

UMWELTDATENBANKEN IM WEB: Workshop-Beiträge und Ergebnisse

**Workshop des Arbeitskreises Umweltdatenbanken
vom 10. und 11. Juni 1999 in Karlsruhe**

Wien, 2000

Impressum

Medieninhaber und Herausgeber: Umweltbundesamt GmbH (Federal Environment Agency Ltd)
Spittelauer Lände 5, A-1090 Wien (Vienna), Austria

Mitherausgeber: Ralf Kramer, Friedhelm Hosenfeld, Veranstalter des Workshops und Sprecher des GI-
Arbeitskreises „Umweltdatenbanken“

Die unverändert abgedruckten Einzelbeiträge geben die Fachmeinung der Autoren wieder.

Druck: Riegelnik, 1080 Wien

© Umweltbundesamt GmbH, Wien; Forschungszentrum Informatik an der Universität Karlsruhe,
Haid-und-Neu-Str. 10-14, 76131 Karlsruhe; 2000
Alle Rechte vorbehalten (all rights reserved)
ISBN 3-85457-538-6

Inhalt

Vorwort

Umweltdatenbanken im Web – Stand der Konzepte und Entwicklungen

Friedhelm Hosenfeld,

1	Überblick	9
2	XML (Extensible Markup Language).....	11
3	Metadaten	13
4	Konzepte verteilter Kataloge	14
5	Erfahrungen und Potential des Umweltdatenkatalogs (UDK)	15
6	Öffentliche Verwaltungen: Erfahrungen und Perspektiven	16
7	Verteilte Systeme.....	17

Kartengestützte und internetbasierte Auskunft über die geographische Verbreitung von Pflanzen und Tieren

Jens Fitzke und Torsten Friebe

1	Ausgangssituation.....	18
2	Fachkonzept.....	19
3	Technisches Design der Anwendungssoftware.....	22
4	Ausblick.....	26
5	Referenzen	27

Regionale Umweltdaten in einem internationalen Projekt: Datenmanagement des trilateralen Monitoring- und Bewertungs- programms TMAP in Schleswig-Holstein und Hamburg

Claus Hoerschelmann

1	Einleitung.....	28
2	Ziele des Monitorings im Wattenmeer.....	29
3	Räumlicher und organisatorischer Rahmen und daraus resultierende Strukturvorgaben für das Projekt.....	29
4	Struktur der TMAP Data Units	30
5	Verwendetes Datenmodell am Beispiel einiger Parametergruppen	31
6	Technische Aspekte	33
7	Zusammenarbeit mit Originatoren der Daten, Grundlagen und praktische Erfahrungen	33
8	Stand der Arbeiten	34
9	Fazit und Ausblick	35
10	Literaturverzeichnis	35

Repräsentation der Struktur biologischer Sammlungen als Grundlage für die Schaffung eines europäischen Metainformationssystems

Anton Güntsch, Andrea Hahn und Walter G. Berendsohn

1	Einleitung.....	37
2	Aus Umfragen gewonnene Information über biologische Sammlungen	39
3	Darstellung von Sammlungsstrukturen.....	42
4	Regelbasierte Verschlagwortung	46
	Literatur	50

Zur Dynamik von Metadaten

Paul Braun

1	Einleitung.....	52
2	Hintergrund.....	52
3	Aufgabenstellung.....	54
4	Herangehensweise	54
5	Schlußfolgerung.....	60
	Literatur	61

Praktische Erfahrungen bei der Datenerschließung für eine Umweltdatenbank

Thomas Dombeck und Klaus Tochtermann

1	Einleitung.....	62
2	Die Umweltdatenbank „Datenbank Grunddaten“	62
3	Datenerschließung in der „Datenbank Grunddaten“	64
4	Anforderungen an Umweltdatenbanken aus Anwendersicht	68
	Literatur	69

Der Datenaustausch für den Alpendatenquellenkatalog

Thomas Heinemeier

1	Einführung	70
2	Die Rahmenbedingungen des ADQK	71
3	Operative Grundsätze	72
4	Darstellung des Datenflusses	73
5	Der Einfluß der Mehrsprachigkeit	78
6	Schlußfolgerung.....	78
7	Danksagung	79
8	Literaturverzeichnis	79

Der Umweltdatenkatalog UDK in Österreich – 5 Jahre Erfahrungen

Legat, Batschi, Hashemi-Kepp, Kruse, Nikolai, Nyhuis, Pultz, Stallbaumer, Swoboda, Zirm

1	Einleitung.....	83
2	Win-UDK 4.0	85
3	Umweltinformationssystem WWW-UDK.....	89
4	Die Befüllung des österreichischen UDK.....	93
5	Umweltthesaurus	96
6	Internationale Zusammenarbeit mit dem ETC/CDS	104
7	Ausblick.....	105
	Literaturverzeichnis.....	106
	Weiterführende Literatur	106

Erfassungsverfahren für Umweltmetadaten im Umweltdatenkatalog in Baden-Württemberg

Thomas Sattler, Ralf Nikolai

1	Einleitung.....	108
2	Einbindung des UDK ins WWW-Angebot der Landesanstalt für Umweltschutz Baden-Württemberg.....	109
3	Metadatenerfassung	110
4	Diskussion	116
5	Ausblick.....	117
6	Literatur	118

Modellgetriebene Integration von Fachverfahren in Informationssysteme der öffentlichen Verwaltung

Ulrike Freitag, Hans-Peter Schreiter, Thomas Schwotzer

1	Einleitung.....	120
2	Beispiel-Projekte.....	122
3	Modellgetriebener Generatoransatz	125
4	Middleware Framework für lightweight-Components <i>Core</i>	130
5	Zusammenfassung und Ausblick	134

Ein Internet-basiertes Umweltdokumentsystem für kleine Nutzergruppen

Klaus Tochtermann, Andreas Kussmaul und Katrin Pursche

1	Einleitung.....	137
2	Anforderungen an das System und prinzipieller Arbeitsablauf.....	138
3	Technische Realisierung.....	142
4	Bewertung und Ausblick	148

Vorwort

In den letzten Jahren haben sich die Techniken und Technologien des World Wide Web als ein wesentliches Mittel für den Zugang zu Datenbanken mit umweltrelevanten Datenbeständen erwiesen. Neben ihrem Inhalt unterscheiden sich Umweltdaten u. a. durch ihren Raum- und Zeitbezug von anderen Daten. Sowohl im Internet, aber auch in Intranets ermöglichen WWW-Techniken einen kostengünstigen Zugang von unterschiedlichen Client-Plattformen. Mit zunehmender Erfahrung stellten sich aber vermehrt auch die Grenzen „klassischer“, beispielsweise CGI-basierter Ansätze heraus. Dies spiegelt sich im zunehmenden Einsatz von Java und CORBA beim Zugriff auf Umweltdatenbanken wider. Aufgrund der zunehmenden Nutzerzahlen erlangen die Probleme von Datenerfassung und -aktualität eine wachsende Bedeutung: Gerade bei einer breiten Nutzung fallen Defizite bei Datenqualität und -quantität verstärkt auf.

Mit dem Workshop „Umweltdatenbanken im Web: Technologische und praktische Herausforderungen“ wurden daher zwei wesentliche Zielrichtungen verfolgt. Zum einen wurden technisch/wissenschaftliche Beiträge vorgestellt, die über neue technologische Möglichkeiten und erste Erfahrungen mit diesen in Verbindung mit Umweltdatenbanken berichten. Stichworte sind hierzu u. a. Java und verwandte Techniken (JDBC, Java RMI, ...) und XML (Extensible Markup Language). Zum anderen wurden Beiträge von Anwendern vorgestellt, die über die praktischen Fragestellungen bei der Erfassung und Aktualisierung von Umwelt(meta)daten berichten. Im vorliegenden Band zum Workshop des Arbeitskreises Umweltdatenbanken der Gesellschaft für Informatik (GI) sind ausgewählte Beiträge des Workshops mit dem gleichen Titel zusammengefaßt, der im Juni 1999 am Forschungszentrum Informatik (FZI) in Karlsruhe, Deutschland, stattfand.

Organisiert war der Workshop in die Sessions Einführung und XML, Metadaten, Konzepte verteilter Kataloge, Erfahrungen und Potential des Umweltdatenkatalogs (UDK), Öffentliche Verwaltung – Erfahrung und Perspektiven – sowie verteilte Systeme. Den Beiträgen vorangestellt haben wir einen aktuellen Überblick über den Stand der Technik bei Umweltdatenbanken im Web. Ergänzende Informationen zu den hier in gedruckter Form vorliegenden Beiträgen finden sich im Web ausgehend von der Homepage des Arbeitskreises Umweltdatenbanken (<http://www.fzi.de/dbs/misc/akudb.html>) sowie unmittelbar unter der URL <http://www.pz-oekosys.uni-kiel.de/akudb>.

Besonders bedanken möchten sich die Herausgeber bei den Mitarbeiterinnen und Mitarbeitern des Umweltbundesamtes in Wien (Lisa Löschl, Rudolf Legat sowie Dr. Johannes Mayer), die die Herausgabe dieses Bandes parallel zu der im Web vor-

liegenden Seite mit den Workshop-Ergebnissen ermöglicht haben. Damit haben wir eine Lösung gefunden, die – ganz im Sinne der auch bei anderen Druckerzeugnissen aktuell zu beobachtenden Trends – zum einen eine adäquate schriftliche Dokumentation der Workshop-Ergebnisse bietet und zum anderen zugleich einen Online-Zugriff auch auf weitergehende Informationen ermöglicht.

Karlsruhe, Kiel im März 2000

Dr. Ralf Kramer, Friedhelm Hosenfeld

Umweltdatenbanken im Web – Stand der Konzepte und Entwicklungen

Friedhelm Hosenfeld¹

1 Überblick

Das Thema *Umweltdatenbanken im Web* wurde im Rahmen eines Workshops anhand verschiedener Aspekte untersucht. Die einzelnen Beiträge können natürlich nur jeweils Ausschnitte des Problemkreises berühren. Sie und die durch sie angestoßenen Diskussionen, die auch Randbereiche der vorgetragenen Themen einbezogen, geben in ihrer Gesamtheit jedoch einen guten Überblick über den konzeptionellen Stand im Bereich Umweltdatenbanken im Web sowie einen Blick in die Zukunft der Entwicklungen in Deutschland und Österreich. Zunächst soll eine kleine Übersicht über die in den folgenden Kapiteln dargestellten Themenblöcke gegeben werden, die wiederum den Diskussionsstand des Workshops kurz referieren.

Mit *XML (Extended Markup Language)* wurde eine noch junge Methodik, die momentan noch sehr Bewegung ist, zu einem der Schwerpunkte gemacht. XML wird zweifellos an Aktualität und Bedeutung gerade für Web-basierte Umweltdatenbanken gewinnen. Dies spiegelt sich auch in verschiedenen der vorgestellten Beiträgen wider, die zumindest mittelfristig die Prüfung einer XML-Einführung ankündigen (Braun, Freitag et al., Fitzke/Friebe, Tochtermann et al., Sattler/Nikolai, alle in diesem Band). Um so wichtiger ist es, bereits in dem Stadium, in dem noch geeignete Einsatzfelder für diesen Standard gesucht werden, die Anwendbarkeit für die Umweltinformatik zu prüfen.

Neben diesem relativ neuen Thema, das bewußt im Rahmen des Workshops etwas fundierter dargestellt wurde, zeigte sich jedoch auch, daß für „Dauerbrenner“ in der Umweltinformatik, wie etwa für den Bereich *Metadaten*, weiterhin Bedarf besteht, sie vor dem Hintergrund aktueller Entwicklungen zu analysieren und zu diskutieren. Nicht nur die Untersuchung neuer Ansätze zum Metadatenmanagement ist interessant. Es ist auch an der Zeit, aus den bisherigen Erfahrungen zu lernen. Das Know-How-Potential, das durch die Erfahrungen mit den aktuellen und auch den bereits als überholt geltenden Realisierungen aufgebaut wurde, scheint noch nicht hinreichend ausgenutzt worden zu sein. Die erarbeiteten De-Facto-Standards, aber auch Kenntnisse über den Umgang mit konzeptionellen Fehlern oder Lehren aus Rückschlägen durch

¹ Lüdemannstr. 31, D-24114 Kiel, email: friedel@hosenfeld.de,
Internet: <http://www.hosenfeld.de/>

mangelnde Akzeptanz scheinen nicht bekannt genug zu sein. Oft werden ähnliche Lösungen neu erarbeitet, ohne bewährte Konzepte zu berücksichtigen und gegebenenfalls anzupassen oder weiterzuentwickeln.

Ein anderes Thema betrifft verteilte Kataloge. Der Bedarf an *Konzepten für verteilte Kataloge* entsteht aus mehreren Gründen: An verschiedenen Stellen sind bereits Datenkataloge und Metadatenkataloge entstanden. Für die Benutzenden ist es sinnvoll, mehrere Kataloge, deren Informationen sich – bezogen auf eine Fragestellung – ergänzen, einheitlich befragen zu können. Wünschenswert ist nicht nur eine einheitliche Bedienung, sondern auch die Abfragemöglichkeit zentral von einer Stelle aus bzw. die Abfrage verschiedener Kataloge von jeder beteiligten Stelle aus. Nicht nur durch historisch gewachsene Gegebenheiten entstehen verteilte Kataloge. Auch aufgrund von Kompetenz- und Aufgabenteilung kann es sinnvoll und notwendig sein, Datenkataloge verteilt zu halten. Naheliegend ist es, die Techniken des Internet in diese Konzeptionen mit einzubeziehen.

Speziell ein System, nämlich der *Umweltdatenkatalog (UDK)*, ist seit einiger Zeit auf dem Weg, einen wichtigen Standard für Metadatenkatalogsysteme im Umweltbereich zu definieren. Hier gibt es in Form des *Virtuellen UDK* bereits eine Realisierung eines verteilten Katalogs, der auf verschiedene Instanzen des UDK zugreift, denn außer für Umweltinformationen des Landes Österreich wird er für das deutsche Umweltbundesamt und zahlreiche weitere Bundesländer eingesetzt (Legat et al., in diesem Band). Auf die Entwicklung des europäischen CDS (Catalog of Data Sources), betrieben durch die Europäische Umweltagentur, hatten die Erfahrungen des UDK ebenfalls erheblichen Einfluß. Aufgrund dieses Stellenwerts wurde dem Austausch über die bisherigen langjährigen Erfahrungen und den zukünftigen Entwicklungsrichtungen eine eigene Session gewidmet.

Während der UDK eine Art Universalsystem darstellt, das Metainformationen austauschbar und allgemein zugreifbar verwaltet, entstehen in den *öffentlichen Verwaltungen* Anforderungen nach maßgeschneiderten Systemen, die Standardapplikationen integrieren und unterstützen. Gewünscht werden einfach zu bedienende Systeme, die teilweise nur für kleinere Benutzergruppen geeignet sind, aber an deren Arbeitsablauf und Tätigkeitsfelder angepaßt sind. Software „von der Stange“ ist oftmals nicht bedarfsgerecht einzusetzen. Bei der Realisierung eigener Lösungen ist das Verwenden proprietärer Techniken trotzdem weitgehend zu vermeiden, um eine Erweiterbarkeit nicht zu erschweren.

Ein letzter Schwerpunkt des Workshops lag auf der Analyse *verteilter Systeme*, deren Aufgabe in der verteilten Datenhaltung umweltwissenschaftlicher Daten besteht, die einheitlich zugänglich sein sollen. Auch hier kann das Web sowohl für die Recherche innerhalb der verteilten Systeme als auch für das Einstellen und Dokumentieren der Daten Lösungswege unterstützen. Gefordert ist ein integrierendes Datenmanagement, das Datenharmonisierung und die Einbeziehung heterogener Datenquellen fördert. Das Web bietet die Möglichkeit, Teillösungen bereits früh in Er-

probungsphasen bereitzustellen und den Nutzenden zur Begutachtung vorzulegen. Rückmeldungen können so bereits im Laufe der weiteren Entwicklung umgesetzt werden.

Der Charakter eines Workshops wie des hier vorgestellten wird nicht nur durch die einzelnen Vorträge deutlich, sondern auch durch die bereits erwähnten fundierten Diskussionen, die teilweise erst die Einordnung der Beiträge in den Gesamtkontext der Entwicklungen ermöglichen. Aus diesem Grund ist der Darstellung der Diskussionen und Anmerkungen ein eigener Beitrag gewidmet.

Diese Zusammenfassung der dort geäußerten Meinungen und Standpunkte gliedert sich der Einfachheit halber nach den thematischen Sessions des Workshops, obwohl sowohl in bezug auf die Vorträge als auch auf die Diskussionen eine klare Aufteilung auf die thematischen Aspekte nicht immer vorgenommen werden kann, denn zu sehr überschneiden sich angesprochene Konzepte und Lösungen.

2 XML (Extensible Markup Language)

An den zukünftigen Einsatz von XML² im WWW werden hohe aber teilweise noch recht diffuse Erwartungen geknüpft. Das Bild von XML ist auch auf Seiten der Anbieter von Umweltinformationen im Internet nicht klar umrissen. So ist es notwendig, sowohl die Fähigkeiten von XML mit den potentiellen Innovationsmöglichkeiten aufzuzeigen, als auch das zu verdeutlichen, was XML nicht bietet und auch die Funktionalitäten, die zwar mit XML realisiert werden können, aber nicht unmittelbar auf XML basieren.

XML kann als weltweites Datenmodell angesehen werden. Die Sprache kann so zum Beispiel sehr gut als globales Austauschformat für Daten eingesetzt werden. Das Problem der Festlegung der Semantik kann durch XML nicht gelöst werden, eine strukturelle Lösung wird aber unterstützt.

Seiteneffekte, die durch XML-basierte Dienste auftreten, werden oft zu Unrecht den Eigenschaften der Sprache XML zugeschrieben. Als Beispiel ist die nutzerabhängige Informationspräsentation zu nennen. Von diesem wichtigen Vorteil in der Übermittlung von Informationen im WWW profitieren einerseits die Benutzer, die auf diese Weise Daten, die auf ihre Interessen abgestimmt sind, in der ihnen gemäßen Form präsentiert bekommen. Andererseits kann der Informationsanbieter gezielt auf die Nachfrage der Benutzenden reagieren. Die Informationen können effektiver vermittelt werden (Beispiele: gewünschte Sprache der Benutzerführung, Komplexität des Aufbaus, Vordefinition von Standardangaben wie Adressen). Die generellen Bedenken gegen die Personalisierung und Verfolgbarkeit von Web-Nutzungen werden weder durch XML verursacht noch dadurch abgebaut. Sicherheitsaspekte müssen auf anderen Ebenen konzeptionell berücksichtigt werden. Die Gefahr der Zensur durch gezielte

² Extensible Markup Language (XML): <http://www.w3c.org/xml/>
The XML Cover Pages: <http://www.oasis-open.org/cover/>

nutzerabhängige Einflußnahme auf die bereitgestellten Informationen ist kein durch XML entstandenes Problem sondern eine inhaltliche Fragestellung, die von den Informationsanbietern zufriedenstellend bearbeitet werden muß.

Für das Angebot von Umweltinformationen im Web muß die Anwendbarkeit der XML-Anfragesprache XQL (XML Query Language³) analysiert werden. XQL bietet die Möglichkeit, mit XML strukturierte Dokumente zu durchsuchen und Anfragen ähnlich wie in Anfragesprachen relationaler Datenbanksysteme zu formulieren. Eine interessante Frage berührt daher die zukünftige Rolle von Datenbanken im Web. Die – je nachdem – als Rückschritt oder Fortschritt beobachtete Entwicklung, daß flache Files Datenbanken wiederum funktional ersetzen, kann nicht bestätigt werden. Datenbanken werden im Web nicht überflüssig, ihre Bedeutung wird im Gegenteil weiterhin zunehmen. Die Aufgaben von dateibasierten Web-Inhalten und Datenbanken im Web lassen sich durch XML besser aufteilen und effektiver kombinieren. Ebenso ist der Aspekt zu erwähnen, daß das Web mit Hilfe von XML (und XQL) deklarativ durchsuchbar gemacht werden kann.

Mit dem XML-„Metadatendefinitionsformat“ RDF⁴ (Resource Description Framework) liegt ein Basis-Standard zur Verarbeitung von Metadaten vor. Damit kann RDF zur automatischen Verarbeitung von Web-Ressourcen verwendet werden⁵.

Die Frage liegt nahe, warum eigens RDF eingeführt wurde, anstatt gleich eine XML DTD (Document Type Defintion) oder SGML (Standard Generalized Markup Language) zu verwenden. Ein Motiv besteht darin, daß ein RDF-Datenschema die Interpretation der Statements des RDF-Datenmodells spezifiziert, anders als eine XML DTD, mit der die Bedingungen für die Struktur eines Dokuments festgelegt werden. Mit RDF können zudem Konsistenzbedingungen auf dem Datenmodell angegeben werden.

Neben der Metadatendefinition besitzen für den Umweltinformationsbereich, speziell zur Darstellung von Geodaten, Vektor-Graphik-Formate, die mittels XML definiert werden, hohe Relevanz.

Ein Grund des großen Datenaufkommens von zusätzlichen XML-Tags, die über das Netz übertragen werden müssen (bei browserseitiger XML-Interpretation), liegt sicher in der Verwendung des universellen ASCII-Formats. Die Überlegung, eine binärcodierte Repräsentation von XML zu schaffen, scheint in die falsche Richtung zu gehen, da das ASCII-Format viele elementare Vorteile aufweist. Interessanter und als denkbare Lösung mit breitem Einsatzgebiet vorstellbar ist eine Online-Kompression und -Dekompression auf der darunterliegenden Protokollebene.

³ Jonathan Robie, R&D Fellow, Software AG, XQL Tutorial,
<http://metalab.unc.edu/xql/xql-tutorial.html>

⁴ XML Resource Guide: RDF, <http://www.xml.com/pub/Guide/RDF/>

⁵ http://www.xml.com/pub/r/RDF_Model_and_Syntax

3 Metadaten

Obwohl die Diskussion über den gesamten Themenkreis *Metadaten* bereits in den letzten Jahren breiten Raum einnahm, scheint das Interesse und auch die Notwendigkeit weiter zu bestehen, sich dieses Themas unter aktuellen Fragestellungen anzunehmen. Das liegt einerseits sicher in den neuesten technischen und konzeptionellen Entwicklungen im Metadatenbereich begründet, andererseits aber auch in dem zunehmenden Bedarf nach dem Einsatz tragfähiger Konzepte.

Inzwischen liegen langjährige Erfahrungen mit dem Einsatz von Metadaten-Konzepten vor, die ausgewertet und für neue Projekte nutzbar gemacht werden können.

Zwei bedeutsame Aspekte, die zunächst gegensätzlich wirken, treten hervor: Einerseits lautet der Appell an diejenigen, die derzeit mit System-Neuentwicklungen befaßt sind, nicht immer wieder das „Rad neu zu erfinden“ und eigene „Standards“ zu erstellen, die sich nur marginal von bestehenden, bewährten unterscheiden. Andererseits ist ganz offensichtlich der Bedarf für verschiedene Metadatenstandards vorhanden. Die Vorstellung, mit *einem* Standard die Aufgaben *aller* Umweltdatenbanken und informationssysteme im Metadatenbereich zu lösen, ist utopisch.

Die Frage, warum es immer wieder zu Neuentwicklungen kommt, die bestehenden und bewährten Standards ähneln, läßt sich unter anderem auch mit dem noch steigerungsfähigen Bekanntheitsgrad der vorhandenen Konzepte beantworten. Eine wichtige Aufgabe von Workshops und Publikationen, wie dem vorliegenden Band, besteht darin, Öffentlichkeit über die aktuellen in der Entwicklung und Anwendung befindlichen Standards und Konzepte herzustellen.

Produkte wie etwa UDK (Umweltdatenkatalog, Legat et al. in diesem Band), CDS (Catalogue of Data Sources), GELOS⁶ oder auch der Dublin Core⁷ Standard dürfen nicht nur Schlagworte bleiben, sondern müssen in bezug auf Anwendbarkeit und Aufgabenbereiche bekannt sein.

Der Kritikpunkt an ausgedehnten, allgemeineren Standards besteht oft in der mangelnden Eignung für sehr spezielle Anwendungsgebiete. Die „große Keule“ allumfassender Vorschriften wirkt einerseits wie das Zielen mit Kanonen auf Spatzen und scheint dann doch noch die speziellen Ziele zu verfehlen, weil die Erfassung der benötigten Feinheiten nicht in dem detaillierten Maße möglich ist. Konflikte treten durch unterschiedliche Schwerpunktsetzung z. B. für Zeitreihen, Geodaten, wissenschaftliche Daten, Kontakt- und Adreßinformationen, Dynamik und Multimedialität auf.

Dieser Konflikt wird in absehbarer Zeit nicht lösbar sein. Eine Annäherung sollte trotzdem stattfinden. Es darf nicht vergessen werden, daß auch die mittlerweile unumstrittenen Standards noch Relikte historischer Entwicklungen und Randbedingungen mit sich herumtragen und daher verbesserungswürdig bleiben.

⁶ Global Environmental Information Locator Service (GELOS): <http://ceo.gelos.org/>

⁷ Dublin Core Metadata Initiative: <http://purl.oclc.org/dc/>

Da die Beibehaltung und Fortführung unterschiedlicher Standards notwendig bleiben wird, lautet eine in der Diskussion geäußerte Forderung, sollte wenigstens die Interoperabilität der verschiedenen Lösungen angestrebt werden. Auch bei Neuentwicklungen von Metadaten-Systemen sollte die Interoperabilität mit bestehenden Standards geprüft und berücksichtigt werden.

Nationale bzw. bundesländer-spezifische Rahmenbedingungen und ähnliche Vorgaben behindern gelegentlich optimale Lösungen, so daß mit suboptimalen Vorgaben entwickelt werden muß (Hoerschelmann, Heinemeier, i. d. Bd.).

Ein anerkanntes Instrument zur Vereinheitlichung und Vergleichbarkeit von (Meta-)Informationen stellen Thesauri dar (Legat et al., i. d. Bd.). Die Thesaurus-Entwicklung in der Umweltinformatik hat in den letzten Jahren Fortschritte gemacht. Dennoch bleiben die Konzepte optimierbar. Nicht immer sind die Fehler in der Zusammensetzung des Thesaurus zu suchen, denn die Einsatzgebiete müssen sich für den gewünschten Thesaurus als geeignet erweisen. Nicht alle Thesauri bestimmter Systeme können wie Allzweckwerkzeuge in komplett unterschiedlichen Zusammenhängen verwendet werden (Dombeck/Tochtermann, i. d. Bd.). Ebenfalls problematisch ist die Übernahme nur von Teilen eines Thesaurus in andere Systeme.

Als hilfreich für den erfolgreichen Einsatz von Thesauri im Rahmen von Informationssystemen erweisen sich Richtlinien zur Verschlagwortung und allgemein zur Metadaterfassung. Auf diesem Gebiet besteht Bedarf nach entsprechenden Erfahrungsberichten und Leitlinien (Dombeck/Tochtermann, i. d. Bd.).

