

Integration von Geodaten in Analytische Informationssysteme: Klassifikation bestehender Ansätze

Michael Kolb

Veröffentlicht in:
Multikonferenz Wirtschaftsinformatik 2012
Tagungsband der MKWI 2012
Hrsg.: Dirk Christian Mattfeld; Susanne Robra-Bissantz



Braunschweig: Institut für Wirtschaftsinformatik, 2012

Integration von Geodaten in Analytische Informationssysteme: Klassifikation bestehender Ansätze

Michael Kolb

Wissenschaftliches Institut für Hochschulsoftware der Universität Bamberg (ihb),
96045 Bamberg, E-Mail: michael.kolb@ceushb.uni-bamberg.de

Abstract

Der Informationswert von Geschäftsdaten in Analytischen Informationssystemen lässt sich durch eine Integration von Geodaten erheblich steigern. In den letzten Jahren entstanden aus diesem Grund vermehrt interdisziplinäre Forschungsbeiträge und Prototypen zur gemeinsamen Analyse von Geschäftsdaten und Geodaten. Momentan existiert jedoch keine geeignete Klassifikation um bestehende und neue Forschungsbeiträge einordnen zu können. In diesem Beitrag werden bestehende Ansätze anhand eines idealtypischen Aufbaus Analytischer Informationssysteme charakterisiert. Nach der Charakterisierung werden durch eine Klassifikation mehrere Varianten für die Integration von Geodaten in Analytische Informationssysteme abgeleitet. Die Unterschiede der Varianten werden abschließend anhand eines Vergleichsrahmens dargestellt.

1 Einleitung

Die Fach- und Führungskräfte einer Organisation sind auf eine umfassende Versorgung mit Informationen angewiesen, um erfolgsrelevante Entscheidungen für ihre Organisation treffen zu können. Analytische Informationssysteme (AIS) besitzen als Anwendungssysteme zur Versorgung von Fach- und Führungskräften mit entscheidungsrelevanten Informationen daher eine große Bedeutung [6]. In den letzten Jahren wurde erkannt, dass Geodaten ein erhebliches Potential für Visualisierungs- und Analysezwecke innerhalb Analytischer Informationssysteme besitzen und einen großen Beitrag zur umfassenden Informationsversorgung der Fach- und Führungskräfte leisten [17]. Durch die Kombination von Karten mit Tabellen oder Diagrammen lassen sich oft neue Erkenntnisse gewinnen und komplexe Zusammenhänge erkennen [4]. Der Informationswert von alphanumerischen Daten über Umsätze, Kosten, Produkte etc. (im weiteren Verlauf als konventionelle Geschäftsdaten bezeichnet) lässt sich durch die „Anreicherung“ mit Geodaten steigern [1]. Während bisher geografische Informationen wie Adressdaten, Postleitzahlen und Gebiets- oder Ortsbezeichnungen zu Analysezwecken oft nur in textueller, beschreibender Form verwendet wurden, sind zunehmend auch die korrespondierenden Geodaten wichtiger Bestandteil von Analytischen Informationssystemen. Dies hängt zum einen mit dem stetig wachsenden Angebot an qualitativ hochwertigen und zum Teil frei verfügbaren Geodatenbeständen zusammen. Organisationen sind nicht mehr darauf angewiesen eigene Geodatenbestände zu erfassen und zu pflegen, sondern können externe Datendienste in Anspruch nehmen.

Zum anderen unterstützen heute zahlreiche Datenbankmanagementsysteme (DBMS) den Umgang mit Geodaten und bieten Funktionen zu deren Bearbeitung an (z. B. PostgreSQL mit PostGIS oder Oracle mit Oracle Spatial). Mittlerweile existieren zahlreiche Ansätze, welche die Funktionen von Online-Analytical-Processing (OLAP) Systemen mit den Funktionen von Geografischen Informationssystemen (GIS) vermischen und somit die Verarbeitung von Geodaten in Analytischen Informationssystemen ermöglichen (vgl. Abschnitt 3). Die propagierten Ansätze behandeln meist detaillierte Problemstellungen und prototypische Implementierungen. Dies liegt vor allem daran, dass Analytische Informationssysteme oft stark auf einen spezifischen Einsatzbereich oder eine bestimmte Domäne ausgerichtet sind und meist einzelfallabhängig konzipiert werden müssen [11]. Trotz der Vielfalt an bestehenden Forschungsbeiträgen und Prototypen zur Integration von Geodaten in Analytische Informationssysteme existiert in der Literatur keine geeignete Klassifikation der unterschiedlichen Ansätze. Eine Klassifikation erlaubt die Einteilung bestehender und neuer Ansätze in Klassen (im weiteren Verlauf als „Integrationsvarianten“ oder kurz „Varianten“ bezeichnet), wodurch die Komplexität des Themenfeldes reduziert wird. Dadurch erhalten Projektverantwortliche für die Erweiterung bestehender Analytischer Informationssysteme um die Nutzung von Geodaten bzw. für den Aufbau neuer Analytischer Informationssysteme mit integrierter Geodatennutzung eine Hilfestellung bei der Auswahl einer geeigneten Integrationsstrategie. Als Grundlage für eine Klassifikation wird in diesem Beitrag eine Charakterisierung bestehender Forschungsansätze anhand eines idealtypischen Aufbaus Analytischer Informationssysteme vorgenommen. Nach der Charakterisierung werden durch eine Klassifikation die Integrationsvarianten identifiziert. Die Varianten werden anschließend anhand eines Vergleichsrahmens gegenübergestellt. Dieser Vergleichsrahmen ermöglicht einen übersichtlichen und strukturierten Zugang zu den Unterschieden der Varianten.

