

Die Neukonzeption des Umweltdatenkataloges

Walter Swoboda¹, Fred Kruse¹, Detlev Nyhuis² und Holger Rousselle²

Abstract

In order to offer public access to environmental data, Germany and Austria have developed a meta information system called Umweltdatenkatalog (UDK). While this system has been introduced since a few years ago, a complete re-design was necessary using modern technologies of software development. The new concept of the UDK is presented in detail addressing both the datamodel and the main functionalities as the UDK-Explorer, the Thesaurus-Navigator, and the GIS-viewer. Furthermore, the update procedure of the decentralized database concept is discussed. In particular, conventional technologies are compared with Web-technologies taking into account that the WWW-access is still very heterogeneous in local administrative institutions. As a prospect a concept of a three level ODBC architecture is presented which strongly reduces the effort of system administrators in client/server architectures.

1 Einleitung

Der Umweltdatenkatalog (UDK) ist ein Metainformationssystem zur Erfassung von Datenbeständen, welche in den Umweltverwaltungen von Deutschland und Österreich vorgehalten werden. Von 1991 - 1995 wurde im Niedersächsischen Umweltministerium das Grundkonzept sowie eine einsatzfähige Software im Rahmen eines vom Bund geförderten Forschungs- und Entwicklungsvorhabens entwickelt (Lessing/Schütz 1994), (Lessing/Schmalz 1994), (Schütz/Lessing 1993). In Deutschland beruht die weitere Entwicklung und Einführung dieses Systems auf einer Verwaltungsvereinbarung, die seit 1.1.1996 in Kraft ist und vom Bund und 13 Bundesländern unterzeichnet wurde. Basierend auf ersten Praxiserfahrungen wurde 1997/98 die Konzeption des Umweltdatenkataloges grundlegend neu überdacht und in einer Windows-konformen Software (32 bit) entsprechend umgesetzt. Gegenstand dieses Beitrages ist die Darstellung der Grundzüge dieser Neukonzeption.

¹ Niedersächsisches Umweltministerium, Koordinierungsstelle UDK,
Archivstraße 2, D-30169 Hannover,
e-mail: udk@mu.niedersachsen.de, Internet: <http://www.mu.niedersachsen.de>

² BVG mbH, Menkestr. 7, D-26419 Schortens
e-mail: nyhuisD@bvg-fri.de, Internet: <http://www.bvg-fri.de>

2 Das Datenmodell des UDK

Der UDK ist zur Erfassung und Recherche von umweltbezogenen Metadaten entwickelt worden. Die Metadaten sollen Auskunft darüber geben, „wer“ über „welche“ umweltbezogenen Daten in welchem Format verfügt. Hierbei kann es sich sowohl um digitale als auch um analoge Datenbestände handeln. Ferner soll recherchiert werden können, wer für welche Fachaufgaben zuständig ist und welche umweltbezogenen Vorhaben oder Projekte durchgeführt werden. Die Metadaten werden in sogenannten UDK-Objekten erfaßt und in einer Datenbank abgelegt. Die Angaben in einem UDK-Objekt beziehen sich auf die Identifizierung des Objektes (Objektname, allgemeine Beschreibung, Verschlagwortung, etc.) sowie auf den Fach-, Raum- und Zeitbezug der zu beschreibenden Daten. Bereits im Rahmen des o.a. FuE-Vorhabens ergab sich die Notwendigkeit, die UDK-Objekte zu klassifizieren, um den unterschiedlichen Datenquellen bei der Beschreibung gerecht zu werden (Lessing et al. 1995). Dieses Klassenkonzept wurde in der Praxis erprobt und im Rahmen der Neukonzeption an die Bedürfnisse der Nutzer angepaßt. Das überarbeitete Klassenkonzept sieht folgende Objektklassen vor:

- Datensammlung / Datenbank
- Dienst / Anwendung / Informationssystem
- Dokument / Bericht / Literatur
- Geo-Information / Karte
- Organisationseinheit / Fachaufgabe
- Vorhaben / Projekt / Programm

Die einzelnen Objektklassen unterscheiden sich im wesentlichen durch unterschiedliche Beschreibungsfelder im Fachbezug. Der UDK 4.0 zeichnet sich durch eine Flexibilität des Datenmodells aus. Über das optional zu installierende Modul UDK-Z (siehe unten) ist es möglich, weitere Objektklassen und Felder zu definieren.