Mit kürzeren Entwicklungszyklen, die nicht nur die Informationssysteme, sondern in einem erheblichen Ausmaß auch Datenformate betreffen, stellt sich die Frage der „Haltbarkeit“ von Daten und Metadaten (Braun, i. d. Bd.). Bei offline archivierten Medien tangiert dies neben der Software, die in der Lage sein muß, Daten weiterhin zu lesen und gegebenenfalls zu konvertieren, unter anderem auch die Hardware. Die Verwendung langlebiger Formate, deren Zukunft absehbar positiv beurteilt wird, ist anzustreben. Daneben muß immer der Aufwand für das Umkopieren und Umformulieren von Daten einkalkuliert werden, sollen diese nicht zu einem bestimmten Zeitpunkt verloren sein.

4 Konzepte verteilter Kataloge

Verteilte Kataloge müssen zudem die mit der heterogenen Granularität von Metainformationen zusammenhängenden Problematiken bewältigen. Der Übergang zwischen Metadaten und den eigentlichen Informationen ist nicht immer klar abgrenzbar, er kann fließend verlaufen. Diese Frage der Grenzziehung bekommt Relevanz, wenn es darum geht, festzulegen, an welcher Stelle welche Information gepflegt und nachgeführt werden muß. Es muß zudem analysiert werden, an welchen Stellen dies überhaupt durchgeführt werden kann. Wenn das Entstehen eines Update-Problems durch konzeptionelle Gegebenheiten bei der verteilten Datenhaltung nicht verhindert werden kann, so

sollte dies zumindest minimiert werden. So stellen die Einrichtung von definierten Update-Mechanismen oder die Bestimmung von Master-Systemen für bestimmte Informationen Wege zum Umgang mit der Problematik dar. Für die Benutzenden der verteilten Systeme muß jedoch transparent bleiben, ob sie auf eine unter Umständen inaktuelle Kopie zugreifen oder ob mit verlängerten Antwortzeiten gerechnet werden muß, wenn auf verschiedene Kataloge gleichzeitig zugegriffen wird.

Metainformationssysteme, die Auskunft über Informationen in anderen Systemen geben, die sie selbst wiederum abfragen (z. B. GEIN 2000: Bandholtz, i. d. Bd.) müssen ähnliche Schwierigkeiten bewältigen: Sie geben – unter Umständen statisch gewartete – Informationen, die in den Bezugssystemen wiederum dynamisch verwaltet werden. Da bei Änderungen der Quell-Systeme kein automatisches Aktualisieren des („übergeordneten“) Metainformationssystems stattfinden kann, kommt es zu Inkonsistenzen zwischen Verweissystem und tatsächlich verwalteter Information. Internet und dem wachsenden Angebot von Web-Zugängen zu Umweltinformationen besteht die Möglichkeit, kostengünstig Informationen verschiedener Quellen miteinander in Beziehung zu setzen. Seitens der Benutzenden wächst verstärkt das Bedürfnis, verschiedene Systeme einheitlich abzufragen. Zur besseren Integration verschiedener Informationen in den einzelnen Umweltdatenbanken und Informationssystemen sowie in übergreifende Metasysteme empfiehlt sich die Berücksichtigung geeigneter einheitlicher Austauschformate. In einigen Bereichen hat sich bereits das SGML-Format etabliert (s.o. und Sattler/Nikolai i. d. Bd.).

5 Erfahrungen und Potential des Umweltdatenkatalogs (UDK)

Der UDK ist mittlerweile ein langjährig erfolgreich eingesetztes Werkzeug (Legat et al., i. d. Bd.), das trotz auch künftig optimierbarer Elemente weite Verbreitung und Anerkennung gefunden hat. Im deutschsprachigen Raum stellt der UDK für Katalogsysteme von Umweltdaten und -adressen einen De-facto-Standard dar.

Aus diesem Grund bietet es sich an, Rückschau zu halten. Es kann festgestellt werden, welche Eigenschaften für den Erfolg maßgeblich waren, welche Entwicklungsrichtungen korrekt eingeschlagen wurden. Für ähnliche Projekte können diese Erfahrungen genutzt werden. Elemente, die optimiert werden können oder die wünschenswert sind, sollten benannt werden.

In Österreich wirkte sich die einheitliche Erfassung in einem gemeinsamen Bundes-UDK positiv auf die Qualität der gewonnenen Informationen aus. Mit dem Virtuellen UDK (Legat et al., i. d. Bd.) besteht jedoch auch die Möglichkeit, verschiedene UDK-Instanzen einheitlich abzufragen.

Eine interessante Frage ergibt sich nicht nur aus der Aufbewahrung und zukünftigen Lesbarkeit von einmal erfaßten Daten. Auch der Bedarf nach älteren Versionen eines UDK kann, unter anderem aus historischen Gründen, entstehen. Andere Motive für diesen Wunsch können in der Wiederherstellung oder Nachvollziehbarkeit von Daten-

wegen und Verantwortlichkeiten liegen. Der Bedarf nach einer kompletten Versionsverwaltung konnte nicht abschließend geklärt werden. Eine einfache Stufe einer solchen ist durch das bereits vorhandene Aktualisierungsflag gegeben.

Auf der Wunschliste an bestehende UDK-Implementierungen steht eine Import-Schnittstelle für eine automatisierte Metadateneinstellung und -aktualisierung. Bisher stehen Stellen, bei denen bereits automatisiert Metadaten anfallen, vor der Wahl, diese entweder manuell erneut in den UDK einzubringen oder selbst Schnittstellen für eine Transformation zu schaffen (Sattler/Nikolai i. d. Bd.). Einheitliche, anpaßbare Lösungen, unter Umständen über Austauschformate (s.o.), würden die Einsetzbarkeit des UDK und die Effizienz der Metadatenbearbeitung steigern.

6 Öffentliche Verwaltungen: Erfahrungen und Perspektiven

Der Bedarf an informationstechnischen Lösungen in der öffentlichen Verwaltung orientiert sich größtenteils an dort vorhandenen und bewährten Verfahren und Vorgängen. Übergreifende IT-Konzepte z. B. im Umweltinformationsbereich müssen einerseits bestehende Fachinformationssysteme und andererseits Standardapplikationen geeignet integrieren.

Einzubindende Fachinformationssysteme, die als unveränderbare Einheit betrachtet werden müssen, werfen Probleme in bezug auf die Richtung des Datenflusses auf: Häufig ist nur der lesende Zugriff erlaubt, Redundanzen durch separate Datenhaltung in verschiedenen Fachinformationssystemen können nicht durch einen Datenabgleich aufgelöst werden. In diesem Fall können beispielsweise Schlüssel-Mapping-Verfahren die Eindeutigkeit und Kombinierbarkeit von Informationen herstellen (Freitag et al., i. d. Bd.). Häufig bedeutet dieses einen manuellen Aufwand, der im Rahmen der Datenpflege anfällt.

Speziell kleine Nutzergruppen benötigen maßgeschneiderte Umgebungen zur Unterstützung der Arbeitsabläufe (Tochtermann et al., i. d. Bd.). Bereits vorhandene Workgroup-Systeme passen nicht, wenn wichtige Spezifika optimal abgebildet werden sollen. Fertige Workgroup-Lösungen orientieren sich eher an Standardabläufen, um universell einsetzbar zu sein. Hier können Eigenentwicklungen, die sich aber vorhandener Standardkomponenten bedienen, die Erfordernisse erfüllen.

In ähnlicher Weise decken einige kommerzielle Informationssystem-Produkte nur einen Teil der gewünschten Anforderungen ab. So wurde das Beispiel eines Systems angeführt, das sowohl im Web als auch auf CD-ROM ohne aufwendige Anpassungsmaßnahmen verfügbar sein soll (Fitzke/Friebe, i. d. Bd.). Bestehende Produkte erfüllen diese Bedingungen nicht ausreichend, so daß die Eigenentwicklung des gewünschten Systems vorgenommen wurde.

7 Verteilte Systeme

Umfassende Aufgabenstellungen, für die thematisch oder geographisch weitgefächerte Datenquellen zugänglich gemacht werden müssen, erfordern häufig verteilte Systeme, deren Informationen dezentral gepflegt und mitunter auch dezentral verwaltet werden. (Güntsch et al., Heinemeier, i. d. Bd.)

Durch die Beteiligung späterer Anwender verteilter Systeme bereits in der Entwicklungsphase, indem ihre Anforderungen erfragt und berücksichtigt werden, können Systeme geschaffen werden, die spezifische (auch historische) Vorgaben erfüllen und rechtzeitig operabel sind.

Dieses Vorgehen vermindert auch die Schwierigkeit, unter denen viele Informationssysteme leiden, die auf die Einbringung von Informationen angewiesen sind: Die Informationslieferanten müssen geeignet motiviert werden, Daten in das System einzubringen und diese gegebenenfalls auch zu aktualisieren. Von großer Bedeutung für den Erfolg solcher Systeme ist die Akzeptanz der Benutzungsschnittstellen und Systemvorgaben durch die Anwendenden (Braun, i.d. Bd.).

Für Umweltdatenbanken – nicht nur in verteilten Systemen, dort aber verstärkt – tritt ein Grundaspekt immer wieder auf: Es handelt sich um die Abbildung und zuverlässige Handhabung von Taxonomien. Durch die Tatsache, daß taxonomische Ordnungen und Bezeichnungen gegenüber der Zeit nicht invariant bleiben, entstehen je nach konkretem Anwendungsgebiet unterschiedliche Gefahren der Inkonsistenz. Auswege aus dem Dilemma sind abhängig von den vorgegebenen Rahmenbedingungen und reichen von komplexen Abbildungsfunktionen, die die Zeitpunkte der Erfassung in das System berücksichtigen, über Synonymklassen bis zu zentralen Festschreibungen einheitlicher Taxonomien innerhalb des Systems (Fitzke/Friebe, Güntsch et al., i. d. Bd.).

Alle Sessions des Workshops sowie weitere Verweise finden sich im Web unter <http://www.pz-oekosys.uni-kiel.de/akudb/>.

Kartengestützte und internetbasierte Auskunft über die geographische Verbreitung von Pflanzen und Tieren

Jens Fitzke¹ und Torsten Friebe²

1 Ausgangssituation

Weltweit existiert eine Fülle von Datenbanken mit Informationen über die geographische Verbreitung von Pflanzen und Tieren. Ihrem jeweiligen Erfassungsbereich entsprechend – lokal, regional, national bis hin zu global – sind in Ihnen Informationen unterschiedlichster Granularität, aber auch unterschiedlichster fachlicher Ausrichtung zu finden. In der Regel handelt es sich um Informationen, die traditionell in Karten- und Atlasform präsentiert werden.

Teile dieser Datenbestände wurden bereits einem weltweiten Nutzerkreis zugänglich gemacht. Dabei überwiegen formulargestützte Ansätze (z. B. FishBase³, Arctic Bird Library des World Conservation Monitoring Centre⁴), mit denen sich ein interaktiver Umgang mit den geographischen Informationen kaum abbilden läßt: An die Stelle von Karten treten Gebietsbeschreibungen oder Listen, beispielsweise von Ländern, in denen eine bestimmte Art vorkommt. Erste Ansätze einer Nutzung interaktiver Karten für Auskünfte aus Verbreitungsdatenbanken finden sich z. B. bei StreamNet⁵ und dem Species Mapper des australischen Umweltministeriums⁶.

Diese Systeme beantworten Fragen wie

1. Wo kommt ein bestimmtes Tier/eine bestimmte Pflanze vor? – und umgekehrt?
2. Welche Tiere/Pflanzen kommen in einem bestimmten Gebiet vor?
3. Welche Eigenschaften charakterisieren ein bestimmtes Tier/eine bestimmte Pflanze?
sowie als Erweiterung mit einer zeitlichen Dimension:
4. Welche Veränderungen hat ein bestimmtes Tier/eine bestimmte Pflanze (in einem bestimmten Gebiet) in einem bestimmten Zeitraum erfahren?

¹ Geographische Institute der Universität Bonn, Meckenheimer Allee 166, D-53115 Bonn, Tel. +49 (0)228 73-2098, Fax -9658, E-mail jens.fitzke@uni-bonn.de

² DV-Beratung und Software-Entwicklung, Zwirner Str.26, D-50678 Köln., Tel. +49 (0)221 93296-06, Fax -03, E-mail tfriebe@compuserve.com

³ <http://www.fishbase.org>

⁴ <http://www.wcmc.org.uk/arctic/data/birds/birds.htm>

⁵ <http://www.streamnet.org/>

⁶ <http://www.environment.gov.au/search/mapper.html>

Es zeigen sich jedoch konzeptionelle Schwächen, die insbesondere in der mangelhaften Integration des Instruments der interaktiven Karte in den Recherche- und Präsentationskontext des jeweiligen Auskunftssystems begründet liegen: Die interaktive Karte dient entweder als Recherche-/Selektions- oder als Präsentationsmedium. Es ist nicht möglich, auf der Grundlage einer Kartendarstellung, die bereits das Ergebnis einer vorhergehenden Recherche darstellt, eine weitere – vertiefende oder erweiternde – Recherche durchzuführen. Komfortable und intelligente Benutzerführung ist jedoch nur durch eine enge Integration von Datenbank-, GIS- und Internet-Technologie zu erreichen.

Den Ausgangspunkt für diesen Beitrag bilden zwei Vorhaben im Rahmen des Projektverbundes GIS Experimental Server⁷, in denen anhand konkreter Vorgaben aus der Praxis Realisierungsmöglichkeiten kartengestützter und internetbasierter Auskunftsmodule zu floristischen und faunistischen Verbreitungsdatenbanken aufgezeigt werden. Es handelt sich um Web-basierte Auskunftsplätze für die floristische Datenbank FLORKART des Bundesamtes für Naturschutz⁸ und für die Datenbank des Global Register Of Migratory Species (GROMS)⁹. Auf der Grundlage dieser Pilotprojekte wird im folgenden ein allgemeines Modell für diese Klasse von Auskunftssystemen entwickelt und zur Diskussion gestellt.

2 Fachkonzept

Aus den o. a. Fragen können Anwendungsfälle (use cases) abgeleitet werden, die bei der objektorientierten Software-Entwicklung in prägnanter Weise den Umgang des Anwenders mit dem System beschreiben. Im Zusammenhang mit oder folgend aus diesen Anwendungsfällen wird die spezifische Weltansicht der Anwender in einem Objektbereichsmodell¹⁰ dargestellt und die daraus resultierenden Mensch-Maschine-Interaktionen abgeleitet.

2.1 Objektbereich

Fundamentales Ordnungselement im Tier- wie im Pflanzenreich ist die Art als Fortpflanzungsgemeinschaft von Individuen. Arten werden nach verschiedenen Kriterien in hierarchisch angeordneten Gruppen (z. B. Familie, Ordnung, Klasse) zusammengefaßt, die Sippen bzw. Taxa (sing. Taxon) genannt werden. Je nach Forschungsstand kann vorausgesetzt werden, daß für jedes Taxon eine eindeutige Bezeichnung existiert, es durch eine Liste von Eigenschaften beschrieben werden kann und daß Be-

⁷ Projekt-Homepage: <http://www.giub.uni-bonn.de/exse/>

⁸ Als Demonstrator bereits verfügbar über <http://www.giub.uni-bonn.de/exse/>

⁹ <http://www.biologie.uni-freiburg.de/data/riede/groms.html>

¹⁰ engl. *domain model*; in Fowler/Scott (1998) mit "Problembereichsmodell" übersetzt – hier wird zum besseren Verständnis die Darstellung des Objektbereichs den Anwendungsfällen vorangestellt.

obachtungsdaten über seine geographische Verbreitung existieren. Dieser Zusammenhang ist in der folgenden Abbildung veranschaulicht:

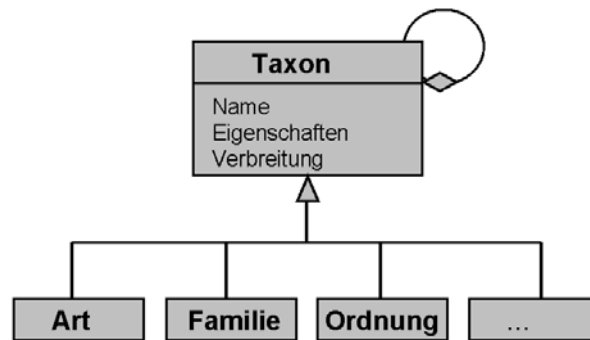


Abbildung 1
Ein einfaches Objektbereichsmodell

2.2 Anwendungsfälle

Es wird gemeinhin empfohlen, Anwendungsfälle in enger Zusammenarbeit mit potentiellen Systemnutzern zu erarbeiten. Bei der Entwicklung von Internet-basierten Auskunftssystemen mit einem – zumindest potentiell – weltweiten Nutzerkreis ist dies kaum möglich. Im vorliegenden Fall wurde daher auf die Zusammenarbeit mit Experten des jeweiligen Fachgebiets zurückgegriffen. Im folgenden werden exemplarisch einige grundlegende Anwendungsfälle aufgeführt¹¹:

1. Taxon-Liste: Zusammenstellung einer Liste von Taxa nach semantischen und/oder geographischen Kriterien
 - a) einfachster Fall: Recherche über einen Suchbegriff (Name oder Namensbestandteil des Taxons)
 - b) Recherche mit Einschränkung über Eigenschaften des Taxons, z. B. Schutzstatus in einer internationalen Konvention oder bestimmte Ansprüche an den Lebensraum (edaphische Verhältnisse bei Pflanzen)
 - c) Recherche mit Einschränkung auf ein bestimmtes Gebiet
 - d) Beliebige Kombinationen der Recherchemöglichkeiten von a), b), c)
 - e) 'Taxon-Browser' als Alternative zur kriteriengestützten Recherche

¹¹Da die Experten bereits eine Vorstellung von den Auskunftssystemen entwickelt hatten, handelt es sich hier eher um Systeminteraktionsbeschreibungen als um die Beschreibung von Anwenderzielen. Dies zeigt das Konstrukt der "Taxon-Liste", die lediglich einen Zwischenschritt zur Erreichung eines Anwenderzielen, nicht aber ein Anwenderziel selbst darstellt.

2. Karte: Kartographische Darstellung der Verbreitung eines Taxons mit Möglichkeiten zur Steuerung der Kartenkomposition (Aktivierung/Priorisierung von einzelnen Ebenen, Auswahl von Basiskarten) und des Kartenausschnitts (Navigationsfunktionen wie Zoom und Pan).
3. Steckbrief: Zusammenfassung der Eigenschaften eines Taxons in einem Steckbrief
4. Spezielle fachliche Auswertungen zur Produktion von thematischen Karten und Berichten.

2.3 Systemdesign

Es kann vorausgesetzt werden, daß der Nutzer an einer bestimmten Tier- bzw. Pflanzengruppe oder an einem bestimmten Gebiet interessiert ist. Im Idealfall kennt er die fachliche Bezeichnung der ihn interessierenden Pflanzen/Tiere und er kann das Gebiet exakt beschreiben bzw. benennen. Dies ist allerdings – insbesondere wenn auch Laiennutzer in die Betrachtung einbezogen werden – nicht immer der Fall: Die aus den Basisanforderungen resultierenden Systemfunktionen erfahren Erweiterungen, die dem Nutzer erlauben, eine korrekte (d. h. für das jeweilige Informationssystem verständliche) Anfrage zu formulieren – dazu gehören:

- Navigation über Metainformationen – hin zu einer eindeutigen Benennung eines Taxons oder hin zu einer Gruppe von Taxa, die bestimmten Kriterien genügen
- Geographische Navigation – hin zu einer eindeutigen Bezeichnung des interessierenden Gebietes

Nach einer erfolgreich ausgeführten Anfrage an das System wird die Treffermenge angezeigt und dem Benutzer ermöglicht, weitere (vertiefende) Fragen an das System zu stellen. Dazu gehören die Komponenten:

- Präsentation der verfügbaren Informationen (Karte, Bericht, Grafik, ...)
- Abfrage/Analyse: Verfeinerung, Vertiefung, Verknüpfung

Die Beobachtungsdaten räumlicher Verbreitung von Pflanzen/Tieren sind i.d.R. nicht als primäre Einzelbeobachtungen, sondern aggregiert in den Datenbanken gespeichert bzw. werden – aus Naturschutzgründen – nur in aggregierter Form oder in Auszügen veröffentlicht.

Jede Verbreitungsdatenbank verfügt über eine (i.d.R.) hierarchische Struktur der Taxa, die sich somit (zumindest innerhalb dieser Datenbank) eindeutig benennen und untereinander zuordnen lassen. Zu den Taxa sind eine Reihe von Attributen gespeichert, die 1) von allgemein fachlich-beschreibender Art sind, 2) die räumliche Verbreitung charakterisieren, 3) den zeitlichen Bezug der fachlichen bzw. räumlichen Informationen beinhalten oder 4) die Informationen aus 1)–3) beschreiben (Metainformationen).

Bei den räumlichen Informationen handelt es sich geometrisch um punkt-, linien- oder flächenhafte Informationen, deren Ausprägung und Qualität von Parametern wie Erfassungsmethode, Erfassungsmaßstab, Aggregations-/Generalisierungsgrad

(Granularität) abhängig ist. Zwei wesentliche Grundtypen des Raumbezuges lassen sich identifizieren:

1. Individuelle Raumbezüge: Ein Beobachtungsdatum hat seine eigene Geometrie (Bsp.: Zugweg des Weißstorchs vom Überwinterungs- zum Sommergebiet)
2. Raumbezugssystem: Viele Beobachtungsdaten sind einer Menge von geometrischen Referenzobjekten zugeordnet (Bsp.: Länderbezogenes Vorkommen einer bestimmten Fledermausart oder Vorkommen einer bestimmten Orchideenart in dem Blattschnittraster der Topographischen Karten 1:25000)

Mit der kartographischen Präsentation dieser Informationen auf Grundlage der beschreibenden Informationen (Metainformationen und Anwendungslogik) können Herkunft, Qualität und andere Metaattribute transparent an den Nutzer weitergegeben werden – man denke beispielsweise an die räumliche Unschärfe der Daten.

3 Technisches Design der Anwendungssoftware

3.1 Anforderungen

Für die Umsetzung des Fachkonzepts in Form eines Web-basiertem Auskunftssystems sind folgende Komponenten notwendig:

- Filtereinstellungen/Anfrageoptionen
- Interaktive Karte
- Ausgabemodul (multifunktional: Text, Tabelle, Grafik etc.)

Die Filtereinstellungen bieten einen umfassenden Überblick über die aktuell gesetzten Anfrageparameter (fachlich, räumlich, zeitlich), die interaktive Karte dient als Navigations-, Präsentations- und Abfragemedium für raumbezogene Informationen und das Ausgabemodul dient der Darstellung aller Arten von fachlichen und beschreibenden Informationen.

Ausserdem werden an die Anwendung folgende Anforderungen gestellt:

- Realisierung einer Web-Applikation sowie auch einer CD-ROM-Anwendung ohne größeren Mehraufwand
- Plattformunabhängigkeit
- einfache und intuitiv handhabbare Benutzerschnittstelle

3.2 Anwendungsarchitektur

Das Konzept der Anwendung basiert auf einer Client-Server-Architektur, die das Ziel verfolgt, eine langfristige Flexibilität in der Systementwicklung und einen höheren Wiederverwendungsgrad dadurch zu erzielen, daß die Präsentationsschicht von der Fachklassenschicht und der Workflowsteuerung entkoppelt ist. Die Abbildung 2

veranschaulicht die Systemarchitektur und zeigt die Zuordnung der einzelnen Bestandteile zu Client und Server.

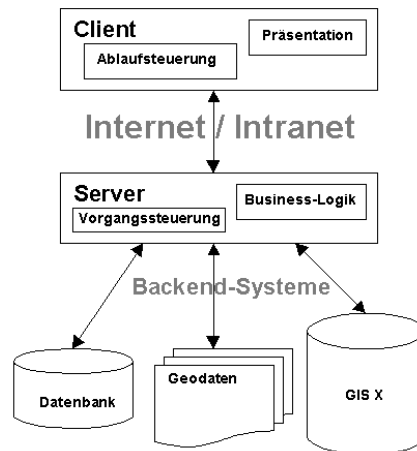


Abbildung 2
Anwendungsarchitektur

Die Workflowsteuerung ist in eine Client- und in eine Server-seitige Komponente aufgeteilt, deren Client-Komponente die Ablaufsteuerung der Benutzerschnittstelle und deren Server-Komponente die Vorgangssteuerung übernimmt. Die Fachklassenschicht wird durch die Business-Logik abgebildet, die die Verbindung zu den Backend-Systemen realisiert.

Die daraus resultierende Mehr-Schicht-Architektur ordnet den einzelnen Schichten folgende Aufgaben zu: Die Datenquellen wie Datenbanken und GIS werden in der Datenhaltungs-Schicht zusammengefaßt. Soll ein weitere Datenquelle zum System hinzugefügt werden, muß für diese ein Adapter, in Form eines Treibers implementiert werden. Die Aufgabe des Treibers besteht darin, die notwendigen Funktionalitäten, die in einer abstrakten Schnittstelle definiert sind, für das System verfügbar zu machen. Die durch die systemabstrahierende Schnittstelle verfügbaren Daten werden im System in einem einheitlichen Modell abgebildet. Funktional werden diese Aufgaben von der Business-Logik übernommen, die dem System somit einen vereinheitlichten Zugriff auf alle vorhandenen Datenquellen ermöglicht. Für die Verarbeitung der Daten ist die Management-Schicht mit der Vorgangssteuerung zuständig. Hier werden die von der Business-Logik bereitgestellten Daten zusammengeführt und für die graphische Darstellung aufbereitet und strukturiert, dies geschieht u. a. durch die Verknüpfung mit eindeutigen Schlüssel. Der Client-seitige Teil der Managementschicht, die Ablaufsteuerung wandelt die Benutzerabfrage, die der Anwender in Eingabemasken

spezifiziert hat, in eine formale Anfrage an das System um, die dann von der Vorgangssteuerung abgearbeitet wird. Das Ergebnis der Anfrage wird an den Client auf Basis eines einheitlichen und standardisierten Kommunikationsprotokolls übertragen und von der Präsentationsschicht auf dem Client in Form von Ein- und Ausgabemasken bzw. in einer 2D-Graphikkomponente dargestellt.

Die verteilte Architektur zielt auf eine Reduzierung der Netzwerkbelastung ab, die Client-seitig dadurch ermöglicht wird, daß die Abfrageergebnisse in einem lokalen Cache des Clients zwischengespeichert werden und dort, ohne auf die Serverkomponente zuzugreifen, bei einer wiederholten Anfrage ausgelesen werden können. Server-seitig wird dies durch die Selektion der notwendigen Daten, deren Homogenisierung durch das Zusammenführen in einem frei wählbaren einheitlichen Koordinatensystem und Reduzierung beispielsweise durch Generalisierung realisiert.

Die Architektur des Systems ermöglicht eine Trennung von Datenmodell, Datenverarbeitung und Präsentation und setzt damit ein gängiges Entwurfsmuster (engl. Design Pattern) der objektorientierten Softwareentwicklung um, das Model-View-Controller (MVC) Paradigma. Entwurfsmuster stellen ein Mechanismus zur Verfügung, um wiederkehrende Probleme bei dem Entwurf von Software zu lösen und helfen bei der Entwicklung von wiederverwendbaren Softwarekomponenten.