2 Analytische Informationssysteme und Geodaten

Zunächst werden die zentralen Begriffe „Analytische Informationssysteme“ und „Geodaten“ erläutert und hinsichtlich der Verwendung in diesem Beitrag abgegrenzt. Unter dem Begriff „Analytische Informationssysteme“ lassen sich Konzepte wie „Data-Warehouse“ (DWH), „Business-Intelligence“, „OLAP“ oder „Data-Mining“ zusammenfassen [6]. Data-Warehouse-Systeme (DWH-Systeme) etablierten sich als Komponente zur Datenhaltung für Analytische Informationssysteme. Sie versorgen die Entscheidungsträger mit vereinheitlichten und historisierten Daten aus organisationsinternen und -externen Quellsystemen und ermöglichen die Analyse des Datenbestandes durch OLAP-Funktionen. Ein idealtypischer Aufbau eines DWH-Systems besteht aus den Ebenen zur Datenerfassung, Datenhaltung und Datenbereitstellung [20]. In der Datenerfassungsebene findet der ETL-Prozess (Extraktion, Transformation, Laden) statt. Die Datenhaltungsebene beinhaltet das eigentliche Data-Warehouse. Die Funktionen der Datenbereitstellungsebene liefern den Datenbestand in multidimensionalen Strukturen für die Präsentationswerkzeuge. Die dem DWH-System nachgelagerte Präsentationsebene und die vorgelagerte Ebene der Datenquellen sind nicht Bestandteil des DWH-Systems. Bezüglich des vorgestellten idealtypischen Aufbaus eines DWH-Systems wird im Folgenden ein DWH-System zusammen mit der Präsentationsebene als Analytisches Informationssystem bezeichnet (vgl. Bild 1).

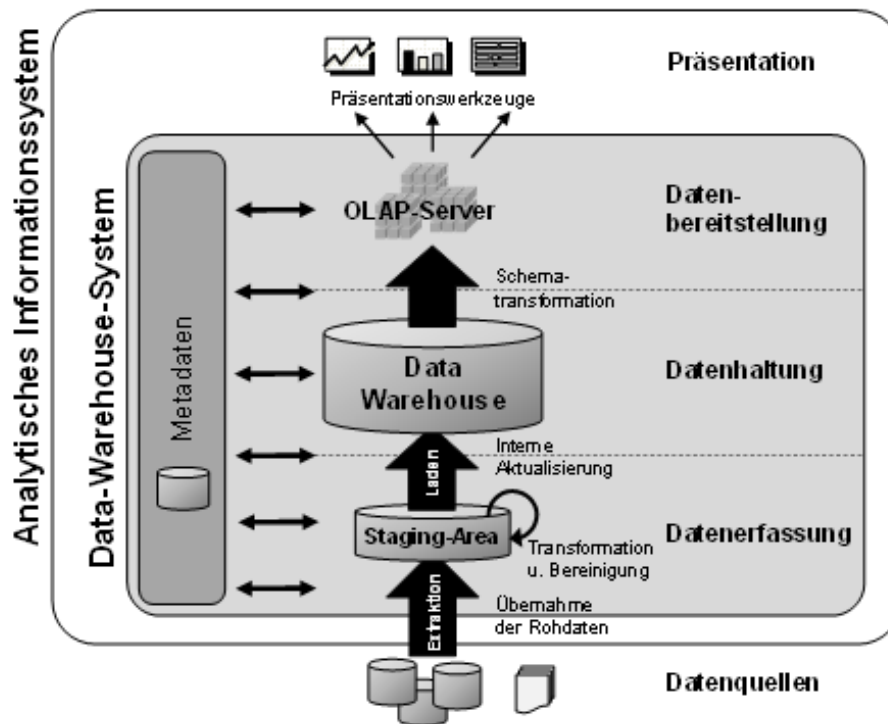


Bild 1: Idealtypischer Aufbau eines AIS (in Anlehnung an [20])

In diesem Zusammenhang stellt sich die Frage, ob auch Geografische Informationssysteme (ggf. erweitert um OLAP-Funktionen) als Analytische Informationssysteme bezeichnet werden können. Geografische Informationssysteme besitzen zwar Analysefunktionen, werden aber aufgrund ihrer Transaktionsorientierung den operativen Systemen zugeordnet [4]. Im Gegensatz dazu steht bei Analytischen Informationssystemen „die Informationsversorgung und funktionale Unterstützung betrieblicher Fach- und Führungskräfte zu Analysezwecken im Vordergrund“ [6]. Deshalb wird in diesem Beitrag eine Erweiterung der operativen Geografischen Informationssysteme um OLAP-Funktionen nicht betrachtet. Die Integration von Geodaten in Analytische Informationssysteme bezieht sich ausschließlich auf den idealtypischen Aufbau eines Analytischen Informationssystems (Bild 1). Dies schließt jedoch eine Verwendung von GIS als Präsentationswerkzeuge in der Präsentationsebene nicht aus.

Nach der Definition und Abgrenzung des Begriffs „Analytisches Informationssystem“ wird der Begriff „Geodaten“ beschrieben. Geodaten (engl. Spatial-Data) sind „Daten über Gegenstände, Geländeformen und Infrastrukturen an der Erdoberfläche, wobei als wesentliches Element ein Raumbezug vorliegen muss“ [16]. Neben textuellen, beschreibenden Attributen müssen Geodaten daher auch Attribute besitzen, die deren Verortung erlauben. Diese Verortung erfolgt durch die Angabe von Koordinaten innerhalb eines bestimmten Koordinatensystems. Die von Geodaten beschriebenen Objekte der Erdoberfläche werden als Geoobjekte bezeichnet. Die Eigenschaften von Geoobjekten lassen sich in die Kategorien geometrische, topologische, thematische, dynamische und dimensionale Eigenschaften einteilen [13]. Die *geometrischen Eigenschaften* beschreiben durch Koordinaten eines bestimmten räumlichen Bezugssystems die Lage und Ausdehnung von Geoobjekten. *Topologische Eigenschaften* beschreiben die Beziehungen von Geoobjekten untereinander. Topologische Beziehungen gelten unabhängig der Geometrie. Liegt beispielsweise für zwei Flächen die topologische Eigenschaft der Nachbarschaft vor, so gilt diese Eigenschaft unabhängig der geometrischen Ausprägung oder des