Kernpunkt des DV-technischen Datenmodells ist die vom System automatisch erzeugte eindeutige Objektidentifikation (OID). Über diesen Primärschlüssel werden alle Relationen zwischen den Objekten hergestellt (Eltern-Kind-Beziehungen, horizontale Verweise). Auf die in früheren Programmversionen zum Aufbau von hierarchischen Strukturen verwendete dekadische Notation wurde aus Gründen der Performanz und mangelnder Benutzerfreundlichkeit verzichtet (Schütz/Lessing 1993).

Neben der Verwaltung von UDK-Objekten enthält der UDK eine einfache Adreßverwaltung. Die Adressen lassen sich hierarchisch beliebig anordnen, so daß auch die Abbildung von Organisationsstrukturen möglich ist. Jedem UDK-Objekt ist mindestens eine UDK-Adresse zugeordnet, so daß der recherchierende Nutzer beim Auffinden eines UDK-Objektes eine Ansprechperson angezeigt bekommt.

3 Der UDK-Explorer

Damit sich der Nutzer einen raschen Überblick über den Inhalt der Datenbank verschaffen kann, verfügt der UDK über ein Navigationssystem (UDK-Explorer), das dem Dateieexplorer von Windows 95 bzw. NT 4.0 nachempfunden ist (siehe Abbildung 1). Diese konzeptionelle Ähnlichkeit wurde bewußt gewählt, um Windows-erfahrenen Nutzern den Umgang mit dem UDK so leicht wie möglich zu machen.

Ein UDK-Objekt bzw. eine UDK-Adresse im UDK-Explorer entspricht einer Datei im Windows-Explorer. Der Unterschied zum Explorer besteht vor allem darin, daß auf das Konzept von Ordnern verzichtet wurde. Statt dessen können die UDK-Objekte bzw. UDK-Adressen selbst hierarchisch strukturiert werden (Strukturbaum). Über ein Vorschau- und ein Eigenschaftsfenster werden dem Nutzer bereits bei der Navigation durch den Datenbestand die wichtigsten Informationen zu jedem UDK-Objekt bzw. zu jeder UDK-Adresse angezeigt. Sind Detailinformationen von Interesse, kann der betreffende Datensatz aus dem Explorer heraus geöffnet und angezeigt werden.

In dem UDK-Explorer ist eine komfortable Suchmaschine für UDK-Objekte und UDK-Adressen integriert. Über eine Auswahlliste werden dem Nutzer die im Datenbestand verwendeten Suchbegriffe angeboten. Wahlweise kann er die Auswahlliste mit Schlagworten befüllen, die über einen Suchindex über alle Textfelder erzeugt wird. Weitere Suchfilter können über den Raum- und Zeitbezug sowie über die Objektklassen gesetzt werden. Schließlich kann noch zusätzlich über die sogenannte Expertensuche feldspezifisch (mit UND/ODER-Verknüpfungen) oder aber über freie SQL-Abfragen gesucht werden.

4 Der Thesaurus-Navigator

Ein integraler Bestandteil des UDK ist der Thesaurus der Umweltbundesämter Berlin und Wien (Umweltbundesamt Berlin/Wien 1997a). Basierend auf dem Produkt THESshow (Umweltbundesamt Berlin/Wien 1997b) wurde für den UDK der Thesaurus-Navigator entwickelt, mit dessen Hilfe der Nutzer von allgemeinen zu immer spezielleren Begriffen (Deskriptoren) navigieren kann (siehe Abbildung 2). Die UDK-Objekte, welche mit Deskriptoren verschlagwortet wurden, werden dem Nutzer bei der Navigation durch das strukturierte Wortgut dynamisch angezeigt.

