

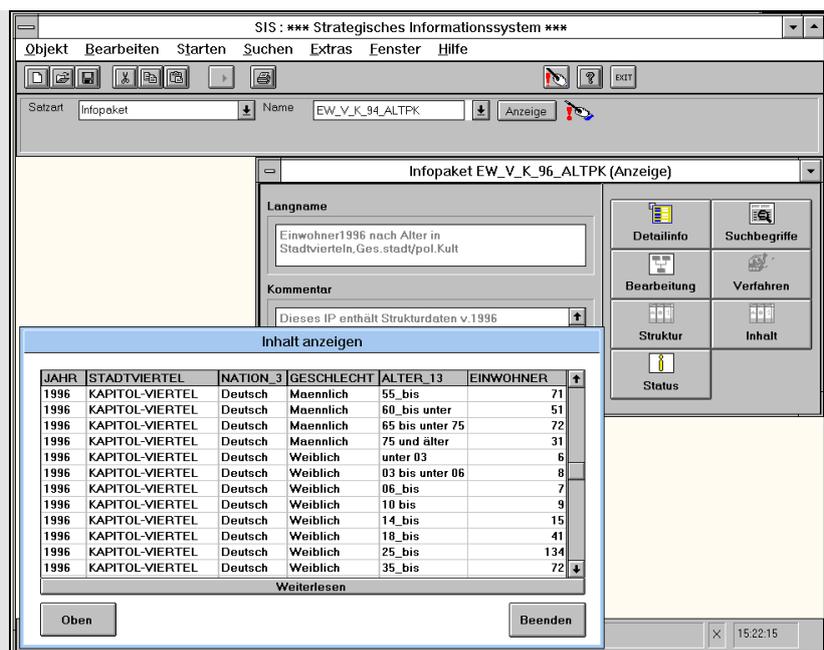
DAS RAUMBEZOGENE DATA WAREHOUSE

Dietmar Hermsdörfer, Alfred Walgenbach
Stadt Köln, Amt für Statistik, Einwohnerwesen und Europaangelegenheiten

Informationsmanagement mit dem Strategischen Informationssystem SIS

Das Amt für Statistik, Einwohnerwesen und Europaangelegenheiten der Stadt Köln hat in den letzten Jahren mit dem *Strategischen Informationssystem SIS* eine Data Warehouse-Lösung entwickelt, die mittlerweile nicht nur in der Stadtverwaltung Köln, sondern auch in anderen deutschen Großkommunen, in Landesverwaltungen und bei privaten Großunternehmen für ein leistungsfähiges Informationsmanagement eingesetzt wird. Dabei gewährleistet die Offenheit des SIS den Einsatz in ganz unterschiedlichen Bereichen der Unternehmen und der Verwaltungen.

Die Idee eines kommunalen Informationsmanagements wurde bei der Stadt Köln schon 1988 - also lange vor der Data Warehouse Diskussion - thematisiert und mündete 1990 in dem Auftrag der Verwaltungsführung ein Informationssystem unter besonderer Berücksichtigung der Belange der planenden Verwaltung, der Verwaltungsführung und der politischen Entscheidungsgremien zu entwickeln. Dieser Auftrag wurde mit der Konzeption des SIS schrittweise umgesetzt und von einer Entwicklungs- und Anwendergemeinschaft verschiedener Großstädte und statistischer Landesämter, unter Federführung der Stadt Köln, getragen.

