart“ sollte kein abgeschottetes Data-Warehouse nur für Geo-Marketing-Anwendungen sein, sondern Teilelement eines Data-Warehouse, welcher für bestimmte Abteilungen oder Funktionen beschaffen wird.

Bei der Problematik der Verlinkung von Geometrie- und Sachdaten ist m.E. zwischen der *georationalen* und der *georelationalen* Speicherung auszuwählen. Sowohl die Stand-alone-Lösung eines GIS-Rechners, als auch die On-the-Fly-Erzeugung von Geometrie per Geocodierung erscheint nicht zweckmäßig, wenn man die Anbindung der geographischen Analysetools in das Data-Warehouse bedenkt, als auch die Komplexität der Geometriedaten, die größere Polygone oder auch Rasterkarten umfassen können.

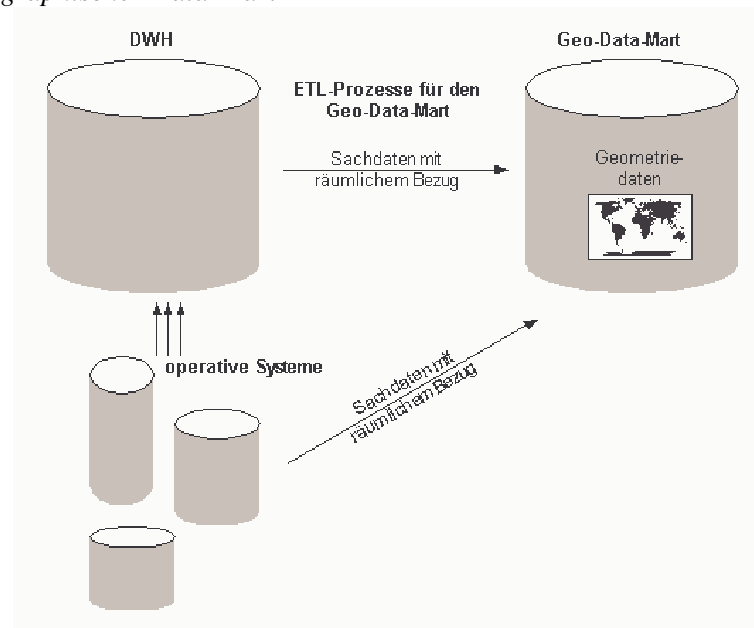
---

<sup>73</sup> Bei der EnBW Vertriebsgesellschaft sind dies je nach Arbeitsanfall ca. 3-5 Mitarbeiter. Bei der Tochtergesellschaft YELLO Strom gibt es eine Abteilung Mikro-Marketing, die aus momentan 8 Mitarbeitern besteht.

Bei der Realisierung einer georelationalen Speicherung ist aber zu beachten, dass eine Datenbank, die mit Middleware-Komponenten ausgestattet ist, zu einer Insellösung führt.<sup>74</sup> Geomarketing-Analysen in Energie-Versorgungsunternehmen zeichnen sich aber weniger durch eine ständige Wiederholung gleicher Vorgänge ab, sondern sind von den aktuellen Vertriebsprojekten abhängig. So werden strategische Analysen wie Werbewirkungsanalysen oder Kundenklassifikation nach Bedarf durchgeführt werden. Eine Interaktion mit den Unternehmensdaten aus unterschiedlichen Bereichen ist daher nötig, so dass eine o.g. Insellösung hier m.E. nicht zum Ziel führt.

Eine Lösung zur Integration von geographischen Daten in ein Data-Warehouse stellt daher der **georationale Data-Mart** dar: eine Teilmenge eines Data-Warehouses die speziell für GIS-Anwendungen bereitgestellt wird, und in dem sowohl Geometriedaten als auch Methoden bereitliegen, Sachdaten an diese anzubinden. Die Geometriedaten werden dabei file-basiert gespeichert, und per Schnittstellen verknüpft („georational“).

Abb. 16: Geographischer Data-Mart



Dieser Data-Mart wird somit speziell für den Subjektbereich „geographische Marketing-Analysen“ bereitgestellt. Die Datenbasis dieses Data-Marts wird teilweise vom unternehmensweiten Data-Warehouse bereitgestellt, was ihn dadurch zu einem *abhängigen* Data-Mart macht, welcher in die Gesamt-Architektur des Data-Warehouse eingebunden ist. Er kann aber auch Daten direkt aus operativen Quellen erhalten. Durch seine spezielle Ausrichtung auf GIS-Fragestellungen, kann er auch eine eigene Software-Architektur und ein eigenes Datenmodell erhalten.

<sup>74</sup> Siehe Ausführungen in Kapitel 4.3

## 5.2 Prozesse

In der ETL-Schicht eines Data-Warehouses<sup>75</sup> sind die Prozesse und Regeln hinterlegt, nach denen operative Daten in das Data-Warehouse aufgenommen werden. Im Falle einer Data-Mart-Lösung können die relevanten Daten entweder direkt aus den operativen Quellen eingespielt werden, oder sie gehen den Umweg über ein zentrales Data-Warehouse.

Der **ETL-Prozess**, also der Prozess des Beladens des Data-Warehouse (bzw. des Data-Marts) lässt sich in mehrere Teilprozesse unterteilen:<sup>76</sup>

- Analyse der Quelldatensysteme
- Extrahieren der Informationen
- Transformation anhand der Data-Warehouse-Konventionen
- Validierung der Informationen
- Integration bzw. Laden in das Data-Warehouse

Der ETL-Prozess ist Kernstück des gesamten Data-Warehouse-Projektes und entscheidet erheblich über dessen weiteren (erfolgreichen) Verlauf. Der Aufwand für diese Phase kann durchaus mit ca. 50% des Gesamtaufwandes veranschlagt werden.<sup>77</sup> Bei den Teilprozessen ist darauf zu achten, dass sie sich alle in das Data-Warehousing als Gesamtprozess einfügen. Dazu gehört z.B., dass die Teilprozesse mit den Data-Warehouse-Konventionen des Unternehmens abgestimmt werden, wie bspw. einem Regelwerk zur Qualitätssicherung, oder dass die zu integrierenden Daten sich in das übergeordnete Datenmodell des Data-Warehouses einordnen lassen.

Bei der Füllung eines „geographischen Data-Marts“ mittels ETL-Prozessen ist besonders den Umständen Rechnung zu tragen, dass die Daten graphischen Objekten zugewiesen werden sollen, und dass räumliche Analysen gefahren werden können, die über reine numerische Analysen hinausgehen.

Der ETL-Prozess kann in drei Phasen gegliedert werden:

### 5.2.1 Analyse und Extraktion der Informationen

Hierbei werden die in dieser stark konzeptionell geprägten Phase die Quelldatensysteme sondiert und katalogisiert. Es werden dann die Daten aus dem Bestand ausgewählt, die für eine Verknüpfung zu den Geometriedaten nötig sind. Des Weiteren werden Daten bestimmt, die für die Durchführung und Aussagekraft von Analysen nötig sind. Dazu werden diese

---

<sup>75</sup> Vgl. Kapitel 2.1.2

<sup>76</sup> Vgl. Kurz (1999), S. 267

<sup>77</sup> Vgl.: Kurz (1999), S. 266

Datenpakete auf der Staging-Area des Data-Warehouse abgelegt, die dort als temporäre Extrakte organisiert sind, die dann weiterverarbeitet werden.

### 5.2.2 Transformation

Die extrahierten Daten müssen nun so aufbereitet werden, dass die Informationen den zuvor definierten Data-Warehouse-Konventionen entsprechen.<sup>78</sup> Dabei werden die Daten auf **Validität** geprüft. Unter einer *syntaktischen Validitätskontrolle* könnte man bspw. einen Check auf Rechtschreibung verstehen. Zu *semantischen Validitätskontrollen* zählen bspw. Wertebereichschecks (Liegt der Monatswert zwischen 1 und 12?) oder Inkonsistenzprüfungen (passt die PLZ mit dem angegebenen Bundesland überein?). Des Weiteren werden in dieser Phase Daten konvertiert oder aufbereitet, um sie so den Anforderungen des Nutzers anzupassen. Hierzu zählen:

- Einfache Rechenschritte wie bspw. Währungsumrechnungen oder Kennzahlen-Berechnung.
- Neu-Definition von Attributen (bspw. durch Dichotomisierung oder Hinzufügen eines Zeitstempels)
- OLAP-Methoden, wie Aggregation oder Klassifizierung von Daten

Sofern eindeutige Geoschlüssel in den Daten noch nicht existieren, werden sie in dieser Phase aus den vorhandenen Daten erzeugt. So könnte bspw. einem Datensatz mit einem regionalen Umsatz die Gemeindekennziffer zugeordnet werden, die in einer weiteren Relation bereitgestellt wird. In dieser Phase würde man auch eine automatische Geocodierung durchführen, wenn man diese durch den ETL-Prozess abwickeln lassen wollte.<sup>79</sup>

### 5.2.3 Integration bzw. Laden in das Data-Warehouse

Ladeprozesse legen die extrahierten und transformierten Daten im Data-Warehouse ab, wobei die Informationen in das Datenmodell eingepasst werden. Hierzu benötigt man Ladesoftware, welche sich im einfachsten Falle aus SQL-Prozeduren zusammensetzen. Diese führen außerdem notwendige Datenbank-Aufgaben durch, die mit der Einfügung oder Änderung von Daten einhergehen wie Neu-Indizierung, Schlüsselerzeugung oder Update der Metadaten im Data Dictionary.

---

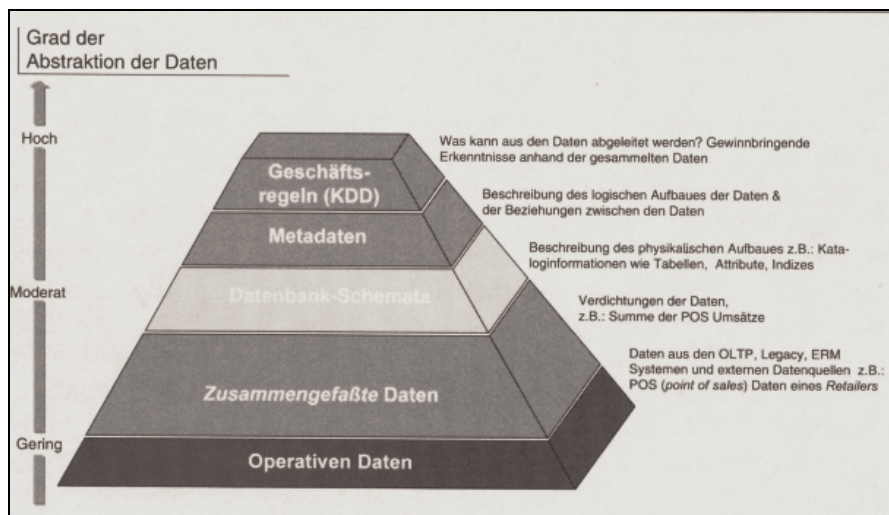
<sup>78</sup> Vgl.: Kurz (1999), S. 268

<sup>79</sup> Vgl.: Kapitel 4.4

### 5.3 Meta-Daten

Die Frage nach der optimalen Ausgestaltung des ETL-Prozesses geht einher mit der Frage eines effizienten Metadaten-Managements, da der ETL-Prozess idealerweise durch Metadaten gesteuert wird. Unter Metainformationen<sup>80</sup> versteht man **Informationen über Informationen**, also Daten, die zur Beschreibung von Daten dienen. Sie bilden daher das „...Inhaltsverzeichnis eines Data-Warehouses...“<sup>81</sup>, und sollen den Benutzern ein schnelles und sicheres Navigieren im Datenbestand des Data-Warehouses ermöglichen. Sie ermöglichen einen gezielten Zugang zu den Daten, in dem sie Informationen über Beschreibung, Katalogisierung und Verwaltung der einzelnen Datensätze anbieten.<sup>82</sup>

Abb. 17: Metadaten in der Datenhierarchie<sup>83</sup>



Metadaten dienen aber nicht nur zur Dokumentation, sondern auch zur Steuerung des informationstechnischen Ablaufes innerhalb eines Data-Warehouses<sup>84</sup>. So sind Daten über

- Zugriffsberechtigungen
- Datenquellen
- Datenziele
- Datentypen
- Transformationsregeln

als Metadaten hinterlegt, welche ETL-Prozesse steuern.