Der hierarchische Thesaurus wird hierbei analog zum Strukturbaum der UDK-Objekte bzw. UDK-Adressen (UDK-Explorer) dargestellt. Aufgrund der Polyhierarchie werden zu jedem Deskriptor die zugehörigen Ober- und Unterbegriffe zusätzlich angezeigt. Bei Bedarf wird dem Nutzer der Einstieg in den Thesaurus über eine alphabetische Liste der Deskriptoren sowie zugehörigen Synonymen angeboten.

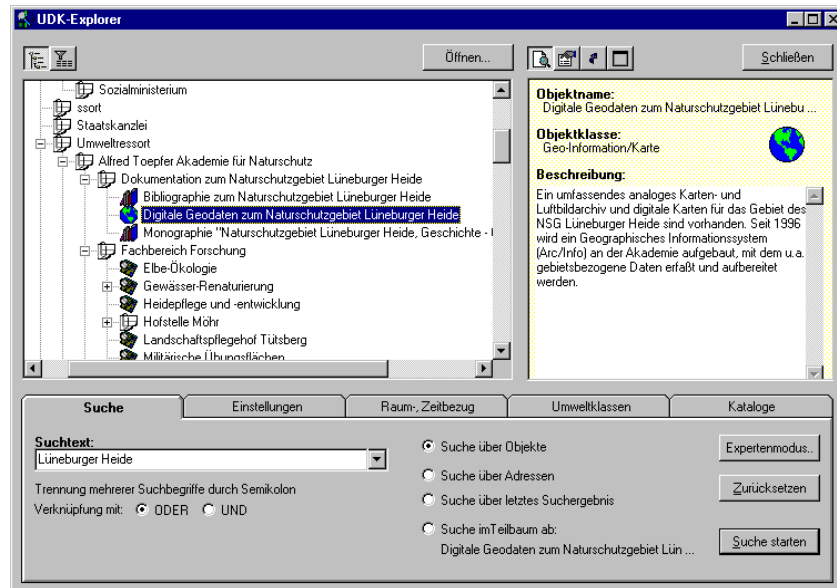


Abbildung 1
Der UDK-Explorer

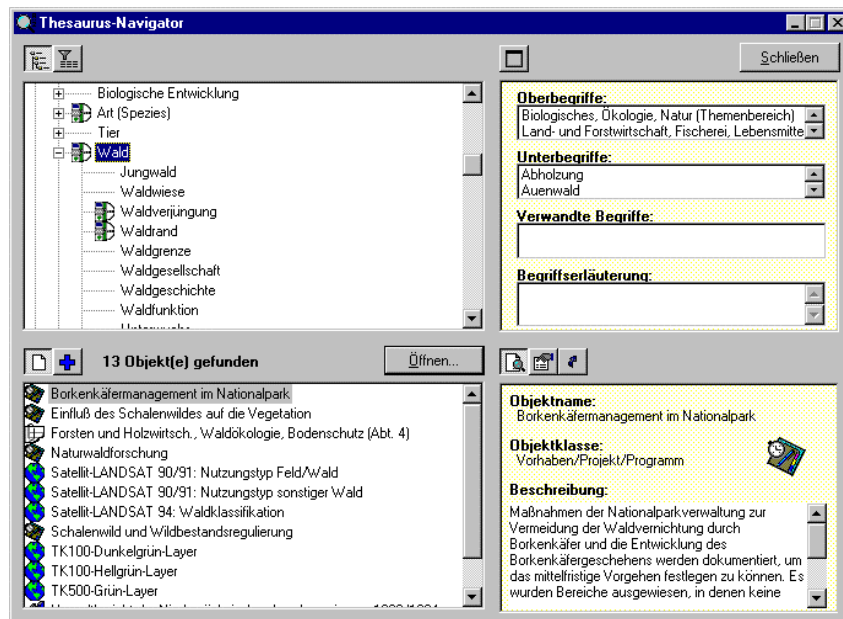


Abbildung 2
Der Thesaurus-Navigator

5 Die Geographische Suche

Beim Anlegen/Editieren eines UDK-Objekts kann der Nutzer den Raumbezug der zu beschreibenden Daten eintragen. Dies erfolgt durch die Zuordnung zur administrativen Einheit (Bundesland/Regierungsbezirk/Kreis/Gemeinde) mit automatischem Eintrag des umschließenden Rechtecks (Gauß-Krüger-Koordinaten, Bessel 3°).

Steht für den recherchierenden Nutzer die Suche nach UDK-Objekten über den Raumbezug im Vordergrund („welche umweltbezogenen Daten werden in meinem Wohnort erfasst?