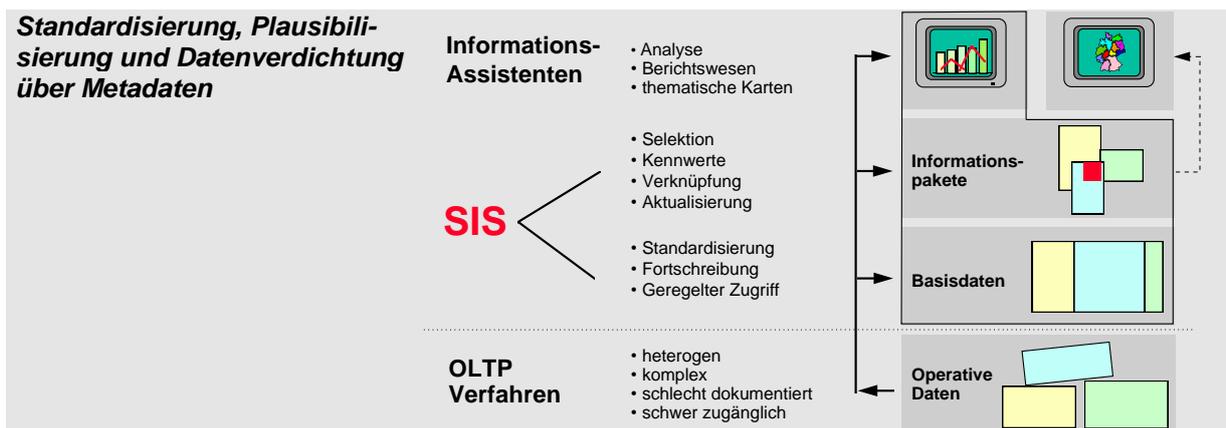


The screenshot shows the SIS interface with a window titled 'SIS : *** Strategisches Informationssystem ***'. The main window displays 'Infopakete' and a specific package 'EW_V_K_96_ALTPK'. A sub-window 'Inhalt anzeigen' shows a table of data for the year 1996, categorized by district, nationality, gender, and age group, with columns for 'JAHR', 'STADTVIERTEL', 'NATION 3', 'GESCHLECHT', 'ALTER_13', and 'EINWOHNER'.

JAHR	STADTVIERTEL	NATION 3	GESCHLECHT	ALTER_13	EINWOHNER
1996	KAPITOL-VIERTEL	Deutsch	Maennlich	55_bis	71
1996	KAPITOL-VIERTEL	Deutsch	Maennlich	60_bis unter	51
1996	KAPITOL-VIERTEL	Deutsch	Maennlich	65_bis unter 75	72
1996	KAPITOL-VIERTEL	Deutsch	Maennlich	75 und älter	31
1996	KAPITOL-VIERTEL	Deutsch	Weiblich	unter 03	6
1996	KAPITOL-VIERTEL	Deutsch	Weiblich	03_bis unter 06	8
1996	KAPITOL-VIERTEL	Deutsch	Weiblich	06_bis	7
1996	KAPITOL-VIERTEL	Deutsch	Weiblich	10_bis	9
1996	KAPITOL-VIERTEL	Deutsch	Weiblich	14_bis	15
1996	KAPITOL-VIERTEL	Deutsch	Weiblich	18_bis	41
1996	KAPITOL-VIERTEL	Deutsch	Weiblich	25_bis	134
1996	KAPITOL-VIERTEL	Deutsch	Weiblich	35_bis	72

Ziel des SIS ist es, Daten aus unterschiedlichen Quellen und mit unterschiedlichen Inhalten für strategische Planungen, sei es in Marketing und Vertrieb, in Controlling, Politikberatung oder für das Verwaltungsmanagement zeitgerecht und nachfrageorientiert nutzbar zu machen. Es sollen also Daten, die i.d.R. für ganz andere Zwecke erhoben wurden, durch Selektionen, Aggregationen, Transformationen und Verknüpfungen zu punktgenauen Informationen verdichtet werden. Dieser Prozess vollzieht sich in SIS hauptsächlich in den folgenden Schritten:

- Das SIS übernimmt Rohdaten aus den operativen Anwendungen (z. B. Abzüge aus automatisierten Verwaltungs- oder Meßverfahren, aber auch Daten aus Umfragen und Ergebnisse von Marktbeobachtungen) und vereinigt diese mit Daten aus anderen Quellen (z. B. der amtlichen Statistik oder aus externen Datenquellen), die keine einheitliche Strukturierung aufweisen, zu einem homogenen Datenbestand (Basisdaten).
- Vor allem aus Gründen der Verarbeitungseffizienz und Performance werden die teilweise sehr umfangreichen und periodisch fortgeschriebenen Basisdaten mittels Verdichtungsfunktionen in SIS zu kompakten multidimensionalen Informationspaketen aufbereitet.
- Diese Informationspakete gliedern die Datenbestände in eine Struktur aus sachlichen, räumlichen und zeitlichen Achsen. Der Anwender kann aus diesen Informationspaketen, für differenzierte Analysen, beliebige Datenkombinationen bilden. So selektiert er, je nach der individuellen Fragestellung, die gewünschten Informationen und schafft durch Transformationen, Selektionen und Aggregationen oder durch Verknüpfung mehrerer Informationspakete, flexibel neue Informationen.



