

Inventur im Data-Warehouse

„Information Overload“. Beständig wächst das Begehren nach einheitlichem Zugriff auf alle relevanten Systeme und Daten. Seit nunmehr über 30 Jahren werden Anstrengungen unternommen, um Wirtschaft und Verwaltung planungs-, controlling- und marketingrelevante Informationen direkt und digital zur Verfügung zu stellen.

Management Information System (MIS), Führungs Information System (FIS), Entscheidungs Unterstützung System (EUS) etc.: Geboten wird eine Reihe ansprechender Werkzeuge. Doch oft wird die Datenhaltung samt Sammlung, Verdichtung und Selektion der Informationen nicht sinnvoll realisiert.

Das Warenlager der Daten

Aufgrund der Ungleichheit operativer Systeme ist eine systematische Zusammenführung der Datenbestände innerhalb eines Unternehmens bzw. einer Verwaltung notwendig. Hier greift das Data-Warehouse-Konzept. Doch was ist das Data-Warehouse? Welchen Nutzwert verspricht es? Lohnen die Investitionen? Und vor allem: Wie und wo sollte der Raumbezug Einbindung finden?

Obwohl naheliegend, ist die Eins-zu-Eins-Übersetzung als Daten-Warenhaus unzutreffend. Selbstbedienung stößt in der Datenwelt rasch an Grenzen. Ein Data-Warehouse gleicht vielmehr einem insbesondere im Versandhandel üblichen Hochregallager. Einer vom Kunden definierten Anfrage, die im Data-Warehouse jederzeit variiert werden kann, folgt die systemgerechte Übersetzung der Anforderung sowie die Datenselektion und -auslieferung. Der Endnutzer orientiert sich – einem Versandhauskatalog ähnlich – an der sogenannten Metadaten-Ebene.

Grundgerüst: Metadaten

Die Metadaten-Steuerung übernimmt nicht nur die Aufgabe eines beschreibenden und standardisierenden Daten-Verzeichnisses, sondern hat auch strukturierenden Charakter. Die Metadaten verkörpern gleichsam das Gerüst, in das

die Daten externer Quellen einfließen. Die Datenschätze gewinnen somit an Transparenz.

Dem Benutzer steht mit dem Metadatenbanksystem ein Informationskatalog sowie eine Navigationshilfe zur Hand. Der Informationskatalog beschreibt die Informationsobjekte (Welche Daten existieren, wo befinden sie sich, in welcher Form liegen sie vor, wann erfolgte das letzte Update, welche Zugriffs-Werkzeuge sind erforderlich? etc.). Die Navigationshilfe, sprich der Browser, unterstützt den Benutzer bei der selbständigen Recherche und Analyse in den Metadaten-Beständen (Themenengrenzung; Veränderung von Objektattributen innerhalb einer Zeitreihe; Beschreibung äquivalenter Zustände – gleich, ähnlich, verschieden – usw.).

Kurzum: Die „Regale“ im Data-Warehouse sind nach Sachgebieten geordnet, das „Warenangebot“ ist kundenorientiert. Der Anwender gelangt an die gewünschten Informationen, und zwar ohne Programmierkenntnisse, ohne spezielles Wissen über Datenbanken-Abfragesprachen und ohne Umwege über die DV-Abteilung.

Lean Management

Die Optimierung der Geschäftsprozesse und der Marktanpassungsfähigkeit wurde zum richtungweisenden Wettbewerbsfaktor. Verschlankeung und Dezentralisierung der Organisationshierarchien ist der Lösungsansatz vieler Unternehmen. Die öffentliche Verwaltung zieht nach. Doch wird die Entscheidungsfindung nach unten delegiert, müssen auch dort Instrumente für operative, taktische und strategische Entscheidungen bereitstehen. Der Weg führt über die Client-Server-Architektur direkt ins Data-Warehouse.