Raumbezugs der beiden Flächen. Die *thematischen Eigenschaften* beschreiben die Attribute des Geoobjekts. Zu einer Ortsteilfläche kann z. B. die Postleitzahl als Attribut angegeben werden. Durch Attribute kann ein Geoobjekt unterschiedlichen thematischen Ebenen zugeordnet werden. Zum Beispiel kann eine Linie gleichzeitig als Geoobjekt „Staatsgrenze“ und auch als Geoobjekt „Bundeslandgrenze“ verwaltet werden. Die *dynamischen Eigenschaften* von Geoobjekten beschreiben deren zeitliche Variabilität. Diese Variabilität gilt sowohl für die geometrischen, topologischen als auch für die thematischen Eigenschaften. Landkreise können sich z. B. im Zeitverlauf in ihrer Größe ändern (geometrische Eigenschaft) oder neue Nachbargebiete erhalten (topologische Eigenschaft). Durch zeitliche Änderungen der beschreibenden Attribute der Landkreisgebiete (z. B. Einwohnerzahl) ändern sich deren thematische Eigenschaften. Auch die *dimensionalen Eigenschaften* von Geoobjekten stehen in direkter Verbindung mit den anderen Eigenschaften des Geoobjektes. Beispielsweise ist die geometrische Darstellung durch zwei- oder dreidimensionale Koordinaten möglich. Die Anzahl der beschreibenden Attribute eines Geoobjekts kennzeichnet dessen thematische Dimension.

3 Klassifikation der Ansätze zur Integration von Geodaten in Analytische Informationssysteme

Als Grundlage für die Klassifikation werden relevante Forschungsbeiträge anhand einer Literaturrecherche in einschlägigen Literaturdatenbanken¹ ermittelt. Geeignete Schlagwörter für die Recherche werden aus den Untersuchungsobjekten „Geodaten“ und „Analytische Informationssysteme“ des Beitrags abgeleitet. Unter dem Begriff „Analytische Informationssysteme“ lassen sich Konzepte wie „Data-Warehouse“ oder „Business-Intelligence“ zusammenfassen [6]. Die Konzeptbegriffe „Data-Warehouse“ bzw. „Business-Intelligence“ sowie der Begriff „Geodaten“ werden zu den Schlagwörtern „Geo-Data-Warehouse“ und „Geo-Business-Intelligence“ kombiniert. Anhand dieser Schlagwörter (unter Verwendung entsprechender Synonyme und Übersetzungen, wie z. B. „Spatial-Data-Warehouse“) wird die Literaturrecherche im Abstract und Titel bestehender Publikationen durchgeführt. Nach positiver Relevanzbewertung des Abstracts erfolgt die inhaltliche Prüfung des Beitrags. Bei fehlender Relevanz wird der Beitrag aus der Ergebnismenge ausgeschlossen. Dies gilt z. B. für Beiträge, die keine Erkenntnisse zur Integration von Geodaten und konventionellen Geschäftsdaten vermitteln (z. B. bei einer Diskussion von ausschließlich Geodaten in einem Spatial-Data-Warehouse) oder für Beiträge, die nur einen Teilbereich des Analytischen Informationssystems abdecken (z. B. Behandlung von ausschließlich konzeptueller Modellierung von Spatial-DWH-Systemen). Da aus den Open-Source-Projekten von *Spatialytics.org* ein erster kommerzieller Spatial-OLAP-Server entstanden ist, werden zusätzlich zu den wissenschaftlichen Publikationen noch die Beiträge von [2] und [3] aufgrund ihrer Relevanz für die Praxis in die Ergebnismenge aufgenommen.

Neben der Datenintegration (Ebene der Datenerfassung und Datenhaltung) ist zur gemeinsamen Analyse der Geodaten zusammen mit den konventionellen Geschäftsdaten auch eine Funktionsintegration von OLAP- und GIS-Funktionen nötig (Ebene der Datenbereitstellung und Präsentation). Die Merkmale Datenerfassung (E), Datenhaltung (H), Datenbereitstellung (B) und Präsentation (P) erscheinen als geeignet, um bestehende Ansätze zu charakterisieren, da zum

¹ ACM Digital Library, CiteSeer, DBLP Computer Science Bibliography, EBSCOhost, IEEE Explore, OPAC der Universitätsbibliothek, ScienceDirect, SpringerLink und Web of Knowledge.

einen der Startpunkt der Datenintegration und zum anderen der Umfang der Funktionsintegration identifiziert und beschrieben werden kann (vgl. Tabelle 1). Bei der Charakterisierung der Datenerfassung wird nicht nach organisationsinternen und -externen Geodatenquellen unterschieden: „An effective enterprise information management strategy must take into account both internal and external data” [15].