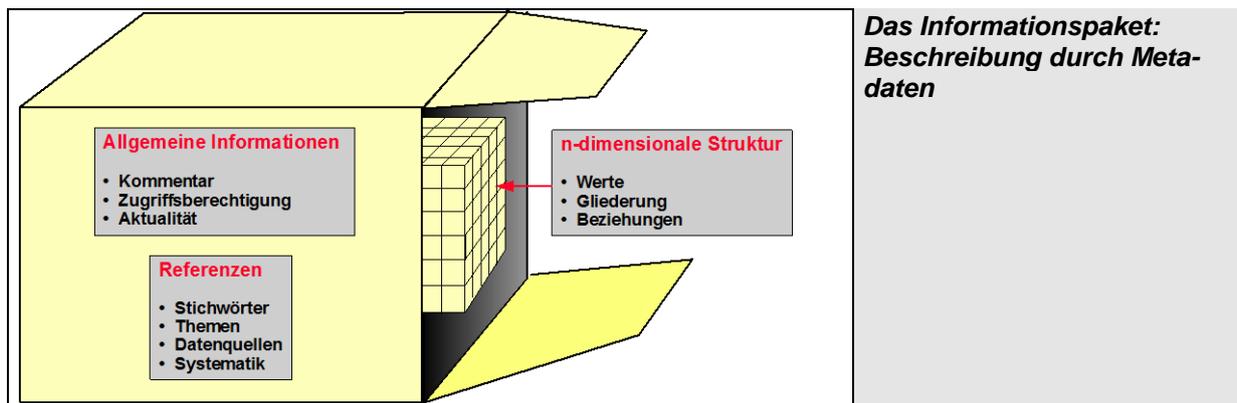
Den unterschiedlichen Datenstrukturen, sowohl der einzelfallbezogenen operativen Verfahren, als auch der tabellenorientierten analytischen Anwendungen, werden auf diese Weise mit dem multidimensionalen SIS gleichermaßen Rechnung getragen.

Genau so offen wie das Datenmodell ist auch das Funktionsspektrum des SIS. Selbst individuelle, komplexe Abfragen an die Datenbank werden ohne Programmieraufwand für den Anwender durchgeführt.

Grundlage dafür ist die Metadatensteuerung des Systems, die alle Daten standardisiert, strukturiert und beschreibt. Sie bildet ein Gerüst, in welches die Daten aus den externen Quellen geladen werden und bietet dem Nutzer vollkommene Transparenz seiner Datenschatze.

Die Metadaten erlauben die Suche nach allen Daten (Retrieval) über ein Thema, die Datenquellen, einzelne Merkmale oder über Ausprägungen. Da die Metadaten nicht nur beschreibenden, sondern auch strukturierenden Charakter haben, kann jeder Anwender frei durch das jeweilige Datenmodell navigieren und die benötigten Informationen, entsprechend seiner Fragestellung, ohne Programmieraufwand zusammenstellen.

Neben der Verknüpfungsmöglichkeit der Daten ist die Einbindung in heterogene Systemwelten mit vernetzten Großrechnern, UNIX-Servern, Workstations und PC's sowie Datenbanken unterschiedlicher Hersteller und Struktur - wie sie in Großkommunen oft vorzufinden sind - sowohl physikalisch als auch logisch realisiert und gleichzeitig die Offenheit für zukünftige Entwicklungen gewahrt.



Die Datenanalyse wird durch zahlreiche Schnittstellen zu den gewohnten Fachanwendungen und Geo-Informationssystemen (z.B. für Verkehr und Umwelt) unterstützt. Die Daten werden über Standardwerkzeuge für Statistik, Tabellenkalkulation, Geschäftsgraphiken oder Kartographie analysiert und transparent visualisiert.