Der Nutzwert dieses IT-Konzepts ist selten direkt monetär faßbar. Nach einer Umfrage des Consultingmultis KPMG in Industrie und Dienstleistungsgewerbe liegen profitable Vorteile vor allem in der gezielten und schnelleren Informationsnutzung zum Zwecke der

Trenderkennung, Verbesserung des Marketings, Erschließung neuer Zielgruppen sowie der Erhöhung der Service-, Vertriebs- und Produktqualität. Wissen ist Macht – aktuelles und punktgenaues Wissen sichert diese Machtposition.

Big Deal Data-Warehouse

Anfang der 90er Jahre griffen nach elementarer Vorarbeit durch IBM verschiedene Hardware- und Softwarehersteller das Data-Warehouse-Konzept auf und bieten dies nunmehr als Dienstleistungspaket an. Zum Einsatz kommen zudem Softwarelösungen zur Unterstützung von Data-Warehouse-Projekten, so in den Sektoren Erfassung und Verwaltung von Metadaten oder Datenmodellierung und -zugriff.

Aus der Angebotsfülle seien beispielhaft genannt Intergraph mit dem in die Jupiter-Technologie eingebundenen Data-Warehouse-Konzept „GeoMedia“ sowie MapInfo mit dem Mapping-Tool „MapOLAP“.

Der Markt verspricht hohe Umsatzsteigerungen. Doch ein Data-Warehouse ist keineswegs eine Standard-Software-Lösung, vielmehr ist es individuell zu konzipieren. Hierbei gilt: „Think big – start small“. Schritt für Schritt planen und aufbauen, denn angesichts des unüberschaubaren Angebots an Analysetools droht „Gigantomanie“. Einen Lösungsweg zeigt z.B. die SAS Institute GmbH, Heidelberg, mit „Rapid Warehousing“ auf: zielgenaue Erstinstallation dort, wo der Informationsbedarf zunächst am größten ist.

Dimensionen der Daten

Operative Datenverarbeitung wertet z.B. die Transaktionen des Tagesgeschäfts aus. Trends lassen sich aber nur mittels analytischer Informationsverarbeitung erkennen: Wo wurde welches Produkt durch welchen Mitarbeiter wie oft in einem bestimmten Zeitraum verkauft? – eine fünfdimensionale Frage. Ein Wechsel in der Perspektive und im Grad der Detaillierung („Drill-Down“) der Datenbetrachtung muß

jederzeit möglich sein. Verlangt wird ein Analyseverfahren für mehrdimensionale Datenstrukturen. Da für diese Aufgabenstellung nicht entwickelt, können relationale Datenbanken diese Anforderungen nur mit Mühe erfüllen. Die Abfragetechnik OLAP (On-Line Analytical Processing) verknüpft jedoch unterschiedlichste Dimensionen, im Falle einer Produktabfrage z.B. Zeit, Ort, Modelltyp, Größe, Alter, Farbe, Käuferschicht etc. Begnügt man sich mit vier bis fünf Dimensionen, können auch weiterhin relationale Datenbanken wie Oracle, Informix, Sybase oder DB2 mit ihrem typischen zweidimensionalen Tabellenmodell eingebunden werden. Virtuell wird daraus ein multidimensionales Datenbankschema kreiert. Datenbanken mit OLAP-Speichertechnik bieten im Data-Warehousing eindeutige Vorzüge.

Objekt im Fokus

Noch unentschieden ist die Diskussion zum Einsatz objektorientierter Datenbanken, wie z.B. ObjektStore, Ontos oder Poet. Hier können mit Begriffen der realen Welt Objekte der realen Welt beschrieben werden. Ohne Zweifel sind die Schwächen relationaler Systeme gerade die Stärken der objektorientierten Datenbanken und umgekehrt.

Informix glänzte Ende 1996 mit der Vorstellung der bislang einmaligen Technologie des objektrelationalen Universal-Server samt DataBlades. DataBlades sind „Plug-

in“-Softwaremodule und ermöglichen die Verarbeitung komplexer, bislang unstrukturierter sogenannter „Rich Data“-Informationen aus Web-Seiten, Bildern, Audio- und Videodokumenten sowie Karten. MapInfo, ESRI und andere widmen sich bereits der Entwicklung entsprechender DataBlades.

Raumbezug – wo bist du?