Metadaten können direkt den Daten beigefügt werden<sup>85</sup> oder auch getrennt von den beschreibenden Daten gehalten werden. Metadaten sind auch prinzipiell nicht an ein Format

<sup>80</sup> von der griechischen Vorsilbe „Meta“ = inmitten, zwischen, hinter, nach; Vgl.: Bill/Zehner (2001), S. 174

<sup>81</sup> Kurz (1999), S. 199

<sup>82</sup> Vgl.: Wegner (2000 I), S. 53

<sup>83</sup> Kurz (1999), S. 206

<sup>84</sup> Vgl.: Kurz (1999), S. 200

<sup>85</sup> z.B. in einem Memo-Feld oder einem Binary Large Object-Feld (BLOB) einer Datenbank-Tabelle

gebunden, sie können als Texte<sup>86</sup>, Grafiken oder Multimedia-Dateien vorliegen. Jedoch ist dies für eine Data-Warehouse-Anwendung nicht sinnvoll. Hierfür werden Metadaten idealerweise tabellarisch organisiert und an einem zentralen Ort im Data-Warehouse hinterlegt, dem **Data Dictionary**<sup>87</sup>. Dies stellt eine eigenständige Komponente des Data-Warehouses dar. Das Data Dictionary ist in Form einer Datentabelle organisiert und beinhaltet ein eigenes Datenmodell. Dieses Data Dictionary wird als **Metainformationssystem** bezeichnet, weil es nicht nur in der Lage ist, dem Nutzer zu zeigen, welche Daten vorhanden sind und was sie bedeuten, sondern auch wo sie abgelegt sind und wie auf sie zugegriffen werden kann. Ein Metainformationssystem steuert also die Informationsflüsse in einem Data-Warehouse.

In Bezug auf geographische Daten stellen Geodaten also nicht zwingend eine räumliche Information dar, sondern sie stellen Informationen *über* räumliche Informationen dar.

**Metadaten beschreiben → Geodaten beschreiben → räumliche Realität**<sup>88</sup>

Laut der DIN-Norm ‚**DIN ENV 12009**‘ von 1997 enthalten Metadaten für Geodaten „...wichtige Informationen, die Nutzern die Brauchbarkeit von Geodaten für Ihre eigenen technischen, rechtlichen und geschäftlichen Anforderungen abzuschätzen erlauben.“<sup>89</sup> Meta-Daten für Geodaten kann man in folgende Kategorien unterteilen:<sup>90</sup>

- geometrisch-räumlicher Bezug  
(z.B. Maßstab, Projektion, Koordinatenlage)
- thematischer Bezug  
(z.B. Inhalt, Thema)
- zeitlicher Bezug  
(z.B. Stand, Gültigkeit, Erwerbszeitpunkt)

Zur konkreten Ausgestaltung von Metadaten für Geoinformation wurden mittlerweile mehrere Standards definiert, bspw.:

- **CSDGM** – **C**ontent **S**tandard for **d**igital **G**eospatial **M**etadata, ein 1994 in den USA für Behörden eingeführtes System
- Die **Catalogue Services** im Rahmen des Open GIS Consortiums, zur Definition eines Austauschformates für Geodaten

---

<sup>86</sup> z.B. ASCII oder XML

<sup>87</sup> auch Repository oder Data Catalog genannt

<sup>88</sup> Wegner (2000 I), S. 37

<sup>89</sup> zitiert in: Wegner (2000 II), S. 513

<sup>90</sup> Vgl.: Wegner (2000 I), S. 40

- Der **Objektartenkatalog** des ATKIS<sup>91</sup> – Amtliches topographisch-kartographisches Informationssystem
- Die **DIN-Norm ,ENV 12657’**,  
welche in Deutschland eine zentrale Rolle einnimmt.

Letztgenannter Standard benutzt mehrere Bereiche um Datensätze eines Geodatenbestands zu beschreiben:<sup>92</sup>

1. Identifikation
2. Beschreibung
3. Qualität (bspw. Herkunft oder Verwendung)
4. Metadaten über Metadaten, um deren Qualität zu beurteilen
5. Raumbezugssystem
6. Ausdehnung (bspw.: 2D- oder 3D-Datenmodell)
7. Datendefinition (... der im Geodatensatz enthaltenen räumlichen Objekte)
8. Klassifizierung (... der im Geodatensatz enthaltenen räumlichen Objekte)
9. Verwaltungs-Metadaten (bspw. Provider oder veröffentlichende Organisation)

Abb. 18: Metadatenelemente nach DIN ENV 12657<sup>93</sup>

Nr. und Bezeichnung der Metadatenelemente	Eintrag <sup>1</sup>	Ausprägungen je Metadatenelement <sup>1</sup>				
		$\Sigma^2$	$\Sigma=Z^3$	$\Sigma=B^4$	$\Sigma=O^5$	$\Sigma=Z/Z+B^6$
1. Datensatzidentifikation	Z	4	1	0	3	1
2. Datensatzüberblick	Z	15	5	1	9	6
3. Datensatz Qualitätselemente	Z	6	0	4	2	4
4. Metadatenverweisungen	Z	5	1	2	2	3
5. Raumbezugssystem	O	12	0	5	7	
6. Ausdehnung	Z	20	3	12	5	15
7. Datenfestlegung	O	25	0	10	15	
8. Klassifizierung	O	14	0	4	10	
9. Verwaltungsmetadaten	O	23	0	10	13	
	Z=5   O=4	124	10	48	66	29

1: Z = zwingend B = bedingt O = optional	2: Summe aller Ausprägungen je Metadaten- element	3: Summe aller zwingenden Ausprägungen je Metadaten- element	4: Summe aller bedingten Aus- prägungen je Metadaten- element	5: Summe aller optionalen Ausprägungen je Metadaten- element	6: Summe aller zwingenden und bedingten Ausprägungen für alle zwin- genden Meta- datenelemente
---	---	---	--	---	---

Metadaten für Geodaten im Data-Warehouse-Kontext sind prinzipiell gleich zu handhaben wie Metadaten für alpha-numerische Daten. Auch wenn das Metainformationssystem eine zentrale Komponente des Data-Warehouse darstellt, sollte für einen Data-Mart ein spezielles Metainformationssystem eingerichtet werden, erst recht, wenn dieser mit Daten arbeitet, die

<sup>91</sup> Dies stellt nur ein „Quasi-Standard“ dar, weil hier räumliche Objektarten definiert sind, mit denen Daten sich thematisch gliedern lassen; Vgl.: Wegner (2000 I), S. 52

<sup>92</sup> Vgl.: Wegner (2000 I), S. 48

<sup>93</sup> Wegner (2000 I), S. 50

in anderen Data-Marts sonst nicht zur Verfügung stehen. In unserem Anwendungsfall eines „geographischen Data-Marts“ wäre dies gegeben. Hier sollte ein Data Dictionary eingerichtet sein, welches speziell auf die Geometrie-Komponente ausgerichtet ist. Um die Daten austauschbar halten zu können, ist es ratsam eine der o.g. Normen für Metadaten einzuhalten. Da das Data Dictionary für die Geometriedaten evtl. auch statische Grafiken beinhalten kann (z.B. Meta-Karten mit Ausschnittskennzeichnungen des eigentlichen Kartenmaterials), sollte das Data Dictionary dies auch vorhalten können, bspw. über den Datentyp *Binary Large Object (BLOB)* welcher ein DBMS bereithält. Hiermit können Daten, die nicht alphanumerisch gespeichert werden können (z.B. Grafiken), als Binärcode in der Tabelle des Data Dictionary hinterlegt werden.

#### **5.4 Data-Mining mit Geodaten**

Damit aus dem Data-Warehouse kein „Data-Graveyard“ wird müssen die Daten in einer Systematik zur Verfügung gestellt werden, um eine geeignete Basis für Data-Mining-Methoden darzustellen. Darunter fasst man Methoden aus diversen Forschungsrichtungen zusammen, die zum Ziel haben, Muster in den Daten zu erkennen.<sup>94</sup>

Die Anwendung von Data-Mining-Methoden erfolgt in Phasen:

- Problemdefinition  
(z.B. Bestimmung der in Zukunft zu erwartenden Bonität eines bestimmten Kunden)
- Definition der benötigten Daten  
(z.B. Kundendaten, allgemeine Bonitätsdaten)
- Exploration der benötigten Daten  
(z.B. Ausreißer und Scheinkorrelationen erkennen)
- Festlegung der Aufgabe  
(z.B. Prognose)
- Festlegung der Methode  
(z.B. Regressionsanalyse)
- Anwendung dieser Methode(n) mit einem zu bestimmenden Werkzeug  
(z.B. SPSS)
- Ergebnisse validieren  
(z.B. statistische Signifikanztests)
- Ergebnisse visualisieren  
(z.B. durch kartographische Darstellung)

---

<sup>94</sup> Vgl.: Kapitel 2.3

- Neues Wissen aus den Ergebnissen generieren  
(z.B. Überarbeitung der Kreditoren-Richtlinien)

Klassische Data-Mining-Methoden beziehen sich fast ausschließlich auf numerische Daten, mit Ausnahmen auch auf Text-Daten. Die Untersuchung von Mustern in graphischen Objekten, findet kaum Anwendung, da „herkömmliche“ Data-Warehouse-Daten alphanumerisch orientiert sind. Will man graphische Objekte, wie Geometrien, in den Data-Mining-Prozess einbinden, gilt es das vorhandene Instrumentarium um geostatistische bzw. raumanalytische Methoden zu erweitern.

Da diese Methoden, genauso wie geographische Daten, einem begrenzten Benutzerkreis zur Verfügung gestellt werden, sind diese „geographischen Data-Mining-Methoden“ idealerweise im Subsystem eines geographischen Data-Mart zu installieren. Dies bietet sich nicht nur konzeptionell an, sondern auch praktisch, da hier auch GIS-Werkzeuge installiert sind, die diese Methoden anbieten.

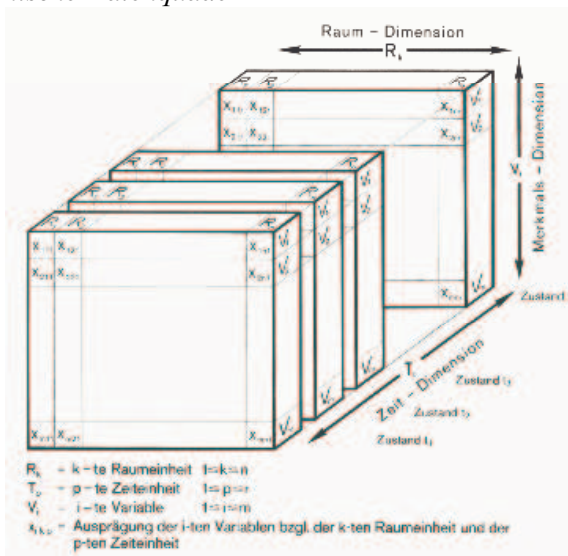
Während man unter dem breit angelegten Begriff der **Raumanalyse** („spatial analysis“) Methoden versteht, die allgemein räumliche Phänomene untersuchen, versteht man unter **Geostatistik** rein statistische Verfahren, die der a-räumlichen Statistik entnommen sind, und auf räumliche Begebenheiten erweitert und angepasst wurden<sup>95</sup>.