“), wird ihm eine geographische Suchmöglichkeit angeboten (siehe Abbildung 3). Mit Hilfe von zoombaren Karten, welche die Grenzen der administrativen Einheiten darstellen, können die in der Datenbank enthaltenen UDK-Objekte über deren Raumbezug lokalisiert werden. Die raumbezogene Suche erfolgt über die Selektion der interessierenden administrativen Einheiten oder über das Aufspannen eines Rechtecks. Es werden entweder alle den selektierten administrativen Einheiten zugeordneten UDK-Objekte oder aber alle UDK-Objekte mit Punkt bzw. Rechteckkoordinaten innerhalb des aufgespannten Rechtecks angezeigt. Analog zum UDK-Explorer erhält der Nutzer die objektspezifischen Grobinformationen über ein Vorschau- und ein Eigenschaftsfenster.

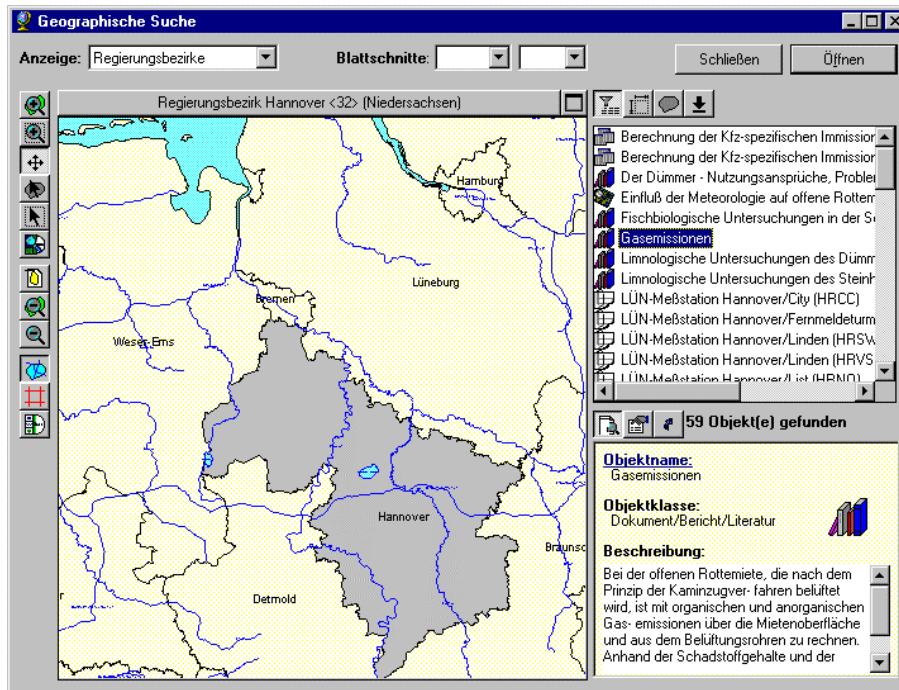


Abbildung 3
Die Geographische Suche

6 Das optionale Modul UDK-Z

Zur technischen Unterstützung des Zentralkatalogadministrators (ZKA) dient das optionale Modul UDK-Z. Es ermöglicht die zentrale Pflege des Datenbestandes eines Kataloges und stellt im wesentlichen folgende Funktionen zur Verfügung:

- Konsistenzprüfung der UDK-Objekte und UDK-Adressen
- Statistische Auswertungen der Kataloge
- Pflege von Auswahllisten
- Festlegung zusätzlicher Attribute und Objektklassen
- Retroindexierung für Thesaurusbegriffe

Ein wesentlicher Bestandteil des UDK-Z Moduls ist das dynamische Anlegen von zusätzlichen Attributen der einzelnen Objektklassen. Dialoggesteuert können freie Attribute angelegt werden, die dann bei der Erfassung der UDK-Objekte zur Verfügung stehen.

7 Die Erfassung der Metadaten: WWW contra konventionelle Technologien

Während die Ersterfassung der UDK-Objekte zumeist von zentraler Stelle organisiert wird, ist für die Pflege des Datenbestands eine dezentrale Struktur vorgesehen. Die Fachleute, welche über die zu beschreibenden Datenbestände verfügen, sollen die angelegten UDK-Objekte selbst ändern und ggf. neue UDK-Objekte hinzufügen können.

Zunächst bietet sich grundsätzlich die Erfassung und Pflege der UDK-Objekte über das WWW an. In der Praxis ist diese Technologie jedoch noch mit erheblichen Problemen verbunden. Zunächst ist in den Umweltverwaltungen der flächendeckende Zugang zum WWW direkt vom Arbeitsplatz noch bei weitem nicht gegeben (in den meisten Fällen ist selbst dann noch kein WWW-Zugang gegeben, wenn ein e-mail- oder X.400-Zugang vorliegt). Auf der technischen Ebene sind das Datenmodell sowie die geforderten Funktionalitäten des UDK für die Erfassung der Metadaten über das WWW noch zu komplex. Während bei konventioneller PC-Software die Bereitstellung umfangreicher Wortlisten (z.B. Thesaurus) ohne Probleme möglich ist, ist dies mit der WWW-Technologie noch immer mit erheblichen Wartezeiten verbunden.

Daher wurde von den Partnern der Bund-Länder-Verwaltungsvereinbarung beschlossen, die Erfassung der Metadaten mit konventioneller PC-Software zu organisieren, bis die oben genannten organisatorischen und technischen Probleme gelöst sind. Die Erfassung erfolgt dezentral über eine lokale Datenbank (stand-alone oder Client-Server). Die erfaßten Daten werden in einem regelmäßigen Aktualisierungszyklus (1-2 mal pro Jahr) an eine zentrale Stelle weitergegeben.

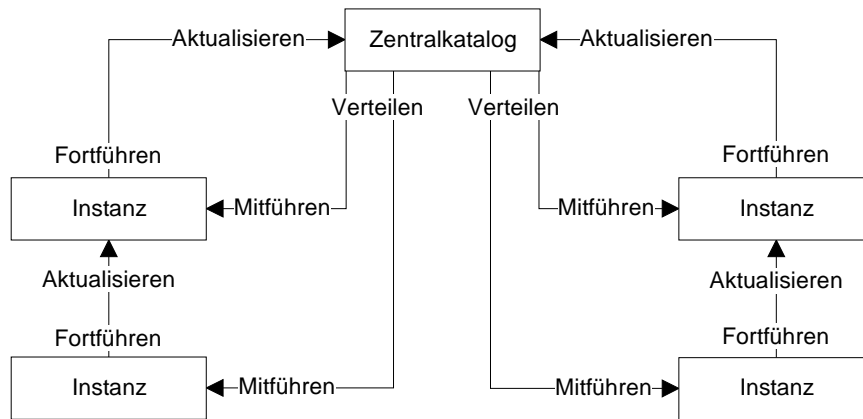


Abbildung 4
Der Aktualisierungszyklus

Technisch ist dieses so realisiert, daß mit jedem Datensatz in der UDK-Datenbank ein Aktualisierungskennzeichen gespeichert wird. Über dieses Kennzeichen ist es somit einfach, veränderte, hinzugefügte oder als gelöscht gekennzeichnete Datensätze zu identifizieren.