Untersucher Ansatz	Charakterisierung nach Merkmalen der Datenerfassung (E), Datenhaltung (H), Datenbereitstellung (B) und Präsentation (P)
Badard, T (2010) [2] und Badard, T; Dubé, E (2009) [3]	<p>E: Integrierte Extraktion und Transformation von konventionellen Geschäftsdaten und Geodaten.</p> <p>H: Verwaltung von Geodaten zusammen mit den konventionellen Geschäftsdaten in integrierten multidimensionalen Strukturen im DWH.</p> <p>B: Spatial-OLAP-Server ermöglicht die Bereitstellung multidimensionaler Daten durch um Spatial-Funktionen erweiterte MDX (Multidimensional Expressions) Anfragen.</p> <p>P: Integration einer Kartografiekomponente zur Darstellung der geografischen Anfrageergebnisse und Nutzerinteraktion zusammen mit Tabellen oder Diagrammen in Dashboards.</p>
Bimonte, S; Wehrle, P; Tchounikine, A; Miquel, M (2006) [5]	<p>E: Transformation verschiedener Geodatenformate in ein einheitliches Format.</p> <p>H: Verwaltung von Geodaten (nur für Kennzahlen) zusammen mit den konventionellen Geschäftsdaten in integrierten multidimensionalen Strukturen im DWH.</p> <p>B: Aggregation von Spatial-Kennzahlen durch benutzerdefinierte Funktionen im Datenbank- und OLAP-Server.</p> <p>P: Integration eines OLAP- und GIS-Web-Clients zu einer Benutzerschnittstelle in der Präsentationsebene.</p>
Escribano, A; Gomez, LI; Kuijpers, B; Vaisman, AA (2007) [8]	<p>E/H: Laden und Verwalten eines GIS-Datenbestandteils (Geodaten) und OLAP-Datenbestandteils (konventionelle Geschäftsdaten) in einem gemeinsamen DWH.</p> <p>H: Beziehungstabellen ermöglichen die Zuordnung von Geobjekten aus einer GIS-Dimension zu den entsprechenden Instanzen einer Dimension der konventionellen Geschäftsdaten über gemeinsame Objekt-IDs.</p> <p>B: Kombination von OLAP- und GIS-Anfragen in einer gemeinsamen Anfrage. Eine Anfrage besteht aus einem GIS- (SQL-ähnliche Anfrage mit geometrischen Operationen) und einem OLAP-Teil (MDX-Anfrage).</p>
Healey, RG; Delve, J (2007) [9]	<p>E: Transformation verschiedener Geodatenformate in ein einheitliches Format und Bereinigung der Geodaten.</p> <p>H: Verwaltung von Geodaten zusammen mit den konventionellen Geschäftsdaten in integrierten multidimensionalen Strukturen im DWH.</p> <p>B: In einem relationalem OLAP-Server werden SQL-Anfragen um Spatial-Funktionen, z. B. zur Abfrage von topologischen Beziehungen, erweitert.</p> <p>P: Visualisierung der Geodaten für eine Anfrage mittels Scalable-Vector-Graphics (SVG). Die SVG werden on-the-fly aus den Geodaten des Data-Warehouse gemäß der Anfrage erzeugt.</p>
Hermsdörfer, D (2005) [10]	<p>H: Getrennt von einem Business-Data-Warehouse für die konventionellen Geschäftsdaten in multidimensionalen Strukturen existiert ein Spatial-Data-Warehouse für die Geodaten.</p> <p>B: Integration der separat bereitgestellten Daten von Spatial-Data- und Business-Data-Warehouse über Informationen zum gemeinsamen Raumbezug (z. B. Adressen, Gebietsschlüssel oder geografische Namen).</p> <p>P: Verwaltung und Analyse der Geodaten im Spatial-Data-Warehouse mittels GIS. Analyse des Business-Data-Warehouse durch eine Business-Intelligence-Plattform. Integrierte Analyse der Daten aus beiden Systemen mit Hilfe eines Semantic-Data-Dictionary (Integration über gemeinsamen Raumbezug).</p>

<p>Kouba, Z; Matoušek, K; Mikšovský, P (2000) [12]</p>	<p>H: Verwaltung der konventionellen Geschäftsdaten in multidimensionalen Strukturen in einem Data-Warehouse-System. Verwaltung der Geodaten getrennt vom Data-Warehouse in einem GIS.</p> <p>B: Bereitstellung der konventionellen Geschäftsdaten durch einen OLAP-Server. Integration der Geodaten aufgrund einer Taxonomie für die Geoobjekte im GIS und korrespondierenden Begriffen der Geografie-Dimension im Data-Warehouse (Verbindung auf Schema- und Instanzebene möglich; keine Unterstützung für Spatial-Kennzahlen).</p> <p>P: Integration von GIS- und OLAP-Subsystemen über ein Integrationsmodul zu einem Gesamtsystem.</p>
<p>di Martino, S; Bimonte, S; Bertolotto, M; Ferrucci, F (2009) [14]</p>	<p>H: Getrennte Verwaltung der konventionellen Geschäftsdaten und der Geodaten.</p> <p>B: Integration der durch einen OLAP-Server bereitgestellten konventionellen Geschäftsdaten mit den Geodaten mittels einer Anwendungskomponente zwischen Präsentations- und Datenbereitstellungsebene zur integrierten Analyse der Daten.</p> <p>P: Visualisierung der Geodaten in einem Geobrowser (z. B. Google Earth™). Bereitstellung der OLAP-Funktionen durch ein webbasiertes OLAP-Tool.</p>
<p>Rivest, S; Bédard, Y; Marchand, P (2001) [17]</p>	<p>H: Realisierung einer getrennten Datenhaltung durch eine multidimensionale Datenbank für die konventionellen Geschäftsdaten und ein GIS für die Geodaten.</p> <p>B: Spatial-OLAP-Server integriert ein GIS (zur Verwaltung der Geodaten und zur Bereitstellung der GIS-Funktionen) mit einem OLAP-Server (zur Bereitstellung der OLAP-Funktionen für die konventionellen Geschäftsdaten).</p> <p>P: Spatial-OLAP-Client integriert ein GIS und einen OLAP-Client zur Präsentation der Geodaten und konventionellen Geschäftsdaten.</p>
<p>Sampaio, MC; de Sousa, AG; Baptista, CdS (2006) [18]</p>	<p>H: Aufbau eines Spatial-Data-Cube als Spatial-objektrelationales-Star-Schema (Spatial-OR-Star-Schema) in einem objektrelationalen DBMS (mit Unterstützung für Geodaten). Gemeinsame Abbildung von konventionellen Geschäftsdaten und Geodaten als Kennzahlen, Dimensionen, Attribute, Hierarchien und Hierarchieebenen im Spatial-OR-Star-Schema.</p> <p>B: Realisierung von Spatial-OLAP-Anfragen auf das Spatial-OR-Star-Schema durch SQL-Anfragen erweitert um Spatial-Operatoren des objektrelationalen DBMS.</p>
<p>da Silva, J; Times, VC; Salgado, AC (2006) [19]</p>	<p>E: Realisierung des ETL-Prozesses für konventionelle Geschäftsdaten und Geodaten durch SQL-Skripte in einem objektrelationalen DBMS (mit Unterstützung für Geodaten).</p> <p>H: Gemeinsame Datenhaltung der konventionellen Geschäftsdaten und Geodaten in einem geografischen Data-Warehouse (GDW; DBMS mit Unterstützung für Geodaten). Geodaten liegen im GDW in normalisierter Form vor und werden den entsprechenden konventionellen Geschäftsdaten in den multidimensionalen Strukturen über gemeinsame Objekt-IDs zugeordnet.</p> <p>B: Eine multidimensionale Abfragesprache GeoMDQL ermöglicht die Verarbeitung von MDX-Anfragen (mit/ohne geografischen Funktionen) an das GDW.</p> <p>P: Gemeinsame Präsentation der Anfrageergebnisse in einem OLAP-Client durch Tabellen-, Diagramm- und/oder Kartendarstellung.</p>
<p>Stefanovic, N; Han, J; Koperski, K (2000) [21]</p>	<p>H: Gemeinsame Anordnung der konventionellen Geschäftsdaten und Geodaten in einem Star-Schema (Spatial-Data-Cube). Spatial-Kennzahlen bestehen aus einer Menge von Objektreferenzen auf Geoobjekte, die mit einer räumlichen Aggregationsoperation verarbeitet werden.</p> <p>B: Roll-up und Drill-down stellen aufwändige OLAP-Operationen auf einem Spatial-Data-Cube dar. Zur Vermeidung von kostenintensiven Ad-hoc OLAP-Operationen werden Methoden zur selektiven Materialisierung von aggregierbaren Geoobjekten in Spatial-Data-Cubes vorgeschlagen.</p>