3.3 Implementierung der Architektur in Java

Für die Implementierung der Architektur in Java-Komponenten wurden folgende Klassenbibliotheken eingesetzt (siehe Abbildung 3).

Die Datenhaltung des Prototyp ermöglicht den Zugriff auf Shape-Dateien und relationale Datenbanken. Für den Zugriff auf Shape-Dateien wurde ein Treiber auf Basis der Spezifikation des Herstellers (ESRI 1998) unter Verwendung der Java API¹² für Ein- und Ausgabedatenströme implementiert. Für den Zugriff auf relationale Datenbanken wurde die Java Database Connectivity (JDBC) API verwendet, die eine Selektion von Daten durch die Standard-Abfragesprache SQL¹³ ermöglicht.

Die Abstraktion der Daten in einem einheitlichen Model wird durch eine Implementierung der Simple Features des OpenGIS-Konsortiums¹⁴ (OGC 1998) in Java realisiert. Die Aufgaben der Management-Schicht werden durch mehrere Servlets übernommen, wobei für jeden Vorgang ein separates Servlet implementiert ist. Servlets besitzen die gleiche Funktionalität wie Java-Applikationen (mit Ausnahme des graphischen Interface) und vereinen zusätzlich die Konzepte von Applets und CGI¹⁵. Servlets laufen in der virtuellen Maschine eines Web-Servers und erweitern dessen Funktionalität.

¹² Application Programming Interface, Programmierschnittstelle zur Nutzung von Software-Funktionen anderer Programme i.w.S. ohne Kenntnis ihrer Implementierung

¹³ Structured Query Language

¹⁴ <http://www.opengis.org>

¹⁵ Common Gateway Interface, Schnittstelle zur Kommunikation von Web-Server mit Anwendungsprogrammen

Präsentation	Java AWT	Client
Management	Java Servlets	
Abstraktion	Simple Features	Server
Datenhaltung	<div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; padding: 2px;">Shapefile Dateisystem</div> <div style="border: 1px solid black; padding: 2px;">RDBMS</div> </div>	

Abbildung 3
Implementierung der Anwendung

Der Client läuft als Java-Applet in der virtuellen Maschine des Webbrowsers und übernimmt Aufgaben der Präsentations- sowie auch der Managementschicht. Die Präsentationsschicht beinhaltet eine graphische Benutzerschnittstelle, die sich aus den Komponenten des Abstract Window Toolkit (AWT) zusammensetzt sowie einer Kartenkomponente, die auf einer Eigenentwicklung eines 2D-Graphikkerns mit Weltkoordinatensystem aufsetzt. Die Ablaufsteuerung nutzt die Ereignissverarbeitung des AWT und ruft über die HTTP¹⁶-Methode POST ein entsprechendes Servlet auf dem Server auf und startet damit einen Vorgang. Die Anfrageergebnisse werden von Datentyp abhängig in einem geeigneten Format übertragen. In der ersten Ausbaustufe wurde für die Geodaten ein proprietäres Datenformat (binär und topologieles) gewählt, das einen einfachen und performanten Datenzugriff ermöglicht. In der zweiten Ausbaustufe wurde die Well-known Binary Representation for Geometry (WKB) der Simple Feature Spezifikation verwendet. Die Sachdaten werden als einfacher Textdatenstrom in Form einer Tabelle übertragen.

Zur Unterstützung von Mehrsprachigkeit sind Beschriftungen der Komponenten der Benutzerschnittstelle sowie die Systemmeldungen für den Benutzer von der Implementierung getrennt und werden in einer separaten Beschreibungsdatei gespeichert.

¹⁶ Hypertext Transfer Protocol

3.4 Standards für die Implementation

Für die Implementierung der Software wird das evolutionäre Prototyping¹⁷ eingesetzt, mit dem Ziel durch inkrementelle und iterative Entwicklungszyklen die vorgegebenen Anwendungsfälle schrittweise in der Anwendung abzubilden, um am Ende des Entwicklungsprozesses ein vollständige Software zu erhalten. Als Vorgehensmodell wurde die vom IBM-Object-Oriented-Technology Center entwickelte Methode in Form eines Arbeitsheftes eingesetzt (IBM 1996).

Als Implementierungsspezifikation für die Simple Features wurde die CORBA¹⁸-Variante als Vorlage genommen (OGC 1998). Die Wahl der Spezifikation für die CORBA Spezifikation ist darin begründet, daß den Spezifikationen für SQL und COM¹⁹ ein klares Konzept für Features fehlt und nur die CORBA-Spezifikation ein durchgängiges Objektmodell bietet (Cuthbert 1999:22).

Der Client ist auf der Basis von Java 1.1 entwickelt worden. Dabei waren die Erfahrungen aus vorherigen Projekten mit Java 1.0 entscheidend. Die Version Java 1.0 ist sehr inkonsistent und noch nicht fehlerfrei, was zur Folge hatte, daß die Software auf den verschiedenen Plattformen ein sehr unterschiedliches Laufzeitverhalten zeigte. Die dem zugrundeliegenden Probleme sind durch den Einsatz von Java 1.1 zum größten Teil behoben. Die Server-Komponente ist ebenfalls in Java 1.1 entwickelt worden und nutzt das Java Servlet Development Kit (JSDK) 2.1, läuft aber in der Virtuellen Maschine von Java 2, da diese für den Einsatz von Enterprise Anwendungen optimiert ist und die Funktionalität des JSDK besser unterstützt.

4 Ausblick

Der skizzierte Demonstrator wird für die Zwecke des Global Register of Migratory Species verfeinert und erweitert. Wesentliches Ziel ist es, die Nutzbarkeit der Simple Features Specification des OpenGIS-Konsortiums am Beispiel des Anwendungsfalls „Auskunft über floristische und faunistische Verbreitungsinformationen“ zu testen. Neben Untersuchungen zur Skalierbarkeit der eingesetzten Technologien hinsichtlich gleichzeitiger Nutzer, Transaktionsvolumen etc. sollen die Einsatzmöglichkeiten von XML (eXtensible Markup Language) als Datenbeschreibungssprache für den Austausch von Daten zwischen Client und Server evaluiert werden.

Darüber hinaus sind Methoden und Techniken zu entwickeln bzw. zu testen, mit denen sich die geometrische Unschärfe der Daten in der kartographischen Darstellung visualisieren läßt. In enger Zusammenarbeit mit den Nutzern des Systems sind Benutzungsschnittstelle und Anwendungslogik auf Tauglichkeit zu prüfen und ggf. anzupassen.

¹⁷Software-Prototypen sind ausführbare Programme, die als erweiterbares Modell für das geplante Software-Produkt dienen.

¹⁸Common Object Request Broker Architecture, Standard der Object Management Group (OMG) für die Kommunikation verteilter Software-Objekte (<http://www.omg.org/corba/>)

¹⁹Component Object Model, Komponentenmodell der Fa. Microsoft

5 Referenzen

- Cuthbert, Adrian (1999): "OpenGIS: Tales from a Small Market Town". In: Vckovski, Brassel und Schek (Hrsg.): *Interoperating Geographic Information Systems*, (Proc. Interop'99, Zurich, March 1999), S. 17–28.
- ESRI Environmental Systems Research Institute, Inc. (Hrsg.) (1998): *ESRI Shapefile Technical Description White Paper*. Redlands, California.
Internet: http://www.esri.com/library/whitepapers/av_lit.html
- Fitzke, Jens (1999): „Praxisrelevante Ergebnisse des OpenGIS-Prozesses“. In: Bock, M., K. Greve und W. Kuhn (Hrsg.): *Offene Umweltinformationssysteme – Chancen und Möglichkeiten der OpenGIS-Entwicklung im Umweltbereich*. Ergebnisse des BLAK UIS Workshops Münster, 4./5. Februar 1999, Münster (IFGI-Prints Bd. 7), S. 185–195.
- Fowler, Martin u. Kendall Scott (1998): *UML konzentriert. Die neue Standard-Objektmodellierungssprache anwenden*. Bonn: Addison-Wesley-Longman.
- Gamma, E. et.al. (1995): *Entwurfsmuster*. Bonn: Addison-Wesley.
- IBM (Hrsg.) (1996): *Developing Object-Oriented Software: An Experience-Based Approach*. Prentice Hall.
- OGC OpenGIS Consortium (Hrsg.) (1998): *OpenGIS® Simple Features Specification For CORBA*. o.O.
Internet: <http://www.opengis.org/techno/specs.htm>
- Oestereich, Bernd (1998): *Objektorientierte Softwareentwicklung - Analyse und Design mit der Unified Modeling Language*. München, Wien: Oldenbourg.
- Pomberger, Gustav u. Günther Blaschek (1996): *Software-Engineering - Prototyping und objektorientierte Software-Entwicklung*. München, Wien: Hanser.
- Wiederhold, Gio (1999): „Mediation to Deal with Heterogeneous Data Sources“. In: Vckovski, Brassel und Schek (Hrsg.): *Interoperating Geographic Information Systems*, (Proc. Interop'99, Zurich, March 1999) S. 1–16.
Internet: <http://www-db.stanford.edu/pub/gio/1999/Interopdocfigs.html>

Regionale Umweltdaten in einem internationalen Projekt: Datenmanagement des trilateralen Monitoring- und Bewertungsprogramms TMAP in Schleswig-Holstein und Hamburg

Claus Hoerschelmann¹

Abstract

The Trilateral Monitoring and Assessment Program TMAP was created to establish an integrated monitoring and assessment of the ecosystem Wadden Sea as a whole. In order to have access to the data of the TMAP, an architecture of data management has been developed. This architecture consists of distributed Data Units with a separated, identical data access. This paper focuses on the works in Schleswig-Holstein and Hamburg, especially data modelling and co-operation between data originators and the data management project.

1 Einleitung

Das europäische Wattenmeer ist ein sensibler Lebensraum, der einer hohen Dynamik unterliegt. Stoffflüsse, Strömungen und Wanderung der unterschiedlichen Tierarten beziehen sich immer auf den gesamten Wattenmeerbereich, so dass auch Umweltbeobachtungen und Schutzmaßnahmen grenzübergreifend durchgeführt werden müssen. Die Wattenmeer-Anrainerstaaten Niederlande, Dänemark und Bundesrepublik Deutschland haben diese Küstenbereiche unter weitreichenden Schutz gestellt, so dass ein Schutzgebietverbund entlang der Küste existiert. Das Monitoring von Umweltparametern wird durch das trilaterale Monitoring- und Bewertungsprogramm TMAP koordiniert. Teil dieser Koordinierung ist der Aufbau eines Datenmanagementsystems in den beteiligten Ländern. Dieses System wird am Beispiel der TMAP Data Unit für Schleswig-Holstein und Hamburg vorgestellt.

¹ Multimar Wattforum, Am Robbenberg, D-25832 Tönning
email: c.hoerschelmann@multimar-wattforum.de

2 Ziele des Monitorings im Wattenmeer

Die Langzeituntersuchungen im Wattenmeer sollen einen Überblick über Entwicklungen im Ökosystem am Beispiel konkreter Fragestellungen bieten, sowie Daten zur Bewertung und Lösung akuter Probleme (Schiffsunglücke, Algenblüten, starke Populationsveränderungen, Verschmutzungsereignisse) bereitstellen.

Im Rahmen des Common Package, der ersten Gruppe von einheitlichen Untersuchungen, werden 28 Parameter aus den Themenbereichen erhoben:

1. allgemeine Daten
2. physikalisch-chemische Daten
3. biologische Daten
4. Daten zur menschliche Nutzung

Zu einzelnen Fragestellungen existieren langfristige, koordinierte Untersuchungreihen, andere Teile des Monitoring befinden sich noch in der Abstimmung innerhalb der trilateralen Gremien.

3 Räumlicher und organisatorischer Rahmen und daraus resultierende Strukturvorgaben für das Projekt

Teil der Vereinbarungen zwischen den Anrainerstaaten ist die regelmäßige Erstellung von Berichten über den Zustand des Wattenmeeres, der Quality Status Reports. Diese Berichte werden von Fachleuten erstellt, die Zugang zu den Daten des Gesamtbereichs haben müssen. Umweltdaten unterliegen der Zuständigkeit der verschiedenen Staaten bzw. Länder, so dass eine zentrale Datenhaltung des Wattenmeermonitorings nicht zu realisieren war. Innerhalb der Anrainerstaaten des Wattenmeeres bestehen zusätzlich unterschiedliche Strukturen zur Verwaltung von Umweltdaten. Sind Datenbanken oder DBMS' vorhanden, so sind dieses unterschiedliche Systeme.

Im Rahmen des durch die EU geförderten Projektes DEMOWAD wurde eine Struktur entwickelt, nach der Umweltdaten in identisch strukturierten, verteilten Datenbanken gehalten werden und ein einheitlich gestalteter Zugriff auf die Daten ermöglicht wird.

Die einzelnen Datenbanken besitzen regionale Zuordnung. Die Entwicklungen der Datenbanken werden zwischen den Anrainerstaaten abgestimmt, das Gemeinsame Wattenmeersekretariat CWSS in Wilhelmshaven koordiniert die Arbeiten. Dieses Rahmenwerk erlaubt eine verteilte Entwicklung der Datenbanken, von der dann alle Teile des Verbundes profitieren.

4 Struktur der TMAP Data Units

Die Data Units basieren auf den Arbeiten des DEMOWAD Projekts. Die Units setzen sich aus einer identischen Datenbankstruktur, einer definierten Schnittstelle und abgestimmten Metadatenätzen zusammen. Innerhalb der Benutzerschnittstelle wird eine password-Authentifizierung vorgenommen, die den Zugriff auf die Ergebnisdaten der Untersuchungen einschränkt. Als Benutzerschnittstelle fungiert ein three-tier Java-Applet, um den Datenbankzugriff unabhängig von verwendeten Datenbanksystem zu gestalten.

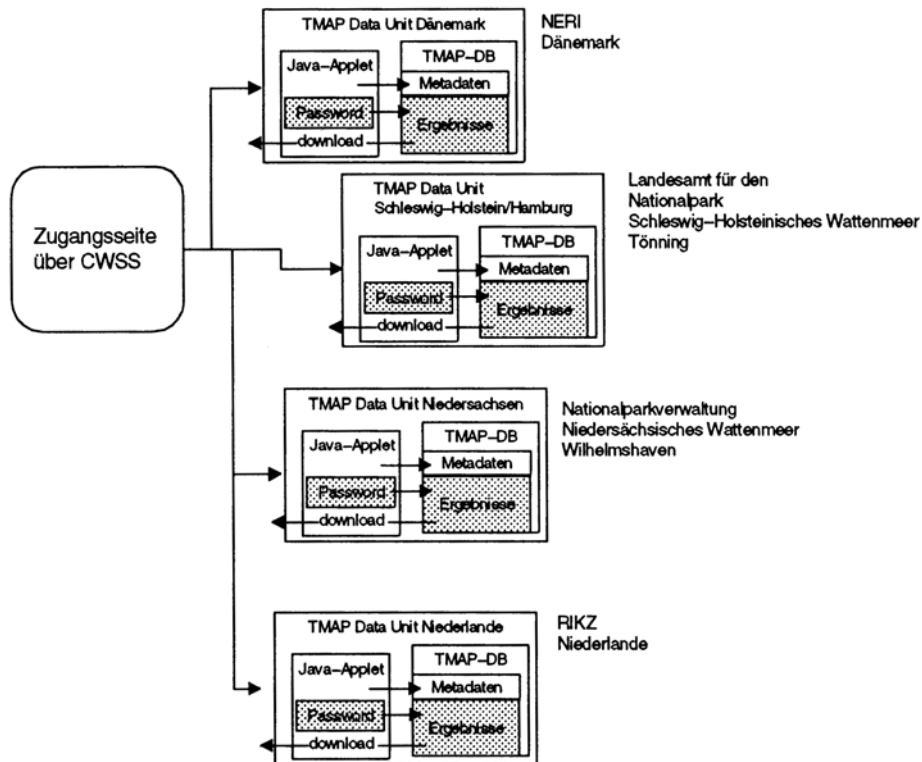


Abbildung 1

Struktur der Datenhaltung des TMAP.

Einheitliche Data Units mit den regionalen Monitoringdaten sind übers www zu erreichen

5 Verwendetes Datenmodell am Beispiel einiger Parametergruppen

Allgemeines

Der unterschiedliche Entwicklungsstand in der Abstimmung und Implementierung der einzelnen Monitoringsvorhaben und die Vorgaben aus dem Projekt DEMOWAD machen die Verwendung eines modularen, subjektbezogenen Datenmodells sinnvoll.

Dieses subject-orientierte Datenmodell mit einheitlichem Metadaten-Block spiegelt zum einen die Trennung der Daten in einen öffentlichen und einen geschützten Bereich wider, zum anderen entspricht dieses dem Aufbau des TMAP, welches sich aus unabhängigen Monitoring-Projekten zusammensetzt.

Innerhalb der einzelnen Parametergruppen wird eine möglichst übereinstimmende Struktur verwendet.

Die primäre Aufgabe der TMAP-Dataunits ist die Datenhaltung, eine Auswertung der Daten oder themenübergreifende Verknüpfungen sind im Rahmen des Projekts nicht vorgesehen. Somit entspricht das Datenmodell für die einzelnen Subjects dem Transferformat, d. h. es werden nur vollständige Ergebnisdatensätze innerhalb einer Auswahl übertragen, ggf. mit den gesamten dazugehörigen Metadatenansätzen.

Metadaten

Der Metadatenbereich des TMAP-Datenmodells beinhaltet die projektübergreifenden Daten, besonderes Gewicht haben hier Angaben zur Datenhaltung, Format der Originaldaten und Ortsangaben. Die aus der Kombination von erfasstem Parameter, Monitoringsprojekt (hier: subject) und Probennahmeort erzeugte search_id stellt die einzige Verbindung zwischen dem Metadatenbereich und den Ergebnisdaten der einzelnen Projekte dar.

Ergebnisdaten am Beispiel der Projekte Phytoplankton und Rastvögel

Die Ergebnisdaten lassen sich jeweils in vier Tabellenbereiche aufteilen:

- Ergebnisnahe Metadaten oder Common Parameter Data mit Tabellen zu verwendeten Methoden, TMAP-relevanten Messgrößen und Artenangaben,
- Tabellen zu Probennahme oder Zählungen,
- Tabellen die die TMAP-relevanten Zähl- oder Messwerte beinhalten und
- Tabellen zu Begleitparametern oder Co-Variablen.

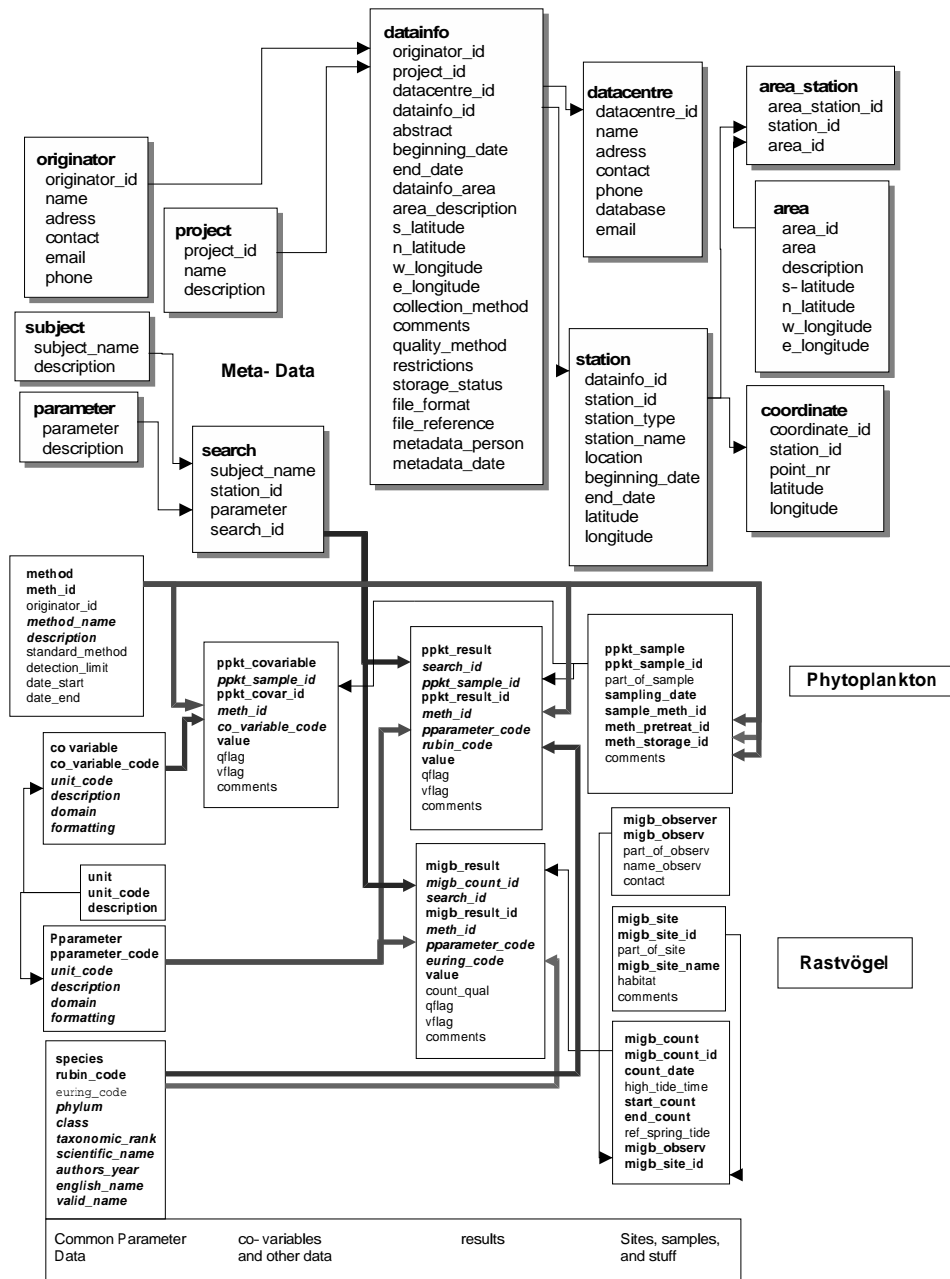


Abbildung 2

Das Datenmodell der TMAP Data-Units mit dem Metadatenblock und Ergebnistabellen zu Phytoplankton und Rastvögeln. Im oberen Bereich die Tabellen der Metadaten, darunter die Tabellen der Ergebnisdaten für Phytoplankton und Rastvögel

6 Technische Aspekte

Im Landesamt für den Nationalpark in Tönning wurde ein Oracle 8 DBMS auf Sun Enterprise Server 450 unter Solaris 2.6 eingerichtet, als HTTP-Server dient ein Apache SSL Web-Server. Die Dateneingabe wird im Landesamt für den Nationalpark vorgenommen, der Transfer der Daten hierhin verläuft zunächst mittels email und ftp, eine Eingabe von Daten durch datenliefernde Stellen ist mittelfristig nicht geplant.

Die Dateneingabe in die Data-Unit wird mittels Routinen des SQL-Loaders vorgenommen, hierbei erfolgt auch eine formelle Überprüfung der Daten. Die inhaltliche Qualitätskontrolle obliegt den Datenoriginatoren.

7 Zusammenarbeit mit Originatoren der Daten, Grundlagen und praktische Erfahrungen

Grundlagen

Die Wattenmeeraanrainerstaaten haben sich mit dem trilateralen Wattenmeerplan der Erklärung von Stade (1997) zum Monitoring des Wattenmeeres verpflichtet. Diese Verpflichtung beinhaltet auch eine koordinierte Datenhaltung entsprechend trilateraler Vereinbarungen.

In Schleswig-Holstein wurde diese Verpflichtung durch einen entsprechenden Kabinettsbeschluß bekräftigt.

Innerhalb der Landesverwaltung besitzt das Landesamt für den Nationalpark Schleswig-Holsteinisches Wattenmeer Funktionen als obere und untere Naturschutzbehörde, hat also administrativen Zugriff auf viele Daten und kann die Vorgaben für Erhebung und Übergabe von Monitoringdaten im Wattenmeer machen. So ist eine Verpflichtung zur Datenübergabe Bestandteil aller Monitoringsverträge im schleswig-holsteinischen Wattenmeer.

Ungeklärt ist bisher die Einbindung von Daten aus Monitoring-Projekten des Bundes in das TMAP-Datenkonzept. Eine Spiegelung entsprechender Datensätze in die räumlich zuständigen TMAP-Datenbanken entspräche dem Prinzip der regional zuständigen TMAP Datenzentren und würde bestehende Strukturen nutzen, ohne dass die »Datenhoheit« verloren wird. Diese Fragen sowie die Finanzierung der nötigen Anpassungen sind in nächster Zeit zu klären.

Zusammenarbeit mit den Originatoren der Daten

Im Bereich des schleswig-holsteinischen Wattenmeeres werden Daten von unterschiedlichen Behörden, Verbänden und Privatleuten erfasst. Originatoren aus dem Bereich der Landesbehörden sind die staatlichen Umweltämter, die Ämter für länd-

liche Räume und das Landesamt für Natur und Umwelt. In das Monitoring-Konzept sind die lokalen Naturschutzverbände eingebunden, der Großteil der Daten zu Vogelzählungen wird hier erhoben. Andere Monitoringprojekte werden von Wissenschaftlern, Forschungsinstituten und dem Nationalparkservice durchgeführt.

Von Seiten dieser Daten-Originatoren besteht reges Interesse an Mitarbeit an der TMAP-Datenbank. Die Wünsche und Ansprüche der Datenerheber an eine trilaterale Datenhaltung wurden frühzeitig berücksichtigt. Hierbei erwies sich insbesondere die Trennung von inhaltlicher und fachlicher Verantwortung (diese bleibt beim Originator) und Zuständigkeit für Datenhaltung und Verwaltung (beim Landesamt für den Nationalpark in Tönning) als sinnvoll.

Häufige Rücksprachen während der Datenmodellierung und die Abstimmung mit Expertengremien sind zwar zeitaufwendig, haben aber eine Atmosphäre produktiver Zusammenarbeit zur Folge, so dass der Datenfluss in die TMAP-Datenbank für den Bereich des hamburgischen und schleswig-holsteinischen Wattenmeeres gesichert ist.

Es werden regelmäßige Monitoring-Workshops in Tönning durchgeführt, um allen Beteiligten einen Überblick über die Untersuchungen zu geben und aktuelle Fragen zu klären. Die Ergebnisse des Monitorings werden in jährlichen Berichten zum Wattenmeermonitoring (Nationalpark Schleswig-Holsteinisches Wattenmeer, 1998) veröffentlicht.

8 Stand der Arbeiten

Von 1998 bis 1999 liefen zwei vom UBA geförderte Projekte in Niedersachsen und in Schleswig-Holstein, um die Datenhaltung für das TMAP in Deutschland einzurichten. Die Projekte waren an der Nationalparkverwaltung Niedersächsisches Wattenmeer in Wilhelmshaven und dem Landesamt für den Nationalpark Schleswig-Holsteinisches Wattenmeer in Tönning angesiedelt. Unterstützt wurden die Arbeiten durch die GKSS in Geesthacht (Hajo Krasemann) und das CWSS in Wilhelmshaven.