Geographische Analysen sind i.d.R. komplexer als a-räumliche Analysen, da die Abfrage von Daten in einem räumlichen Kontext ungleich komplizierter sind. So ist eine Abfrage wie „*Welche Punkte liegen innerhalb des Polygons*“ mit Standard-Abfragesprachen wie SQL nicht möglich. So erweitern geographische Analyse-Methoden die Daten um eine weitere, die räumlichen Dimension:

---

<sup>95</sup> Bill/Zehner (2001), S. 221 & S. 112

Abb. 19: Der geographische Datenquader<sup>96</sup>



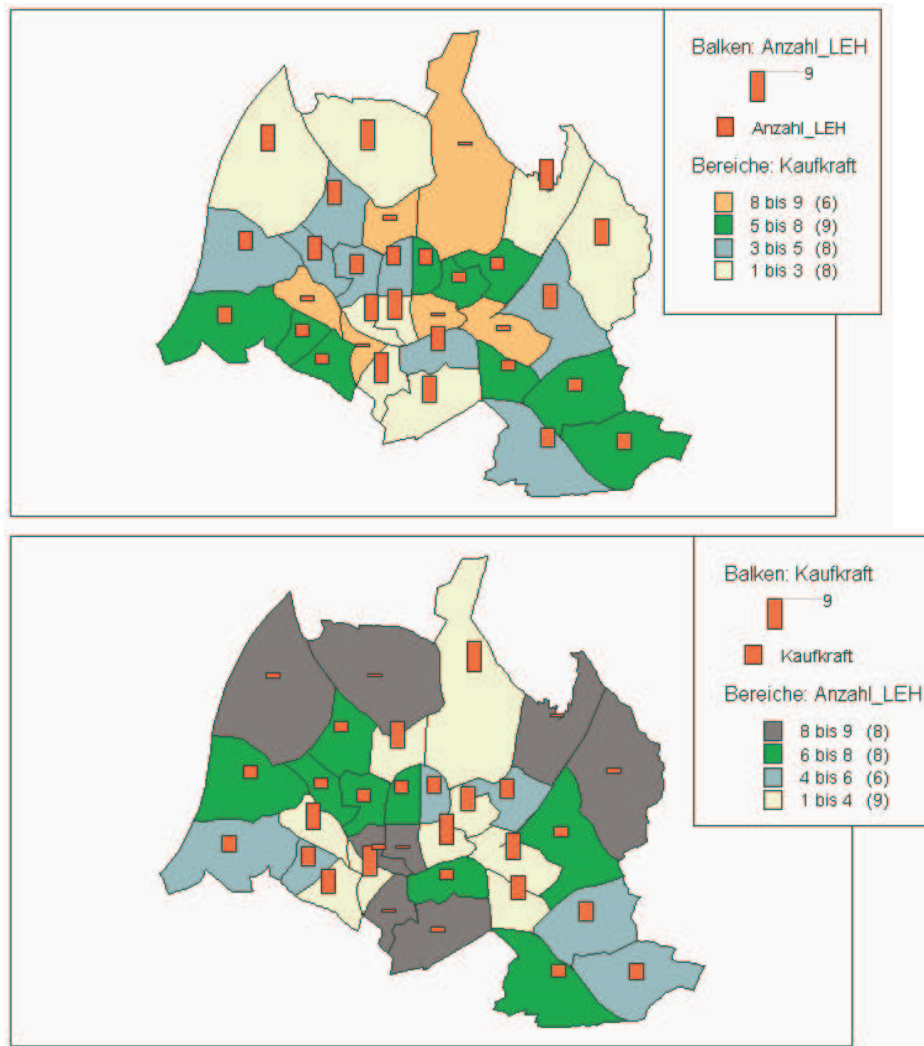
Die in Kapitel 2.3 aufgeführten Data-Mining-Aufgaben oder Anwendungsklassen lassen sich größtenteils auf räumliche Objekte anwenden, wie z.B. Klassifikations- oder Clustermethoden. Dabei werden Variablen aber nicht aufgrund Ihrer numerischen Ausprägung, sondern aufgrund Ihrer räumlichen Gegebenheit verarbeitet.

Eine starke Anwendung im Data-Mining-Kontext wird aber auf der graphischen Analyse liegen, da Geodaten eine visuelle Orientierung besitzen. Dies bedeutet, dass Muster in den Daten weniger über Kennzahlen oder Tabellen dargestellt werden, sondern in graphischer Form, v.a. über Karten.

Über aufbereitetes Kartenmaterial können räumliche Datenmuster erkannt und analysiert werden, die in einem reinen alpha-numerischen Kontext untergehen würden. Dabei ist es auch möglich mehrere Ebenen von Informationen darzustellen, die sonst über mehrere Tabellen verteilt wären, und erst gedanklich in Verbindung gebracht werden müssten. In einem Kartenmodul können die Informationsebenen, ähnlich einem OLAP-Würfel, ein – und ausgeblendet werden.

<sup>96</sup> Bahrenberg, Giese, Nipper (1999), S. 22

Abb. 20: OLAP-Funktionalität mit Karten



Im oberen Beispiel in Abb. 20 ist die Kaufkraft-Information durch die Choroplethen dargestellt und die Anzahl der LEH<sup>97</sup>-Betriebe durch die Balkendiagramme (obere Darstellung). Durch „Drehen des OLAP-Würfels“ (Data Slicing) werden die Darstellungsarten vertauscht (untere Darstellung).

### 5.5 Das (Geo-)Data-Warehouse als Bestandteil eines Marketing-Systems

Um eine effiziente Data-Warehouse- bzw. Data-Mart-Lösung aufzubauen, muss in jeder Phase des Data-Warehousing-Prozesses das Umfeld berücksichtigt werden, in der das System eingebettet wird. Dazu zählen hauptsächlich Quellsysteme, Organisationsstruktur und die Aufgaben die ein Data-Warehouse bearbeiten soll. Nur so ist der effiziente Einsatz einer Data-Warehouse-Lösung gewährleistet.

Gerade im Marketing, im Sinne einer zielgerichteten Marktbearbeitung, kann ein effizienter Einsatz eines Data-Warehouses und seiner Information entscheidend über die Marktperformance sein. Wie in keinem anderen Unternehmensbereich entscheiden Qualität

<sup>97</sup> Lebensmittel-Einzelhandel

und Geschwindigkeit der Informationsbereitstellung über den Markterfolg. Aus diesem Grund hat sich der Begriff des **Computer Aided Marketing Managements (CAMM)** etabliert.<sup>98</sup> Darunter versteht man computergestützte Systeme, die operative und strategische Prozesse im Marketing-Management unterstützen und steuern helfen. Hier ist vor allem die Bereitstellung von relevanten Daten zu nennen. Insofern fügt sich ein Data-Mart, der für Geomarketing-Fragestellungen optimiert ist, in ein unternehmensweites CAMM-System ein.

Durch die Integration Marketing-relevanter Daten wird der oben beschriebene „Geo-Data-Mart“ mit Leben gefüllt. Daten, die im Geomarketing-Kontext zu einer optimalen Marktbearbeitung integriert sein sollten, sind vielfältiger Natur. Diese Daten sind in erster Linie kundenbasiert, d.h. in letzter Instanz sollten sie auf den Kunden als Analyseobjekt ableitbar sein.

Im Bereich der **internen Sachdaten**<sup>99</sup> wäre dies v.a.:

- Stammdaten der Kunden
- Kaufakte der Kunden (in Mengen und monetär gemessen)
- Daten über Kundenbonität
- Daten über Direkt-Marketing-Aktionen beim Kunden
- Daten über allgemeine Marketing-Aktivitäten im Gebiet des Kunden
- Umsatzkennziffern

Im Bereich der **externen Sachdaten**<sup>100</sup> wäre dies v.a.:<sup>101</sup>

- Kaufkraftdaten und andere sozioökonomische Daten
- Soziodemographische Daten (z.B. Wohnbevölkerung, Bildung)
- Volkswirtschaftliche Indikatoren (z.B. BIP)
- Umsatzkennziffern der gesamten Branche
- Bestandsgrößen, die den speziellen Markt betreffen, der abgebildet werden soll. (z.B. Anzahl der Haushaltsanschlüsse für den Strommarkt)

Im Bereich der **Geometriedaten** sind zu nennen:

- Vektor-Geometrien (Polygone) über Standard-Gebietseinheiten (z.B. Kreise, Gemeinden)
- Vektor-Geometrien (Polygone) über unternehmensspezifische Gebietseinheiten, z.B. selbst definierte Vertriebsgebiete
- Vektor-Geometrien (Linien) über das Straßennetz für Entfernungsmessungen (z.B. Navtech- oder TeleAtlas-Daten)
- Vektor-Geometrien (Punkte) für geokodierte Adressdaten (z.B. Kundenstandorte)

---

<sup>98</sup> Vgl.: Schürmann (1997), S. 90

<sup>99</sup> d.h. alpha-numerische Sachdaten, die innerhalb des Unternehmens aus operativen Datenquellen zur Verfügung stehen.

<sup>100</sup> d.h. alpha-numerische Sachdaten, die vom Unternehmen aus externen Quellen zugefügt wurden.

<sup>101</sup> Vgl.: Kastin (1995), S. 333

- Rasterkarten für kartographische Zwecke, wie georeferenzierte Stadtpläne, Übersichtspläne oder Straßenkarten

Zusammengefasst lässt sich sagen: Eine ideale Kombination an Daten für ein „Geo-Data-Mart“ sollte den Kunden **identifizieren** (interne Sachdaten), ihn **klassifizieren** (externe Sachdaten), und ihn **räumlich darstellen** (Geometriedaten). Diese Kombination an Daten soll den Marketing-Manager befähigen, raumrelevante bzw. kundenrelevante Planungen und Analysen durchzuführen. Sind diese Daten als Datenbasis eines Data-Mart organisiert, ermöglicht dies einen sicheren Zugriff und effizientere (da vordefinierte) Analysemöglichkeiten im Vergleich zu einer Datenhaltung außerhalb eines Data-Warehouses.

## 6 Fallbeispiel: Penetrationsanalyse für den Strom-Endkundenmarkt

### 6.1 Aufgabenstellung

Mittels einer Penetrationsanalyse lässt sich die Ist-Situation des Unternehmens in den Absatzmärkten feststellen. Ziel ist es dabei, für ein definiertes Gebiet die Versorgung mit einem bestimmten Produkt festzustellen. Die **Marktdurchdringung bzw. Penetration** stellt dabei dar, inwieweit das ausgeschöpfte Potential dem Gesamtpotential entspricht. Diese Kennzahl wird als Quotient zwischen 0 und 1, bzw. als Prozentgröße zwischen 1% und 100% dargestellt. Ein Wert von 0 bedeutet dabei keinerlei Durchdringung des Marktes, während ein Wert von 1 eine vollkommene Dominanz des betrachteten Produktes im definierten Marktgebiet darstellt. Nach einer vorher zu bestimmenden Grenze wird schließlich eine Über- bzw. Unterversorgung diagnostiziert.