William H. Immon, amerikanischer DV-Berater und Vater des Data-Warehouse-Konzepts bringt es auf den Punkt: Die enthaltenen Daten müssen themenbezogen, zeitlich variabel (Zeitreihen), beständig (nur Datenerweiterung, keine -veränderung) und vereinheitlicht sein. Doch wo steckt bei all dem der Raumbezug?

Traditionelle Präsentationen wie Tabellen und Diagramme aller Art versagen bei der Darlegung räumlicher Zusammenhänge. Vertriebsgebiets-Reorganisation, Verteilung potentieller Kunden oder die Optimierung des Strom- oder Telekommunikationsnetzes aufgrund von Daten zur Einwohnerdichte erfordern Analyse- und Darstellungsmedien, wie sie allein Geo-Informationstechnologien bieten können. Nahezu 85% aller Daten tragen einen – wenn auch nicht immer offensichtlichen – Raumbezug! Dieses Potential gilt es im und mit dem Data-Warehouse auszuschöpfen. In Wirtschaft und Verwaltung eröffnen sich somit ungeahnt neue Entscheidungskriterien.

Geocodierung

Zunächst müssen die Daten jedoch über einen Raumbezug identifizierbar sein. Die entsprechende Geocodierung kann an verschiedenen Punkten der Daten-Transformation stattfinden, vor oder während des Einlesens der operativen Daten in das Data-Warehouse. Ist jedoch eine Vervielfältigung der Daten vorgesehen, muß die Geocodierung, um Redundanzen auszuschließen, vor der Daten-Replikation stehen.

Kommunen im Wandel

Die öffentliche Hand propagiert ihr neues Leitbild „Kommunalverwaltung als öffentliches Dienstleistungsunternehmen“. Die Umwandlung führt zum Aufbau einer unternehmensähnlichen dezentralen Organisationsstruktur.

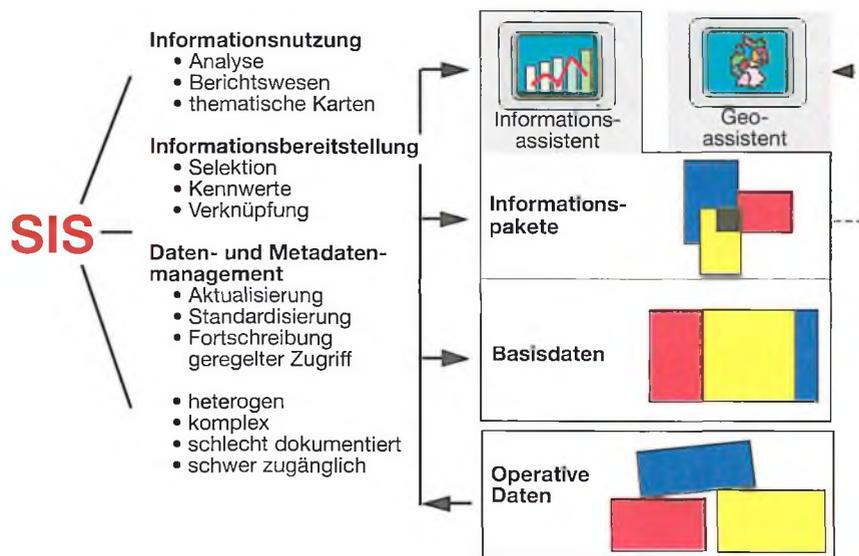
So implementiert beispielsweise der DV-Verband Baden-Württemberg seit 1994 ein landeseinheitliches strategisches Abfrage- und Auswertungsinstrument namens „Kommunales Informations Management“ (KIM). Doch so ausgereift die Steuerungsmodelle dieses Data-Warehouse mit den Elementen Finanz-, Personal-, Einwohner-, Kraftfahrzeugwesen, Gewerberegister usw. auch sein mögen, es mangelt an der effizienten Einbindung der räumlichen Komponente.