Innerhalb dieser Teilprojekte wurde eine Arbeitsteilung der Datenmodellierung vorgenommen. Da beide Projektteile unterschiedliche Systeme der Datenhaltung nutzen, konnte so die systemunabhängige Funktionalität der jeweiligen Entwicklungen der Datenmodelle und der Zugangsmodalitäten überprüft werden. Diese systemunabhängige Funktionalität ist wichtig, da die TMAP Data Units in den einzelnen Ländern unter verschiedenen Systemkonfigurationen laufen.

Im Rahmen der Projekte wurden Datenmodelle für biologische und physikalisch-chemische Daten des Monitoring erstellt und zum Teil implementiert. Die Entwicklung eines Modells zur Datenhaltung GIS-basierter Daten ist abgeschlossen, eine Implementierung konnte noch nicht stattfinden.

Andere Parametergruppen befinden sich im Prozess der trilateralen Abstimmungen, teils fehlen noch Vorgaben aus übergeordneten Monitoringprogrammen (JAMP).

Der Prototyp des Datenzugangs aus dem DEMOWAD-Projekt (Demowad-Unit, CGI-Script) musste durch eine neue Anwendung ersetzt werden, dieses Java Applet wurde durch Brockmann Consult / Scicon erstellt und ermöglicht den Zugang zu den Datenbanken (<http://cwss.www.de/TMAP/Data-Unit/Data.html>).

Die Neuentwicklung des Zugangsmoduls zu den Data Units und der unerwartet hohe Zeitaufwand bei den Abstimmungen innerhalb der Teilprojekte hat zu Verzögerungen im Zeitablauf geführt, so dass die Datenmodelle der TMAP-Data Units noch nicht vollständig implementiert werden konnten.

9 Fazit und Ausblick

Im Rahmen des Projekts »Datenmanagement für das trilaterale Monitoring und Bewertungsprogramm TMAP« konnten funktionsfähige Data Units in Tönning und in Wilhelmshaven erstellt werden, die Datenhaltung und Zugangsmechanismen zu den Monitoring-Daten des Wattenmeeres bereithalten.

Die Entwicklungen der Datenmodelle sind weitgehend abgeschlossen, als nächstes müssen die Datenbanken weiter mit den Daten des Monitoring gefüllt werden. Dieses und die Erstellung und Implementation der restlichen Datenmodelle sind Aufgaben der nächsten Projektphase. Eine weitere Aufgabe ist die Einbindung von Daten aus Monitoring-Projekten des Bundes, nach Klärung von Fragen der Datenhaltung stehen hier auch die Anpassungen und Erweiterungen der Data Units an.

Als vorteilhaft hat sich die Verwendung einer modularen Struktur der Datenbanken und der Datenmodelle erwiesen, so konnte in den Partnerprojekten verteilt entwickelt werden, ohne dass die Datenmodelle den anderen Datenbanken angepasst werden mussten.

10 Literaturverzeichnis

- INTEGRATED MONITORING PROGRAM OF THE WADDEN SEA ECOSYSTEM. Report of the Trilateral Monitoring Expert Group (TMEG). 1993, Common Wadden Sea Secretariat, Wilhelmshaven
- Landesamt für den Nationalpark Schleswig-Holsteinisches Wattenmeer (Hrsg.): Wattenmeermonitoring 1998, Schriftenreihe des Nationalparks Schleswig-Holsteinisches Wattenmeer, Tönning
- Life Nature Project DEMOWAD: Data Handling, Final Report on the Definition Phase, October 1996, Common Wadden Sea Secretariat, Wilhelmshaven
- Marencic et al: THE TRILATERAL MONITORING AND ASSESMENT PROGRAM (TMAP). Expert Workshop 1995/1996. Wadden Sea Ecosystem No. 6. 1996

- Pommerencke, M.; Boer, H.; Larsen, H. S.; Luerßen, G.; Sandbeck, P.: Networking and Harmonisation of International Monitoring Data in: W. Geiger, A. Jaeschke, O. Rentz, E. Simon, Th. Spengler, L. Zilliox, T. Zundel (Hrsg.): *Umweltinformatik 1997 / Informatique pour l'Environnement 1997*, „Umweltinformatik aktuell“, Band 15
- STADE DECLARATION – Trilateral Wadden Sea Plan. Ministerial Declaration of the Eighth Governmental Conference on the Protection of the Wadden Sea, Stade 22 October 1997, Common Wadden Sea Secretariat, Wilhelmshaven
- Trilateral Monitoring and Assessment Group, 1997: TMAP Manual. The Trilateral Monitoring and Assessment Program (TMAP). Common Wadden Sea Secretariat, Wilhelmshaven

Repräsentation der Struktur biologischer Sammlungen als Grundlage für die Schaffung eines europäischen Metainformationssystems

Anton Güntsch, Andrea Hahn und Walter G. Berendsohn¹

1 Einleitung

Biologische Sammlungen (Lebenssammlungen, Präparate, Substanzsammlungen und Beobachtungsdatenbanken) sind die wichtigsten „Auskunftssysteme“ für rezente und ausgestorbene Organismen auf der Erde und bilden damit eine wertvolle Informationsquelle für Wissenschaft und Wirtschaft (Campbell 1998). Im Umweltbereich dient die in ihnen enthaltene Biodiversitätsinformation unter anderem als Grundlage für Gutachten (z. B. Umweltverträglichkeitsprüfung und Landschaftsplanung), für das Monitoring von Umweltveränderungen und für den Naturschutz (Berendsohn 1998).

Um Zugang zu Objekten oder Beobachtungsdaten biologischer Sammlungen zu erlangen, kann derzeit nur sehr eingeschränkt auf übergreifende Informationssysteme zurückgegriffen werden, so zum Beispiel im Bereich Zoologischer Gärten (ISIS²) oder pflanzengenetischer Ressourcen (IPGRI³). In den meisten Fällen ist es nötig, zunächst die Sammlungsträger zu identifizieren und direkt mit ihnen in Kontakt zu treten. Die Unterschiede im Erschließungsstand der Sammlungen lassen die Implementierung eines gemeinsamen europäischen Informationssystems, das den Zugriff bis auf die Ebene aller einzelnen Sammlungsobjekte gestattet, für die nächsten Jahre als nicht realistisch erscheinen.

Ein pragmatischer Ansatz für die Schaffung einer gemeinsamen Schnittstelle ist die Implementierung eines Metainformationssystems, das Sammlungsinformation auf verschiedenen Abstraktionsebenen vorhalten kann und abfragbar macht (Abbildung 1). Im einfachsten Fall kann diese Information nur aus dem Namen und der Adresse sowie einer groben Kategorisierung bestehen, wenn detailliertere Daten nicht vorliegen. Gleichzeitig sollen Sammlungsbeschreibungen aber auch wesentlich exakter wiedergegeben werden können, zum Beispiel, wenn genau bekannt ist, welche taxonomischen Gruppen (Arten, Pflanzenfamilien, Ordnungen von Tieren) in einer Sammlung vor-

¹ Botanischer Garten und Botanisches Museum Berlin-Dahlem, Abteilung für Biodiversitätsinformatik und Laboratorien, Königin-Luise-Str.6-8, D-14191 Berlin,
email: a.guentsch@mail.bgbm.fu-berlin.de, Internet: <http://www.bgbm.fu-berlin.de/biocise>

² <http://www.worldzoo.org/default.htm>

³ <http://198.93.227.125/Institute/intro.htm>

handen sind. Wenn für Sammlungen der Zugriff auf die einzelnen Objekte vorgesehen ist, sollte es möglich sein, daß Anfragen direkt an das betreffende Sammlungsinformationssystem weitergeleitet werden. Eine adäquate Repräsentation von Sammlungsstrukturen ist notwendige Voraussetzung für die Implementierung eines solchen Informationssystems (Powell 1998).

Im folgenden sollen die Ergebnisse einer Umfrage unter europäischen sammlungs-tragenden Institutionen dargestellt und darauf aufbauend ein Entwurf für eine hierarchische Repräsentation biologischer Sammlungen umrissen werden. Anhand eines Beispiels wird die Möglichkeit der automatischen Verschlagwortung von Sammlungsmetadaten skizziert.

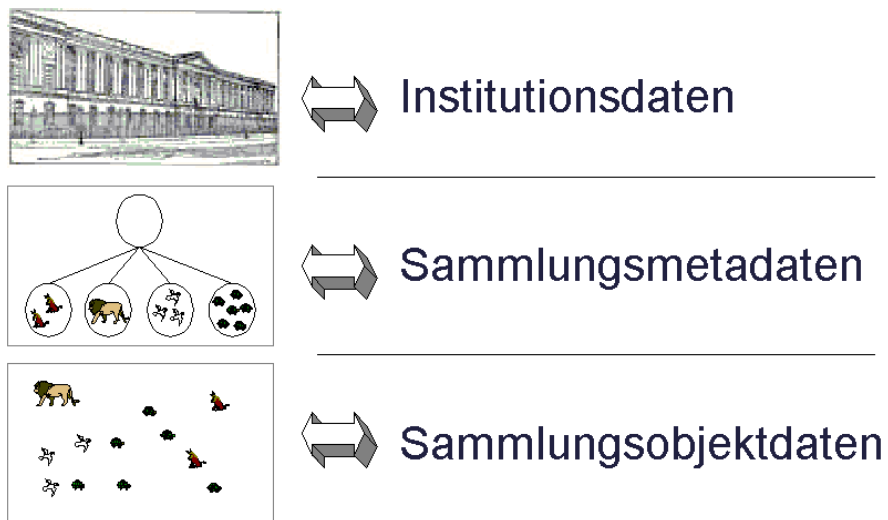


Abbildung 1
Ebenen des Informationszugangs zu biologischen Sammlungen

2 Aus Umfragen gewonnene Information über biologische Sammlungen

BioCISE ("A Biological Collection Information Service for Europe – Resource Identification"; Berendsohn 1997a) ist ein von der Europäischen Union (GD XII) gefördertes interdisziplinäres Projekt unter der Beteiligung von 10 EU Staaten und Israel. Das Ziel der gegenwärtigen Projektphase (August 1997 bis Dezember 1999) ist es, die Grundlagen für die Implementierung eines europäischen Informationssystems für biologische Sammlungen zu schaffen. Basierend auf einer umfassenden Umfrage und den Erkenntnissen der ersten Projektphase (1993-1996) (Berendsohn 1997b), in welcher ein umfassendes Informationsmodell für biologische Sammlungen entwickelt wurde, werden europäische Sammlungen analysiert und katalogisiert. Der resultierende Sammlungskatalog wird in einem relationalen Datenbank Management System gepflegt und ist über das WWW öffentlich zugänglich⁴.

2.1 Merkmale und Besonderheiten der Umfragedaten

Die in der Umfrage erhobenen Daten beschreiben Trägerschaft und Inhalt biologischer Sammlungen. Ein Schwerpunkt liegt auf der Untersuchung und Katalogisierung von Sammlungsinformationssystemen. Das in der Umfrage abgedeckte Spektrum von Sammlungstypen ist sehr weit: Botanische und zoologische, aber auch mikrobiologische Lebendsammlungen; naturhistorische und sonstige Präparatesammlungen; Beobachtungsdatenbanken floristischer und faunistischer Felduntersuchungen und spezialisierte Sammlungsformen wie Tierstimmenarchive oder Sammlungen von Naturstoffisolaten. Entsprechend ist eine Anleitung der befragten Institutionen im Ausfüllen der Fragebögen in den jeweiligen Spezialbereichen nur sehr eingeschränkt möglich, und für viele Bereiche ist die Unvollständigkeit der gewonnenen Daten charakteristisch.

Das Niveau der informationstechnischen Erschließung der einzelnen Sammlungen reicht von nicht inventarisierten bis zu vollständig durch Datenbanken verwalteten Sammlungen mit aufwendigen Schnittstellen für das World Wide Web. Korrespondierend mit der organisatorischen Struktur der sammlungstragenden Institutionen treten in diesem Gesamtspektrum u. a. die folgenden typischen Fälle auf:

1. Die Sammlung ist weitgehend elektronisch erfasst, einzelne Teilsammlungen wurden allerdings nicht einbezogen.
2. Die Sammlungserfassung wurde in einzelnen Bereichen einer Institution unabhängig begonnen, beispielsweise in den Zuständigkeitsbereichen verschiedener Kuratoren eines naturhistorischen Museums. Die resultierenden Teildatenbanken sind nur bedingt korrespondierend, die Dokumentation des gesamten Sammlungsbestandes ist lückenhaft.

⁴ <http://www.bgbm.fu-berlin.de/biocise/DataBase/default.htm>

3. Die Dokumentation des Sammlungsbestandes liegt weitgehend noch als Kartei-system vor, eine elektronische Erfassung ist oft aus finanziellen Gründen in nächster Zeit nicht oder nur in Teilsammlungen vorgesehen.

Vor allem der Anteil an Sammlungsträgern ohne elektronische Datenerfassung macht die Notwendigkeit deutlich, zwischen Sammlungsinformationen und Sammlungsinformationssystemen zu unterscheiden. Eine angemessene Repräsentation der Sammlungsinhalte ist in der Regel nur unter Einbeziehung der nicht elektronisch dokumentierten Bestände möglich. Bei der Sammlungserhebung werden daher drei grundlegende Kategorien von Sammlungsinformation erfragt:

- **Sammlungsträger:** Institutionelle Adresse des Sammlungshalters und Selbstdarstellung, gegebenenfalls organisatorische Struktur von Teilsammlungen, Expertise im Bereich Sammlungsdatenbanken, Kooperationspartner.
- **Sammlung:** Gegenstand der Sammlung und Einschränkungen (beispielsweise zeitlich oder geographisch), Anzahl der Objekte, Konservierungsstatus / Beschaffenheit des Objekts (z. B. lebend, Fotos, Beobachtungsdaten).
- **Sammlungsdatenbank:** Name und Akronym, Software und System, Urheberschaft, Administration, ggf. Dokumentation, Umfang der Abdeckung der genannten Sammlungen durch die Datenbank.

Kategorische Sammlungsbezeichnungen, wie zum Beispiel „Botanischer Garten“, geben meist nur einen Hinweis auf Teile der Sammlung. In der Regel sind die Sammlungsobjekte verschiedenen Kategorien zugeordnet: Botanische Lebendsammlungen oder Beobachtungsdaten beispielsweise sind oft begleitet von Belegsammlungen in Herbarien, entsprechendes gilt im Bereich zoologischer Sammlungen. Diese Teilung ist in Datenbanken nicht immer eindeutig repräsentiert und muß den unterschiedlichen Sammlungskonzepten z. B. eines Botanischen Gartens und eines Herbariums (Sammlung konservierter Pflanzen) häufig sinngemäß zugeordnet werden.

Eine weitere Besonderheit ist die Existenz von Teilsammlungen, deren Zugehörigkeit zum Sammlungskonzept vom Sammlungsträger als bekannt vorausgesetzt und daher nicht gesondert angeführt wird (z. B. Samenlager in botanischen Gärten), oder die aus historischen Gründen in einer Gesamtsammlung verblieben sind, obwohl sie nach sachlichen Kriterien einer anderen Sammlung besser zugeordnet wären (z. B. eine Teilsammlung rezenter mikroskopischer Algen, insbesondere Diatomeen, in der Sammlung fossiler Invertebraten des Naturkundemuseums Berlin (Lazarus/Jahn 1998)). Zur Erschließung solcher verborgener Information, die zur Zeit nur sehr eingeschränkt zugänglich ist, ist eine genaue Analyse von Sammlungshierarchien notwendig.

2.2 Gegenwärtige World Wide Web Schnittstelle des BioCISE Sammlungskatalogs

Soweit die Informationen zur Veröffentlichung freigegeben wurden, sind die erfaßten Sammlungsinformationen und Experten in einer dynamisch generierten Ausgabe zugänglich⁵.

Als Modellsystem für eine mögliche Kopplung von Informationssystemen über biologische Sammlungen dient die Zusammenarbeit mit dem BIODIV⁶ Projekt in Belgien (Güntsch/Vander Velde 1998), das auf nationaler Ebene eine vergleichbare Zielsetzung verfolgt: Über den Sammlungskatalog von BioCISE sind auch die von BIODIV erfaßten Sammlungsinformationen zugänglich (Abbildung 2). Eine einfache Protokollvereinbarung erlaubt die Integrierung derjenigen Informationen in die BioCISE Datenbank, die für die Generierung des gemeinsamen Sammlungskataloges benötigt werden. Anfragen an den Sammlungskatalog von BioCISE werden über einen dynamisch generierten Hyperlink direkt an die entsprechenden Seiten im Informationssystem von BIODIV weitergeleitet (Abbildung 2).

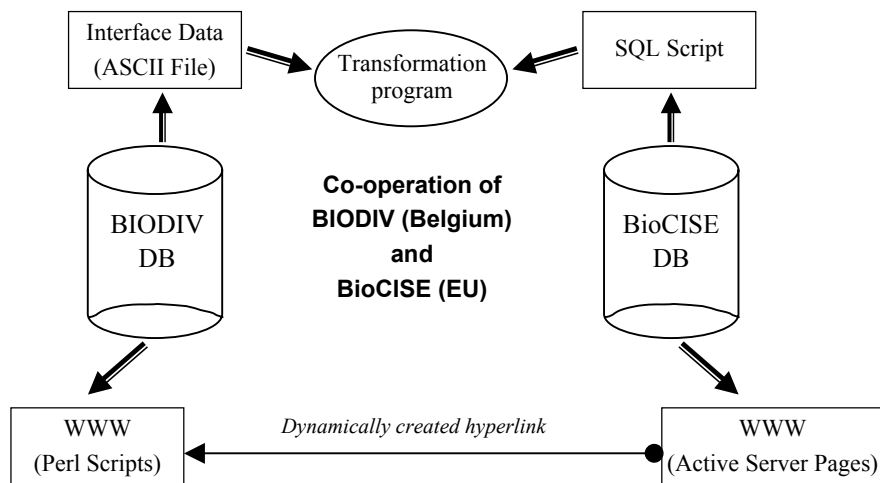


Abbildung 2

Integration belgischer Sammlungen in den BioCISE Sammlungskatalog

⁵ <http://www.bgbm.fu-berlin.de/biocise/>

⁶ <http://www.br.fgov.be/biodiv/>

2.3 Integration europäischer Sammlungen durch Metainformation

Um einen Zugang zu Sammlungsinformationen zu ermöglichen, sind grundsätzlich zwei Wege vorstellbar: Die Vernetzung von Sammlungsdatenbanken auf der Ebene der Sammlungsobjekte als bottom-up-Verfahren, oder die Beschreibung der Sammlungsinhalte (top-down). Die umfassende Integration einer großen Anzahl von Sammlungen auf europäischer Ebene ist innerhalb überschaubarer Zeit nur durch Metainformationen zu erreichen: Die Heterogenität der Datenbankanwendungen und die Notwendigkeit, auch bisher nicht elektronisch erfaßte Sammlungen zugänglich zu machen, empfehlen dieses Vorgehen. Die Informationstiefe ist damit zunächst teilweise stark eingeschränkt, Sammlungen sind aber umgreifend erfaßbar. Notwendige Bedingung für die Einbeziehung einer Sammlung sind Informationen zu ihrem Namen und ihrer Lokalität. Eine Vernetzung mit bestehenden Datenbanken ist aber nicht ausgeschlossen, so daß in entsprechenden Fällen auch Informationen bis auf die Ebene der Sammlungsobjekte zugänglich sind.

3 Darstellung von Sammlungsstrukturen

Eine notwendige Voraussetzung für die Schaffung eines benutzerfreundlichen und flexiblen Sammlungsinformationssystems ist die geeignete Repräsentation der zu verarbeitenden Objekte, in diesem Falle also der biologischen Sammlungen. Das zu entwickelnde Modell muß dabei so einfach strukturiert sein, daß es sich durch Algorithmen rechnerisch erfassen läßt. Andererseits darf es die Wirklichkeit nicht so vereinfachen, daß wesentliche Eigenschaften biologischer Sammlungen aus dem Modell nicht mehr ablesbar oder berechenbar sind.

In diesem Abschnitt werden die grundlegenden Anforderungen an das Modell erläutert und ein geeigneter Rahmen für die interne Repräsentation angegeben. Besondere Bedeutung hat dabei, daß die zu verarbeitenden Daten vorwiegend aus Fragebögen stammen werden und damit Eigenschaften aufweisen, welche die Modellbildung stark beeinflussen.

3.1 Hierarchische Organisation biologischer Sammlungen

Biologische Sammlungen sind ganz überwiegend hierarchisch aufgebaute Strukturen aus Teilsammlungen, die wiederum weiter unterteilt sein können. Die Kriterien, nach denen solche Aufteilungen vorgenommen werden, sind vielfältig. So kann ein naturhistorisches Museum zum Beispiel aus den Teilsammlungen Fossilien, Insekten und Schausammlung bestehen. Das bedeutet, die Teilsammlungen sind nach der Form (Fossilien), taxonomisch (Insekten) und nach der Motivation (Schausammlung) gebildet worden. Diese können dann wieder Teilsammlungen enthalten, die nach anderen Kriterien (z. B. geographisch) gebildet werden.

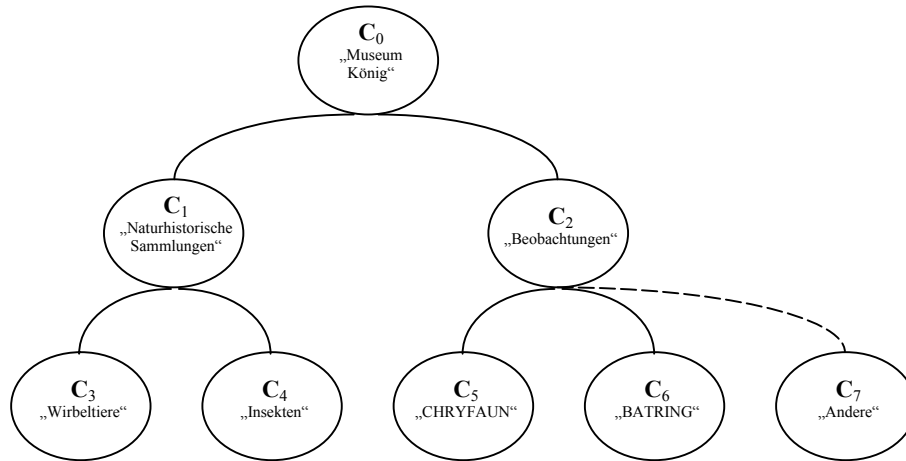


Abbildung 3

Hierarchische Struktur der Sammlung eines naturhistorischen Museums

Eine adäquate Repräsentation sollte Sammlungen daher hierarchisch als Bäume darstellen. Jeder Knoten des Baumes repräsentiert dabei eine Teilsammlung, also eine Menge von Sammlungsobjekten. Abbildung 3 zeigt den Aufbau der Sammlung eines größeren naturhistorischen Museums. Die Sammlung besteht aus den naturhistorischen Sammlungen und einer Menge von Datenbanken, die Feldbeobachtungen enthalten⁷. Die naturhistorischen Sammlungen lassen sich wieder in die Teilsammlungen „Wirbeltiere“ und „Insekten“ aufspalten.

Da die Daten aus Fragebögen stammen und somit häufig nicht klar ist, ob diese vollständig ausgefüllt wurden, muß eine Möglichkeit geschaffen werden, diese Unsicherheit auszudrücken. Hierfür wird ein Knoten mit der Bezeichnung „other“ eingeführt, der besagt, daß möglicherweise noch weitere Sammlungen existieren. Im obigen Beispiel wurde ein solcher Knoten an den Knoten C₂ angehängt, um anzuzeigen, daß weitere Beobachtungsdatenbanken existieren könnten.

⁷ Ein wichtiges Ergebnis des CDEFD Projektes war, daß physikalische Sammlungsobjekte und Objekte, die lediglich als Datensätze in Beobachtungsdatenbanken gehalten werden, weitgehend mit dem gleichen Modell beschrieben werden können (Berendsohn, W. G. 1997b, Berendsohn, W.G. et al. 1999). Dieses Erkenntnis gilt erst recht für die Repräsentation ganzer Sammlungen.

3.2 Die Beschreibung von Teilsammlungen

Nachdem im letzten Abschnitt skizziert wurde, wie sich hierarchische Sammlungsstrukturen als Bäume darstellen lassen, soll nun gezeigt werden, wie die einzelnen Teilsammlungen beschrieben werden können. Die Erfahrungen aus der gegenwärtigen Projektphase und die bisherigen Ergebnisse der "Collection description working group" (Powell 1998), die sich – basierend auf dem Dublin Core (Anon. 1999) – mit der Spezifikation von beschreibenden Attributen für Sammlungen allgemein (also auch nicht biologische Sammlungen) beschäftigt, haben ergeben, daß die folgenden Attributbereiche für die Charakterisierung biologischer Sammlungen relevant sind:

- **Administration (Admin):** Attribute, mit denen die Verwaltung der Sammlung beschrieben wird. Relevant ist hier zum Beispiel die Frage, ob ein Sammlungsinformationssystem eingesetzt wird oder ob zumindest ein herkömmliches Karteikartensystem vorhanden ist.
- **Sammlungshalter (Owner):** Attribute, welche die Institutionen und Personen beschreiben, die für die Sammlung verantwortlich sind.
- **Zeitliche Abdeckung (Time):** Denkbare zeitliche Bezüge sind der Zeitraum, in dem die Objekte im Feld gesammelt wurden, die erdgeschichtliche Zeit (z. B. bei Fossilien) und der Zeitraum, in dem die Sammlung aufgebaut wurde.
- **Räumliche Abdeckung (Area):** Beschreibung der geographischen Herkunft der Sammlungsobjekte. Der Bezug kann dabei sehr unterschiedlich sein und reicht von administrativen Bezügen (z. B. Nordrhein-Westfalen) über historische Bezüge (z. B. „Preußen“) bis hin zu geomorphologischen Bezügen (z. B. „Atlantikküste“).
- **Form der Objekte (Shape):** Attribute, mit denen der Zustand der enthaltenen Sammlungsobjekte beschrieben wird, also zum Beispiel „lebend“ in einem Botanischen Garten, „konserviert“ in einem Herbarium oder „Beobachtung“ in einer Kartierungsdatenbank.
- **Taxonomie (Taxon):** Taxonomische Klassifikation der Sammlung. Eine geeignete Behandlung von Synonymen und unterschiedlichen taxonomischen Auffassungen muß dabei gewährleistet sein.
- **Einschränkungen (Constraints):** Häufig unterliegen Sammlungen speziellen Einschränkungen, die mit den vorher angegebenen Attributbereichen nicht ausreichend beschrieben werden können. Im botanischen Bereich könnte dies zum Beispiel die Beschränkung auf Heilpflanzen sein.
- **Sammlungsgrund (Purpose):** Beschreibung der Motivation für die Existenz der Sammlung. Mögliche Gründe sind zum Beispiel „Wissenschaft“, „Ausstellung“ oder „Ausbildung“.