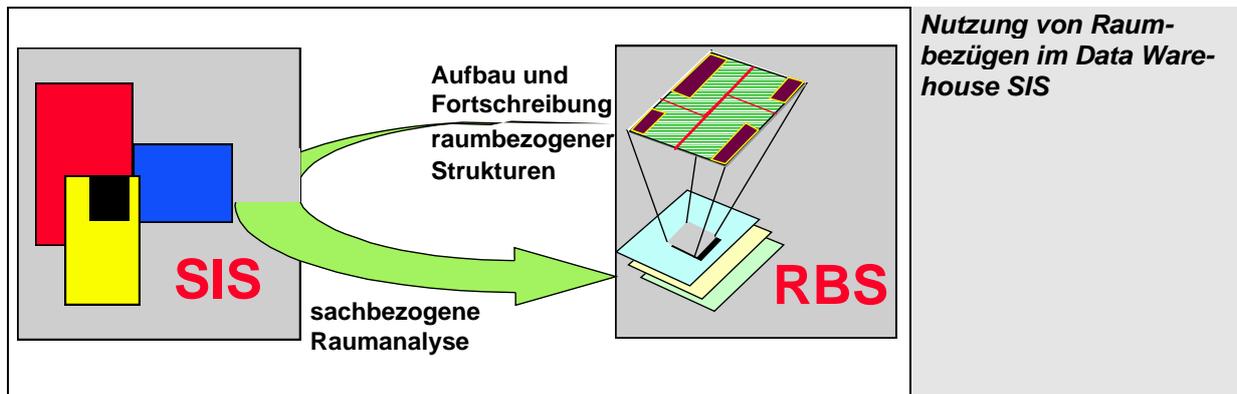
Der Raumbezug im Data Warehouse SIS

Seit Beginn der SIS-Entwicklung wird bei der Stadt Köln der Raumbezug als ein zentraler Bestandteil eines bereichsübergreifenden Informationsmanagements angesehen. Fast alle ökonomischen und administrativen Bereiche, aber auch die gesellschaftlichen Grundfunktionen Wohnen, Arbeiten, Einkaufen, Erholen, Bilden sind nämlich räumlich wirksam und führen zu einem sehr differenzierten und i.d.R. nicht zufälligen Verteilungsmuster innerhalb des Stadtgebietes. In vielen planenden Bereichen einer Stadtverwaltung (Stadt-, Schul-, Sozialplanung etc.) aber auch in strategischen Bereichen privater Unternehmen (Marketing, Vertriebssteuerung etc.) ist die Kenntnis solcher Verteilungen und räumlich wirksamer Prozesse von großem Interesse.

Um den Raumbezug für ein bereichsübergreifendes Informationsmanagement nutzbar zu machen wurde in den letzten Jahren das statistische Raumbezugssystem RBS aufgebaut. Ziel des RBS ist die flächendeckende räumliche Gliederung des Kölner Stadtgebietes und der Region in unterschiedlichen Differenzierungsgraden.

Dazu wurde als Basis die Kommunale Gebietsgliederung aufgebaut und fortgeschrieben. Diese strukturiert das Stadtgebiet in Stadtbezirke, Stadtteile und Stadtviertel, sowie kleinräumig in Blöcke, die wiederum weiter differenziert werden. Aber auch das komplette Straßennetz Köln und die ca. 141.000 Gebäudeadressen der Stadt sind im RBS verortet, attributiert und mit Beziehungen zu den anderen Objekten versehen. Mittlerweile steht so ein Geo-Datenbestand mit rund 250.000 Objekten zur Verfügung, der nicht nur für operative Verwaltungsaufgaben genutzt wird, sondern seine eigentlichen Qualitäten in der raumbezogenen Analyse unterschiedlicher Themenbereiche entwickelt.

Dies ist möglich, weil die Geo-Objekte mitsamt ihren Beziehungen für die Selektion und Aggregation von Sachdaten auf fast beliebige höherrangige räumliche Strukturen im Data Warehouse SIS genutzt werden können.