Tabelle 1: Charakterisierung von Forschungsbeiträgen zur Integration von Geodaten in AIS

Aus den Ergebnissen der Charakterisierung ist erkennbar, dass ein wesentlicher Unterschied in der Art und Weise der Datenhaltung von Geodaten und konventionellen Geschäftsdaten liegt. Während die konventionellen Geschäftsdaten in den Beispielen stets in einem Data-Warehouse gespeichert werden, erfolgt die Verwaltung der Geodaten zum Teil in Geografischen Informationssystemen oder Geodatenbanken außerhalb des DWH-Systems ([10], [12], [14] und [17]). Demgegenüber steht eine gemeinsame Verwaltung der konventionellen Geschäftsdaten zusammen mit den Geodaten im DWH-System ([2], [3], [5], [8], [9], [18], [19] und [21]). Diese Trennung bzw. Zusammenführung der Daten hat folglich auch Einfluss auf die Datenbereitstellung und Präsentation zu Analysezwecken. Bei einer Datenhaltung der Geodaten außerhalb des DWH-Systems verfolgen die untersuchten Ansätze verschiedene Techniken, um eine gemeinsame Analyse in der Präsentationsebene zu ermöglichen. Zum Beispiel werden die Anfrageergebnisse des OLAP-Servers mit Geodaten in einem Geobrowser [14] oder Geografischen Informationssystem [12] in der Präsentationsebene zusammengeführt. Bei Ansätzen mit Geodaten in den multidimensionalen Datenstrukturen eines Data-Warehouse erfolgt diese Zusammenführung schon beim Aufbau der Anfrage für einen Spatial-OLAP-Server (z. B. durch eine erweiterte MDX-Anfrage) ([2], [3] und [19]). Zusammenfassend kann die DWH-interne bzw. DWH-externe Geodatenhaltung als Klassifikationsmerkmal „Geodatenhaltung“ identifiziert werden.

Bei der gemeinsamen Verwaltung der konventionellen Geschäftsdaten zusammen mit den Geodaten im DWH-System liegt ein weiterer markanter Unterschied vor. Hier kann eine Abgrenzung dahingehend getroffen werden, ob die Geodaten einen ETL-Prozess durchlaufen oder nicht. Der ETL-Prozess der Geodaten wird in der Mehrzahl der untersuchten Beiträge recht oberflächlich betrachtet. Nur [2], [3], [9] und [19] berücksichtigen in ihren Prototypen einen ETL-Prozess mit Funktionen zur Geodatenverarbeitung. Die anderen Ansätze laden lediglich die Geodaten, zum Teil mit einer Transformation in ein einheitliches Datenformat [5], in das Data-Warehouse. Die notwendigen einheitlichen Objektschlüssel (z. B. Gebietsschlüssel, Adressdaten oder geografische Namen) zur Verknüpfung mit den konventionellen Geschäftsdaten müssen in diesen Fällen bereits korrekt vorliegen. Zusammenfassend kann die Durchführung eines ETL-Prozesses für Geodaten als Klassifikationsmerkmal „ETL-Prozess“ identifiziert werden.

Mit den ermittelten Klassifikationsmerkmalen „Geodatenhaltung“ und „ETL-Prozess“ werden folgende drei Integrationsvarianten (Klassen) ermittelt, in die bestehende und neue Forschungsansätze eingeteilt werden können:

- Die Geodaten durchlaufen die Datenerfassungsebene (den ETL-Prozess) und werden in der Datenhaltungsebene des DWH-Systems zusammen mit den konventionellen Geschäftsdaten verwaltet und durch den OLAP-Server bereitgestellt (vgl. I in Bild 2). Diese Variante I wird im weiteren Verlauf als **Integration von Geodaten ab der Datenerfassungsebene** bezeichnet.
- Die Geodaten werden den konventionellen Geschäftsdaten in der Datenhaltungsebene des DWH-Systems zugeordnet, d.h. sie erfahren keine Transformationen und Bereinigungen im Rahmen des ETL-Prozess. Die Geodaten werden in die Datenhaltungsebene geladen, zusammen mit den konventionellen Geschäftsdaten verwaltet und durch den OLAP-Server bereitgestellt (vgl. II in Bild 2). Diese Variante II wird im weiteren Verlauf als **Integration von Geodaten ab der Datenhaltungsebene** bezeichnet.

- Die Geodaten werden erst in der Präsentationsebene den bereitgestellten konventionellen Geschäftsdaten des DWH-Systems zugeordnet. Die Geodaten durchlaufen nicht den ETL-Prozess, werden nicht in der Datenhaltungsebene des DWH-Systems verwaltet und auch nicht durch den OLAP-Server bereitgestellt (vgl. III in Bild 2). Diese Variante III wird im weiteren Verlauf als **Integration von Geodaten ab der Präsentationsebene** bezeichnet.