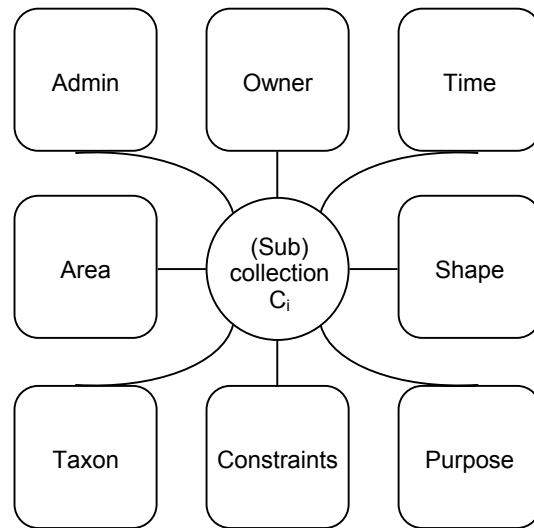


Abbildung 4
Charakterisierung biologischer Sammlungen

Eigenschaften von Teilsammlungen müssen dabei quantifizierbar sein, da sie sich jeweils nicht notwendigerweise auf die gesamte Teilsammlung beziehen müssen. Die Verwendung des Vokabulars

$$V = \{\text{"all"}, \text{"mainly"}, \text{"some"}\}$$

hat sich dabei als ausreichend erwiesen.

Die genaue Spezifikation der relevanten Attribute wird Gegenstand der weiteren Projektarbeit sein und mit Hilfe des vorliegenden umfangreichen Datenbestandes europäischer Sammlungen validiert werden. Im folgenden Abschnitt wird gezeigt, wie bereits mit wenigen Attributen Sammlungsbeschreibungen so implementiert werden können, daß mit einem regelverarbeitenden System Schlagwörter automatisch zugewiesen werden können.

4 Regelbasierte Verschlagwortung

Im Hinblick auf die zu schaffende Benutzerschnittstelle für ein Sammlungsinformationssystem kommt der Verschlagwortung der einzelnen Sammlungen eine wichtige Rolle zu. Schlagwörter bilden eine gemeinsame Quelle für Benutzer und Indexierer und minimieren somit die Unsicherheit bei der Systembenutzung (van der Vet & van Bakel 1998). Dabei geht es wegen der inhomogenen informationstechnischen Erschließung biologischer Sammlungen nicht um die automatisierte Verschlagwortung der Datenquellen selbst, wie bei Nikolai et al. (1998), sondern um die Abbildung von Sammlungsmetadaten auf Schlagwörter. Neben der manuellen Vergabe bietet sich die Möglichkeit, Schlagwörter auf Regeln basierend automatisch zu vergeben, was folgende Vorteile hat:

1. Die Semantik der Schlagwörter ist durch die Regeln explizit definiert.
2. Es ist „interessanter“, die Zeit, die für die Verschlagwortung nötig wäre, auf die Formulierung von Regeln zu verwenden.
3. Ändert sich die Semantik von Schlagwörtern, oder kommen neue Schlagwörter hinzu, so muß nicht die gesamte Menge der Dokumente (in unserem Falle der Sammlungs-Metainformation) neu verschlagwortet werden.

4.1 Interne Repräsentation von Sammlungen mit Prolog

Sowohl für die Implementierung des in Abschnitt 3 beschriebenen Rahmens für die Beschreibung biologischer Sammlungen als auch für die Formulierung von Beispielergebnissen zur Schlagwortvergabe haben wir uns für die auf der Prädikatenlogik erster Stufe basierende Programmiersprache Prolog entschieden, da sich regelverarbeitende Systeme mit Prolog als Prototypen schnell implementieren lassen⁸.

Eine Sammlung läßt sich durch das folgende Prolog-Fakt repräsentieren:

collection(CollectionID,	/* Eindeutiger Schlüssel der Sammlung */
Name,	/* Name der Sammlung */
InstitutionPt,	/* Zeiger auf die verwaltende Institution */
NumberOfUnits,	/* Anzahl der Sammlungsobjekte */
IsPartOfPt,	/* Zeiger auf die übergeordnete Sammlung */
shape(Status, ShapeLst),	/* Form der Sammlungsobjekte */
taxon(Status, TaxonLst),	/* Taxonomische Klassifikation */
area(Status, AreaLst),	/* Geographischer Bezug */
time(Status, TimeLst),	/* Temporaler Bezug */
constraint(Status, ConstraintLst),	/* Einschränkung des Sammlungsgebietes */
purpose(Status, PurposeLst),	/* Grund der Sammlung */
admin(Status, AdminLst)).	/* Sammlungsverwaltung */

⁸ Praktische Einführungen in Prolog werden in Sterling & Shapiro (1986) und Hanus (1987) gegeben. Die Grundlagen logischer Programmierung werden in Lloyd (1984) erläutert.

Diejenigen Attribute, für die Mehrfacheinträge zulässig sind, werden durch die Struktur

$$\text{Attributname}(\text{Status}, \text{AttributLst}), \quad \text{Status} \in \{1,2,3,4\}$$

wiedergegeben, wobei *AttributLst* eine (möglicherweise leere) Liste von Tripeln (*Attr, Qualifier, Number*) ist. *Attr* bezeichnet dabei eine Ausprägung für das jeweilige Attribut, *Qualifier* ist ein Element von V wie in Abschnitt 3.2 definiert, und *Number* gibt die Anzahl der Objekte an, auf welche die jeweilige Eigenschaft zutrifft. Der Eintrag für Status bedeutet jeweils:

- 1: Es existieren Einträge.
- 2: Es existieren keine Einträge, weil das zugehörige Attribut auf einer anderen Ebene des Baumes vergeben wird.
- 3: Es existieren keine Einträge, weil nicht bekannt ist, ob das zugehörige Attribut anwendbar ist.
- 4: Die Sammlung ist bezüglich des Attributes nicht eingeschränkt (sammelt z. B. ein naturhistorisches Museum Objekte aus der ganzen Welt, dann würde dies durch den Eintrag `area(4, [])` repräsentiert).

Als Beispiel für die Darstellung mit Prolog Fakten soll eine kommerzielle Rebsortensammlung dienen, welche aus einer Freiland- und einer Samensammlung besteht. Die Samensammlung besteht wiederum aus einer Wildsorten- und einer Rebsortensammlung. Die zugehörigen Fakten werden automatisch aus einer relationalen Datenbank, welche die Metadaten verwaltet, generiert, wobei fehlende numerische Einträge mit -1 wiedergegeben werden.

```

Collection(0, 'Rebsortensammlung', 188,
  -1,
  -1,
  shape(1, [(dormant, some, 16000)
            , (living_out, some, -1)]),
  taxon(1, [(plant, all, -1)]),
  area(3, []),
  time(2, []),
  constraint(2, []),
  purpose(1, [(commercial, mainly, -1)]),
  admin(1, [(cis, some, -1)])
).

Collection(1, 'Samensammlung', 188,
  16000 ,
  0 ,
  shape(1, [(dormant, all, 16000 )]),
  taxon(1, [(plant, all, 16000 )]),
  area(3, []),
  time(2, []),
  constraint(2, []),
  purpose(1, [(commercial, mainly, -1)
              , (conservation, some, -1)]),
  admin(1, [(cis, mainly, 15000 )])
).

Collection(2, 'Wildartensamen', 188,
  -1,
  1 ,
  shape(1, [(dormant, all, -1)]),
  taxon(1, [(plant, all, -1)]),
  area(3, []),
  time(2, []),
  constraint(2, []),
  purpose(1, [(conservation, all, -1)]),
  admin(1, [(cis, all, -1)])
).

collection(3, 'Rebsortensamen', 188,
  -1,
  1 ,
  shape(1, [(dormant, all, -1)]),
  taxon(1, [(plant, all, -1)]),
  area(3, []),
  time(2, []),
  constraint(2, []),
  purpose(1, [(commercial, all, -1)]),
  admin(1, [(cis, all, -1)])
).

collection(4, 'Freilandkultur', 188,
  -1,
  0 ,
  shape(1, [(living_out, all, -1)]),
  taxon(1, [(plant, all, -1)]),
  area(3, []),
  time(2, []),
  constraint(2, []),
  purpose(1, [(commercial, all, -1)]),
  admin(3, [])
).

```


4.2 „Botanischer Garten“ als Beispiel für automatische Verschlagwortung

Als Beispiel für die automatische Verschlagwortung von Sammlungen soll nun der Begriff „Botanischer Garten“ dienen, der dem Benutzer eines Sammlungsinformationssystems sicher als vordefinierte Anfrage zur Verfügung gestellt würde. Welche Eigenschaften charakterisieren einen Botanischen Garten?

Notwendig ist sicherlich zumindest eine Teilsammlung mit lebenden Pflanzen in einer Gartenkultur. Es sollte aber auch gewertet werden, wenn diese Sammlung öffentlich zugänglich ist und wissenschaftlich genutzt wird. Der folgende Programmausschnitt realisiert diese natürlichsprachliche Regel als Prolog Klausel. Dabei werden den Sammlungen Punkte bezüglich des Schlagwortes „Botanischer Garten“ zugewiesen. Sammlungen ohne lebende Pflanzen in Gartenkultur bekommen 0 Punkte. Alle anderen bekommen 50 Punkte und jeweils 10 weitere Punkte für den Nutzen Ausstellung und Wissenschaft.

```
checkKeyword(CollID, 'Botanical Garden', Score) :-
  isDescribedBy(taxon,plant,[all,mainly,some],_,CollID),
  isDescribedBy(shape,living_out,[all,mainly,some],_,CollID),
  shouldBeDescribedBy(purpose,exhibition,[all,mainly,some],_,CollID, S1),
  shouldBeDescribedBy(purpose,science,[all,mainly,some],_,CollID, S2),
  Score is 50 + S1 + S2.

checkKeyword(_,_,0).
```

Die Anwendung dieser Regel auf 10 ganz verschiedenartige Sammlungsbeschreibungen führte zu folgendem Ergebnis, wobei die Punktzahl jeweils das Maximum der Punktzahlen der jeweiligen Teilsammlungen ist:

Sammlung	Punkte
Sammlungen des Instituts für Rebenzüchtung	50
Sammlungen des Zoologischen Gartens Hof	70
Dutch plant collections, Boskoop	50
Sammlung von Conjugaten-Kulturen, Hamburg	0
Blumeninsel Mainau: Park, Garten und Forst	70
Sammlungen der „Fasanerie“ Wiesbaden	70
Botanischer Garten und Botanisches Museum Berlin-Dahlem	70
Museum für Haustierkunde am Institut für Tierzucht und Tierhaltung	0
Station de Recherches Agronomiques, Citrus collections	60
Sammlungen des Forstbotanischen Gartens der Stadt Köln	70

Offensichtlich ist hier der Begriff „Botanischer Garten“ recht gut abgebildet worden, obwohl die Sammlungen beschreibenden Attribute noch äußerst einfach gehalten sind. Auffällig ist die hohe Punktzahl, die dem Zoologischen Garten Hof und der Sammlung der „Fasanerie“ Wiesbaden zugewiesen wurde. Diese resultiert aus der Tatsache, daß beide Sammlungen über eine Teilsammlung verfügen, welche alle in der Regel formulierten Kriterien für einen „Botanischen Garten“ erfüllt. Ein „menschlicher“ Indexierer hätte sich hier möglicherweise durch die Bezeichnungen „Zoologischer Garten“ und „Fasanerie“ täuschen lassen. Die resultierenden Punkte können auf der Ebene einer Benutzerschnittstelle genutzt werden, um die Ergebnisse von Suchanfragen zu sortieren.

4.3 Zusammenfassung und Ausblick

Diese Arbeit liefert einen Beitrag zur Frage nach der geeigneten Repräsentation von Metadaten für biologische Sammlungen. Dabei wurden – basierend auf den Erfahrungen des BioCISE Projekts – weniger die konkreten sammlungsbeschreibenden Attribute untersucht, als vielmehr ein Rahmen angegeben, mit dem sich hierarchisch aufgebaute Sammlungen, von denen häufig nur unvollständige und ungenaue Informationen vorliegen, einfach darstellen lassen. Es wurde gezeigt, wie sich dieser Rahmen leicht in der logischen Programmiersprache Prolog implementieren läßt. Darauf aufbauend wurde am Beispiel des Schlagwortes „Botanischer Garten“ gezeigt, wie eine automatische Abbildung von Sammlungsmetadaten auf Schlagworte realisiert werden kann.

Die weitere Arbeit wird sich nun auf die exakte Spezifikation der sammlungsbeschreibenden Attribute und die Verarbeitung von parallelen Bezugssystemen, wie sie z. B. im Bereich der geographischen Daten vorkommen, konzentrieren. Der regelbasierte Ansatz zur automatischen Schlagwortzuweisung sollte anhand weiterer Regeln z. B. „Zoologischer Garten“, „Genbank“ oder „Kartierungsprojekt“ näher untersucht werden. Hierfür sollte ein Akquisitionsmodul implementiert werden, welches die Eingabe neuer Regeln auch für in der Programmierung unerfahrene Sammlungsexperten möglich macht.

Literatur

- Anon. (1999): The Dublin Core: A Simple Content Description Model for Electronic Resources. Dublin Core Metadata Initiative (ed.) [<http://purl.oclc.org/dc/index.htm>].
- Berendsohn, W. G. (1997a): The BioCISE project: Resource development for a Biological Collection Information System in Europe. In: European Science Foundation, International Workshop “New Directions in Systematics“. Hersonissos October 1997; Abstract Volume.
- Berendsohn, W. G. (1997b): A Common Datastructure for European Floristic Databases (CDEFD). In: Hoeveler, A. & Cresti, M. (ed.), Biotechnology (1992-1994), Final Report Vol.1. Luxembourg.

- Berendsohn, W. G. (1998): Datenstrukturforschung und international vernetzte Umweltinformation auf der Ebene der Organisation. In: Hoppe, J., Helle, S. & Krasemann, L. (Hrsg.): Praxis der Umweltinformatik Band 7 – Vernetzte Umweltinformation. Marburg: 33-45.
- Berendsohn, W.G., Anagnostopoulos, A., Hagedorn, G., Jakupovic, J., Nimis, P.L., Valdés, B., Güntsch, A., Pankhurst, R.J. & White, R.J. (1999): A comprehensive reference model for biological collections and surveys. *Taxon* 48: 511-562.
- Campbell, P. (1998): 101 uses for a dead bird. *Nature* 394 (6689): 105.
- Güntsch, A. & Vander Velde, A. (1998 [Nov]): Collaboration of BIODIV and BioCISE [<http://www.bgbm.fu-berlin.de/biocise/TheProject/IntroCollab.htm>].
- Hanus, M. (1987): Problemlösen mit Prolog. Stuttgart.
- Lazarus, D. & Jahn, R. (1998): Using the Ehrenberg Collection. *Diatom Research* 13(2): 273-291.
- Lloyd, J. W. (1984): Foundations of Logic Programming. Berlin, Heidelberg, ...
- Nikolai, R., Koschel, A., Kramer, R. & Sattler, T. (1998): Automatisierung der Metadatenaktualisierung am Beispiel des Umweltdatenkatalogs UDK. In: Hoppe, J., Helle, S. & Krasemann, H.L. (Hrsg.). Praxis der Umweltinformatik Band 7 – Vernetzte Umweltinformation. Marburg: 141-158.
- Powell, A. (ed.) (1998 [Oct]): Metadata – Collection Level Description (Collection description working group – Work in progress) [<http://www.ukoln.ac.uk/metadata/cld/wg-report/>].
- Sterling, L. & Shapiro, E. (1986): The Art of Prolog – Advanced Programming Techniques. Cambridge, London.
- van der Vet, P. E. & van Bakel, B. (1998): Combining linguistic and knowledge-based engineering for information retrieval and information extraction. In: Proceedings of the 14th Twente Workshop on Language Technology (TWLT14).

Zur Dynamik von Metadaten

Paul Braun¹

1 Einleitung

Informationen über Forschungsdaten erlangen eine Schlüsselfunktion wenn letztere keine einheitliche Struktur besitzen, unterschiedlichen Typs sind oder die Nutzer der Daten aus verschiedenen Fachrichtungen stammen. Im Bereich des Umweltschutzes sind gemeinhin alle drei Bedingungen erfüllt. Metadaten bilden dann die Voraussetzung für das fachübergreifende Arbeiten mit Umweltforschungsdaten. Um deren Eignung für bestimmte Fragestellungen zu erkennen, verlangen gerade Umweltdatenbanken kennzeichnende Informationen über die enthaltenen Daten (Günther et al. 1994).

Dabei ist man versucht, Metadaten als statisches Element einer Umweltdatenbank zu betrachten. Der folgende Beitrag soll aufzeigen, wie in der Praxis Metadaten durchaus eine Dynamik entwickeln können, welche zu Beginn eines Umweltdatenbankprojekts nicht erwartet worden war.

2 Hintergrund

Mit der Einrichtung des Sonderforschungsbereichs 299 (SFB 299) „**Landnutzungskonzepte für periphere Regionen**“ entstand ein fachbereichsübergreifendes Großforschungsprojekt im Bereich Umweltsicherung an der Universität Gießen. Der SFB 299 hat sich zum Ziel gesetzt, eine Methodik zu entwickeln, um regionalspezifische und nachhaltige Optionen der Landnutzung zu erarbeiten und zu bewerten. Stellvertretend für die Vielzahl der Regionen in Deutschland werden die Naturräume Lahn-Dill-Bergland sowie Dilltal, im Westen des Bundeslandes Hessen, untersucht (Abb. 1). Diese Region zeichnet sich durch einen besonders starken Rückgang ackerbaulich genutzter Flächen aus. Damit verbunden ist die Frage, welche Möglichkeiten der Landnutzung dort zukünftig bestehen und wie sie wahrgenommen werden (Frede und Bach 1998). Neben naturräumlich/ökologischen, kulturhistorischen und landwirtschaftlichen Rahmenbedingungen werden sozioökonomische und rechtliche Faktoren die weitere Landentwicklung dieser Region beeinflussen.

¹ Justus-Liebig-Universität - FB Agrarwissenschaften und Umweltsicherung
Institut für Pflanzenbau und Pflanzenzüchtung II - Biometrie und Populationsgenetik -
Ludwigstr. 27, 35390 Gießen, Fon: 0641/99 37554, Fax: 0641/99 37549
email: Paul.Braun@agrar.uni-giessen.de

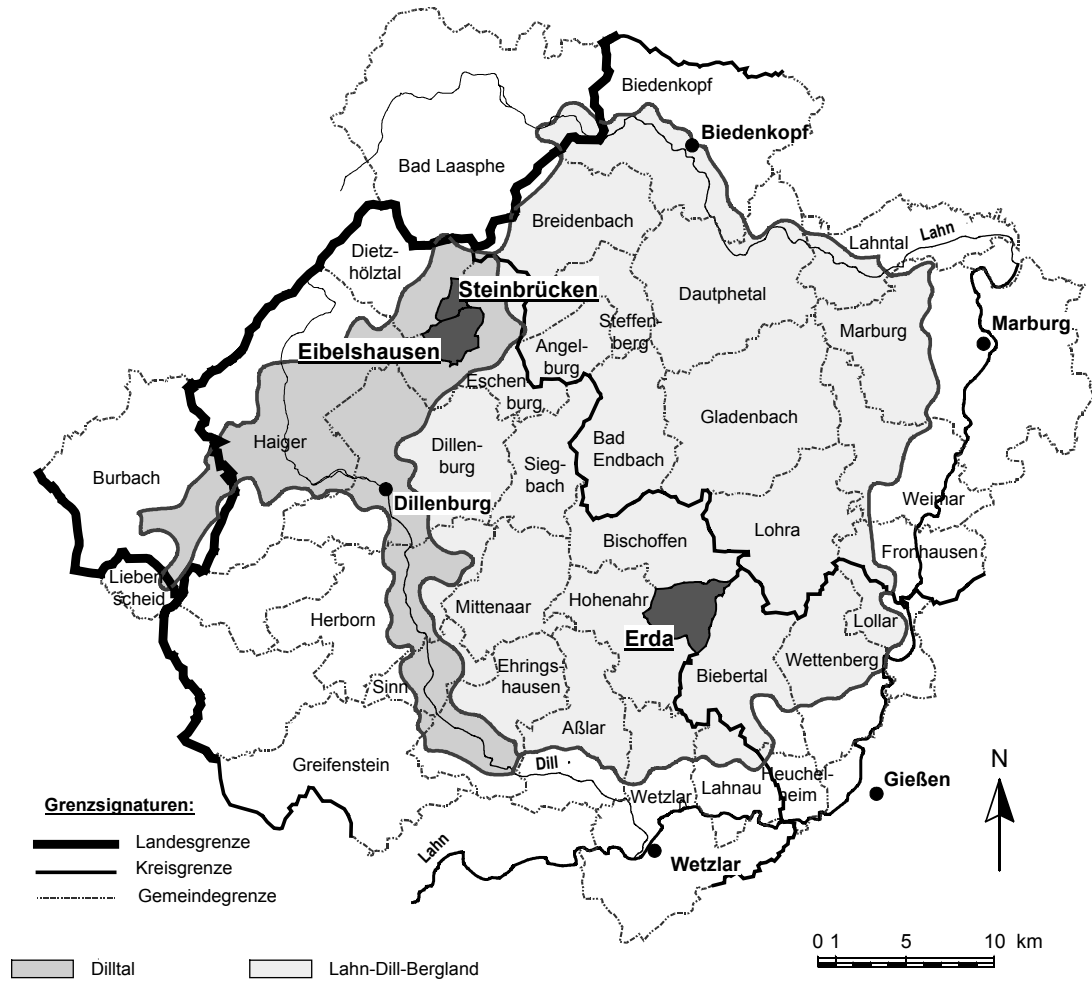


Abbildung 1

Natur- und politikräumliche Gliederung der Untersuchungsregion
des Sonderforschungsbereichs 299 „Landnutzungskonzepte für periphere Regionen“.

Nach Antragstellung im Jahr 1996 wurde der SFB für das darauffolgende Jahr bewilligt und am ersten Januar 1997 eingerichtet. Ein eigenständiges Teilprojekt übernahm die zentrale Datenverwaltung der Forschungsergebnisse aus dem SFB. Dies unterstreicht die Wichtigkeit, die dem Aufbau einer Umweltdatenbank beigemessen wird. Die erste Antragsphase läuft Ende 1999 aus. Ein Fortsetzungsantrag ist eingereicht.

3 Aufgabenstellung

Aufgrund des erwarteten starken Datentransfers zwischen den 20 Teilprojekten des SFB 299 wurde von Projektbeginn an ein zentrales Datenmanagement betrieben. Dieses sollte im wesentlichen den Austausch von Daten fördern, deren Archivierung ermöglichen und die Datensicherung gewährleisten. Aufbauend auf den aus dem Forschungsprojekt entstehenden Datenbanken könnte in der zweiten Projektphase ein Umweltinformationssystem entwickelt werden, das die teilprojektübergreifende Datenverarbeitung erleichtern würde. Aus dem hohen Stellenwert der Datenhaltung und der fachübergreifenden Zusammenarbeit im Projekt folgte die Notwendigkeit eines Metadatenmanagements. Mit Blick auf den SFB und seine Dauer mußten die Metadaten drei Aufgaben erfüllen:

1. Der kurzfristige **Datentransfer** innerhalb einer Projektphase von drei Jahren ist durch Metadaten zu erleichtern. So soll sich jeder berechtigte Datennutzer schnell und einfach in den Daten eines fremden SFB Teilprojekts zurecht finden.
2. Langfristig, also über mehrere Projektphasen hinweg, ist sicherzustellen, daß Datennutzer, die nicht zeitgleich mit den Autoren dieser Daten im SFB arbeiten, diese problemlos **auswerten** können.
3. Aus **Archivierungsgründen** und zum Zwecke der Einordnung der Ergebnisse in den Projektfortschritt sind Metadaten zu generieren, welche die Herkunft der Daten und deren Intention dokumentieren.

Diese Aufgaben schienen, basierend auf Voruntersuchungen, einfach zu lösen.

4 Herangehensweise

Bereits in der Antragsphase zum SFB 299 fand eine **Erkundung** hinsichtlich der zu erwartenden Daten statt. Dazu wurden alle teilnehmenden Arbeitsgruppen befragt, was für Daten aus ihrem Projekt anfallen würden, welche Struktur diese vermutlich besäßen, und wie groß der entsprechende Speicherbedarf demnach sein müßte. Dabei zeigte sich, daß vornehmlich mit relationalen Daten aus Meßreihen zu rechnen war. Daneben wurden aggregierte, relationale Daten sowie Computerprogramme angekündigt.

4.1 Metadatendynamik

Ein erster **Metadatenrahmen** wurde daraus unverzüglich abgeleitet. Dieser Metadatenrahmen wurde einfach gehalten und bestand aus einer Quellangabe sowie maximal 5 Schlüsselwörtern. Beim Datentransfer sollten diese Angaben in zwei getrennten Zeilen vor den eigentlichen Daten stehen. Aufgrund dieser Erkundung wurde der zentrale SFB-Server dimensioniert und ein relationales Datenbankmanagementsystems (DBMS) gesucht, welches ein Textmodul anbot. Die Entscheidung fiel auf Oracle, welches zu güns-

tigen Konditionen erhältlich war. Gleichwohl wurde bei der Entscheidung für Oracle als DBMS im Auge behalten, daß mit Fortgang des Projektes auch Daten anfallen könnten, die besser durch einen objektorientierten als einen relationalen Ansatz modelliert würden.

Metadatenkonzept (Handout): Für das Datenmanagement selbst galten die üblichen Probleme von Umweltdatenbanken entsprechend. Hinzu trat, daß es sich um ein junges Projekt handelte, in dem Daten- und Informationsaustausch in diesem Umfang neu waren. Eine angemessene Informationstradition hatte sich bis dahin noch nicht bilden können. Um die Notwendigkeit der Datendokumentation möglichst früh in den Teilprojekten zu verankern, wurden deshalb zu Projektbeginn im Januar 1997 erste Maßgaben zu Datentransfer und -qualität an die Arbeitsgruppen des SFB herangebracht. Hierbei wurde die Metadatenproblematik herausgestellt und ein Handout an alle SFB Teilnehmer verteilt, welches die formalen Voraussetzungen der Metadaten beschrieb. Diese Angaben fußten auf der Erkundung im Jahr 1996 und dem daraus abgeleiteten Metadatenrahmen. Kernforderung dieses Metadatenkonzepts war, Daten so zu dokumentieren, daß sie jederzeit und ohne Rückfragen von Projektneulingen bzw. anderen Arbeitsgruppen auswertbar sein würden. Zu diesem Zweck war für jede Dateneinheit (Datenobjekt, Bezugsdaten, Daten) ein Metadatensatz durch das verantwortliche Teilprojekt zu erstellen. Dieser sollte jedem Datenobjekt voranstellen. Mit diesen Metadaten würden folgende Ziele erreicht werden:

Transfer: Metadaten sollen die Verstehbarkeit von Daten im laufenden Betrieb gewährleisten. Damit sind sie eine unabdingbare Voraussetzung für den Datentransfer und das gemeinsame Arbeiten an den gleichen Datensätzen zwischen den Teilprojekten.

Archivierung: Metadaten müssen die Information bereitstellen, die notwendig ist, um zu bestimmten Themen Daten aufzufinden.