Im Datenkontext liegt hier die Besonderheit, dass hier Daten aus zwei Quellen miteinander verglichen werden. Das selbst **ausgeschöpfte Potential** sind interne Daten, wie z.B. interne Umsatz- oder Absatzdaten. Das **Gesamtpotential** sind Daten, die branchenweit erhoben werden müssen, also externe Marktteilnehmer mitberücksichtigt. Diese Daten werden aus demographischen Daten oder branchenspezifischen Daten gewonnen.

Im hier aufgeführten Beispiel soll eine Penetrationsanalyse in Bezug auf die *mit Strom versorgten Privathaushalte* aufgezeigt werden. Das ausgeschöpfte Potential wären somit die Privathaushalte, die bereits Strom des Unternehmens beziehen, während das Gesamtpotential durch die Haushalte dargestellt werden, die Strom beziehen, egal von welchem Anbieter. Da Strom ein lebenswichtiges Produkt ist, welches zur absoluten Mindestausstattung jeden Haushaltes zählt, kann man die Anzahl der Anschlüsse mit der Anzahl der Haushalte gleichsetzen. Man kann daher einen abgerechneten Hausanschluss als einen Kundenhaushalt betrachten.

Eine Penetrationskennziffer würde sich hiernach definieren als:

$$Penetration = \frac{Kundenhaushalte}{Gesamthaushalte}$$

wobei *Kundenhaushalte* das bereits ausgeschöpfte Potential darstellt, und *Gesamthaushalte* alle verfügbaren Stromanschlüsse (=Haushalte).

Man kann verschiedene Penetrationsstufen unterscheiden, welche sich v.a. auf den Nenner beziehen. So kann man **alle Haushalte** miteinbeziehen (Stufe 1), nur die Haushalte, die **potentiell** als Kunden in Frage kommen (Stufe 2), bzw. die Haushalte die schon das Produkt **eines anderen Herstellers** besitzen (Stufe 3).<sup>102</sup> Da im Energiemarkt jeder Haushalt einen

---

<sup>102</sup> Vgl.: Nitsche (1998), S. 152

Stromanschluss besitzt und ihn nutzt, d.h. jeder als potentieller Stromkunde in Frage kommt, wird diese Unterscheidung hier nicht getroffen.<sup>103</sup>

Stellt man diese Penetrationskennziffer für jede Gebietseinheit dar, erhält man die geographische Verteilung dieser Kennziffer. Die Untersuchung und Darstellung dieser Verteilung stellt dann die eigentliche Penetrationsanalyse dar, die Antwort auf die Frage gibt, wo die „weißen Flecken“ liegen, also in welchen Gebieten regionale Marketing-Maßnahmen von Nutzen wären.

Als Vorarbeit für eine Penetrationsanalyse können eine Gebietsstrukturanalyse und eine Kundenstrukturanalyse vorangestellt werden.

Bei einer **Kundenstrukturanalyse** wird der Zähler des o.g. Quotienten betrachtet, also die schon bestehenden Kundenbeziehungen. Diese können nach diversen Strukturmerkmalen klassifiziert werden, die durch Erhebungen gewonnen werden. Auch die genaue Definition des Kundenstandortes bspw. durch Geocodierung fällt unter den Begriff der Kundenstrukturanalyse.<sup>104</sup>

Im Rahmen einer **Gebietsstrukturanalyse** wird das zu definierende Marktgebiet näher betrachtet. Um dieses „Abschöpfungsgebiet“ zu ermitteln, müssen objektive Kriterien angewandt werden, die eine Raumeinheit als zugehörig betrachtet oder nicht. Im Falle einer Standortanalyse kann man von einem Standort ausgehen, der das Zentrum eines Abschöpfungsgebietes bildet. In unserem Beispiel ist dies nicht gegeben. Da Strom zu den Kunden geliefert wird, und die Kosten der Lieferung nicht maßgeblich an die Entfernungsüberbrückung gebunden ist, kann man nicht vom „Marktgebiet“ eines Anbieters oder womöglich eines Kraftwerkes sprechen. Strom wird nicht im eigentlichen Sinne „geliefert“, sondern ist im Netz vorhanden. Soweit es der rechtliche Rahmen zulässt<sup>105</sup>, kann ein Energie-Versorgungsunternehmen jeden Haushalt mit Strom versorgen, der am Stromnetz angeschlossen ist. Insofern ist eine Gebietsstrukturanalyse hier nicht an einen Standort gebunden, sondern kann auf jedes frei zu definierende Marktgebiet aufgebaut werden.

## 6.2 Data-Warehouse-Modell

Anhand dieses in Kapitel 6.1 beschriebenen einfachen Beispiels soll die typische Vorgehensweise der Data-Warehouse-Nutzung beschrieben werden. Das Beispiel soll zeigen, wie die Schritte von der Evaluierung der operativen Daten bis hin zur Visualisierung

---

<sup>103</sup> Die sehr geringe Anzahl an Selbstversorger (z.B. durch private Solaranlagen) sei hier vernachlässigt.

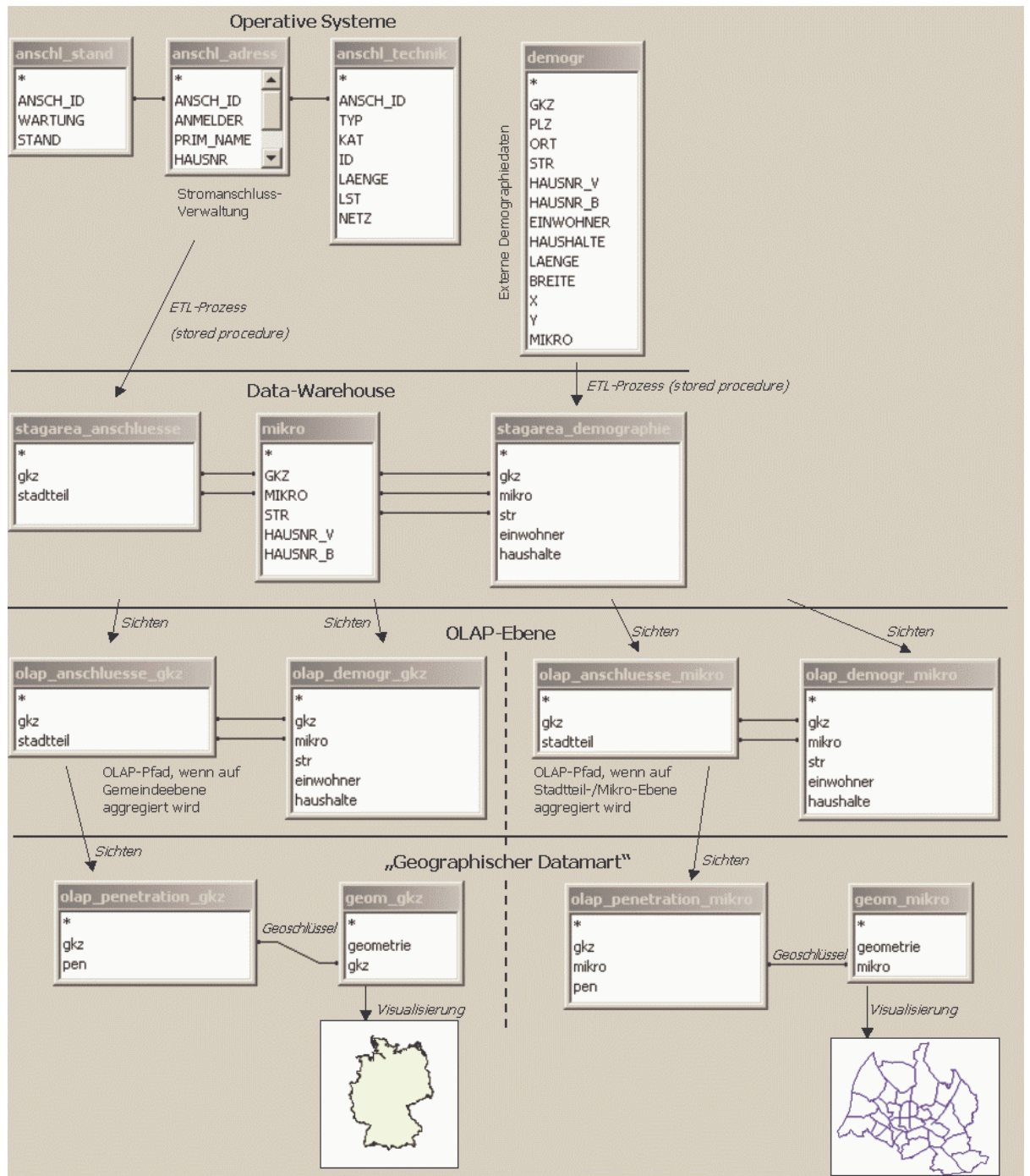
<sup>104</sup> Nitsche (1998), S. 151

<sup>105</sup> Ein deutsches Energie-Versorgungsunternehmen darf in Deutschland jeden Kunden versorgen, ebenfalls in den meisten westeuropäischen Ländern, wobei es hier jedoch Einschränkungen gibt, je nach Liberalisierungsfortschritt der Energiemärkte.

der mittels OLAP aufbereiteten Daten ablaufen. Als DBMS wurde hier Microsoft SQL Server 2000 gewählt und als Desktop-GIS MapInfo professional 5.5.

Vorausgreifend soll hier das Data-Warehouse-Modell der betroffenen Daten dargestellt werden:

Abb. 21: Data-Warehouse-Modell für einen „geographischen Data-Mart“ (Auszug)



### 6.3 Ladeprozesse

Die erste Station eines Data-Warehouse-Prozesses besteht in der Auswahl und Evaluierung der operativen Quelldaten-Systeme.<sup>106</sup> In unserem Beispiel für eine Penetrationsanalyse werden folgende operativen Daten benötigt:

- Daten über **Stromanschlüsse der Haushalte**.

Diese eher technisch orientierten Daten kommen aus den Unternehmensteilen, die die Anschlüsse der Kunden technisch und abrechnerisch betreuen. Um für einen „geographischen Data-Mart“ nutzbar zu sein, sollte in den Daten Attribute vorhanden sein, die das Anlegen eines Geoschlüssels ermöglichen. In diesem Falle ist es der Ortsname und der Straßenname, der eine eindeutige geographische Zuordnung zulässt.