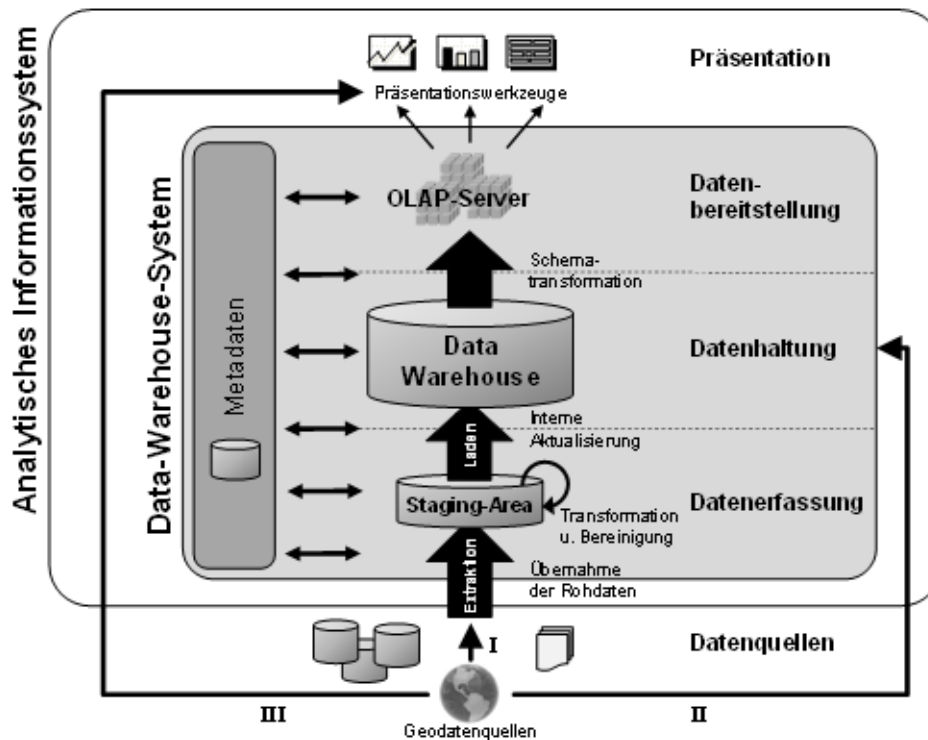


Bild 2: Varianten zur Integration von Geodaten in AIS (in Anlehnung an [20])

4 Vergleich der Varianten zur Integration von Geodaten in Analytische Informationssysteme

Für den Vergleich der Integrationsvarianten I, II und III werden exemplarisch einige Kriterien anhand des idealtypischen Aufbaus Analytischer Informationssysteme [20] und der obigen Charakterisierung bestehender Ansätze ausgewählt. Dabei werden nur Kriterien gewählt, die eine Bedeutung für die Datenintegration von Geodaten in Analytische Informationssysteme und eine Bedeutung für die OLAP-GIS-Funktionsintegration besitzen. Die Kriterien werden in die Kategorien Datenerfassung, Datenhaltung, Datenbereitstellung und Präsentation eingeteilt. Um den Evolutionsstand der Varianten vergleichen zu können, werden noch Kriterien zur Praxis-tauglichkeit und zum Realisierungsaufwand ergänzt. Der Vergleichsrahmen erhebt keinen Anspruch auf Vollständigkeit. Er erlaubt jedoch eine strukturierte Gegenüberstellung der Varianten anhand markanter Unterschiede. In der nachfolgenden Tabelle wird das Ergebnis des Vergleichs der Integrationsvarianten dargestellt. Bei den Kriterien wird geprüft, ob diese für die Varianten I, II und III erfüllbar, zum Teil erfüllbar oder nicht erfüllbar sind. Im Anschluss an die Tabelle wird die Gegenüberstellung anhand der einzelnen Kriterien im Detail erläutert.

Kategorie	Kriterium	Variante		
		I	II	III
Datenerfassung	Unterstützung mehrerer Geodatenquellen und -formate	✓	(✓)	(✓)
	Syntaktische u. semantische Bereinigung der Geodaten	✓	–	–
Datenhaltung	Integrierte Historisierung und Versionierung	✓	(✓)	–
	Zentrale Datenhaltung (Single-Point-of-Truth)	✓	✓	–
	Effiziente Verwaltung großer Geodatenbestände	✓	✓	–
Datenbereitstellung	Integrierte, multidimensionale Datenstrukturen	✓	✓	–
	Einsatz eines Spatial-OLAP-Servers	✓	✓	–
	Maßnahmen zur Steigerung der Anfrageperformance	✓	✓	(✓)
Präsentation	Visualisierung von Kennzahlen anhand Geodaten	✓	✓	✓
	Kombination von OLAP-Funktionen mit GIS-Funktionen	✓	✓	(✓)
Reifegrad	Praxistauglich	(✓)	(✓)	✓
	Forschungsbedarf	✓	✓	✓
Realisierung	Niedriger Komplexitätsgrad bei der Realisierung	–	–	✓
	Geeignet für Erweiterung bestehender Systeme	(✓)	(✓)	✓

Legende: ✓ = Kriterium erfüllbar, (✓) = Kriterium zum Teil erfüllbar, – = Kriterium nicht erfüllbar

Tabelle 2: Gegenüberstellung der Integrationsvarianten anhand eines Vergleichsrahmens