Auswertung: Im Hinblick auf eine langfristige Verwertbarkeit enthalten die Metadaten Informationen, die es erlauben auch noch im Nachhinein sinnvoll mit den Versuchsergebnissen zu arbeiten.

Als Nebenziel sollen die Metadaten bewirken, daß die Nutzer bzw. die Datensender eine Art von **Datenhygiene** entwickeln, in der Weise, daß sie redundante Daten weglassen und die Daten in eine leicht les- und verstehbare Form bringen.

Entsprechend diesem Metadatenkonzept wurden folgende Angaben gefordert:

WER hat das Datenfile erstellt (Kennung) ?

WAS ist der Inhalt des Datenfiles?

WO wurden die Daten erhoben?

WIE erfolgte die Datenerhebung?

Noch wurden den Arbeitsgruppen Spielräume beim Umsetzen dieses Konzepts für ihre spezifischen Daten gelassen, da verschiedentlich erkennbar wurde, daß die nach der Erkundung erwarteten Datenobjekte doch nicht die einzigen bleiben würden.

Diese Offenheit erlaubte es, die Metadaten im Wechselspiel mit den Teilprojekten fortzuentwickeln.

Anhand dieser Vorgaben konnten bereits früh Daten standardisiert auf den zentralen SFB-Server Biber abgelegt werden. Das Handout bzw. Metadatenkonzept löste den Metadatenrahmen ab, der bis dahin galt. Während der Metadatenrahmen ausschließlich auf der Befragung der SFB Teilnehmer im Vorfeld der Antragstellung fußte, war das Metadatenkonzept bereits eine Fortentwicklung dieser Struktur, die nach Gesprächen mit Arbeitsgruppen während der gemeinsamen Beschaffung oder dem Aufbau von Hard- und Software entstand.

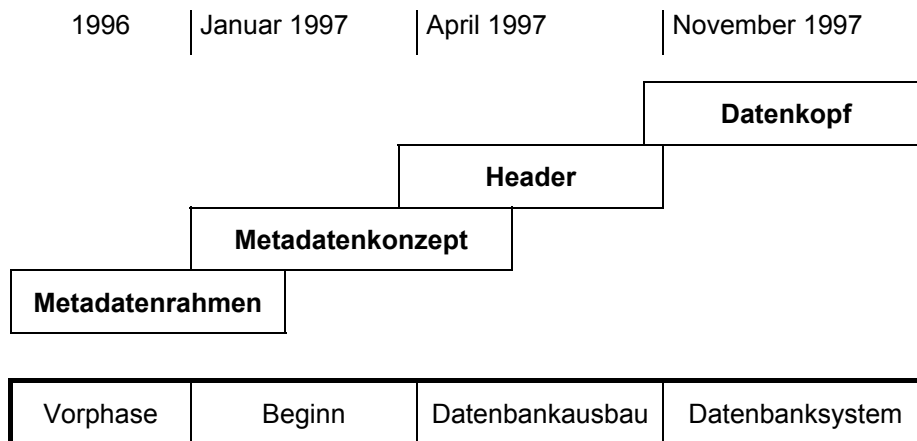


Abbildung 2

Zeitliche Abfolge der Metadatenentwicklung in der Umweltdatenbank LADIB des Sonderforschungsbereichs 299.

Header: Im April 1997 mußte das Metadatenkonzept erneut angepaßt werden, da in diesem Monat Daten aus einem mit dem SFB thematisch verwandten, jedoch bereits abgeschlossenen Projekt, für den SFB bereitgestellt wurden (Abb. 2). Es handelte sich um thematische Karten aus dem Lahnprojekt, die mit dem Geographischen Informationssystem ERDAS verarbeitet wurden. Waren im Metadatenkonzept lediglich inhaltliche Festlegungen getroffen und deren Reihenfolge abgestimmt worden, mußte nunmehr eine Konkretisierung und auch Erweiterung der Metadaten erfolgen. Daraus entstand ein Header, der künftig jedem Datenbankobjekt vorausgehen sollte. Zusätzlich wurde erstmals bestimmt, daß der Header eine eigenständige Datei zu einem Datenobjekt sein würde. Bei der Entwicklung dieses Headers mußten mögliche Weiterentwicklungen eines Datenbanksystems beachtet werden. Daher wurde der Header bereits so strukturiert, daß er Erweiterungen aufnehmen konnte.

Seitens der Teilprojekte war man sich von Projektbeginn an einig über die Notwendigkeit eines Datenmanagements. Doch durfte die daraus resultierende Mehrar-

beit für die Teilprojekte nicht zu groß werden. Der Header blieb daher einfach, kurz und auf allen Plattformen zu verarbeiten.

```

DATEIKOPF                                     !                                     !
1234567890123456789012345678901234567890123456789012345678901234567890
KENNUNG*****!++++++KENNUNG
  Autor/in:      Dirk Wenzel
  Teilprojekt:   D_7
  Antragsperiode: 1
  erstellt:      01.01.1998
  Transfer:      01.03.1998
  Dateinamen:    d2_001a #
  Anmerkungen:   dies ist nur ein Beispiel
WAS*****!++++++WAS
  Kurzüberschrift: Selbstbild der Landwirte. (max. 60 Zeichen bzw. 60 Z.)
  Kurzbeschreibung
  des Dateiinhalts: Diese Datei enthaelt Sekundärdaten zur Entwicklung des
                    Bildes von Landwirtschaft bei Landwirtschaftsfamilien und
                    dessen Dynamik frueher und heute. (300 Zeichen, 300 Z.) #
WO*****!++++++WO
  Naturraum:     Gladenbacher Bergland (Lahn-Dill-Bergland), Dilltal (120 Z.)
  Land:          Hessen, Nordrhein-Westfalen (60 Z.)
  Kreis:         Angelburg, Bad Endbach, etc. (180 Z.)
  Gemeinde:      Alle Gemeinden des Kreises wurden befragt (120 Z.) #
  Gemarkung:     nicht relevant (120 Z.)
  Flur:          nicht relevant (120 Z.)
  Stueck:        nicht relevant (120 Z.)
  Punkt:         nicht relevant (120 Z.)
  Topographische K.: 1:25000 #
VERKNUEPFUNG*****!++++++VERKNUEPFUNG
  mit Teilprojekt: D1, D3, D4, A1, E1
  zu Datei:        d2_001b, d2_001c, d2_002, d3_005d ... (120 Z.)
  abgeleitet aus: Primärdaten - Dateien; a1_103
AUSWERTUNG*****!++++++AUSWERTUNG
  Versuchsanlage: Befragung, Zufallstichprobe geschichtet nach Gemeinden
                  (120 Z.)
  Auswertung:     Rangkorrelation, Varianzanalyse (1 faktoruell, ordinal), ...
                  (120 Z.) #
  Prozesstufe:    Sekundärdaten
  Rasterelement: 25 x 25 Meter
  Dateiformat:    ASCII
BEISPIELDATENSATZ**!++++++BEISPIELDATENSATZ
  raster rechts hoch datum id var1 var2 var3 var4 var5 v1 v2
  1517 1105 3603820.13 5793801.08 050697 ab34 2.53 395.2 86.3 56 83540 42 38
  v3 v4 v5 q1 q2 q3 q4 n1 n2 text1
  12 9 7 ... 4 1 2 9 ... j n ... Regen ... (max. 255 Z.) #
  Erlaeuterung:
  raster          : Spalte 1-9 (s 1-9), Angabe des Rasterelements
  rechts          : s 11-20, Rechtswert, Gauss-Krueger-Koordinate
  links          : s 22-31, Hochwert, Gauss-Krueger-Koordinate
  datum          : s 32-38, Zeitpunkt der Rohdatenerhebung
  id             : s 40-43, Identifikation des Datensatzes
  var1           : s 45-49, Ak-Besatz, verhaeltnisskaliert (v)
  var2           : s 51-55, Wohngrundstueck (v) #
  var3           : s 57-61, Betriebsgroesse in ha (v)
  var4           : s 63-65, Milchvieh (v)

```

Abbildung 3
Variablen im Datenkopf von LADIB mit Beispielen.

Datenkopf: Parallel zur Entwicklung des Datenbanksystems LADIB, welches die Forschungsergebnisse aus dem SFB enthielt, wurde der Header zu einem Datenkopf ausgebaut (Abb. 2). Im Herbst des ersten Versuchsjahres erhielten SFB-Angehörigen ein Muster des Datenkopfes (Abb. 3). Zusätzlich wurde er in das Datenbanksystem

LADIB eingestellt. Dieser, auch gegenwärtig noch gültige Datenkopf besteht aus einem festen Bereich, dem Kern, der unabhängig von der Art der Daten immer anzugeben ist (KENNUNG, WAS). Ein zweiter Teil ist variabel, dergestalt, daß er bei bestimmten Bezugsdaten wegfallen kann (WO, VERKNÜPFUNG, AUSWERTUNG, BEISPIELDATENSATZ). Die zweigeteilte Datenhaltung, bei der Metadaten getrennt von den Bezugsdaten bleiben, wurde beibehalten. Mit diesem Aufbau ist der Datenkopf seit Ende 1997 Bestandteil des Datenbanksystems LADIB. Die Angaben im Datenkopf wurden umfangreicher verglichen mit der Vorläuferversion, dem Header. Dies war eine Folge der Zunahme verschiedener Bezugsdatentypen einerseits und dem Wunsch nach teilprojektübergreifender Verständlichkeit andererseits. Die grundsätzliche inhaltliche Struktur wurde aus dem Header übernommen.

Obschon der Datenkopf sich im laufenden Betrieb bewährte, traten neue Aspekte auf, welche in die Weiterentwicklung des Datenkopfes einfließen:

1. Die kurze Erläuterung datenspezifischer Eigenheiten (Abb. 3, Beispieldatensatz) ist zu starr. Für relationale Daten bewährt sich diese Form zwar. Für Texte ist dieser Aufbau aber beispielsweise nicht mehr in allen Teilen notwendig.
2. Spezifische Projektdaten erfordern Anpassungen des Datenkopfes. Die Anforderungen sind aber noch so verschieden, daß gegenwärtig von Fall zu Fall entschieden werden muß, was wegfallen kann oder zusätzlich angegeben werden muß.

Die genannten Aspekte führen zur internen Weiterentwicklung des Datenkopfes. Um die Anwender nicht zu verunsichern und um abzuwarten, wie wichtig verschiedene Anpassungen sind, werden Änderungsvorschläge des Datenkopfes zunächst gesammelt. Auf diese Weise kann eine stufenweise Anpassung des Datenkopfes erfolgen. Die grundlegende Struktur einer Kernmetadaten-datei welche um projektspezifische Anpassungen ergänzt wird, bleibt weiterhin gültig.

4.2 Formale Anforderungen an Metadaten

Das Datenmanagement führt im laufenden Betrieb eine individuelle Kontrolle der Metadaten und der zugehörigen Daten durch. Diese Aktivität trägt zur Verbesserung des Datenkopfes bei, ist aber zeitaufwendig. Der Vorteil eines derart **nicht**-automatisierten Vorgehens liegt darin, Denkanstöße aus den Fehlern beim Ausfüllen eines Datenkopfes zu erhalten. Beispielsweise erwies sich die Frage nach Dateinamen als fehlerträchtig (Abb. 3, Dateinamen). Als Folge daraus wird fortan die Angabe des Metadatei- und Bezugsdateinamens gefordert. Um auch weiterhin den Datenkopf im Wechselspiel mit den Teilprojekten zu verbessern, wird die Automatisierung der Metadatenverarbeitung noch zurückgehalten. In der zweiten Projektphase wird aber die Eingabe der Metadaten auch über WWW-Techniken angeboten. Ab diesem Zeitpunkt liegt soviel Erfahrung beim Datenmanagement vor, daß die Dynamik der Metadaten minimal bleibt. Dann kann auch die laufende Veränderung des Datenkopfes über eine Eingabemaske sinnvoll werden. Solange die Teilprojekte noch geringe Erfahrung mit dem Da-

tentransfer besitzen, muß deren Disziplin beim Erstellen von Metadaten überwacht werden.

4.3 Langfristige Lesbarkeit von Daten

Neben formalen Forderungen an den Datenkopf sind auch Aspekte langfristiger Lesbarkeit unterschiedlicher Formate zu beachten. Mit Blick auf die mögliche Laufzeit des Projekts (4 x 3 Jahre) und der Nutzbarkeit der Ergebnisse über die SFB-Dauer hinaus ist zu entscheiden, welches Datenformat die Metadaten sowie die Bezugsdaten künftig haben sollen. Zunächst leitet sich daraus die Forderung ab, im Datenkopf die zur Ergebnisdatenverarbeitung genutzte Software und Rechnerplattform anzugeben. Für die Metadaten erscheint der ASCII-Standard nach wie vor am verlässlichsten und wird hier beibehalten. Der Mangel an Formatierungsmöglichkeiten in ASCII-Formaten wird bei weitem durch dessen problemlose Portabilität aufgewogen. Inwieweit HTML-Formate hier für eine Ablösung sorgen könnten ist ungewiß. Für die Formate der Datendateien selbst stellt sich dieselbe Frage. Die in der Tabelle 1 aufgeführten Formate sind gegenwärtig für LADIB vorgesehen.

Tabelle 1: Vorschlag für Datenformate, um deren langfristige Lesbarkeit sicherzustellen.

Format	Datenart
ASCII	Meßwert
HTML, RTF	Text
PS	Grafik
TIFF	Karte

Darüber hinaus wird alte Software aufbewahrt, um notfalls auch überholte Formate zu lesen. Inwieweit auch alte Hardware vorgehalten werden muß, auf der entsprechende Software noch lauffähig ist, wird zu prüfen sein. Mit Blick auf eine langfristige Datenverfügbarkeit, scheint dies gegenwärtig ebenfalls ein gangbarer, wenn auch umständlicher Weg.

4.4 Anbindung an das WWW

Eine Schnittstelle von LADIB zum WWW wurde im Frühjahr 1998 generiert (Braun et al. 1998). Seither können sich Nutzer von LADIB dessen Datenbestand mit Hilfe eines Browsers betrachten. Zu jedem Datenbankobjekt ist ein kennzeichnendes Abstract lesbar. Dieses wird aus den Metadaten abgeleitet und gibt einen schnellen Eindruck vom Inhalt des jeweiligen Objekts. Der Zugriff auf die Daten und Metadaten ist jedoch nur für angemeldete Nutzer möglich.

Von Anwenderseite wurde der Wunsch geäußert, daß bei thematische Karten neben dem Abstract auch ein GIF-Bild dieser Karte als Metainformation verfügbar gemacht wird. Diese Anregung wird umgesetzt werden, führt aber zu einem zweiten Metadatenformat neben ASCII.

5 **Schlußfolgerung**

Das interdisziplinäre Umweltforschungsprojekt SFB 299 benötigt einen flexiblen und einfachen Datenaustausch zwischen den Teilprojekten. Um allen beteiligten Arbeitsgruppen diesen Transfer greifbar und einsichtig zu machen, wurde früh in der Projektlaufzeit ein zentrales Datenmanagement etabliert. Sehr schnell mußte dieses Vorgaben bezüglich des Datentransfers und der Metadaten machen. Da die Anforderungen aus den Teilprojekten hinsichtlich der Datenhaltung projektspezifisch variierten sowie einer zeitlichen Dynamik unterlagen, war die Festlegung der erforderlichen Metadaten schwierig. Dennoch erfolgten entsprechende Vorgaben zügig, damit der Datentransfer von Projektbeginn an in standardisierter Weise vonstatten gehen konnte. Allerdings mußte dabei in Kauf genommen werden, daß die Metadaten zunächst einer gewissen Dynamik unterliegen würden. Wie der vorstehende Artikel zeigt, war diese Dynamik anfangs des Projekts besonders groß. Eine allzu starre Festlegung bei Projektbeginn wäre aber den Erfordernissen aus den Teilprojekten zuwider gelaufen. Ebenso hätten zu detaillierte Vorgaben die Akzeptanz des standardisierten Datentransfers wenigstens in Mißkredit gebracht. Das Metadatenmanagement muß aber ein Vertrauensverhältnis zu den Nutzern aufbauen, weil die Güte einer Metadatenbank von deren Mitarbeit abhängt. Voreiliges und unflexibles Handeln ist deswegen kontraproduktiv. Auch künftig werden die Metadaten im SFB-Projekt eine, wenn auch geringe, Dynamik aufweisen, denn sie dokumentieren einen offenen Umweltdatenpool.

In der Vergangenheit und teilweise auch noch heute findet man sich im Umweltbereich vor großen bereits existierenden Datenmengen, die es mittels Metainformation zu dokumentieren und nutzbar zu machen gilt. In diesen Fällen ist der Datenpool abgeschlossen und die Daten selbst erzwingen keine Dynamik der Metadaten mehr. Allenfalls externe Bedürfnisse können eine Überarbeitung der Metadaten herbeiführen. Unter diesen statischen Bedingungen lassen sich Datenstrukturen besser beschreiben bzw. modellieren als bei einem offenen Datenpool, der ständig erweitert wird. Neben der Dynamik der Bezugsdaten erwiesen sich in den jungen Teilprojekten des SFB 299 auch die Nutzeranforderungen an die Umweltdatenbank als dynamisch, weil sie noch nicht ausgereift waren. Daher scheint ein dynamisches Metadatenkonzept die geeignete Form, um den Anforderungen aus dem SFB 299 gerecht zu werden. Um nicht wiederholt mit neuen Metadatenstrukturen an die Anwender herantreten zu müssen, erwies es sich als ein praktikabler Weg themaspezifische Metadaten an einen Metadatenkern anzuhängen.

Metainformationen oder Metadaten werden durch die Art und Dynamik der Daten, welche sie dokumentieren, beeinflusst. Bleiben diese Bezugsdaten statisch, bleiben es gemeinhin auch die Metadaten. Sind sie dagegen dynamisch so sind es die Metadaten gleichfalls, wenn auch abgeschwächt. Die Qualität einer Metadatenstruktur ist meines Erachtens daran zu erkennen, wie stark es die Dynamik von Bezugsdaten abfängt bzw. puffert.

Die Dynamik der Metadaten betrifft deren Inhalt, deren Anordnung sowie deren Format. In diesem Beitrag wurden insbesondere die ersten beiden Aspekte angesprochen. Längerfristig muß das Datenmanagement im SFB weitere Datenformate finden, die für unterschiedliche Datenarten verbindlich sind, wie auch für Metadaten. Die künftige Verwaltung der wachsenden Metadatenbank des SFB erfordert geeignete Werkzeuge. Augenblicklich versucht das SFB-Datenmanagement die Möglichkeiten des Umweltdatenkatalogs (Schütz & Böhm 1994, Legat et al. 2000) und von XML (Pott & Wielage 1999) für die Bedürfnisse des SFB abzuwägen.

Literatur

- Braun, P., Becker, H. & Köhler, W. 1998: Informationsmanagement im SFB 299: „Landnutzungskonzepte für periphere Regionen“. – Berichte der Gesellschaft für Informatik in der Land-, Forst- und Ernährungswirtschaft, Band 11, 6-9.
- Frede, H.-G., Bach, M. 1998: Leitbilder für Agrarlandschaften. Z. Kulturtechnik Landentwicklung 39, 117-120.
- Günther, O., Radermacher, F. J. & Riekert, W.-F. 1994: Umweltmonitoring: Modelle, Methoden und Systeme. In: Umweltinformatik. (Hrsg.: Page, B. & Hilty, L. M.), 51-78. – Oldenbourg Verlag, München.
- Legat, R., Batschi, W.-D., Hashemi-Kepp, H., Kruse, F., Swoboda, W., Nikolai, R., Nyhuis, D., Pultz, S., Stallbaumer, H. & Zirm, K. 2000: Der Umweltdatenkatalog UDK in Österreich – 5 Jahre Erfahrungen. – (in diesem Band) <http://www.pz-oekosys.uni-kiel.de/akudb/>
- Pott, O. & Wielage, G. 1999: xml. – Markt & Technik Buch und Software-Verlag GmbH, München.
- Schütz, T. & Böhm, R. 1994: Die Datenstrukturierung des Metainformationssystems Umweltdatenkatalog. In: Umweltdatenbanken. (Hrsg.: Kremers, H.), 245-258. – Metropolis-Verlag, Marburg.

Praktische Erfahrungen bei der Datenerschließung für eine Umweltdatenbank

Thomas Dombek und Klaus Tochtermann¹

1 Einleitung

Seit langem stehen in Deutschland eine Vielzahl von Einzelinformationen zum Zustand der Umwelt auf Bundes- und Länderebene zur Verfügung, die z. B. vom Umweltbundesamt (UBA) in Form von Berichten wie „Daten zur Umwelt“ regelmäßig publiziert werden. Die darin enthaltenen Berichtselemente werden im Projekt „Umwelt 2000“ elektronisch erfaßt, inhaltlich erschlossen und in einer zentralen Datenbank systematisch archiviert. Vor diesem Hintergrund beauftragte das UBA das Fraunhofer-Institut für Systemtechnik und Innovationsforschung (ISI) in Karlsruhe, das Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW) in Ulm, das Deutsche Institut für Wirtschaftsforschung (DIW) in Berlin und das Institut für Energie- und Umweltforschung (IFEU) in Heidelberg mit der Erstellung der „Informationsgrundlage für das Umweltpräsentationssystem des Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit (BMU) im Rahmen des Globalansatzes EXPO 2000“. Als Beitrag des BMU zur EXPO 2000 bezweckt „Umwelt 2000“ die flächendeckende Gesamtdarstellung des Umweltzustandes in Deutschland.

2 Die Umweltdatenbank „Datenbank Grunddaten“

Die Umweltdatenbank „Datenbank Grunddaten“ (DBG 1999) dient zur Unterstützung der Berichterstellung im UBA. Sie ist ein Arbeitsinstrumentarium des Berichterstatters, mit dem die Verwaltung aller Berichtselemente (Sachdaten) erleichtert wird. Die Betonung liegt auf alle, da dort sowohl vom BMU bzw. von den Ressorts freigegebene, nicht freigegebene als auch „nur“ mit dem UBA und weiteren Bundesbehörden abgestimmte Berichtselemente eingestellt werden (Abb. 1). Im Kontext der „Datenbank Grunddaten“ besteht ein Berichtselement aus zwei Teilen, der Sachdatei und den Rahmeninformationen (Metainformation). Anders als in reinen Metainformationssystemen werden in dieser Datenbank – neben den beschreibenden Informationen – auch die einzelnen Sachdaten selbst verwaltet. Die Rahmeninformationen sind un-

¹ Thomas Dombek und Klaus Tochtermann, Forschungsinstitut für anwendungsorientierte Wissensverarbeitung Ulm, Helmholtzstr. 16, 89081 Ulm
Email: {dombeck, tochterm}@faw.uni-ulm.de

abdingbare Voraussetzung für die Abschätzung, ob die Inhalte zweier Berichtselemente zu einem Thema identisch sind, um sie fallweise für eine verdichtete Zustandsbeschreibung vereinen zu können.

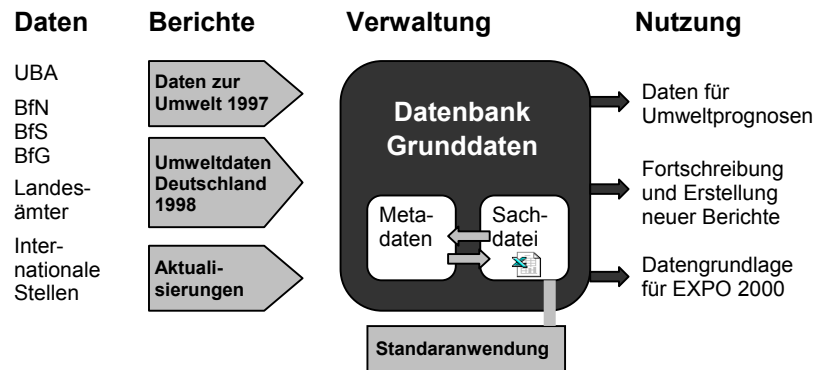


Abbildung 1

Inhaltlicher Kontext der „Datenbank Grunddaten“
(UIQUER: Umweltinformationsquellregister des UBA)

2.1 Technische Realisierung

Als Anwendung von Microsoft Access läßt sich die „Datenbank Grunddaten“ nahtlos in die Bürokommunikationsumgebung (Office 97) der Benutzer integrieren. Anhand der Rahmeninformationen kann über verschiedene Suchtechniken ein Berichtselement ausgewählt werden und die zugeordnete Sachdatei direkt in deren Standardanwendung zur Anzeige oder inhaltlichen Bearbeitung geöffnet werden. Um den Datenzugriff auch über externe Metainformationssysteme zu ermöglichen, werden derzeit Schnittstellen zum Umweltdatenkatalog (UDK 1999) und dem Umweltinformationsquellregister UIQUER des UBA entwickelt. Für die Recherche stehen vier unterschiedliche Suchtechniken zur Verfügung:

- Suche im Bericht, strukturell gegliedert nach den Berichts- und Kapitelebenen der Datenbank;
- Suche im Thesaurus, hier kann nach den Sachdateien zugeordneten Suchbegriffen aus dem Thesaurus recherchiert werden;
- Suche nach in der Datenbank vorgegebenen Themenbereichen;
- Freie Suche nach Begriffen in allen Metadatenfeldern, dabei können mehrere Felder verknüpft werden.

2.2 Der aktuelle Stand der Anwendung

Bisher wurden sämtliche Berichtselemente der vom UBA veröffentlichten Berichte „Daten zur Umwelt 1997“ und „Umweltdaten Deutschland 1998“ sowie weitere aktualisierte Zeitreihen und Umweltprognosen für ausgewählte Bereiche in die Datenbank aufgenommen. Das System wird seit mehreren Jahren im UBA entwickelt. Zur Zeit steht den Mitarbeitern des UBA die aktuelle Version 3.1 über das Intranet zur Recherche zur Verfügung. Darüber hinaus werden die Inhalte der „Datenbank Grunddaten“ von den Entwicklern des Umweltpräsentationssystems „Umwelt 2000“ für die EXPO 2000 genutzt. Inzwischen sind ca. 1500 Sachdateien in der Datenbank archiviert und inhaltlich erschlossen worden. Derzeit wird die Datenbank auch im Rahmen des Werkzeugkastens HUDA (Tochtermann et al. 1998) als zentrales Datenverwaltungssystem eingesetzt.

3 Datenerschließung in der „Datenbank Grunddaten“

Im Rahmen des Projektes „Umwelt 2000“ stellt das FAW Ulm die dem UBA bereits vorliegenden Sachdaten in die „Datenbank Grunddaten“ ein und erschließt diese Daten durch Rahmeninformationen. Diese werden in einem manuellen Verfahren erhoben und über ein Eingabeformular erfaßt, das gleichzeitig zur Anzeige der Rahmeninformationen bei der Recherche dient (Abb. 2). Damit erhalten auch die Nutzer einen Schreibzugriff auf die Rahmeninformationen.