Abb. 22: Tabellenstruktur des operativen Systems „Anschlussverwaltung“ (Auszug)

```

SELECT *
FROM   anschl_adress INNER JOIN
       anschl_stand ON anschl_adress.ANSCH_ID = anschl_stand.ANSCH_ID INNER JOIN
       anschl_technik ON anschl_adress.ANSCH_ID = anschl_technik.ANSCH_ID
    
```

ANSCH_ID	ANMELDER	PRIM_NAME	HAUSNR	SEK_NAME	gkz	WARTUNG	STAND	TYP	KAT	ID	LST	NETZ
1	Anonymous	Schwetzing Str	57	<NULL>	08212000	31.12.2001	44836,69	12	5	53178637	95	3
2	Anonymous	Kaiserslauterner St	44	<NULL>	08212000	31.12.2001	43779,22	12	5	53178845	46	3
3	Anonymous	Kaiserslauterner St	16	<NULL>	08212000	31.12.2001	87530,47	12	5	53178860	2	3
4	Anonymous	Alter Postweg	92	<NULL>	08212000	31.12.2001	15831,88	12	5	53178862	82	3
5	Anonymous	Kaiserslauterner St	32	<NULL>	08212000	31.12.2001	45650,44	12	5	53178867	44	3
6	Anonymous	Beuthener Strasse	59	<NULL>	08212000	31.12.2001	30629,07	12	5	53178945	81	3
7	Anonymous	Beuthener Strasse	82	<NULL>	08212000	31.12.2001	15458,03	12	5	53178946	6	3
8	Anonymous	Beuthener Strasse	42	<NULL>	08212000	31.12.2001	81244,62	12	5	53178948	53	3
9	Anonymous	Beuthener Strasse	36	<NULL>	08212000	31.12.2001	48348,42	12	5	53178949	22	3
10	Anonymous	Beuthener Strasse	86	<NULL>	08212000	31.12.2001	76285,61	12	5	53178950	74	3

Diese Daten werden benötigt, um die Anzahl der bereits angeschlossenen Haushalte zu ermitteln,<sup>107</sup> welche somit das bereits ausgeschöpfte Potential darstellen.

In unserem Beispiel betrachten wir aus der Datenbank *Anschlussverwaltung* drei Relationen:

- *Anschl\_Technik*: Hier finden sich technische Daten, wobei eine eindeutige *Ansch\_ID* den Schlüssel darstellt.

<sup>106</sup> Die in diesem Fallbeispiel benutzten Daten wie demographische- oder technische Daten sind vom Autor künstlich generiert, und entsprechen nicht den realen Gegebenheiten.

<sup>107</sup> Wie oben beschrieben setzen wir einen Stromanschluss für einen Haushalt voraus

- *Anschl\_Stand*: Hier finden sich Daten zu Wartung, Ablesezeitpunkt und –stand. Auch hier findet sich das Schlüsselfeld *Ansch\_ID*
- *Anschl\_Adresse*: Hier sind die für das Beispiel interessanten Adressdaten gespeichert. Diese helfen zur Identifikation, in welcher Gebietseinheit sich der Zähler befindet. Diese Relationen lassen sich über das gemeinsame Schlüsselfeld *Ansch\_ID* miteinander verknüpfen.

- Externe **demographische** Daten

Diese Daten werden von externen Datenlieferanten zugekauft, um sie dann mit dem Unternehmensdaten zu verknüpfen. Für unsere Analyse werden die Daten benötigt, um das Gesamtpotential zu ermitteln, dem das bereits ausgeschöpfte Potential gegenüber gestellt wird.

Abb. 23: Tabellenstruktur der externen Demographiedaten (Auszug)

The screenshot shows the SQL Server Enterprise Manager interface. The table 'demogr' is selected, and its columns are listed in the left pane. The main pane shows the table structure with the following columns: STR\_ID, GKZ, PLZ, ORT, STR, HAUSNR\_V, HAUSNR\_B, EINWOHNER, HAUSHALTE, LAENGE, BREITE, X, and Y. The data is displayed in a grid format.

STR_ID	GKZ	PLZ	ORT	STR	HAUSNR_V	HAUSNR_B	EINWOHNER	HAUSHALTE	LAENGE	BREITE	X	Y
1	08212000	76131	Karlsruhe	Adenauerring	0	0	39	21	30293	176481	8,414635	49,022504
2	08212000	76131	Karlsruhe	Ahaweg	0	0	29	14	30252	176469	8,403273	49,019114
3	08212000	76131	Karlsruhe	Albert-Nestler-Str.	0	0	0	0	30391	176472	8,442038	49,020054
4	08212000	76131	Karlsruhe	Am Fasanengarten	0	0	8	3	30318	176452	8,421556	49,014373
5	08212000	76131	Karlsruhe	Am Künstlerhaus	0	0	170	80	30296	176433	8,415435	49,009232
6	08212000	76131	Karlsruhe	Am Sportpark	0	0	6	2	30406	176500	8,446029	49,027844
7	08212000	76131	Karlsruhe	An der Fasanengarten	0	0	11	5	30272	176469	8,408954	49,019204
8	08212000	76131	Karlsruhe	August-Schwall-Str.	0	0	108	70	30326	176426	8,423996	49,007202
9	08212000	76131	Karlsruhe	Berckmüllerstr.	0	0	479	308	30367	176443	8,435407	49,011883
10	08212000	76131	Karlsruhe	Bernhardstr.	0	0	259	168	30318	176435	8,421646	49,009602
11	08212000	76131	Karlsruhe	Brettener Str.	0	0	93	46	30401	176468	8,444649	49,018873
12	08212000	76131	Karlsruhe	Brohrainstr.	0	0	61	37	30365	176447	8,434607	49,012953
13	08212000	76131	Karlsruhe	Brunnenstr.	0	0	15	6	30300	176435	8,416665	49,009762
14	08212000	76131	Karlsruhe	Buntestr.	0	0	221	153	30324	176428	8,423316	49,007742

In unserem Beispiel halten wir in einer externen Datenbank u.a. eine Relation *demogr*, welche allgemeine demographische Daten (hier Einwohnerzahlen und Haushalte auf Straßenebene) hält. In unserem Beispiel sind die Daten straßenscharf aufgesplittet, d.h. die Strassen-ID bildet den niedrigsten Geoschlüssel. In unserem Beispiel soll es nicht nur

möglich sein, die Daten auf Gemeindeebene zu betrachten, sondern auch auf Stadtteilebene. Hierzu wurden beispielhaft **Mikrogebiete**<sup>108</sup> definiert.

Um selbstdefinierte Mikrogebiete betrachten zu können, benötigt man aber eine eindeutige Zuordnung, welche Strassen welchen Mikrogebieten zugeordnet sind. Dies kann in Form einer Pärchentabelle organisiert sein, welche bereits im Data-Warehouse vorhanden ist:

Abb. 24: Tabellenstruktur der Pärchentabelle „Mikrogebiet vs. Strassen“ (Auszug)

STR_ID	GKZ	MIKRO	STR	HAUSNR_V	HAUSNR_B
1	08212000	0018	Adenauerring	0	0
2	08212000	0022	Ahaweg	0	0
3	08212000	0018	Albert-Nestler-Str.	0	0
4	08212000	0025	Am Fasanengarten	0	0
5	08212000	0023	Am Künstlerhaus	0	0
6	08212000	0018	Am Sportpark	0	0
7	08212000	0023	An der Fasanengar	0	0
8	08212000	0025	August-Schwall-Str.	0	0
9	08212000	0025	Berckmüllerstr.	0	0
10	08212000	0025	Bernhardstr.	0	0
11	08212000	0024	Brettener Str.	0	0
12	08212000	0025	Brohrainstr.	0	0

Mit Hilfe dieser Datei lassen sich die Strassen aus der Demographie-Datenbank den vordefinierten Mikrogebieten zuordnen.

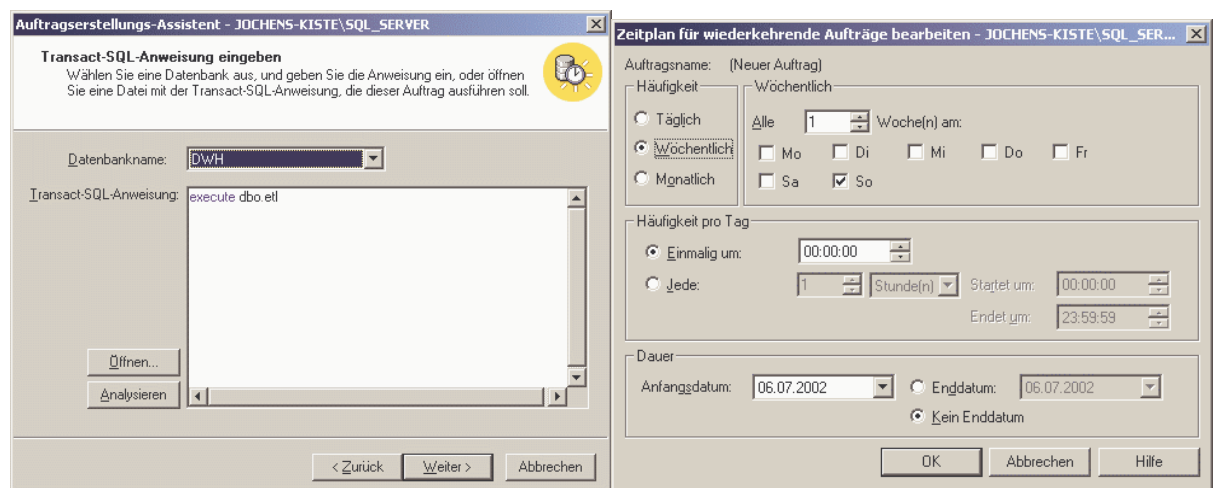
Das Laden der operativen Daten erfolgt über eine *stored procedure*. Dies sind Prozeduren, die SQL-Code enthalten, und die direkt in der Datenbank abgespeichert werden, also deren Bestandteil sind.

```
CREATE PROCEDURE dbo.etl AS
(SELECT gkz,stadtteil INTO stagarea_anschluesse
FROM anschlussverwaltung.dbo.anschl_adress)
(SELECT d.gkz,m.mikro,d.str,d.einwohner,d.haushalte INTO stagarea_demographie
FROM dwh.dbo.mikro AS m,demographiedaten.dbo.demogr AS d
WHERE d.str_id=m.str_id)
GO
```

<sup>108</sup> Die in diesem Fallbeispiel benutzen Mikrogebiete wurden vom Autor frei definiert und entsprechen nicht real-existierenden statistischen Einheiten.

Mit dieser Prozedur werden die Daten aus den operativen Systemen in das Data-Warehouse geladen. In der ersten *SELECT*-Anweisung werden die Daten der Anschlüsse in das Data-Warehouse übernommen, wobei sich die Tabelle *stagarea\_anschluesse* konzeptionell in der Staging Area des Data-Warehouse befindet. Es werden nur die Attribute übernommen, die für eine Bestimmung der Anzahl der Haushaltsanschlüsse pro Gebietseinheit nötig sind. In der zweiten *SELECT*-Anweisung werden die Daten aus der externen Demographiedatenbank geladen und sogleich mit der Pärchentabelle *mikro* verknüpft, um gleich im ETL-Prozess die Zuordnung zu den Mikrogebieten zu erhalten, die in der originalen Demographie-Datenbank nicht enthalten war. Auch hier befindet sich die Zieltabelle *stagarea\_demographie* in der Staging Area des Data-Warehouses. Eine *stored procedure* kann zu bestimmten Zeiten automatisch vom System gestartet werden. Damit ist es möglich einen ETL-Prozess nach bestimmten Kriterien automatisch zu starten.

Abb. 25: Auftragsstellungs-Assistent (SQL-Server 2000)



## 6.4 OLAP-Prozesse

Sind die Daten nun extrahiert und im Data-Warehouse gespeichert, werden diese in der **OLAP-Schicht** des Data-Warehouses nun nach den Anforderungen der Analyse aufbereitet.

In unserem Beispiel benötigen wir aus den Daten nun folgende Informationen:

- Wie viel Haushalte (=Stromanschlüsse) gibt es in einer bestimmten Gebietseinheit?
- Wie viel Haushalte (=Stromanschlüsse) in dieser Gebietseinheit werden bereits von uns versorgt?
- Wie hoch ist die Penetrations-Kennziffer in dieser Gebietseinheit?

Als Gebietseinheiten sollen **bundesdeutsche Gemeinden** und selbstdefinierte **Mikrogebiete** in ausgewählten Großstädten herangezogen werden.