Ganz anders bei der – auch im internationalen Vergleich – fortschrittlichsten Data-Warehouse-Lösung im Amt für Statistik der Stadt Köln: dem „Strategischen Informationssystem“ (SIS). Durch die Prozeßintegration eines GI-Systems wurde eine neue Qualität des Data-Warehousing erreicht und der erste Schritt zur metadaten gesteuerten raumbezogenen „Information Factory“ vollzogen.

Data-Warehouse Stadt Köln

SIS trägt kompromißlos der Forderung Rechnung, daß standardisierte und plausibilisierte Daten und Datenschlüssel, ein mehrdimensionales Datenmodell sowie ein umfassendes Metadaten-System bereitzustellen sind. Zudem besitzt SIS ein hohes Maß an Hersteller- und Plattformunabhängigkeit. Schon die dynamische Entwicklung der DV-Technik verlangt nach dieser Systemoffenheit.

Unstrukturierte Rohdaten aus operativen Datenquellen (Umfragen,



Die Data-Warehouse-Lösung SIS

Marktbeobachtungen, Einwohner-, Kfz-Meldewesen etc.) werden übernommen und im SIS zu einem homogenen Datenbestand (Basisdaten) vereinigt. Diese umfangreichen und periodisch fortgeschriebenen Basisdaten werden über Aggregationsfunktionen zu kompakten, multidimensionalen Informationspaketen aufbereitet. Gegliedert in eine Struktur aus sachlichen, räumlichen und zeitlichen Achsen kann der Anwender aus diesen Infopaketen differenzierte Analysen vornehmen. Daten-selektion, -aggregation, -transformation und -verknüpfung sind über die Metadaten-Steuerung beliebig miteinander kombinierbar.

Das Raumbezugssystem

Unsere individuellen räumlichen Wahrnehmungen variieren erheblich. Nicht anders unterscheiden sich die räumlichen Vorstellungen von Verkehrs-, Umweltplanern oder Sozialdezernenten. Gefragt ist also ein themenneutrales Raumbezugssystem (RBS), realisiert mittels eines Geo-Informationssystems. Grundlage des räumlichen Datenmodells der Stadt Köln ist die kleinräumige Gebietsgliederung in Stadtbezirke, Stadtteile, Blöcke, Straßen etc., die zur räumlichen Verknüpfung vielfältiger Sachinformationen genutzt werden kann. Derzeit werden rd. 250.000 Geo-Objekte samt ihrer geometrischen Ausprägung der Topologie und zugehörigen Beschreibung (Attribute) im RBS über ein GIS gepflegt. Entscheidend für einen Analyseprozess ist die Vielzahl der zugeordneten Beziehungen untereinander. Auch der Aufbau von Ad-hoc-Beziehungen über sogenannte Verschneidungen sind möglich. So bleiben selbst Nachfragen aus der Privatwirtschaft (z.B. Verschneidung eines Vertriebsgebiets mit der Einwohnerstruktur einer Raumeinheit) oder aus der Forschung (z.B. Ausbreitungswolke einer Schadstoffimmission) nicht unbeantwortet.

Sach- und Raumdaten getrennt

Aufbau und Fortschreibung der räumlichen Objekte samt ihrer Attribute erfolgen im GIS, die eigentliche Sachdaten-Haltung übernimmt das SIS. Die Pflege der multidimensionalen Sachdaten im GIS würde lediglich „Tabellenfriedhöfe“ produzieren. Damit keine Dateninkon-

Für die Erde auf der Erde der Marktplatz Nr. 1

geotechnica



Internationale
Fachmesse und
Kongreß für
Geowissenschaften
und Geotechnik



Köln, 13.-16. Mai

1997

Die geotechnica, das Forum der Konzepte, Verfahren, Produkte und Dienstleistungen der Geowissenschaften und Geotechnik, stellt die konkreten Anwendungen in den Vordergrund. Hier präsentiert sich das internationale Angebot, hier ist der Informations- und Beschaffungsmarkt, hier können Investitionsentscheidungen mit hoher Sicherheit vorbereitet und getroffen werden. Für Sie – aus der privaten Wirtschaft und der öffentlichen Verwaltung – ist deshalb der geotechnica-Termin in Köln ein Muß.