Datenerfassung: Für eine umfassende Informationsversorgung von Entscheidungsträgern müssen in einem Analytischen Informationssystem meist mehrere organisationsinterne und -externe Datenquellen berücksichtigt werden. Durch die Existenz zahlreicher, zum Teil proprietärer Geodatenformate (z. B. ESRI Shape-File oder MapInfo-Interchange-Format) und Geodienste (z. B. Web-Feature-Service) ist eine *Unterstützung mehrerer Geodatenquellen und -formate* von großer Bedeutung. Durch den ETL-Prozess besteht bei der Variante I die Möglichkeit, die entsprechenden Schnittstellen für unterschiedliche Formate anzubieten und die Daten in ein einheitliches, internes Geodatenformat zu überführen. Bei den anderen beiden Varianten müssen die operativen Vorkomponenten ein einheitliches Datenformat liefern. Die *syntaktische und semantische Bereinigung der extrahierten Geodaten* betrifft die geometrischen, topologischen, thematischen und dimensional Eigenschaften (vgl. Abschnitt 2) der durch Geodaten beschriebenen Geobjekte. Mögliche Transformationsschritte sind beispielsweise die Überführung unterschiedlicher Raumbezugssysteme der Geodaten in ein gemeinsames Raumbezugssystem oder die Überprüfung und Korrektur der topologischen Eigenschaften der Geobjekte. Ist aufgrund der Semantik der Anwendungsdomäne bekannt, dass sich Geobjekte immer berühren müssen aber nie überlappen dürfen (z. B. bei administrativen Gebietsflächen), so müssen die topologischen Beziehungen überprüft und ggf. durch Anpassung der geometrischen Eigenschaft wiederhergestellt werden. Diese Aufgaben werden bei Variante I im ETL-Prozess durchgeführt. Bei den anderen beiden Varianten muss die syntaktische und semantische Korrektheit bereits in den operativen Vorkomponenten vorliegen.

Datenhaltung: Durch eine *integrierte Historisierung und Versionierung* der Geodaten und konventionellen Geschäftsdaten werden zeitbezogene Datenbestände im Data-Warehouse aufgebaut. Je nach Historisierungsstrategie werden bei Änderungen der geometrischen, topologischen, thematischen und/oder dimensional Eigenschaften von Geobjekten neue zeitbezogene Datenbestände erzeugt. Da bei Variante III die Geodaten getrennt von den konventionellen

Geschäftsdaten im Data-Warehouse vorgehalten werden, ist keine integrierte Historisierung und Versionierung möglich. Durch die getrennte Verwaltung der Datenbestände kann auch das Kriterium einer organisationsweiten zentralen Datenhaltung im Sinne eines *Single-Point-of-Truth* [11] bei Variante III im Gegensatz zu den anderen beiden Varianten nicht erfüllt werden. Geodatenbestände besitzen oft ein beträchtliches Datenvolumen. Durch Maßnahmen wie Indexvergabe oder materialisierte Sichten [21] kann bei den Varianten I und II eine *effiziente Verwaltung großer Geodatenbestände* im Data-Warehouse erreicht werden.

Datenbereitstellung: Eine *integrierte multidimensionale Datenstruktur* von konventionellen Geschäftsdaten und Geodaten (z. B. als Star- oder Snowflakeschema) sowie der *Einsatz eines Spatial-OLAP-Servers* sind nur bei Variante I und II mit einer gemeinsamen Verwaltung der Daten in einem Data-Warehouse erfüllbar. Ähnliches gilt für *Maßnahmen zur Steigerung der Anfrageperformance*. Bei der gemeinsamen Anfrage der Geodaten und konventionellen Geschäftsdaten können die Bestandteile des Anfragestatements optimiert ausgeführt werden. Ferner besteht die Möglichkeit, durch Caching-Funktionen des Spatial-OLAP-Servers die Anfrageperformance zu verbessern. Bei der Variante III müssen Maßnahmen zur Steigerung der Anfrageperformance sowohl bei der Bereitstellung der Geodaten als auch bei der Bereitstellung konventioneller Geschäftsdaten vorgesehen werden.

Präsentation: Eine *Visualisierung von Kennzahlen anhand Geodaten* ist bei allen drei Integrationsvarianten erfüllbar, sofern Geodaten und konventionelle Geschäftsdaten einander zugeordnet werden können (z. B. durch einheitliche Gebietsschlüssel). Zur *Kombination von OLAP-Funktionen mit GIS-Funktionen* ist bei der Variante III eine Synchronisation der einzelnen Nutzerinteraktionen erforderlich, da in der Regel verschiedene Präsentationswerkzeuge miteinander kombiniert werden. Zum Beispiel muss ein Drill-down in einer Kartenkomponente auch eine entsprechende Aktion in einem zugehörigen Bericht auslösen, et vice versa [4]. Ad-hoc-Analysen und eine erweiterte Integration von OLAP- und GIS-Funktionen sind mit Variante III nur mit großem Aufwand erfüllbar. Diese Anwendungsfälle erfordern ein mächtiges Präsentationswerkzeug für Geodaten (z. B. GIS) und eine hinreichende Interaktion mit den OLAP-Präsentationswerkzeugen.

Reifegrad: Für die Variante III existieren bereits mehrere kommerzielle Produkte. Diese Variante wird daher grundsätzlich als *praxistauglich* eingestuft. Zu den beiden anderen Varianten existieren nur wenige kommerzielle Entwicklungen. Insbesondere die Entwicklung eines praxistauglichen Spatial-OLAP-Servers mit entsprechenden Client-Anwendungen stellt eine große Herausforderung dar. Aktuell existiert nur ein kommerzieller Spatial-OLAP-Server, der aus einem Open-Source-Projekt hervorgegangen ist ([2] und [3]). Für alle drei Varianten ist zur sukzessiven Weiterentwicklung weiterhin *Forschungsbedarf* vorhanden.

Realisierung: Da bei den Varianten I und II ein Spatial-OLAP-Server, eine integrierte multidimensionale Datenhaltung und ggf. ein Spatial-ETL-Werkzeug erforderlich sind, liegt hier ein relativ hoher *Komplexitätsgrad zur Realisierung* vor. Demgegenüber steht eine relativ einfache Realisierung der Variante III, da eine Integration lediglich „außerhalb“ des DWH-Systems in der Präsentationsebene erfolgt. Aus diesem Grund eignet sich diese Variante auch besonders für die *Erweiterung bestehender Analytischer Informationssysteme*.