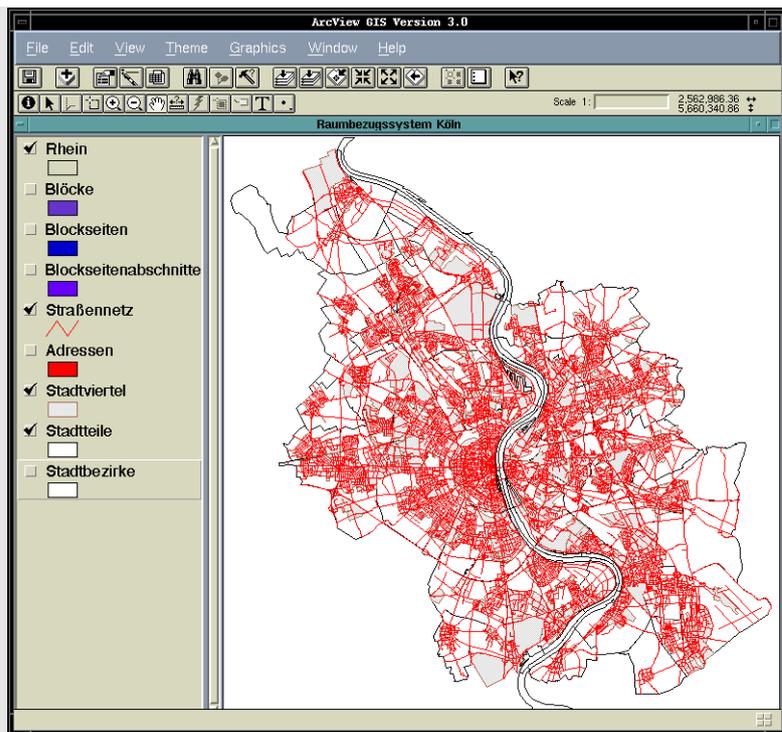
Systemtechnisch wird dieser wechselseitige Zugriff durch die Integration eines Geo-Informationssystems in das SIS realisiert. Mit dem RBS werden raumbezogene Strukturen für das SIS aufgebaut und fortgeschrieben, die über die Verbindung zu den SIS-Sachdaten zur sachbezogenen Raumanalyse genutzt werden können. Alle Änderungen an Geo-Objekten (Geometrie, Attribute, Beziehungen) werden ausschließlich im RBS vorgenommen und dann im SIS nachgezogen. So ist gewährleistet, daß beide Systemkomponenten immer auf einen konsistenten Datenbestand zugreifen.

GIS-Integration mit ESRI Produkten

Das statistische Raumbezugssystem RBS wurde zunächst auf der Basis des Geo-Informationssystems GRADIS-GIS der Fa. strässle entwickelt und aufgebaut. Durch den Konkurs der Fa. strässle wurde eine Migration des RBS auf eine neue GIS-Plattform notwendig. Die Stadt Köln hat in der Fa. ESRI einen neuen Entwicklungs- und Vertriebspartner gefunden, der die Data Warehouse Konzeption des SIS unterstützt und die Integration des Raumbezuges auf eine neue Qualitätsstufe stellen wird. Im wesentlichen waren folgende Gründe für die nun begonnene Zusammenarbeit mit ESRI ausschlaggebend:

- Ein transparentes Migrationskonzept in drei Stufen sichert die Produktion des RBS und der Weiterentwicklung zum raumbezogenen Data Warehouse.
- Die ESRI-Produkte lassen sich optimal in das Data Warehouse Konzept der Stadt Köln einbetten.
- Die vielfältige Kooperationsbeziehungen zu den ESRI-Anwendern Umweltamt, Amt für Straßen und Verkehrstechnik sowie Gas-, Elektrizitäts und Wasserwerke in Köln erleichtern ein bereichsübergreifendes Informationsmanagement.

ArcView mit RBS-Daten



Migration in drei Stufen

Die Weiterentwicklung der Geo-Integration soll durch verschiedene metadaten-gesteuerte Geo-Assistenten auf der Basis von SDE und SIS erfolgen. Die Migration wird dabei schrittweise vorangetrieben.