Der inhaltliche Teil der Rahmeninformationen gliedert sich in Fach-, Raum- und Zeitbezug sowie ein kurzes Abstract, das bei der Recherche separat angezeigt werden kann. Weitere Felder beinhalten Angaben über die Informationsquelle, die Art der Datenerhebung sowie den Status und die Zugriffsrechte auf die jeweilige Sachdatei. Durch eine Liste von Querverweisen können inhaltliche Verknüpfungen zu verwandten Themen gesetzt werden. Bei der Verschlagwortung ist eine Auswahl aus etwa 3500 Begriffen eines integrierten Thesaurus zu treffen. Insgesamt bestehen die Rahmeninformationen aus ca. 40 verschiedenen Feldern. Um das optimale Wiederfinden der Berichtselemente zu gewährleisten, müssen die Rahmeninformationen nach einer einheitlichen Strategie erhoben werden.

Dokumentverwaltung

Titel: **Emissionen nach Emittentengruppen in Deutschland 1990 bis 1994**

Thema >

Inhalt: Absolute Jahresemissionsbilanzen für die Stoffe: Kohlendioxid, Stickstoffoxide, Schwefeldioxid, Kohlenmonoxid, Ammoniak, Distickstoffoxid, Staub, Flüchtige organische Verbindungen NMVOC und Methan in Mio.t bzw. kt. für die Jahre 1990, 1991, 1992, 1993 und 1994 in Deutschland, gegliedert nach Emittentengruppen;

Info -bestand (UIQUER): Emissionen ausgewählter Schadstoffe

Fachbezug >

Komponente/Stoff: CO2, NOx, SO2, CO, NH3, N2O, Staub, NMVOC, CH4

Erhebungsgröße: Menge pro Jahr

Maßeinheit: Mio.t, kt

Raumbezug >

Art der Verwaltungseinheit: Staat

Bezeichnung d. Verwalt.-einheit: Deutschland

Bezeichnung d. naturräuml. Einheit: Deutschland

Zeitbezug >

Zeitraum: 01.01.90 bis 31.12.94

*** Referenz-Version ***

Sachdatei:

Darstellungsform: Graphik

Datenerhebung

Info.-Quellen

Status

Querverweise

Anmerkungen

Thesaurus

Zuordnung: **Daten zur Umwelt 1997 // Luft // Grafiken**

Datensatz: 1 von 28 (Gefiltert)

Abbildung 2
Formular zur Eingabe und Anzeige der Rahmeninformationen

3.1 Inhaltliche Erschließung komplexer Umweltdaten

Eine grundsätzliche Problematik bei der inhaltlichen Erschließung von Umweltdaten besteht in deren heterogenem Charakter. Derart vielschichtige Informationen lassen sich nur bedingt durch ein homogenes System von Rahmeninformationen beschreiben. Auch die Verschiedenheit der in den Berichten enthaltenen Datenformate wie Texte, Tabellen oder Grafiken erschweren eine konsistente Erfassung der Rahmeninformationen. Ein einheitlicher Fachbezug – etwa bei umfangreichen Tabellen mit mehreren untersuchten Komponenten oder bei rein beschreibenden Textelementen – läßt sich kaum noch herstellen. Manche Elemente haben weder einen Zeit- noch einen Raumbezug, wie z. B. die Darstellung bedrohter Tierarten.

Insbesondere bei der inhaltlichen Erschließung durch Dritte bedarf es zum einen der Abstimmung mit dem Datenlieferanten, zum anderen aber auch einer gezielten Evaluierung durch die Nutzer des Systems. Im Interesse eines optimalen Rechercheergebnisses müssen die Rahmeninformationen sowohl im gesamten System konsistent erfaßt sein, als auch eine ausreichend genaue Beschreibung der einzelnen Sachdatei liefern. Häufig stehen beide Ziele im Widerspruch und der Erhebende muß im Einzelfall abwägen.

Damit ist eines der Hauptprobleme angeschnitten: die inhaltliche Abhängigkeit der Rahmeninformationen vom subjektiven Verständnis des Erhebenden. Arbeiten mehrere Personen an der Erfassung der Rahmeninformationen, so unterscheiden sich deren Einträge oft um Details und erschweren die spätere Datenrecherche. Freie Feldeinträge fallen meist individuell sehr unterschiedlich aus. Hier sind Vorgaben, welche die Datenbank in Form von Auswahllisten liefert, von Vorteil. Angaben etwa zur Erhebungsgöße oder der verwendeten Maßeinheit lassen sich damit weitgehend vereinheitlichen. Allerdings sollten solche Listen erweiterbar sein, da sich der Bedarf erst bei der Erfassung zeigt.

Als problematisch hat sich die Zuordnung von Informationsquellen im Fall der Erfassung von Rahmeninformationen durch Dritte erwiesen. Aus den Sachdateien selbst geht meistens keine Quelle hervor, und es erfordert langwierige Nachforschungen, den Datenlieferanten bzw. Autor des jeweiligen Dokumentes zu ermitteln. Diese Zusatzinformationen sollten nach Möglichkeit innerhalb der Sachdatei vermerkt oder durch den Datenlieferanten mitgeliefert werden.

Um auf inhaltliche Schwierigkeiten beim Erfassen der Rahmeninformationen zu reagieren, wurde im Fall der „Datenbank Grunddaten“ ein Katalog verfaßt, der die wichtigsten Kriterien für die einzelnen Feldeinträge dokumentiert. Dort wird z. B. festgelegt, ob chemische Formeln oder Trivialnamen zur Beschreibung der untersuchten Komponenten verwendet werden, bzw. die räumliche Auflösung der Darstellungsgebiete definiert. Solche einheitlichen Vorgaben reduzieren den manuellen Aufwand bei der Erfassung der Rahmeninformationen und sichern die Konsistenz der Einträge, wenn unterschiedliche Personen an der Erfassung beteiligt sind. Da der Katalog speziell auf die Anforderungen der „Datenbank Grunddaten“ zugeschnitten ist, lassen sich seine Aussagen nur eingeschränkt auf andere Systeme übertragen. Letztlich zeigt sich die Qualität der Rahmeninformationen aber erst während der Recherche und der Arbeit an der Datenbank.

3.2 Verschlagwortung mittels Thesaurus

Die Verschlagwortung der Sachdaten erfolgt in der „Datenbank Grunddaten“ über eine Liste von Suchbegriffen (Deskriptoren), die dem deutschen Teil des GEMET-Thesaurus (General European Multilingual Environmental Thesaurus) der EEA (European Environment Agency) entnommen sind (Umweltthesaurus, 1999). Momentan umfaßt die Liste etwa 3500 Begriffe aus allen Umweltbereichen (Abb 3).

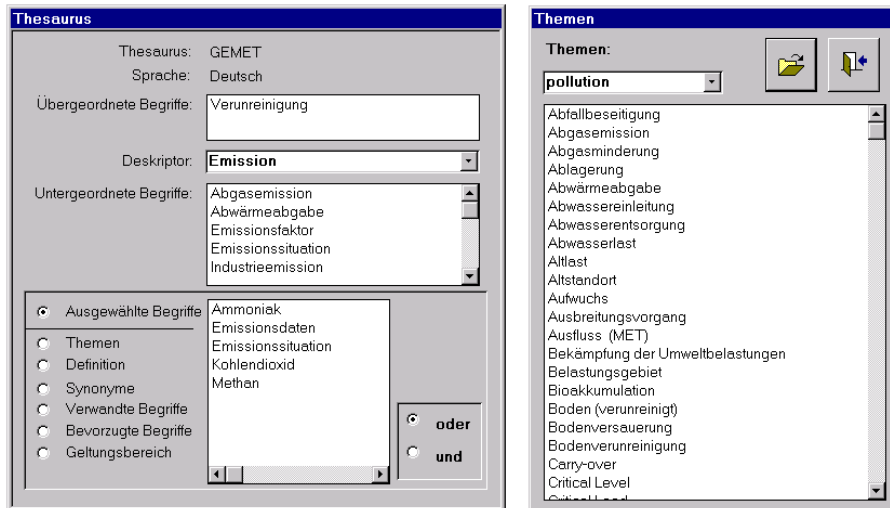


Abbildung 3
Maske zur Auswahl der Suchbegriffe aus dem Thesaurus

Allein die Verwendung eines leistungsfähigen Thesaurus garantiert allerdings noch keine optimale Verschlagwortung. Folgende Problemstellungen treten bei der Verschlagwortung durch fest definierte Thesauri häufig auf:

- Wichtige Begriffe werden aufgrund der Fülle an Daten nicht gefunden;
- Genau auf den Inhalt zutreffende Begriffe fehlen im Thesaurus;
- Die Relevanz der Begriffe wird vom jeweiligen Benutzer individuell unterschiedlich interpretiert;
- Es besteht Unklarheit über die optimale Breite der Begriffsauswahl;
- Eine Serie von durchgängig zutreffenden Begriffen kann nicht als Ganzes zugeordnet werden.

Durch daraus resultierende Unvollständigkeiten und Abweichungen bei der manuellen Verschlagwortung wird das Ergebnis der Recherche oft beeinträchtigt. Aus Anwendersicht zu begrüßen wäre hier ein Werkzeug zum leichteren Auffinden relevanter Begriffe, etwa durch die Einteilung in Kategorien oder eine hierarchische Gliederung des Thesaurus. Denkbar ist auch eine "Memory-Funktion" des Systems, um wiederholt verwendete Begriffe mehreren Datensätzen gleichzeitig zuordnen zu können. Wird – wie in diesem Fall – ein vorgegebener Thesaurus verwendet, so sollte er in Absprache mit den Benutzern an die speziellen Anforderungen der Datenbank angepasst werden. Häufig gesuchte Begriffe, die im Thesaurus nicht vorkommen, könnten auf diese Weise ergänzt werden.

4 Anforderungen an Umweltdatenbanken aus Anwendersicht

Der Verfasser der Rahmeninformationen spielt eine zentrale Rolle für die Leistungsfähigkeit eines Informationssystems, da Qualität und Homogenität der Rahmeninformationen entscheiden, ob die Recherche erfolgreich verläuft oder nicht. Im Interesse der Qualitätssicherung sollte das System den Vorgang der Erfassung beschreibender Informationen (Rahmeninformationen bzw. Metainformationen) möglichst effizient unterstützen. Eine verstärkte Automatisierung der Erfassung sichert die Einheitlichkeit der Einträge, vor allem wenn mehrere Personen beteiligt sind, und wirkt sich damit positiv auf Rechercheergebnisse aus. Häufig treffen Teile der beschreibenden Informationen auf mehrere inhaltlich verwandte Dokumente gleichzeitig zu. In diesem Fall muß sichergestellt werden, daß die entsprechenden Einträge identisch sind. Liefert eine Umweltdatenbank dazu kein geeignetes Werkzeug, so ist der manuelle Aufwand erheblich, und es treten leicht Abweichungen auf. Sinnvoll wäre die Option, mehrere Datensätze zum direkten Vergleich gegenüberzustellen oder eine Funktion zur Übernahme von „Informationspaketen“ in unterschiedliche Datensätze. Eine weitere Möglichkeit besteht darin, übergeordnete beschreibende Informationen zu definieren, die dann mehreren ausgewählten Datensätzen zugeordnet werden.

Ein entscheidender Punkt ist die detaillierte Abstimmung der Zugriffsrechte auf die Datenbank. In der Regel teilt sich der Benutzerkreis in Datenlieferanten und Datennutzer, die vorwiegend recherchieren. Die Arbeit an den Rahmeninformationen sollte auf erstere beschränkt bleiben. Je nach Beschaffenheit des Systems empfiehlt sich auch eine weitergehende Unterteilung. Änderungen an der Struktur der Rahmeninformationen – etwa die Ergänzung von Auswahllisten – unterliegen sinnvollerweise einer zentralen Kontrolle.

In jedem Fall ist eine nachträgliche Qualitätskontrolle der Rahmeninformationen durch fachkompetente Dritte empfehlenswert. Dieser Vorgang kann erheblich erleichtert werden, wenn es möglich ist, die Umweltdaten (Berichtselemente im Falle der „Datenbank Grunddaten“) automatisch nach gleichen oder ähnlichen Rahmeninformationen gruppieren zu können. Andernfalls ist es erforderlich, jedes Berichtsdokument zusammen mit seinen beschreibenden Informationen in den Abstimmungsprozeß zu geben. Bei einem Datenbestand von mehreren Tausend Berichtselementen führt dies zu einem erheblichen Aufwand.

Literatur

DBG (1999): Dokumentation der Datenbank Grunddaten Version 3.1. Dr. Lippke und Dr. Wagner GmbH, Berlin.

Tochtermann, K., Riekert, W.-F., Kadric, L., Kramer, R., Schmidt, R., Geiger, W., Peter, N., Reißfelder, M., Doberkat, E.-E., Sobotka, H.-G., Keitel, A., Zitzmann, W., Schütz, T., Burkhardt, J., (1998): HUDA: A toolbox for environmental report production. Tagungsband 12. GI Symposium Umweltinformatik, Metropolis Verlag.

UDK (1999): http://www.uvm.baden-wuerttemberg.de/www-udk/RPT_DATA/about_d.html.

Umweltthesaurus (1999): <http://www.umweltbundesamt.de/uba-datenbanken/thes.htm>

Der Datenaustausch für den Alpendatenquellenkatalog

Thomas Heinemeier¹

1 Einführung

Die Alpenkonvention verpflichtet die Teilnehmerparteien zur nachhaltigen Entwicklung der Alpenregion. Ein Alpenbeobachtungs- und -informationssystem (ABIS) wurde eingerichtet mit dem Auftrag, die Information zu erarbeiten, welche die Umsetzung dieser Verpflichtung zu verfolgen erlaubt. Hierfür bedarf es einer breiten Datenbasis aus so unterschiedlichen Bereichen wie Umwelt, Wirtschaft, Soziales, Kultur, Forschung. Die Internationalität, die Vielzahl der hiervon betroffenen Einrichtungen, ihre unterschiedliche institutionelle Einbettung in den Ländern sowie die Einbeziehung regionaler Datenquellen macht es unmöglich, die Realisierbarkeit eines Informationsproduktes hinreichend abzuschätzen.

Im Rahmen des ABIS wird daher ein Datenquellenkatalog erstellt, der als Verweisinstrument angelegt ist. Die in ihm erhaltene Metainformation dient der Kenntnisnahme und Lokalisierung vorhandener Daten. Die Nutzer sollen aufgrund der angebotenen Information entscheiden können, wie die allgemeine Datenlage im Hinblick auf eine gegebene Aufgabe aussieht, um eine Arbeitsrichtung festlegen zu können. Hierzu ist eine möglichst vollständige Abdeckung der Datenquellen erforderlich. In einem nächsten Schritt können dann detaillierte Metainformationen über die Spezifikation der Daten, welche zur Bewertung der Verwendbarkeit der Datenquelle erforderlich sind, erhalten werden ausgehend von in der Metainformation bezeichneten Bezugsquellen. Insbesondere im internationalen Umfeld sind solche Angaben wichtig, um die Vergleichbarkeit von Daten beurteilen zu können.

Der Alpendatenquellenkatalog (ADQK) muß mithin einen erheblichen Bedarf an Metainformation decken. Um dies zu erreichen, wird auf die Zusammenarbeit mit bestehenden oder im Aufbau befindlichen Katalogen gesetzt. Diesem Herangehen stehen die wohl dokumentierten Schwierigkeiten, mehrere Systeme dauerhaft zu verbinden und unterschiedliche Organisationen zu koordinieren, entgegen. Erfahrungen ähnlicher Unterfangen deuten darauf hin, daß ein verteiltes Metainformationssystem konzipiert als offenes System und internationale Metainformationsstandard verwendet ein günstiges Zielkonzept darstellt. Das Angebot an Datenhalter, Metainformationen ihrer Datenbestände durch bereitgestellte Instrumente selbst führen zu können wird gemeinhin als wichtiger Anreiz zur Mitarbeit angesehen. Auch die größere, öffentliche

¹ Umweltinstitut, GD Gemeinsame Forschungsstelle, Europäische Kommission. TP 321; I-21020 ISPRA. thomas.heinemeier@jrc.it

Sichtbarkeit der geleisteten Arbeit dank der Webtechnik sind als der Zusammenarbeit förderlich betrachtet worden.

Der vorliegende Beitrag stellt die bislang angetroffenen Herausforderungen und erreichten Ergebnisse dar im Hinblick darauf, inwieweit die Erwartungen erfüllt wurden. Das nachfolgende Kapitel stellt die Rahmenbedingungen vor, unter denen der ADQK errichtet wird. Der dritte Abschnitt illustriert den Datenfluß, worauf ein Kapitel zur Rolle der Mehrsprachigkeit diesbezüglich folgt. Schlußfolgerungen beschließen den Artikel.

2 Die Rahmenbedingungen des ADQK

2.1 Die Bezugsquellen

Alle Partner des ABIS betreiben eigene Datenquellenkatalog (DQK) Aktivitäten zur Zeit des Aufbaus des ADQK. In Deutschland gibt es einen nationalen Katalog ebenso wie Bundesländeraktivitäten, von denen in erster Linie Bayern für die Alpenkonvention relevant ist als einziges Bundesland, das im Alpenkonventionsgebiet liegt. Frankreich führt einen nationalen Katalog, der durch thematisch spezialisierte Kataloge sowie in den zum Teil im Konventionsgebiet gelegenen Rhônes-Alpes und Provences-Alpes-Côtes d'Azur (PACA) durch Kataloge regionaler Daten ergänzt wird. Der italienische nationale Katalog befindet sich im Aufbau, während vereinzelt Kataloge in Alpenregionen bestehen. In Liechtenstein und Monaco werden Kataloge eingeführt. Österreich führt einen nationalen Katalog, der auch weitgehend Bundesländerdatenquellen beinhaltet. In der Schweiz steht ein föderaler DQK vor dem Abschluß, wobei regionale Daten auch dezentral vorliegen werden. Slowenien hat einen Katalog digitaler umwelrelevanter Daten veröffentlicht.

Die thematische Erschließung durch diese Kataloge ist hinsichtlich des Bedarfes des ABIS unvollständig und je nach Partner unterschiedlich ausgeprägt (Grolimund 1996). Die sich hierdurch ergebende Notwendigkeit ergänzender Erhebungen wird an dieser Stelle allerdings nicht weiter verfolgt. Die französischen und italienischen nationalen Kataloge zielen darauf ab, lediglich solche Datenquellen zu verzeichnen, die gesamtstaatlich vorhandene Daten beinhalten. Auch die Konzepte unterscheiden sich zum Teil grundsätzlich: Während in Frankreich lediglich regelmäßig anfallende Datenquellen aufgenommen werden, ist dies in anderen Ländern, wie zum Beispiel der Schweiz nicht der Fall. Der ADQK verfolgt ebenfalls ein weitgefaßtes Konzept, wobei allerdings die Möglichkeit vorgesehen ist, die konzeptionellen Unterschiede bei Abfragen und Austausch zu berücksichtigen.

Alle Kataloge enthalten einen gewissen Anteil an Daten, die für das ABIS relevant sind. Die Gesamtmenge der Daten variiert offensichtlich mit dem Anteil des Landes am Gebiet der Alpenkonvention. Dort, wo nationale Kataloge lediglich landesweit vorliegende Daten umfassen, ist die Zahl der Kataloge, die erschlossen werden müssen, höher.

Es kann nicht davon ausgegangen werden, daß die Metainformationen in den Quellkatalogen in allen Fällen so beschrieben sind, daß sie den Anforderungen des ABIS genügen. Die Beschlagwortung soll insbesondere erreichen, daß die Themen der Konvention erfaßt sind und konzeptionelle Unterschiede durch geeignete Filter deutlich berücksichtigt werden können. Eine besondere Rolle spielt hierbei auch die geforderte mehrsprachige Erschließbarkeit der Information, welche durch die Verwendung eines einheitlichen gemeinsamen Thesaurus erreicht werden kann, der allerdings in den Ursprungskatalogen nicht zur Anwendung kommt.

Die Europäische Union (EU), Liechtenstein, Monaco und die Schweiz verwenden Produkte des Datenquellenkatalogs (CDS) der Europäischen Umweltagentur (EUA). Italien wird zumindest das CDS Format führen, das von einigen Regionen bereits verwendet wird. Slowenien hat seinen Katalog derzeit auf einer Vorabnorm des Europäischen Normungsausschusses zu geographischen Informationssystemen gegründet, wird allerdings auch den CDS im Rahmen seiner EUA Aktivität einführen. Deutschland auf der nationalen Ebene und Österreich führen das Umweltdatenkatalog (UDK) Format mit einer Schnittstelle zum CDS. Frankreich und Rhône-Alpes haben zur Zeit noch ein Format ohne CDS Schnittstelle, während PACA den CDS verwendet. Bayern hat ein eigenes Produkt ebenfalls ohne CDS Schnittstelle.

Die Ausgangssituation ist somit gekennzeichnet durch eine beträchtliche Vielfalt an Handlungsträgern und Übergangsstellen zwischen unterschiedlichen Katalogen. Offensichtlich bedingt dies hohe Koordinierungskosten mit je nach Land unterschiedlich hohen Anfangsinvestitionen.

3 Operative Grundsätze

Das ABIS besteht aus einem sogenannten Kommunikationszentrum (KomZ) pro Teilnehmerpartei. Diese KomZ haben den Auftrag, die alpenrelevante Information der nationalen Beobachtungssysteme in das ABIS einzubringen sowie die Produkte des ABIS in ihrem Land zu verbreiten. Diese Struktur kann ergänzt werden durch eine Koordinationseinheit, welche bei einem der KomZ angesiedelt ist, aber auch außerhalb des ABIS liegen kann. Die Koordinationseinheit (KE) hat hierbei in erster Linie die Aufgabe diejenigen Arbeitselemente zu erledigen, welche Voraussetzungen für die Arbeiten aller KomZ schaffen. Sie ist allerdings keineswegs eine zentrale Instanz, welche insbesondere die Arbeiten der KomZ in ihrem nationalen Rahmen beaufsichtigt, finanziert oder gar durchführt.

Das ABIS hat seine Aufgaben mit den Mitteln zu erfüllen, welche die Trägerorganisationen der KomZ bereits zur Verfügung haben. Die Ausstattung des Systems erlaubt es mithin nicht selbst gestaltend zu wirken. Es muß sich vielmehr nach den Situationen in den jeweiligen Ländern richten. Diesem Umstand trägt die Vorgabe Rechnung, die Standards, Verfahren und Produkte der EUA zu verwenden. Die Erwartung bestand, daß diese von den Vertragsparteien im Rahmen der EUA Aktivitäten eingeführt

bzw. übernommen werden würden. Insbesondere im Hinblick auf andere internationale Aktivitäten ist geboten, Doppelarbeit zu vermeiden.

Die Alpenkonvention kennt vier Amtssprachen: Deutsch, Französisch, Italienisch und Slowenisch. Das ABIS ist daher gehalten seine Informationsprodukte in diesen Sprachen zugänglich zu machen. Hinzu kommt das Englische, um den Informationsaustausch mit der EUA und seinen Benutzern zu gewährleisten.

4 Darstellung des Datenflusses

4.1 Die Ebene des ABIS

Jedes KomZ muss die Daten aus seinem Zuständigkeitsbereich auswählen, geeignet beschreiben und in dem gemeinsamen Austauschformat bereitstellen. Das KomZ wird zum Metainformationsverwalter für seine nationale Komponente. Diese gibt es an die Koordinationseinheit weiter.

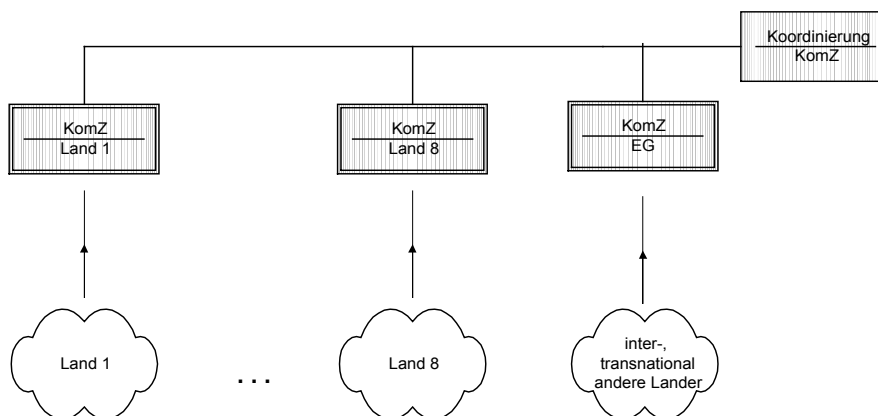


Abbildung 1
Zusammenarbeit bei Erfassung nationaler Kataloge

Die KE faßt die einzelnen nationalen Komponenten zusammen. Diese Zusammenfassung kann zu gewissen Änderungen an den nationalen Komponenten infolge eventuell verbesserter Harmonisierung und Aggregation führen. Die nationalen Komponenten gehen dann an die KomZ zurück. Die KomZ führen die nationalen Komponenten nach, in Abhängigkeit der Aktualisierung der national vorhandenen Kataloge durch die jeweiligen Dienststellen.

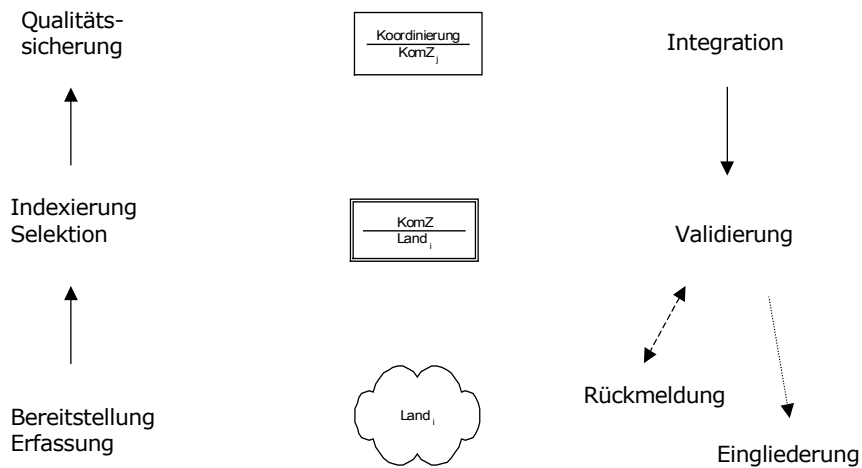


Abbildung 2
Aufgabenverteilung

Die Übereinstimmung der ursprünglichen Metainformationen mit den Datenquellen wird als von dem Metainformationsverwalter des Ursprungskataloges sichergestellt betrachtet. Das KomZ beurteilt lediglich die Relevanz für das ABIS und nimmt eine zusätzliche auf das ABIS abgestimmte Indexierung vor. Der Grad der Harmonisierung in der Beschreibung, Aggregation und thematischen Breite über die Komponenten sind wesentliche Elemente der gemeinsamen Qualitätssicherung. Veränderungen, die im Zuge der Integration vorgenommen werden, müssen vom zuständigen KomZ validiert werden. In einigen Ländern wird hierzu auch der Datenhalter miteinbezogen, der gelegentlich auch vorgeschlagene Änderungen in seinem Katalog eingliedern mochte. Die nationalen Komponenten der integrierten, validierten Datenbasis gehen an die KomZ zurück. Es ist daher erforderlich, daß für jeden Datensatz der jeweilige Status bekannt ist, die Berechtigten kontrolliert und bei Veränderungen ihrer Daten informiert werden. Da das Vorgehen zwischen den Ländern verschieden ist, ist eine flexible Statusverfolgung nötig.