Sie werden erwartet von forschenden Universitäten bis zu marktorientierten Produzenten, von jungen, innovativen Geo-Wissenschaftlern bis zu eingeführten, erfolgreichen Geo-Unternehmen. Die Ausstellungen auf der geotechnica ermöglichen für Sie die Brücke zu Praxis und Business. Parallel zur Fachmesse findet am 14. und 15. Mai der geotechnica-Kongreß statt, der unter dem Generalthema „Umweltgerechte Nutzung unserer Ressourcen an der Schwelle zum nächsten Jahrtausend“ steht.

Go Geo - Business!



Alfred-
Wegener-
Stiftung

Weitere Informationen:
KölnMesse, Abteilung 213, Postfach 210760, D-50532 Köln, Tel. (0221) 821-0,
Fax. (0221) 821-2574, Internet: <http://www.koelnmesse.de/geotechnica>

► Bitte senden Sie mir weitere Informationen zur geotechnica '97:

- Allgemeine Besucherinformationen
- Kongreßprogramm mit Anmeldeunterlagen
- Ihr Angebot mit günstigen Reisearrangements

Firma _____

Name _____

Straße _____

PLZ, Ort _____

Telefon _____

Köln Messe

GeoBIT - Feb.

sistenzen zwischen beiden Komponenten entstehen, werden räumliche Veränderungen, z.B. Umbenennungen von Straßen oder Neugliederung von Wahlbezirken, ausschließlich vom RBS (GIS) aus angestoßen und im SIS automatisch nachgeführt.

Schnittstellen

Leistungsfähige Fortschreibungsschnittstellen sind auch Basis für die im GIS sehr problematische Historienverwaltung. Inaktuelle Daten werden im SIS mit einem Zeitstempel versehen („gültig bis ...“) und weiterhin vorgehalten – die Grundlage für Zeitreihenanalysen.

Bidirektionale Analyseschnittstellen erlauben hingegen sowohl die räumliche Selektion vom RBS aus (über die Geometrie oder über Attribute und Beziehungen) als auch die Eingrenzung des Abfragethemas über Sachdaten aus dem SIS.

Flexible, sachbezogene Raum-analyse

Die entwickelte Applikation SIS-GIS unterstützt die Auswahl eines Gebietes nach frei definierten Regeln (Abfragepanel), die Analyse des Gebietes durch Verknüpfung mit Sachdaten (Mengenpanel) sowie die kartographische Darstellung der Ergebnisse (thematisches Kar-

tierpanel). Diese Funktionspanels sind voneinander unabhängig und speichern jeweils das Abfrageverfahren, welches so beliebig auf andere Teilräume angewandt oder auf ähnliche Fragestellungen einfach angepaßt werden kann.

Die flexible Zusammenführung von Sachdaten und Raumbezügen ermöglicht nicht nur die Abfrage vorhandener Daten, sondern generiert über die Verknüpfung und Verdichtung völlig neue Informationen.

Stark in der Gemeinschaft

SIS wurde von einer Entwicklungsgemeinschaft sieben deutscher Großstädte und vier Statistischer Landesämter unter Federführung der Stadt Köln und in Zusammenarbeit mit der Software AG, Darmstadt, realisiert. Verbreitung findet SIS mittlerweile auch in der Privatwirtschaft (u.a. Generaldienst Post, Investitionsbank Berlin, Telekom). Die Integration der Geo-Komponente, die derzeit in Köln auf ARC/INFO / ArcView umgestellt wird, bietet für viele Bereiche sinnvolle Einsatzmöglichkeiten. Sie soll daher in Zukunft auch bei anderen Anwendern Einsatz finden.

Köln wagt derzeit mit dem Projekt InterGIS sogar den nächsten Schritt zur intranet-gestützten Kommunikation zwischen unterschiedlichen Geodaten- und SIS-

Servern innerhalb der kommunalen Großorganisation.