Stufe 1: Sicherstellung der Produktion mit ARC/INFO und ArcProjekt

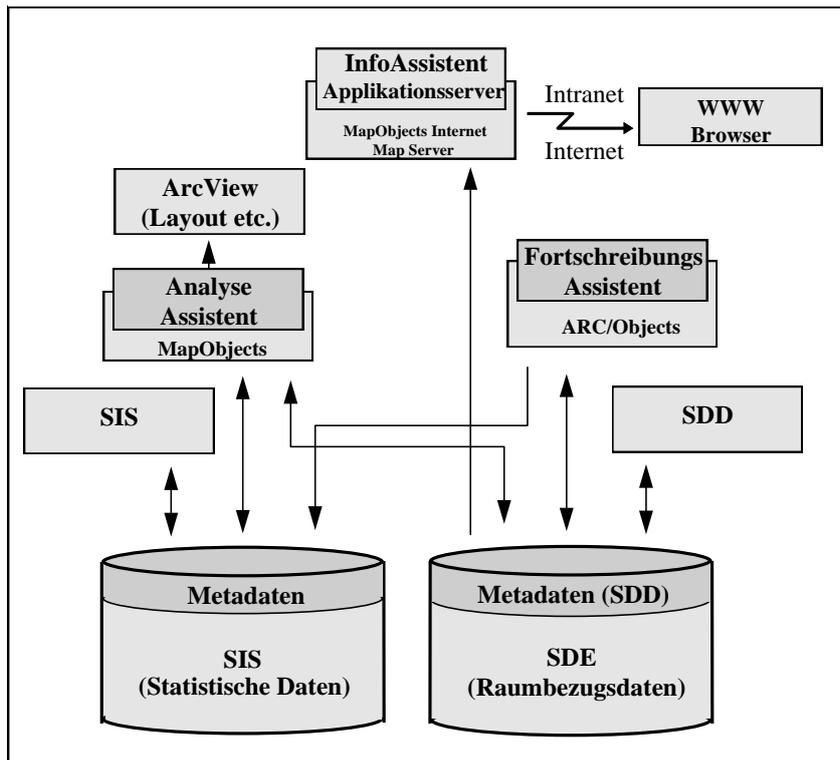
In einem ersten Schritt wird die derzeitige, mit dem Strategischen Informationssystem SIS gekoppelte, Produktion des RBS auf ARC/INFO umgestellt. Die Applikation wird mit Hilfe der ArcProjekt Entwicklungsumgebung realisiert. Neben den bekannten GIS-Funktionen sorgt sie dafür, daß die Fortschreibungs- und Analyse-schnittstellen zwischen SIS und RBS mit den entsprechenden Datenstrukturen versorgt und somit auch Daten aus dem SIS analysiert werden können. Einen wichtigen Schritt stellt die Modellierung und Übernahme der Geo-Daten des RBS dar.

Stufe 2: Geo-Assistenten auf der Basis von SDE

In einem zweiten Schritt wird auf der Basis der Spatial Database Engine SDE der metadatengesteuerte Zugriff auf Geo- und Sachdaten ermöglicht. Dabei wird das Semantic Data Dictionary (SDD) als anwendungsorientierte Beschreibungsschicht über SDE gelegt, die nicht nur die Geo-Daten selbst dokumentiert, sondern auch verschiedene Arten von Mengen, Abfragen und kartographischen Präsentationen zur Wiederverwendung abspeichert. Ein eigener SDD-Client dient dem Systemverwalter zur Erzeugung und Pflege von anwendungsspezifischen Geo-Datenmodellen. Änderungen im SDD wirken sich unmittelbar auf die Geo-Assistenten aus. Erscheint z.B. ein neues Attribut mit seinem Klartextnamen in der Oberfläche muß der Programmcode nicht geändert werden.

Der Fortschreibungs-Assistent unterstützt die metadatengesteuerte Pflege nicht nur der Geo-Daten unter SDE, sondern aktualisiert auch alle raumbezogenen Informationen im SIS. Die Fortschreibungsfunktionalität kann flexibel auf Änderungen im Datenmodell angepaßt werden. Plattform für den Fortschreibungs-Assistenten wird ARC/Objects sein.