Hierzu werden im Data-Warehouse **Sichten** (Views) erstellt, die es erlauben, die geladenen Daten für bestimmte Zwecke aufzubereiten und zu verändern, ohne dass neue physische Tabellen erstellt werden müssen. Bestimmte Attribute werden als Sicht in einer virtuellen Tabelle zusammengefasst und aufbereitet. In einem ersten Schritt sollen die Daten, die auf Straßenebene vorliegen auf die Gemeinde hochaggregiert werden. Dies geschieht anhand der Gemeindekennziffer (GKZ).

**VIEW *olap\_anschluesse\_gkz*:**

```
SELECT gkz, COUNT(gkz) AS n FROM etl_anschluesse GROUP BY gkz
```

**VIEW *olap\_demogr\_gkz*:**

```
SELECT gkz, SUM(einwohner) AS n_einw, SUM(haushalte) AS n_hh  
FROM etl_demographie GROUP BY gkz
```

Mit der ersten *Select*-Anweisung wird aus der Tabelle mit den Stromanschlüssen eine neue Sicht namens *olap\_anschluesse\_gkz* nach der Gemeindekennziffer **aggregiert**. Dies bedeutet, dass jede GKZ dabei nur noch einmal auftritt. Zusätzlich ist die Anzahl der Datensätze der ursprünglichen Tabelle eingefügt, die diese GKZ besitzen. Diese Anzahl entspricht der Anzahl der Stromanschlüsse dieser Gebietseinheit, hier der Gemeinde.

Abb. 26: Ergebnistabelle der Sicht *olap\_anschluesse\_gkz*

gkz	n
08211000	5231
08215100	235
08214100	356
08213201	112
08221000	9568
08221100	554
08221200	657
08222300	1257
08225200	1235
08231000	10253
08232100	1223
08232800	2481
08232900	1298
08532100	6587
08532300	6598
08532500	2136
08212000	11250
08225300	5268
08532700	1268
08532900	2578

Auch die zweite *Select*-Anweisung aggregiert eine neue Sicht *olap\_demogr\_gkz* aus der Tabelle der Demographiedaten nach der GKZ. Hier werden aber die Anzahl der Einwohner und der Haushalte mit aggregiert. Somit wurde ermittelt, wie viel Einwohner und Haushalte eine Gemeinde im Aggregat besitzt.

Abb. 27: Ergebnistabelle der Sicht *olap\_demogr\_gkz*

	gkz	n_einw	n_hh
▶	08215007	26708	15215
	08215009	41124	21352
	08215017	38486	20532
	08215025	3096	1652
	08215039	5416	2954
	08215040	2413	1965
	08215064	12424	6212
	08215090	8997	4652
	08215094	1578	895
	08215096	15981	8965
	08215097	14589	7862
	08212000	276571	150074
	08215021	7282	3825
	08215029	4950	2652
	08215046	13417	7869
	08215047	5403	2896
	08215059	10017	6012
	08215066	12370	7210
	08215082	4425	2354
	08215084	12192	7025
	08215089	8720	5012

Sollen die Daten nicht auf Gemeindeebene, sondern nur auf Mikrogebiets Ebene aggregiert werden, werden andere Sichten herangezogen, welche die selben Daten verarbeiten, aber einen anderen Blickwinkel auf die Daten erlauben:

***VIEW olap\_anschluesse\_mikro:***

```
SELECT gkz,stadtteil,COUNT(stadtteil) AS n FROM etl_anschluesse  
GROUP BY gkz,stadtteil
```

***VIEW olap\_demogr\_mikro:***

```
SELECT gkz,mikro,SUM(einwohner) AS n_einw,SUM(haushalte) AS n_hh  
FROM etl_demographie GROUP BY gkz,mikro
```

Auch diese Sichten aggregieren die Daten auf eine höhere Gebietsebene, hier auf die Ebene der Mikrogebiete.

Abb. 28: Ergebnistabelle der Sicht *olap\_anschluesse\_mikro*

gkz	stadtteil	n
08212000	0018	468
08212000	0011	434
08212000	0004	148
08212000	0006	229
08212000	0031	577
08212000	0009	213
08212000	0002	481
08212000	0007	173
08212000	0025	474
08212000	0019	328
08212000	0012	216
08212000	0015	278
08212000	0028	378
08212000	0021	799
08212000	0023	296
08212000	0013	438
08212000	0020	521
08212000	0024	267
08212000	0026	434
08212000	0014	234
08212000	0027	281
08212000	0016	216
08212000	0029	225
08212000	0022	367
08212000	0010	107
08212000	0017	226
08212000	0030	385
08212000	0005	488
08212000	0008	425
08212000	0001	739
08212000	0003	405

Abb. 29: Ergebnistabelle der Sicht *olap\_demogr\_mikro*

gkz	mikro	n_einw	n_bh
08212000	0018	13722	7451
08212000	0011	10693	5525
08212000	0004	1328	582
08212000	0006	4439	2127
08212000	0031	21670	12285
08212000	0009	3625	1680
08212000	0002	9510	4890
08212000	0007	6838	4023
08212000	0025	15365	9051
08212000	0019	6155	3059
08212000	0012	3322	1867
08212000	0015	3338	1711
08212000	0028	16382	9417
08212000	0021	18333	9850
08212000	0023	6247	3521
08212000	0013	6140	3061
08212000	0020	11052	5869
08212000	0024	7277	3705
08212000	0026	11861	6582
08212000	0014	2842	1409
08212000	0027	13732	7647
08212000	0016	3246	1808
08212000	0029	4922	2797
08212000	0022	9532	5361
08212000	0010	3357	1820
08212000	0017	2121	1153
08212000	0030	13567	8213
08212000	0005	11750	6173
08212000	0008	6935	3329
08212000	0001	14391	7376
08212000	0003	12814	6704

Um in einem letzten Schritt die Penetrationskennziffer für jede Gebietseinheit zu ermitteln, wird die Relation mit den aggregierten Anschlussdaten und die Relation mit den aggregierten Demographiedaten mittels der GKZ bzw. der Kennziffer des Mikrogebietes verknüpft.

**VIEW olap\_penetration\_gkz:**

*SELECT d.gkz, (a.n/d.n\_hh)\*100 AS pen FROM olap\_demogr\_gkz AS d,  
olap\_anschluesse\_gkz AS a WHERE d.gkz=a.gkz*

**VIEW olap\_penetration\_mikro:**

*SELECT d.gkz,d.mikro,(a.n/d.n\_hh)\*100 AS pen FROM olap\_demogr\_mikro AS d,  
olap\_anschluesse\_mikro AS a WHERE d.mikro=a.stadtteil*

Mit der Formel  $(a.n/d.n\_hh)*100$  AS pen wird der Quotient aus der Anzahl der Stromanschlüsse und der Anzahl der Haushalte in ein neues Attribut pen geschrieben, welches die Penetrationsziffer als Prozentwert darstellt. Diese Sicht wird nun als virtueller Extrakt im Data-Warehouse gehalten und steht somit zur Visualisierung via GIS-Komponente zur Verfügung.

Abb. 30: Ergebnistabelle der Sicht olap\_penetration\_mikro

gkz	mikro	pen
08212000	0018	6,28103610253657
08212000	0011	7,8552036199095
08212000	0004	25,4295532646048
08212000	0006	10,766337564645
08212000	0031	4,6967846967847
08212000	0009	12,6785714285714
08212000	0002	9,83640081799591
08212000	0007	4,30027342779021
08212000	0025	5,23699038780245
08212000	0019	10,7224583197123
08212000	0012	11,569362613819
08212000	0015	16,2478082992402
08212000	0028	4,01401720293087
08212000	0021	8,11167512690355
08212000	0023	8,40670264129509
08212000	0013	14,3090493302842
08212000	0020	8,87715113307207
08212000	0024	7,2064773279352
08212000	0026	6,59374050440596
08212000	0014	16,6075230660043
08212000	0027	3,67464365110501
08212000	0016	11,9469026548673
08212000	0029	8,04433321415803
08212000	0022	6,84573773549711
08212000	0010	5,87912087912088
08212000	0017	19,6010407632264
08212000	0030	4,68769024716912
08212000	0005	7,90539445974405
08212000	0008	12,7665965755482
08212000	0001	10,0189804772234
08212000	0003	6,04116945107395

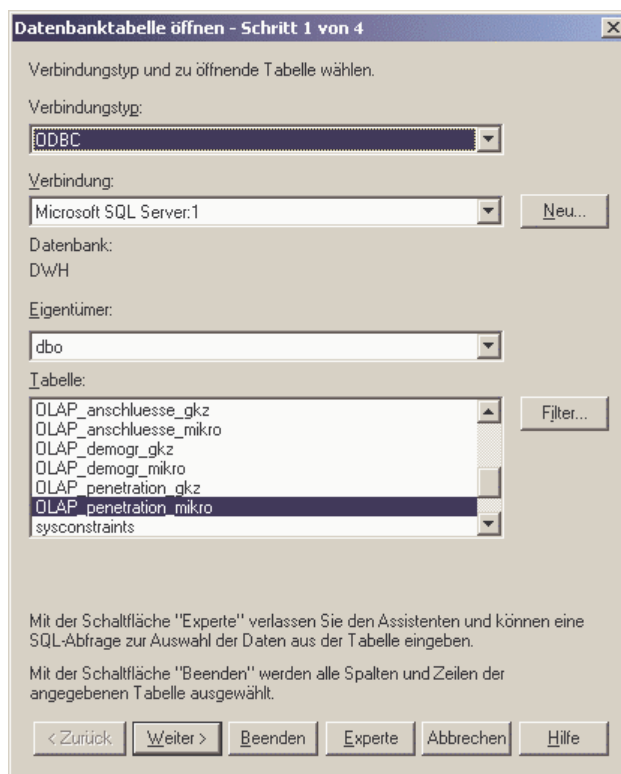
## 6.5 GIS-Integration

Um die Daten in einem GIS zu visualisieren und zur weiteren Bearbeitung verfügbar zu machen, muss das GIS auf die Daten zugreifen können. Wie oben gezeigt, sind diese Daten für die jeweiligen Analyseebenen in Sichten organisiert. Der Nutzer soll nun eine Sicht aus dem Data-Warehouse mit einem GIS visualisieren und bearbeiten können.

### 6.5.1 Einrichten der Datenquelle

Um eine Tabelle oder Sicht eines DBMS in einem externen Programm zu öffnen, muss dafür eine ODBC-Quelle, der sogenannte *Data Source Name* eingerichtet werden.<sup>109 110</sup>

Abb. 31: Einrichten der ODBC-Datenquelle (MapInfo professional 5.5)



Damit ist es möglich ein Datenbankobjekt in einem GIS einzubinden. Dieser Vorgang darf nicht mit einem Import verglichen werden: Die Daten bleiben im DBMS vorhanden und sind von dort aus editierbar. Es wird im GIS lediglich eine Datei *olap\_penetration\_mikro.TAB* erstellt, welcher sich auf den erstellten *Data Source Name* bezieht. Dort wiederum ist der Verweis auf die Sicht *olap\_penetration\_mikro* enthalten.