5 Fazit und Ausblick

In den letzten Jahren entstanden vermehrt Forschungsbeiträge und Prototypen zur Integration von Geodaten in Analytische Informationssysteme. Um die Komplexität dieses interdisziplinären Forschungsgebiets zu reduzieren, klassifiziert dieser Beitrag bestehende Forschungsansätze in die drei Varianten

- Integration von Geodaten ab der Datenerfassungsebene,
- Integration von Geodaten ab der Datenhaltungsebene und
- Integration von Geodaten ab der Präsentationsebene.

Durch diese Klassifikation können bestehende und zukünftige Ansätze den drei Integrationsvarianten zugeordnet werden. Aus der Klassifikation ergeben sich aber auch weitere neue Forschungsfragen. Zum Beispiel gilt es zu prüfen, wie ein konzeptueller Modellierungsansatz für DWH-Systeme gestaltet bzw. erweitert werden muss, um die einzelnen Integrationsvarianten zu unterstützen. Ferner bietet die Integration von Geodaten ab der Datenhaltungsebene ausreichend Potential für vielversprechende neue Ansätze. Zum Beispiel können Geodaten in multidimensionalen Strukturen über serviceorientierte Architekturen ausgetauscht werden. In diesem Bereich existieren bereits erste Vorschläge [7], deren Weiterentwicklung interessante zukünftige Forschungsaufgaben darstellen.

6 Literatur

- [1] Apel, D; Behme, W (2010): Datenintegration – Ein Prozess zur Verbesserung der Datenqualität. In: Chamoni, P; Gluchowski, P (Hrsg.), *Analytische Informationssysteme. Business-Intelligence-Technologien und –Anwendungen*. Springer, Berlin.
- [2] Badard, T (2010): Developing geospatial business intelligence solutions. <http://emag.geospatialtoday.com/Login.aspx?issue=issue24&page=35>. Abgerufen am 10.09.2011.
- [3] Badard, T; Dubé, E (2009): Enabling geospatial Business Intelligence. http://geosoa.scg.ulaval.ca/~badard/article-tbadard-osbr_2009-long_version-enabling_geospatial_bi.pdf. Abgerufen am 10.09.2011.
- [4] Bédard, Y; Rivest, S; Proulx, MJ (2006): Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective. In: Wrembel, R; Koncilia, C (Hrsg.), *Data Warehouses and OLAP: Concepts, Architectures and Solutions*. IRM Press, Hershey.
- [5] Bimonte, S; Wehrle, P; Tchounikine A; Miquel M (2006): GeWOLap: A Web Based Spatial OLAP Proposal. In: Meersman, R; Tari, Z; Herrero, P (Hrsg.), *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*. Springer, Montpellier.
- [6] Chamoni, P; Gluchowski, P (2010): Analytische Informationssysteme – Einordnung und Überblick. In: Chamoni, P; Gluchowski, P (Hrsg.), *Analytische Informationssysteme. Business-Intelligence-Technologien und –Anwendungen*. Springer, Berlin.
- [7] Dubé, E; Badard, T; Bédard, Y (2009): XML Encoding and Web Services for Spatial OLAP Data Cube Exchange: an SOA Approach. *Journal of Computing and Information Technology* 17(4):347-358.

- [8] Escribano, A; Gomez, LI; Kuijpers, B; Vaisman, AA (2007): Piet: a GIS-OLAP Implementation. In: ACM (Hrsg.), *DOLAP '07*. Lissabon.
- [9] Healey, RG; Delve, J (2007): Integrating GIS and data warehousing in a Web environment: A case study of the US 1880 Census. *International Journal of Geographical Information Science* 21(6):603-624.
- [10] Hermsdörfer, D (2005): *Generische Informationsmodellierung*. Wichmann, Heidelberg.
- [11] Kemper, HG; Mehanna, W; Unger, C (2006): *Business Intelligence – Grundlagen und praktische Anwendungen*. Vieweg, Wiesbaden.
- [12] Kouba, Z; Matoušek, K; Mikšovský, P (2000): On Data Warehouse and GIS Integration. In: Ibrahim, MT; Küng, J; Revell, N (Hrsg.), *Database and Expert Systems Applications (DEXA)*. Springer, London.
- [13] de Lange, N (2006): *Geoinformatik in Theorie und Praxis*. Springer, Berlin.
- [14] di Martino, S; Bimonte, S; Bertolotto, M; Ferrucci, F (2009): Integrating Google Earth within OLAP Tools for Multidimensional Exploration and Analysis of Spatial Data. In: Filipe, J; Cordeiro, J (Hrsg.), *Enterprise Information Systems, 11th International Conference (ICEIS 2009)*. Springer, Berlin.
- [15] Petschulat, S (2010): Other People's Data. *Communications of the ACM* 53(1):53-57.
- [16] Professur für Geodäsie und Geoinformatik der Universität Rostock (2011): Geodaten. <http://www.geoinformatik.uni-rostock.de/einzel.asp?ID=762>. Abgerufen am 18.09.2011.
- [17] Rivest, S; Bédard, Y; Marchand, P (2001): Toward Better Support for Spatial Decision Making: Defining the Characteristics of Spatial On-Line Analytical Processing (SOLAP). *Geomatica* 55(4):539-555.
- [18] Sampaio, MC; de Sousa, AG; Baptista, CdS (2006): Towards a Logical Multidimensional Model for Spatial Data Warehousing and OLAP. In: ACM (Hrsg.), *DOLAP '06*. Arlington.
- [19] da Silva, J; Times, VC; Salgado, AC (2006): An Open Source and Web Based Framework for Geographic and Multidimensional Processing. In: ACM (Hrsg.), *Proceedings of the 2006 ACM symposium on Applied computing (SAC)*. Dijon.
- [20] Sinz, EJ; Ulbrich-vom Ende, A (2010): Architekturen von Data-Warehouse-Systemen. In: Chameni, P; Gluchowski, P (Hrsg.), *Analytische Informationssysteme. Business-Intelligence-Technologien und –Anwendungen*. Springer, Berlin.
- [21] Stefanovic, N; Han, J; Koperski, K (2000): Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. *IEEE Transactions on Knowledge and Data Engineering* 12(6):938-958.