Für einen korrekten Aktualisierungszyklus ist es erforderlich, daß dieser über den hierarchischen Aufbau der Instanzen geführt wird (siehe Abbildung 4). Das heißt, eine untere Instanz erzeugt Dateien, welche dann an die jeweils übergeordnete Instanz weitergereicht werden (Fortführen). Bei der übergeordneten Instanz werden diese Daten eingelesen (Mitführen).

Im Anschluß werden hier, wie oben, die Daten wiederum fortgeführt und an die übergeordnete Instanz weitergereicht. Die Weiterreichen der Daten von einer Instanz zur jeweils übergeordneten wird so oft ausgeführt, bis die zentrale Stelle erreicht ist.

An zentraler Stelle (Zentralkatalog) werden die Änderungen aller UDK-Instanzen gesammelt und der Datenbestand aktualisiert.

Anschließend werden Dateien erzeugt (Verteilen), welche an alle UDK-Instanzen weitergereicht werden und dort eingelesen werden (Aktualisieren).

Die Weitergabe erfolgt über e-mail oder über den Austausch von Disketten. Die Aktualisierung über e-mail erfolgt automatisiert, so daß nach Ansicht der Autoren eine echte Alternative zur Erfassung über das WWW erreicht wurde.

Der automatische Aktualisierungszyklus ist technisch wie folgt realisiert:

Ein sogenannter „UDK-Agent“ durchsucht zyklisch die Datenbank nach veränderten, hinzugefügten und als gelöscht markierten Datensätze. Diese Datensätze

werden verpackt und - mit einem eindeutigen Identifier (Absender-ID + Laufende Nummer) versehen - an die übergeordnete Instanz verschickt. Der „UDK-Agent“ der übergeordneten Instanz überprüft kontinuierlich den Eingangsbereich des angeschlossenen e-mail-Systems. Der UDK-Agent nimmt dann automatisch, die sich aus den eintreffenden Sendungen ergebenden Datenbankänderungen vor. Nach dem gleichen Verfahren werden die entsprechenden Datensätze an alle untergeordnete Instanzen wieder verteilt (siehe Abbildung 5).

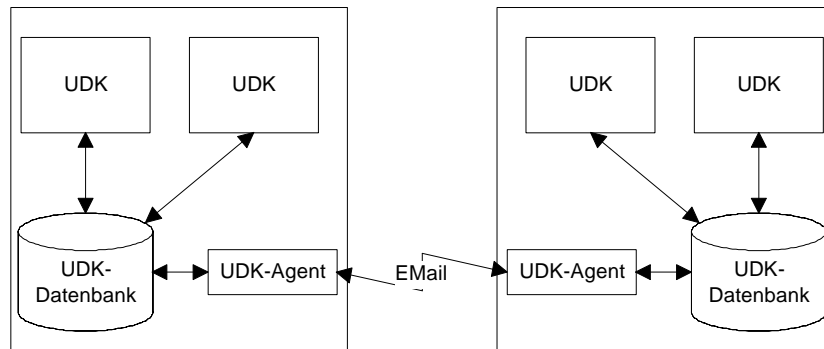


Abbildung 5
Der automatische Aktualisierungszyklus

Um Inkonsistenzen zu vermeiden, muß bei diesem Verfahren darauf geachtet werden, daß die e-mail-Sendungen einer untergeordneten Instanz vollständig und in der korrekten Reihenfolge in die Datenbank der übergeordneten Instanz eingespielt werden. Durch die fortlaufende Nummerierung wird dies gewährleistet, so daß z.B. „verlorengegangene“ e-mails automatisch neu angefordert werden.