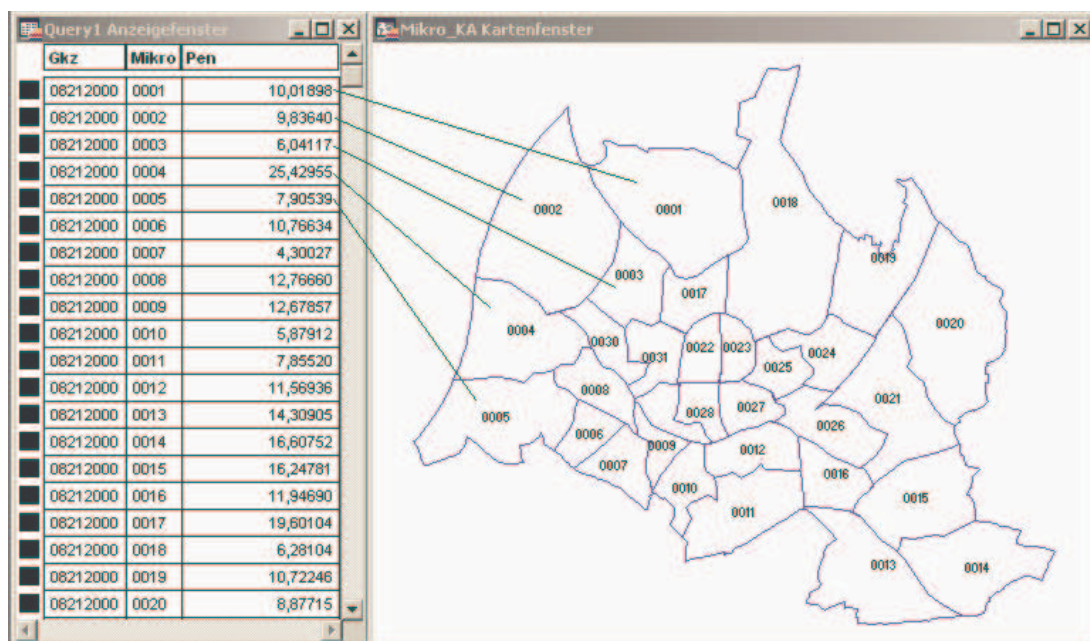
### 6.5.2 Verknüpfung der Attributdaten mit den Geometrien

Bei einer georationalen Speicherung werden die Daten in einem DBMS gehalten (in unserem Beispiel in den erstellten Sichten), während die Geometrien bereits in einem GIS existieren. Diese beiden Relationen müssen nun miteinander verknüpft werden, damit den Attributdaten die entsprechende Geometrie zugeordnet werden kann. Dazu benötigt man ein gemeinsames Feld in den beiden Relationen, den sogenannten **Geoschlüssel**. In unserem Beispiel ist dies die Gemeindekennziffer bzw. die Kennziffer für das Mikrogebiet.

<sup>109</sup> Siehe Kapitel 4.5

<sup>110</sup> In diesem Beispiel wurde dem „alten“ ODBC dem Vorzug gegenüber dem aktuellen OLE DB gegenüber, da das verwendete Desktop-GIS „MapInfo professional 5.5“ keine OLE DB-Unterstützung bietet.

Abb. 32: Verknüpfung der Sachdaten-Relation mit der Geometrie-Relation



Damit der Nutzer diese Verknüpfung nicht bei jeder Analyse neu aufbauen muss, kann diese in einem Makro (hier MapBasic) hinterlegt werden, die diese Verknüpfung automatisch aufbaut:

```

OPEN TABLE "olap_penetration_mikro.TAB" INTERACTIVE
OPEN TABLE "geom_mikro.TAB" INTERACTIVE
SELECT * FROM olap_penetration_mikro, geom_mikro
  WHERE olap_penetration_mikro.Mikro= geom_mikro.Mikro INTO sql_mikro
MAP FROM sql_mikro

```

Mit den ersten beiden Befehlen werden die beiden Relationen geöffnet.

*olap\_penetration\_mikro.tab* ist die MapInfo-Datei, die auf die ODBC-Quelle zugreift; *geom\_mikro.tab* enthält die reine Geometrie. Mit der SQL-Anweisung werden die beiden Relationen über das gemeinsame Attribut *Mikro* miteinander verknüpft und mittels der *map from*-Anweisung als Karte visualisiert.

Damit stehen die Data-Warehouse-Daten nun in einem GIS zur Verfügung und sind räumlichen Einheiten zugeordnet. Hier können sie nun weiterverarbeitet oder lediglich visualisiert werden, wenn das GIS „nur“ als standardisiertes Reporting-Tool genutzt werden soll. In diesem Falle könnte im obigen Code noch eine Anweisung zur Erstellung einer thematischen Karte hinterlegt sein.