Der Analyse-Assistent ermöglicht die Verarbeitung von SDE- und SIS-Daten. Die vom SIS bereitgestellten multidimensionalen Infopakete bieten die Möglichkeit beliebige Sichten zu bilden und vermeiden somit einen „Tabellenfriedhof“. Über Beziehungen können beliebige Aggregationen von Sachdaten vorgenommen werden. Die Definition von Abfragen kann abgespeichert, editiert und immer wieder genutzt werden. Im Analyse-Assistenten aufbereitete Daten und erstellte kartographische Präsentationen können in ArcView weiterverarbeitet werden, z.B. zu einem Kartenlayout. Plattform für den Analyse-Assistenten wird MapObjects sein.



Systemkomponenten

Stufe 3: InfoAssistent im Internet

In einer dritten Stufe wird die schnelle und komfortable Bereitstellung von Informationen im Internet/Intranet über den Info-Assistenten ermöglicht. Dabei dient der Analyse-Assistent nicht nur der Aufbereitung von Geo- und Sachdaten, sondern es können auch vordefinierte kartographische Präsentationen und sach- bzw. raumbezogene Abfragen dem Info-Assistenten zur Verfügung gestellt werden. Plattform für den Info-Assistenten wird der Map Objects Internet Map Server sein.

EU-Projekte als Finanzierungsbasis

Die oben beschriebenen Arbeiten werden z. T. im Rahmen von EU-Projekten realisiert. Im EU-Projekt ENTRANCE wird das SDD als Basis für alle Geo-Assistenten konzipiert. Die Entwicklung des Fortschreibungs- und der Analyse-Assistenten ist Teil des EU-Projektes EUROSOCPE, der Aufbau des Info-Assistenten Inhalt des EU-Projektes ENTIRE. Da die Bereitstellung von Informationen über Internet/Intranet-Technologie eine hohe Priorität hat, befindet sich zudem, unter der Abkürzung GALA, ein neues EU-Projekt zu dieser Thematik in der Beantragungphase.

DATA WAREHOUSE

Der Data Warehouse Begriff entstand in Anlehnung an die Aufgaben automatisierter Hochregallager, in die Waren (Daten) strukturiert eingelagert und über einen Bestellkatalog (Metadaten) zielgerichtet abgerufen oder als Warenpakete (Infopakete) für den Kunden zusammengestellt werden.

Als Schöpfer dieses Begriffs gilt W. Inmon, der ein Data Warehouse als eine Sammlung themenorientierter, integrierter, nichtflüchtiger und zeitbasierter Daten definiert, die die Informationsbedürfnisse von Managern erfüllen sollen.

Kern des Data Warehouse ist eine multidimensionale Datenstruktur, die den mehrdimensionalen Fragestellungen des Managements gerecht wird und von diesem ohne tiefgreifende DV-Kenntnisse abgerufen werden kann. Im SIS werden diese Datenstrukturen Infopakete genannt. Sie enthalten neben den eigentlichen Daten (Werten) beschreibende und strukturierende Informationen (Metadaten), die für die Navigation durch das Data Warehouse und für die flexible Analyse wichtig sind.

KOMMUNALE GEBIETSGLIEDERUNG

Die Kommunale Gebietsgliederung wird in vielen deutschen Kommunen als räumliche Grundlage für die Verwaltungsorganisation, Planungsunterstützung und Statistikproduktion aufgebaut. Sie ist die inhaltliche Basis für das Raumbezugssystem RBS. Das wichtigste Element stellt die Adresse dar. Dieser sind fast alle Informationen zuzuordnen. Eng verbunden mit der Adresse ist das Straßennetz.

Ferner existiert eine hierarchische Struktur vom Stadtbezirk über den Stadtteil bis zu mehreren Detailierungsebenen innerhalb der sehr kleinräumigen Blockstruktur. Diesen Grundelementen sind eine Vielzahl von raumbezogenen Gebieten, wie z. B. Postleitzahlgebiete, Stimmbezirke, Schuleinzugsbereiche, Verkehrszellen, etc., zugeordnet.

Stadt Köln

**Amt für Statistik, Einwohnerwesen und Europaangelegenheiten
Athener Ring 4
50765 Köln**

Ansprechpartner: Alfred Walgenbach

Telefon: +49-(0)221-221 18 55

Telefax: +49-(0)221-221 17 05

E-Mail: Alfred.Walgenbach@X400.STATISTIK-

EINWOHNERWESEN-

EUROPA KOELN ION DE