8 Die technische Umsetzung

Die Software zur Erfassung der Metadaten wurde mit Visual Basic 5.0 entwickelt. Als Betriebssystem wird Windows 95 oder Windows NT 4.0 vorausgesetzt. Die Datenhaltung erfolgt in einer Access-Datenbank (stand-alone) oder in einer zentralen Datenbank (über ODBC). Grundsätzlich werden alle ODBC-fähigen Datenbanken unterstützt. Im Einsatz sind Microsoft-SQL-Server, Oracle, Informix und Ingres.

Die Grundzüge der Systemarchitektur ist in Abbildung 6 wiedergegeben.

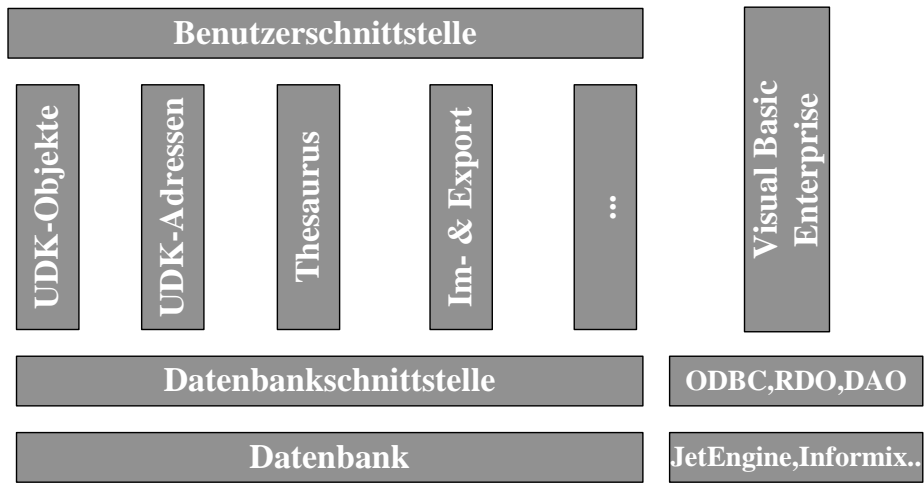


Abbildung 6
Systemarchitektur UDK 4.0

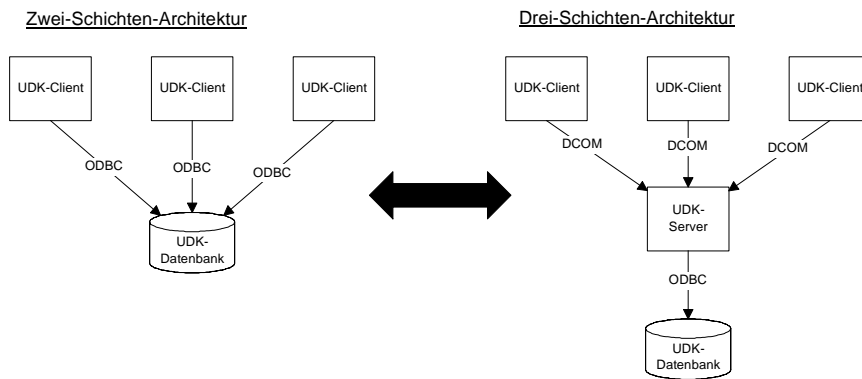


Abbildung 7
Vergleich der Zwei- und Drei-Schichten-Architektur

9 Ausblick

9.1 Drei-Schichten-Architektur

Üblicherweise werden Zwei-Schicht-Architekturen mit dem Begriff „Client/Server“ identifiziert. Die Datenbankzugriffe sind mit den Geschäftsregeln und der Benutzerschnittstelle gekapselt. Der Datenbankzugriff erfolgt also jeweils direkt von den einzelnen (UDK-) Clients über eine ODBC-Schnittstelle auf die Datenbank (siehe Abbildung 7).