Abb. 33: Beispiel für Penetrationsanalyse für GKZ-Gebiete

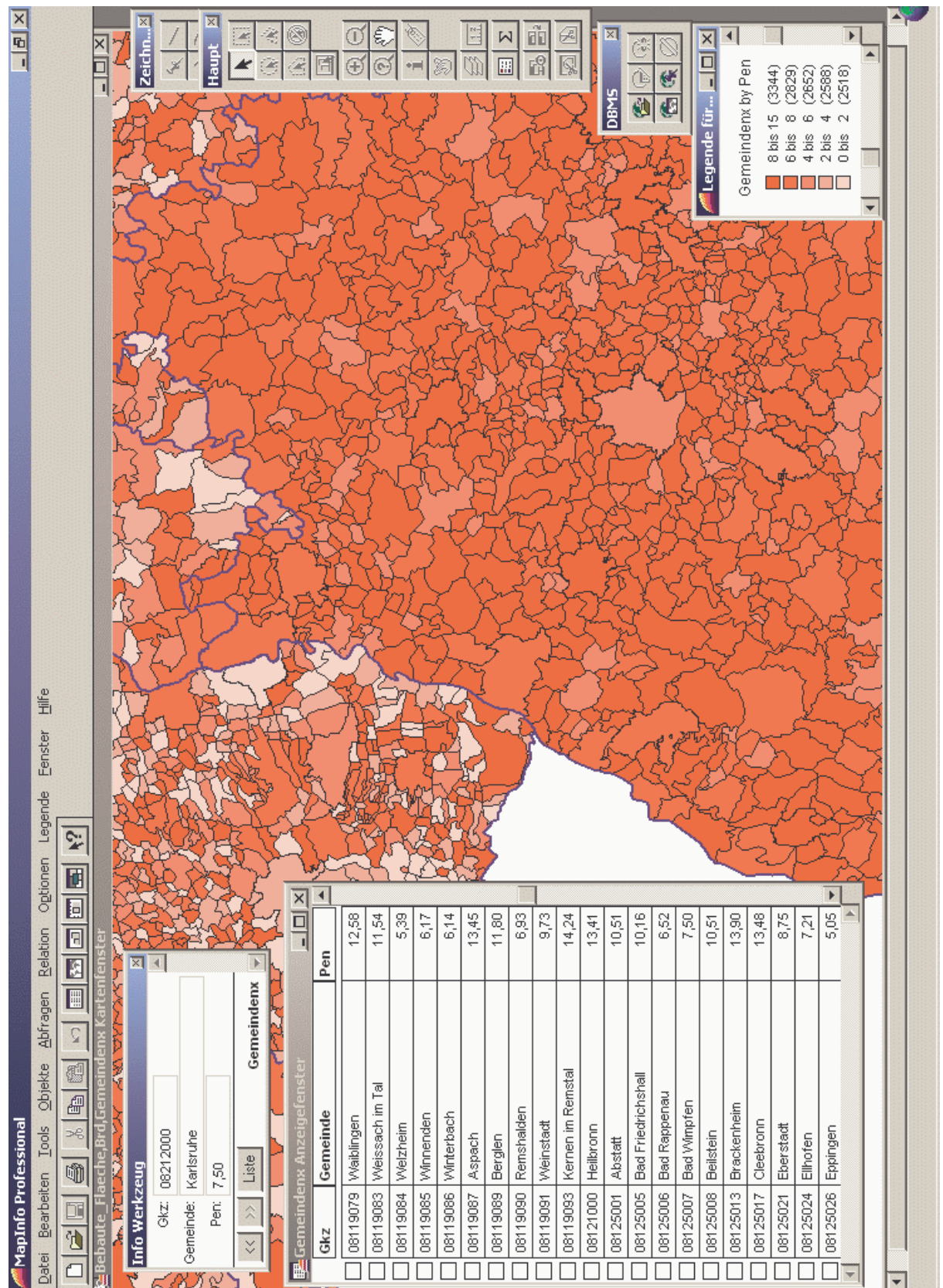


Abb. 34: Beispiel für Penetrationsanalyse für Mikro-Gebiete

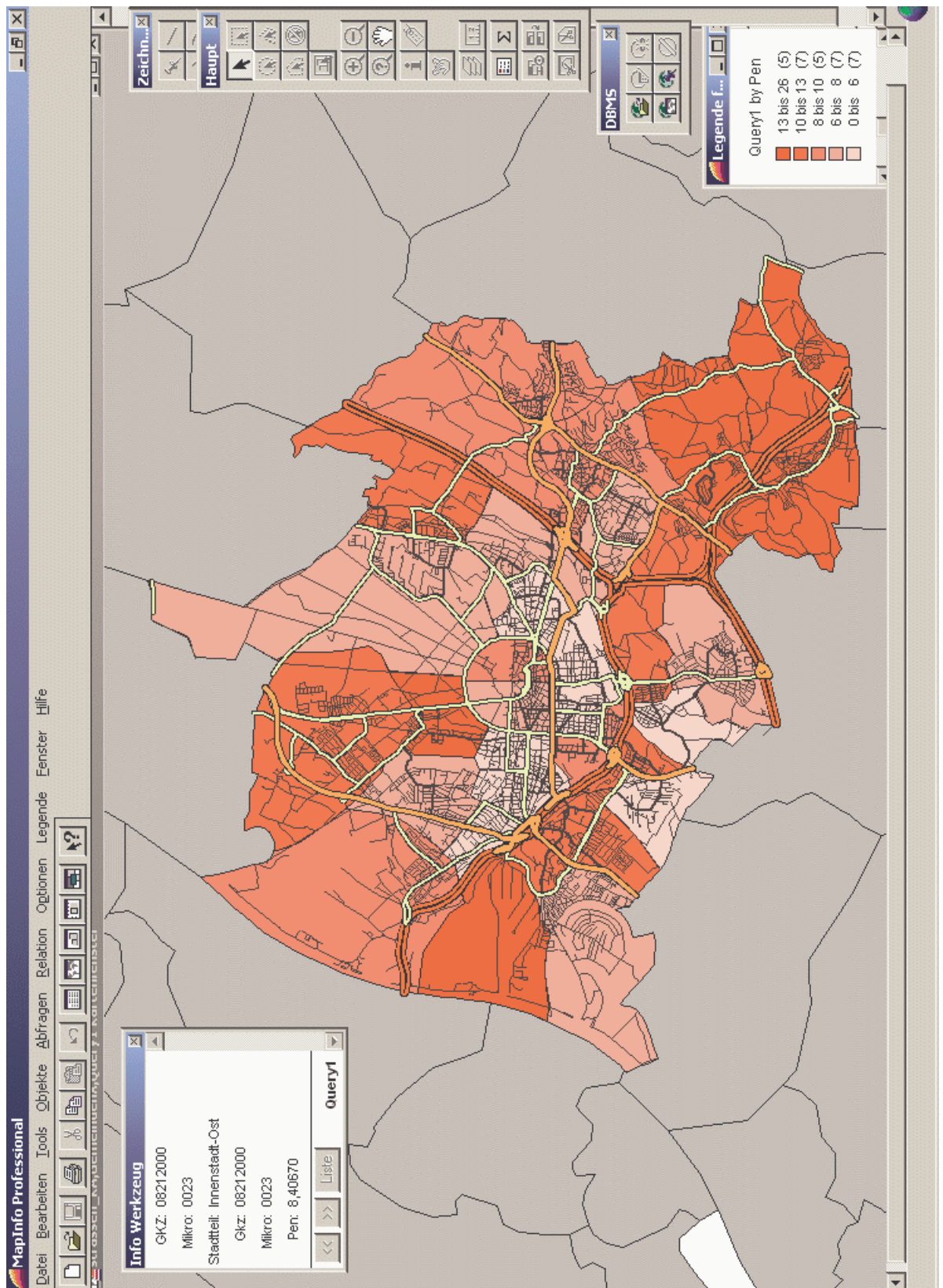
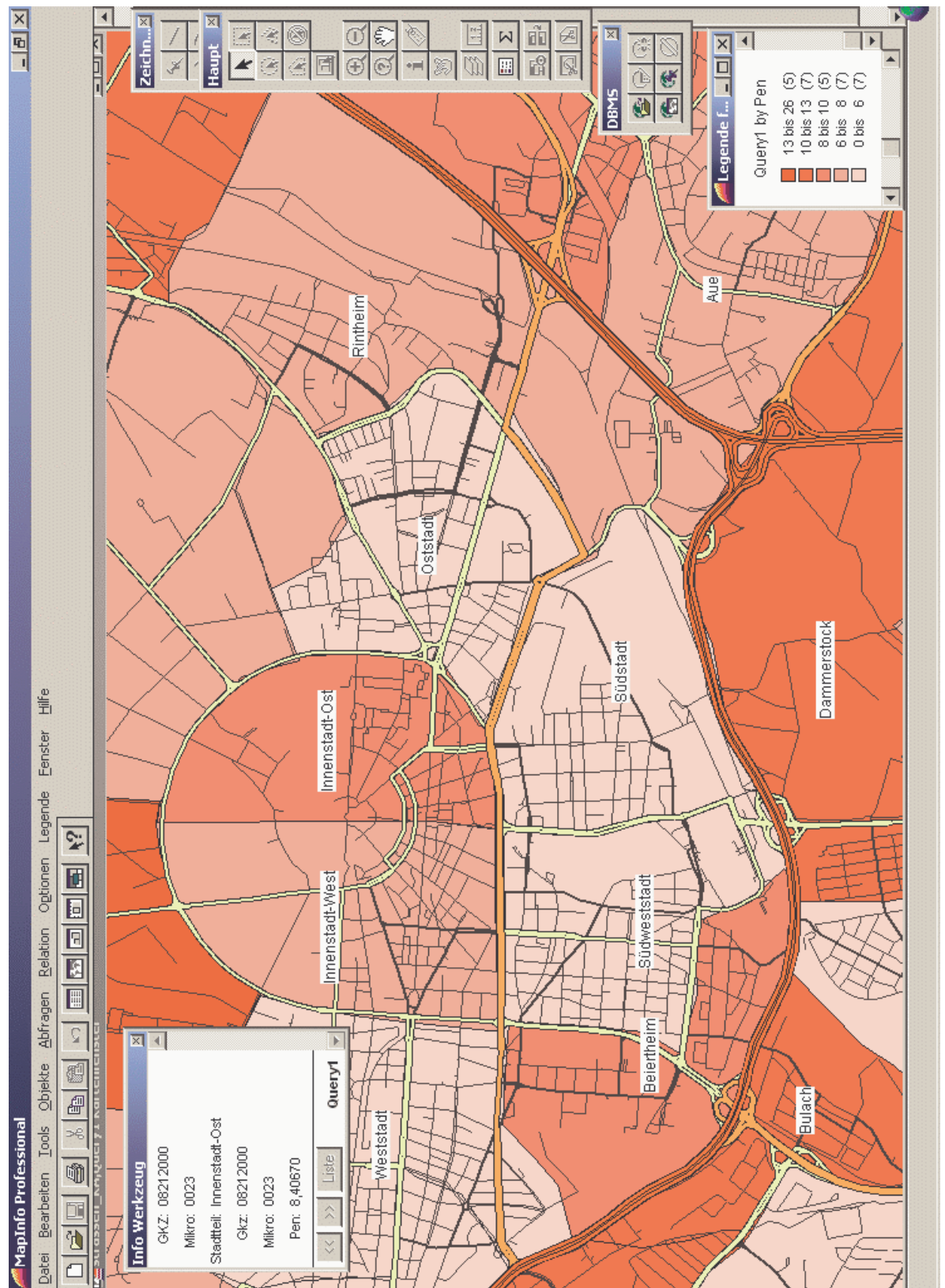


Abb. 35: Beispiel für Penetrationsanalyse für Mikro-Gebiete (Auszug)



## Literaturverzeichnis

- Allgayer, Florian  
Energieversorger: Fußball statt Farbenlehre, in:  
Media & Marketing, Heft 1-2/2002, München (2002)
- Alpar, Paul  
Niedereichholz, Joachim  
(Hrsg.)  
Einführung zu Data-Mining, in:  
Data Mining im praktischen Einsatz, Braunschweig,  
Wiesbaden (2000)
- Bahrenberg, Gerhard  
Giese, Ernst  
Nipper, Josef  
Statistische Methoden in der Geographie  
Bd. 1: Univariate und bivariate Statistik,  
Stuttgart, Leipzig (1999)
- Bill, Ralf  
Zehner, Marco L.  
Lexikon der Geoinformatik  
Heidelberg (2001)
- Bissantz, Nicolas  
Hagedorn, Jürgen  
Mertens, Peter  
Data-Mining , in: Mucksch H./ Behme, W.:  
Das Data Warehouse- Konzept, Wiesbaden (1998)
- Breitner, Christoph  
Herzog, Uwe  
Mülle, Jutta  
Schlösser, Jörg (Hrsg.)  
Data Warehousing, Karlsruhe (1996)
- Chamoni, Peter  
Gluchowski, Peter  
On-Line Analytical Processing (OLAP), in:  
Mucksch, H. / Behme, W.: Das Data Warehouse-  
Konzept, Wiesbaden (1998)
- Christmann, Alfred  
Asselborn Herbert  
Fuchs, Toni  
Walgenbach, Alfred  
Strategisches Informationssystem SIS – Die Data  
Warehouse-Lösung der Stadt Köln, in:  
Mucksch, H. / Behme, W.: Das Data Warehouse-  
Konzept, Wiesbaden (1998)

- Czeranka, Marion Business Geographics und Geomarketing als Schlüssel zur unternehmenseigenen Schatztruhe, in: Fally, M. / Strobl, J.: Business Geographics, Heidelberg (2000)
- Czeranka, Marion Seminarunterlagen zum ZGIS-Seminar: „GeoMarketing“ vom 07.07.2000
- Degen, Reinhard Der skalierbare Datamart, in: Martin, W. (Hrsg.): Data Warehousing, Bonn (1998)
- Fitzke, Jens GIS-Pool – eine Managementkomponente für  
Greve, Klaus Geodaten in komplexen Informationen, in:  
Lessing, Rolf Strobl, J / Blaschke, T. / Griesebner, G. (Hrsg.):  
Müller, Markus Angewandte geographische Informations-  
verarbeitung XII, Heidelberg (2000)
- Fleischmann, Robert Oracle Spatial Option – Theorie und Praxis, in:  
Kaiser, Konrad Fally, M. / Strobl, J.: Business Geographics, Heidelberg  
(2000)
- Hippner, Hajo Der Prozess des Data-Mining im Marketing, in:  
Wilde, Klaus D. Hippner, H. / Küsters, U. / Meyer, M. / Wilde, K.:  
Handbuch Data-Mining im Marketing,  
Braunschweig, Wiesbaden (2001)
- Höhn, Reinhard Der Data-Warehouse-Spezialist – Entwurf, Methoden  
und Umsetzung eines Data Warehouse, München (2000)
- Holthuis, Jan Der Aufbau von Data-Warehouse-Systemen –  
Konzeption, Datenmodellierung, Vorgehen,  
Wiesbaden (1998)
- Hummeltenberg, Wilhelm Data Warehousing: Management des Produktionsfaktor  
Information, in: Martin, W. (Hrsg.): Data Warehousing,  
Bonn (1998)

- Joos, Gerhard  
Zur Qualität von objektstrukturierten Geodaten,  
Schriftenreihe im Studiengang Geodäsie und  
Geoinformation der Universität der Bundeswehr  
München, Heft 66, Neubiberg (2000)
- Kapp, Josef  
Vierheilig, Norbert  
Die richtige IT-Strategie für den Wettbewerb im  
Energiemarkt, in: ew Fachthema, Jg. 100, Heft 20-21  
Berlin (2001)
- Kastin, Klaus S.  
Marktforschung mit einfachen Mitteln,  
München (1995)
- Keller-Giessbach, Dieter  
Marketing und Vertrieb mit GIS – Geomarketing als  
Wettbewerbsvorteil im liberalisierten Energiemarkt,  
in: Fally, M. / Strobl, J.: Business Geographics,  
Heidelberg (2000)
- Kiehn, Werner  
Kunden gesucht – Regionale Vertriebssteuerung anhand  
von Raumsegmentierung, in: GeoBit – Zeitschrift für  
raumbezogene Informations-technologie, 10/2001,  
Heidelberg (2001)
- Kirchner, Joachim  
Online Analytical Processing, in:  
Martin, W. (Hrsg.): Data Warehousing, Bonn (1998)
- Kuhnt, Dietmar  
Der europäische Energiemarkt zwischen Wettbewerb und  
Regulierung, in: VIK-Mitteilungen, Heft 6-2001, Essen  
(2001)
- Kurz, Andreas  
Data Warehousing – Enabling Technology  
Bonn (1999)
- Linder, Wilfried  
Geo-Informationssysteme  
Berlin, Heidelberg (1999)
- Lusti, Markus  
Data Warehousing und Data Mining,  
Berlin, Heidelberg (1999)

- Martin, Wolfgang Data Warehouse, Data Mining und OLAP: Von der Datenquelle zum Informationsverbraucher, in: Martin, W. (Hrsg.): Data Warehousing Bonn (1998)
- Martin, Wolfgang Der Data Mining-Prozeß, in: Martin, W. (Hrsg.): Data Warehousing Bonn (1998)
- Mentzl, Ronald  
Ludwig, Cornelia Das Data-Warehouse als Bestandteil eines Database Marketing-Systems, in: Mucksch, H. / Behme, W.: Das Data Warehouse-Konzept, Wiesbaden (1998)
- Meyer, Matthias Data-Mining im Marketing: Einordnung und Überblick, in: Hippner, H. / Küsters, U. / Meyer, M. / Wilde, K.: Handbuch Data-Mining im Marketing, Braunschweig, Wiesbaden (2001)
- Mittlboeck, Manfred Seminarunterlagen zum ZGIS-Seminar: „GeoDatabase in der Praxis“ vom 27.02.2002
- Monadjemi, Peter Datenbank-Programmierung mit Visual Basic 6, München (1999)
- Nakhaeizadeh, Gholamreza  
Reinartz, Thomas  
Wirth, Rüdiger Wissensentdeckung in Datenbanken und Data-Mining: Ein Überblick, in: Nakhaeizadeh, G.: Data Mining – Theoretische Aspekte und Anwendungen, Heidelberg (1998)
- Nallet, Pierre OLE DB, München (2001)
- Nitsche, Martin Micro-Marketing: Daten-Methoden-Praxis Wien (1998)