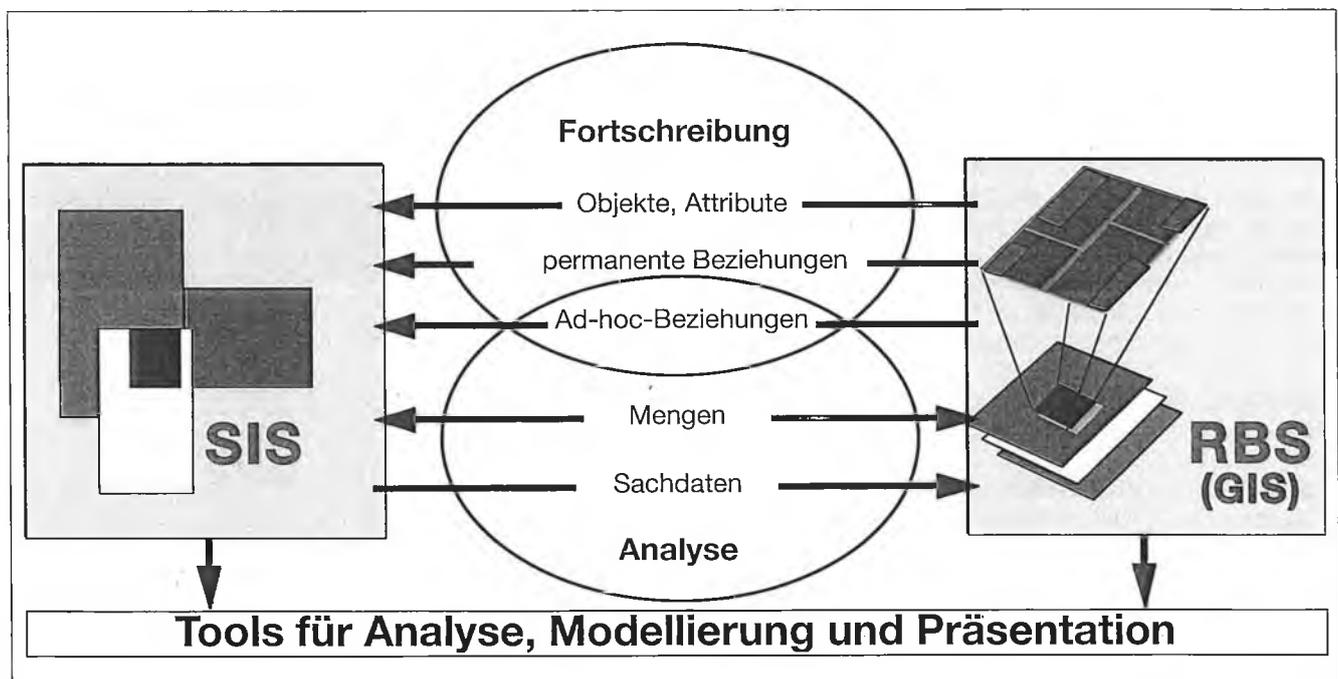
Nicht nur eine Frage des Geldes

Die Realisierung des Data-Warehouses SIS in Köln schlug bislang mit 6 Mio. DM zu Buche; 2 Mio. DM waren über EU-Fördermittel abgedeckt.

Im allgemeinen entfallen bis zu 80% der anfallenden Arbeiten bei Aufbau und Pflege eines Data-Warehouses auf die Datenorganisation, -strukturierung und -fortführung.

Die finanzielle Seite kann jedoch auch mit positiven Impulsen aufwarten. „Werden das Data-Warehouse als Lagerhaus betrachtet und Informationen als Ware angesehen, deren Beschaffung mit Kosten verbunden ist, darf diese nicht umsonst zu haben sein. Allerdings fehlt dafür derzeit noch eine breite Lobby“, so Data-Warehouse-Experte Dr. Harry Mucksch, European Business School Schloß Reichartshausen.

Die Data-Warehouse-Euphorie bleibt mehr als berechtigt. Das Data-Warehouse wird sich etablieren – anders als viele konzeptionell unausgereifte Management Support Systeme der 60er bis 80er Jahre, die nicht selten in Resignation endeten.



Wechselseitiger Daten- und Metadaten austausch