Als Erweiterung zur Zwei-Schicht-Architektur wird eine dritte Schicht eingeführt (UDK-Server). Dadurch können komplexe Programmteile von der Client- und der Server-Seite zusammengezogen und an zentraler Stelle vereint werden. Die (UDK-) Clients greifen nicht mehr direkt auf Datenbankfunktionen zu, sondern kommunizieren nur mit der Server-Komponente. Datenbankupdates werden zentral aus diesem Server heraus gestartet. Alle Plausibilitätsprüfungen und die gesamte Transaktionslogik sind somit auf dieser Komponente verankert. Eine aufwendige Administration der ODBC Schnittstelle bei jedem (UDK-) Client entfällt, statt dessen ist dies nur noch zentral auf dem Server notwendig.

9.2 Der WWW-UDK

Die Veröffentlichung des UDK erfolgt bereits jetzt im WWW (Kramer et al. 1997). Bisher ist der WWW-UDK unter folgenden Adressen verfügbar:

Baden-Württemberg:

<http://www.uis-extern.um.bwl.de/cgi-bin/www-udk/umlenkung.sh/willkommen.sh>

Österreich:

<http://udk.bmu.gv.at>

Niedersachsen:

<http://www.mu.niedersachsen.de>

Die Anpassung des WWW-UDK auf das im Rahmen der Neukonzeption überarbeitete Datenmodell ist für die zweite Jahreshälfte 1998 vorgesehen.

9.3 UDK und GEIN 2000

Der Bund erstellt zur Zeit das Umweltinformationsnetz Deutschland (German Environmental Information Network / GEIN 2000), welches Umweltinformationen des Bundes und der Länder bereitstellen soll (Seggelke/Lessing 1996). GEIN 2000 soll im Rahmen der Weltausstellung Expo 2000 präsentiert werden. Es ist vorgesehen, daß der UDK eine zentrale Komponente als Metainformationssystem einnimmt.

Literaturverzeichnis

- Kramer, R., Nikolai, R., Koschel, A., Rolker, C., Lockemann, P., Keitel, A., Legat, R., Zirm, K. (1997): WWW-UDK: A Web-based Environmental Metainformation System, in: ACM SIGMOD Record, Bd. 26, Nr. 1, S. 16-25, New York
- Lessing, H., Schütz, T. (1994): Der Umwelt-Datenkatalog als Instrument zur Steuerung von Informationsflüssen, in: Informatik für den Umweltschutz, 8. Symposium, Metropolis-Verlag, Marburg
- Lessing, H., Schmalz, R. (1994): Der Umwelt-Datenkatalog Niedersachsens, in: Umweltinformationssysteme in der öffentlichen Verwaltung, R. v. Decker's Verlag, Heidelberg
- Lessing, H., Günther, O., Swoboda, W. (1995): Ein objektorientiertes Klassenkonzept für den Umwelt-Datenkatalog (UDK), in: Informatik für den Umweltschutz, 9. Symposium, Metropolis-Verlag, Marburg
- Seggelke, J., Lessing, H. (1996): Globales Umweltinformationsnetz: Eckpunkte, Chancen und Gefahren, in: Informatik für den Umweltschutz, 10. Symposium, Metropolis-Verlag, Marburg
- Schütz, T., Lessing, H. (1993): Metainformation von Umwelt-Datenobjekten - Zum Datenmodell des Umwelt-Datenkataloges Niedersachsens, in: Informatik für den Umweltschutz, 7. Symposium, Springer-Verlag, Berlin
- Umweltbundesamt Berlin, Umweltbundesamt Wien (Hrsg.) (1997): Thesaurus des Umweltdatenkatalogs (UDK-Thesaurus), Bd. 1-3, Berlin, Wien
- Umweltbundesamt Berlin, Umweltbundesamt Wien (Hrsg.) (1997): THESshow, CD-ROM, Berlin, Wien