Die nationalen Stellen geben bereits Metainformationen in anderen Rahmen als dem ABIS an eine Reihe internationaler Einrichtungen weiter. Dies trifft insbesondere in den Ländern zu, bei denen die Funktionen des Nationalen Fokuspunktes (NFP) der EUA und KomZ von der selben Einrichtung wahrgenommen werden. Das KomZ der Europäischen Gemeinschaften (EG) bringt diese Daten aus den internationalen Quellen, vor allem der EUA, in das System ein. Die nationalen Stellen haben somit nur noch diejenigen alpenbezogenen Daten einzubringen, welche nicht von der EUA erfaßt werden. Ein Fluß vom ABIS zu EUA ist hingegen nicht vorgesehen, denn die EUA ist von ABIS-spezifischen Kriterien nicht berührt. Eine derartige Aufteilung

soll die nationalen KomZ entlasten. Es sollte allerdings erreicht werden, daß dem KomZ bei der Selektion erkenntlich ist, welche Informationen bereits über die internationale Schiene in das System eingeführt werden.

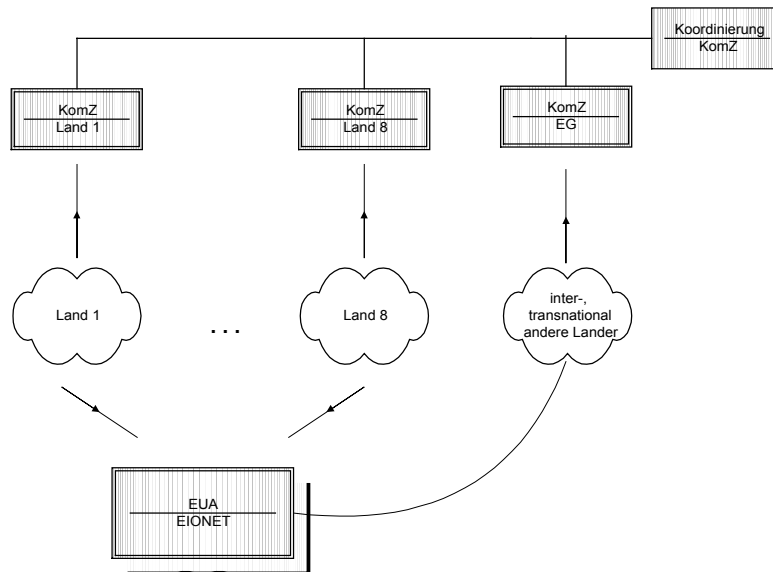


Abbildung 3

Zusammenarbeit bei Erfassung über nationale und europäische Kataloge

Andererseits kann das KomZ der EG eine Teilmenge des ADQK an andere transnationale Einrichtungen im Alpenraum weitergeben. Dies konnte z. B. der Fall der Arbeitsgemeinschaft Alpen-Adria sein. Hier besteht ebenfalls ein Projekt, einen DQK einzuführen und Interesse an Übernahme der EUA Produkte. Dieser wurde auf regionaler Aggregationsebene liegen und durch Datenmengen nicht-alpiner Regionen vervollständigt werden.

4.2 Die Ebene der KomZ

Die Erstellung der nationalen Komponente des ADQK bedarf der Selektion der alpenrelevanten Daten aus dem Bestand eines Kataloges. Die Erschließung mehrerer Kataloge durch das KomZ kann sich hierbei durchaus als notwendig erweisen, damit die erforderliche regionale Granularität und thematische Spezialisierung erreicht werden kann. Wie in Abschnitt 1.1 geschildert variiert die Situation diesbezüglich von Land zu Land. Ein komplexer Fall ist in Abbildung 4 schematisiert.

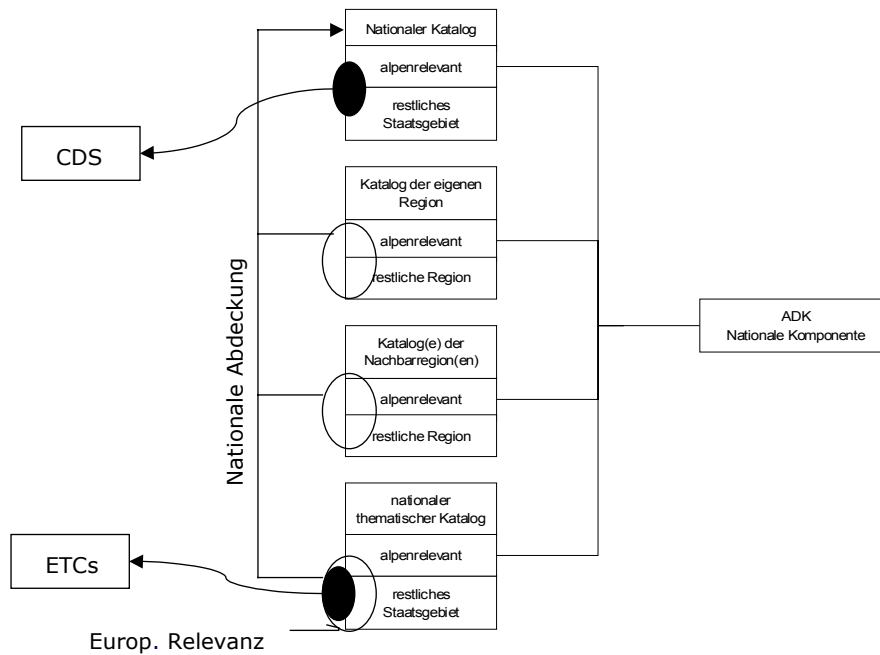


Abbildung 4
Fallbeispiel der Zusammenarbeit auf nationaler Ebene

Ein KomZ, das, z. B., in einer regionalen Einrichtung angesiedelt ist, kann alpenrelevante Informationen in mehreren Katalogen finden, die unterschiedliche räumliche oder thematische Aufgaben erfüllen. Regionale Kataloge (einer Alpenregion) geben Daten an den nationalen Katalog weiter, wenn diese Daten auch von anderen Regionen zur nationalen Abdeckung des Themas gemeldet werden. Aus diesem gelangen entweder über die EUA, wenn sie von europäischer Relevanz sind, oder direkt über das KomZ Daten in das ABIS. Das KomZ sieht sich somit der Situation gegenüber, regionale Daten, nationale und gegebenenfalls international bereits in das System eingebrachte Daten abzugleichen. Die gleichzeitige Überlagerung thematischer Kataloge, die zumindest in einem Land ebenfalls regional verteilt gehalten werden, kompliziert die Situation weiter.

4.3 Interoperabilität

Innerhalb des ABIS wird als Metainformationsformat nur der CDS verwendet, da dies als das gemeinsames Bezugs- und Austauschformat vereinbart wurde. Es sind die KomZ auf nationaler Ebene, welche sich mit der Heterogenität der Ursprungskataloge auseinandersetzen müssen. Die Menge der Metainformationen, die zur Zeit bereitgestellt

wird, zeigt deutlichen Zusammenhang mit der Heterogenität auf: Diejenigen, welche Daten beisteuern, sind KomZ, deren wichtigster Ursprungskatalog das CDS Format übernommen hat. Bei ihnen besteht der Aufwand in erster Linie in der Auswahl relevanter Datensätze sowie der Gewährleistung der ABIS-spezifischen Beschlagwortung. Auch eine landesinterne Koordination über mehrere Katalogbetreiber ist in diesen Fällen nicht erforderlich gewesen.

Als Haupthindernis zur Übertragung in den anderen Ländern hat sich der Mittelbedarf für den Übergang von Fremdformaten in das CDS Format erwiesen. Die Übertragungsmethode hängt von der Aufwandsabschätzung des KomZ ab: bei kleinen Katalogen wäre eine manuelle Wiederaufnahme vertretbar gewesen, während bei größeren die Erstellung eines Formatkonvertierers zum halbautomatischen Transfer notwendig gewesen ist. Es hat sich auch gezeigt, daß in Einrichtungen, die zum Zeitpunkt der Kontaktaufnahme seitens des KomZ keinen rechnergestützten Katalog besaßen, die Möglichkeit die Produkte des ETC/CDS einzuführen, sei es in Eigenregie oder sich auf das KomZ abstützend, gerne aufgegriffen wurde.

4.4 Implementation

Die Implementation des ADQK gründet sich auf die Produkte des Europäischen Thematischen Zentrums zum CDS (ETC/CDS). Die Informationen sollen öffentlich angeboten werden über das World Wide Web, während die Nachführung derzeit noch auf eine Anwendung in Microsoft Umgebung angewiesen ist. Die ursprüngliche Planung sah vor, daß jedes KomZ seine nationale Komponente mittels des WebCDS (Kazakos et al. 1998) anböte. Benutzerabfragen, welche Informationen aus mehreren nationalen Komponenten betrafen, sollten an die Kataloge der KomZ verteilt werden und mit einer einzigen Ergebnisliste erfüllt werden, so daß der Nutzer nicht die gleiche Abfrage an mehrere Komponenten richten müßte. Die Ressourcen, welche die Partner effektiv zur Verfügung hatten, führten dazu, daß über den Durchführungszeitraum ein verteiltes System nicht errichtet werden konnte.

Als Zwischenschritt wird daher eine einzelne gemeinsame Einrichtung des WebCDS an der KE betrieben, damit die vorhandenen nationalen Komponenten wie vorgesehen veröffentlicht werden können. Die Nachführung der Komponenten wird den KomZ dadurch erleichtert werden, daß diese internetgestützt unter Verwendung von Java Technologie erfolgen wird. Soweit diese Möglichkeit direkt von Datenhaltern oder -verlegern ergriffen wird, bedürfen die Datensätze der Kontrolle durch das zuständige KomZ, welche in diesen Fällen über das Internet erfolgt.²

² Das Deutsche Polardatenverzeichnis führt eine ähnliche Organisation durch (Reinke 1997).

5 Der Einfluß der Mehrsprachigkeit

Die Anforderung, die Information mehrsprachig anzubieten dient nicht nur den Endnutzern des Kataloges, sondern kann auch die Bereitschaft von Datenhaltern Metainformationen beizutragen begünstigen. Die Erfahrung zeigt, daß der Rückzug auf eine Sprache – üblicherweise Englisch – praktisch nicht ausreicht.

Die systematische Erstellung von Übersetzungen ist wegen der damit verbundenen Kosten vom ABIS abschlägig beschieden worden. Die aus Nutzersicht primäre Funktionalität, nämlich eine Abfrage in der eigenen Sprache formulieren und dennoch relevante Datenquellen aus anderen Sprachbereichen lokalisieren zu können, wird durch die auf den Allgemeinen Europäischen Mehrsprachigen Umweltthesaurus (GEMET) der EUA gestützte Abfragemodalität realisiert. Sollte ein Nutzer näheres Interesse an einer identifizierten Datenquelle haben, kann er diese vom Datendistributor erhalten bzw. eine Übersetzung bedarfsgerecht anfertigen lassen.

Die Verwendung des GEMET erhöht den Aufwand zur Erstellung des ADQK. Die Beschlagwortung des Ursprungskataloges muß mit dem Vokabular des GEMET abgestimmt werden, um eine ausreichende Abdeckung bei der Stichwortsuche zu erreichen. Dies wäre auch bei einem verteilten heterogenen System der Fall. Der Abgleich müßte hier allerdings von dem Betreiber des Ursprungskataloges vorgenommen werden, der idealerweise den GEMET in seinem System einführen würde.

Eine weitergehende Alternative würde sich auf die Verwendung von Maschinenübersetzung stützen; die Qualität könnte sich als ausreichend erweisen zumal dank der Unterstützung durch GEMET, damit der Benutzer zumindest die Relevanz der fremdsprachlichen Datenquelle abschätzen kann, bevor er sich um nähere Information oder gar die Datenquelle selber bemüht. Die Vorbereitungen zu einem – letztlich nicht zustande gekommenen – Forschungsprojekt in diesem Sinne haben allerdings bereits gezeigt, daß der Gleichbehandlung der Sprachpaare zwischen den Konventionssprachen Schwierigkeiten entgegenstünden.

Die Wirklichkeit einer solchen Anforderung wird illustriert durch die Arbeitsgemeinschaft Alpen-Adria, die jeden Datensatz in den vier Sprachen der Arbeitsgemeinschaft erfaßt – ohne Englisch. Fleuti [Fleuti 1993, 374] weist darauf hin, daß eine derart weitgehende Berücksichtigung der Mehrsprachigkeit auch Konsequenzen insbesondere für die Harmonisierung der Metainformation mit sich bringt.

6 Schlußfolgerung

Die Verwendung des internationalen Standards für Metainformationen CDS innerhalb des ABIS hat sich positiv auf die Bereitstellung der Information ausgewirkt. Die ersten Metainformationen, welche in den ADQK eingelesen werden, haben ihren Ursprung dort, wo die KomZ sich auf CDS basierte Systeme stützen können. Die Übernahme des Formates, auch innerhalb der Länder, und somit die Breite der erfaßten Information ist eindeutig dadurch gefördert worden, daß Softwareanwendungen, die

das CDS Format implementieren, den Datenhaltern zur Erleichterung ihrer eigenen Arbeit zur Verfügung standen.

Die Erwartung der Migration bestehender Katalogsysteme zu einem neuen, sei es auch international verwendeten Metainformationsformat hat sich in den meisten Fällen als unrealistisch erwiesen. Die Erstellung von Schnittstellen zwischen den Formaten zeichnet sich in diesen Fällen als Kompromißlösung ab, deren baldige Verwirklichung in einigen Ländern zur weiteren Füllung des ADQK führen wird.

7 Danksagung

Ich danke den Sprechern und ADQK Referenzpersonen der Kommunikationszentren für die Informationen zu Situation und Erfahrungen auf nationaler Ebene sowie den Gutachtern des Arbeitskreistreffens für ihre förderlichen Anmerkungen.

8 Literaturverzeichnis

- Fleuti, G. (1993): Switzerland's Statistical Databank, in: European Communities – Commission, Statistical Meta Information Systems, Proceedings of the conference, Luxembourg, 2 to 4 February 1993, Luxembourg, Amt für amtliche Veröffentlichungen der Europäischen Gemeinschaften, ISBN 92-826-0478-0.
- Grolimund, P. (1996): Datenquellenkatalog für den Vollzug der Alpenkonvention und deren Protokolle, Eine Zusammenfassung der Ergebnisse eines Pilotprojektes, Interner Bericht an die Arbeitsgruppe Alpenbeobachtung der Alpenkonvention.
- Kazakos, W., Kramer R., Nikolai R., and Rolker C. (1998): WebCDS – A Java-based Catalogue System for European Environment Data, in: Proceedings International Workshop on Issues and Applications of Database Technology, Berlin, Juli 1998, S. 482-490.
- Reinke, M., Marx, B. (1997): Polar Data Directory Systems. In: Lautenschlager, Michael; Reinke, Manfred (Hrsg); Climate and Environmental Database Systems. Dordrecht: Kluwer; S. 189-197, ISBN 0-7923-9832-7.

Die verteilte Suche in GEIN 2000 (G2K)

Thomas Bandholtz¹

Die Ist-Aufnahme bei den G2K-Informationsanbietern zeigte, daß bereits heute ein umfangreiches und qualitativ ernst zu nehmendes öffentliches *Internet-Angebot* für die hier untersuchte Umweltinformation besteht. Dies wird in der Projektlaufzeit, insbesondere bis zur EXPO 2000, in noch kaum vorhersehbarem Maße ausgedehnt und qualitativ restrukturiert werden. Da auch der Anteil der Internet-Nutzer bei den Anwendern entsprechend steigt, ist der direkte Zugang des „interessierten Laien“ zur Umweltinformation dabei, auf einem unvorhergesehenen Weg und in unvorhergesehenem Ausmaß wahr zu werden.

Dieser Zugang zu dieser Information lief bisher an den *Metadatenkatalogen* vorbei. Mit dem UDK 4.0 sind Internet-Verweise bewußt in das Datenmodell (und die Interaktion) aufgenommen worden. Bisher bieten drei Länder den UDK auf ihren öffentlichen Web-Seiten an und etablieren ihn als lokale „Suchmaschine“. (Am konsequentesten scheint dies Karlsruhe (LfU) zu betreiben, und das übrigens mit der Version 3.02. Hier gab es auch eine Anwenderbefragung zur Beurteilung des UDK).

Doch der Strukturbaum des UDK und die „Baumstruktur“ von HTML-Dateien in WebSites sprechen nicht dieselbe Sprache. Der UDK ignoriert die Suchmuster des Internet und wird im selben Maße vom Internet ignoriert. Dies kann jeder leicht überprüfen, indem er mit einer gängigen Suchmaschine z. B. nach Ozon sucht und allein darauf achtet, ob und an welcher Stelle Bund- oder Länderbehörden in der Trefferliste auftauchen. Das Ergebnis ist eher deprimierend. Ein UDK ist in einer solchen Trefferliste noch nie aufgetaucht.

Das hat seinen Grund: Allgemein geht das Internet vom Zugang zu Informationen aus und versucht diese zu strukturieren, während der UDK von einer Struktur ausgeht und versucht, den Zugang zu ihren Elementen herzustellen.

Der UDK hat ja in den frühen 90ern als reiner Katalog begonnen, der sich in seiner Struktur an die Organigramme von Ministerien orientierte. 1993/4 bemühte er sich um ein „Kommunikationsmodul“, noch auf LAN/WAN-Technik basierend, mit dem parametrisierbare Abfragen auf die Fachdatenbanken ermöglicht werden sollten – Visionen, aus denen geradewegs die aktuellen Erwartungen an ein „Z39.50-Profil“ hervorgegangen zu sein scheinen (die allerdings heute nicht vom UDK vorgetragen werden).

Der *WWW-UDK* (seit Version 3.0) beschränkt sich auf eine *Internet-Oberfläche*. Dahinter liegt derselbe UDK. Mit der Version 4 werden erstmals ausdrücklich URLs

¹ email: thomas@sema.de

(Unified Resource Locator, in etwa = „Internetadressen“) als Verweise zu UDK-Objekten zugelassen.

Es gibt allerdings auch heute und in Zukunft Umweltinformation, die gar nicht öffentlich zugänglich gemacht wird, sei es aus Personenschutzgründen, sei es wegen der Priorität des Betriebsgeheimnisses (Einleiterdaten), sei es, weil sie nicht hinreichend validiert vorliegt. Hier ist die Position des „Users“, daß er erfahren möchte, welche Informationen es gibt und wen er danach befragen kann. Es ist schließlich nicht ausgeschlossen, daß er solche Daten im einzelnen einsehen kann, wenn er ein berechtigtes Interesse nachweist.

Jedes Metainformationssystem verfügt hier über ein *Alleinstellungsmerkmal*: Keine Web-Suchmaschine kann Informationen finden, auf die sie nicht zunächst zugreifen kann. Der UDK (im „klassischen Gewand“ der Versionen bis einschließlich 3) kann dies (und *nur* dies), indem er Personen befragt, die von den nicht-allgemein-zugänglichen Informationen wissen. Das kann eine Suchmaschine nicht.

Die Krux des UDK ist seine vollkommene Abhängigkeit von eben diesen Personen (Informationsträgern), wenn es darum geht, Metainformation zu erfassen. Die Internet-Suche arbeitet vergleichsweise provokativ: bevor sie mit einem Informationsanbieter spricht, hat sie bereits die Informationen selbst bis in alle Winkel durchsucht, sie konnte sich den Quellcode einzelner Seiten anschauen und Hyperlinks aufzeichnen, bis sie jede Seite aus allen Richtungen kennt. (Von der eigentlichen Information hat sie allerdings bis dahin noch nichts verstanden). Der Informationsanbieter entwickelt zwangsläufig eine ganz andere Motivation, als wenn man ihn schlicht fragt, über welche Informationen er denn verfügt.

Es entwickelt sich also die Idee einer Vereinigung des *strukturorientierten* UDK mit der *faktenorientierten* Internet-Suchmaschine.

An dieser Stelle müssen wir uns die Frage gefallen lassen, welche Rolle denn eigentlich die *Fachdatenbanken* für G2K spielen, denn sie sind sozusagen der geplante „Dritte im Bunde“ (neben Internet und UDK). Tatsache ist, daß kein einziges Bundesland den direkten Zugriff auf Fachdatenbanken angeboten hat, während alle an einer Integration ihres Internet-Angebots sehr interessiert waren.

Interessiert an direkten Schnittstellen zwischen G2K und Fachdatenbanken zeigte sich allein der Bund, jedenfalls in den Fällen, wo eine eigene Internet-Präsentation (noch) nicht verfügbar ist. Man wünscht sich dort von G2K eine Visualisierung, und dafür nimmt man den Datenbankzugriff eher in Kauf, als das man ihn eigentlich selbst wünscht.

Die Probleme einer Visualisierung, die nicht unmittelbar beim „Datenherren“ (wäre „Datenherrin“ korrekt?) entsteht, sind im Leserkreis dieses Konzepts hinlänglich bekannt. Es bestehen da sicher auch gewisse Eitelkeiten auf beiden Seiten, tatsächlich ist die fachlich korrekte Interpretation der Daten weder den Daten selbst, noch den heute verfügbaren Metadaten allein zu entnehmen. Hier liegt die Schnittstelle zwischen Umweltbeobachtung und Informationsmanagement. Das Informationsmanagement

arbeitet viel formaler und abstrakter, fixiert auf seine Systemmodelle. Die Umweltbeobachtung steht dagegen immer mit einem Bein in der Realität, weil sie sie *beobachtet*. Sie erkennt auch den startenden Lastwagen (alter Diesel) unmittelbar neben dem Meßcontainer, obwohl der im Systemmodell eigentlich gar nicht vorgesehen war.

G2K kann keine universelle Visualisierungsmaschine sein. Die Visualisierung ist immer speziell. Wenn eine Harmonisierung von Länder- und Bundesdaten in einer einzigen thematischen Karte möglich ist, dann ist dies das Ergebnis einer langjährigen fachlichen und administrativen Abstimmung, aber nicht die Knopfdruck-Leistung² eines besonders schlauen Computer-Programms.

Damit geht konform, daß G2K eigentlich weniger daran interessiert ist, viel Aufwand in die Übertragung von Fachdatensätzen zu stecken, mit deren Präsentation es dann allein gelassen wird. Eine Vergleichbarkeit von z. B. Meßnetzdaten zwischen den Bundesländern ist auf der Ebene der Daten selbst nicht gegeben.

G2K sollte die Entstehung von geeigneten Visualisierungen in Fachregie in jeder erdenklichen Hinsicht fördern. Dies erscheint erfolgversprechender als eine technische Allgemeinlösung in G2K selbst.

Dennoch beinhaltet die Kommunikation mit Fachdatensätzen einen besonderen Reiz für die anspruchsvolleren unter den „interessierten Laien“. Und nicht zuletzt ist ja auch jeder Informationsanbieter selbst wieder Suchender und möchte mit den Fachkollegen anderer administrativer Einheiten kommunizieren. Dies möchte G2K unterstützen (und davon lernen).

Der Fokus liegt dennoch auf dem interessierten Laien, und der steht einem „ausgewachsenen“ Fachdatensatz wahrscheinlich eher hilflos gegenüber.

Dennoch kann die direkte Kommunikation mit Fachdatenbanken nicht aus dem weiteren Konzept ausgespart werden: der Experte soll gleichberechtigt unterstützt werden, und: in gewisser Weise treten uns auch die Metainformationssysteme wie Fachdatenbanken gegenüber.

G2Ks verteilte Suche umfaßt also

- WebSites (Visualisierungen) des Bundes und der Länder,
- Metadaten, insbesondere die Umweltdatenkataloge des Bundes und der Länder,
- ausgewählte Fachdatenbanken ohne heute bestehende eigene Internetpräsentation.

Diese drei Kategorien werden hier als Domänen bezeichnet. Sie unterscheiden sich inhaltlich und methodisch hinsichtlich:

- ihrer Verbreitung und Einheitlichkeit,
- der auf sie anzuwendenden Suchverfahren,
- der in ihnen verbreiteten Protokolle,
- des angebotenen Zugriffs auf die angebotene Information,
- der durch sie vertretenen Betrachtungsebene, indem sie zum Teil aufeinander verweisen.

² “Press the Button - We Do the Rest” - damit warb Kodak vor Jahrzehnten für den Kauf einer Kamera .

Der Umweltdatenkatalog UDK in Österreich

5 Jahre Erfahrungen

Legat¹, Batschi², Hashemi-Kepp³, Kruse⁴, Nikolai⁵, Nyhuis⁶, Pultz⁷,
Stallbaumer⁸, Swoboda⁴, Zirm¹

1 Einleitung

Die Suche nach Daten zu umweltbezogenen Themen gestaltet sich für Fachleute, insbesondere aber für die interessierte Öffentlichkeit nicht immer einfach, vor allem, wenn nicht bekannt ist, ob die gesuchten Daten überhaupt erhoben wurden, von wem sie erhoben wurden und wo man sie bekommen kann. Um diesem Informationsbedürfnis gerecht zu werden, wurden in den letzten zehn Jahren in vielen Staaten Umweltdatenkataloge aufgebaut. Diese Metainformationssysteme über umweltrelevante Datenbestände enthalten wichtige Hinweise über die Verwendbarkeit und den Zugriff auf die Daten, wie fachliche Beschreibung, fachlicher Kontext, Raum- und Zeitbezug sowie Angaben zur Zuständigkeit, Verfügbarkeit und Aktualität.

Auch die Gesetzgeber in den Staaten Europas unterstützen aus demokratiepolitischen Gründen die Entwicklung, den Zugang zu Umweltdaten so einfach wie möglich zu gestalten und darüber hinaus eine aktive Umweltinformationstätigkeit der Behörden zu entwickeln.

In Österreich wurde, im Rahmen der Umsetzung der Richtlinie 90/313/EWG des Rates der Europäischen Gemeinschaften vom 7. Juni 1990 über den freien Zugang zu Informationen über die Umwelt, das Umweltinformationsgesetz (UIG 1993) beschlossen. Dieses sieht im Sinne einer modernen und offenen Umweltverwaltung

¹ Umweltbundesamt Wien, Thesaurus-Koordinierungsstelle, Spittelauer Lände 5, A-1090 Wien, udk@ubavie.gv.at, <http://udk.ubavie.gv.at>

² Umweltbundesamt Berlin, Bismarckplatz 1, D-14193 Berlin, wolf-dieter.batschi@uba.de, <http://www.umweltbundesamt.de>

³ MUVIS Umweltsysteme, Liesingtalstr. 74, A-2384 Breitenfurt, helmut.hashemi-kepp@muvis.telecom.at,

⁴ Niedersächsisches Umweltministerium, UDK-Koordinierungsstelle, Archivstr. 2, D-30169 Hannover, udk@mu.niedersachsen.de, <http://www.mu.niedersachsen.de/udk/>

⁵ Forschungszentrum Informatik, Haid-und-Neustr. 10-14, D-76131 Karlsruhe, nikolai@fzi.de, <http://www.fzi.de>

⁶ BVG, Menkestraße 7a, Hillers-Eck, D-26419 Schortens, nyhuisd@bvg-fri.de

⁷ Bundesministerium für Umwelt, Jugend und Familie, Stubenbastei 5, A-1010 Wien, silvia.pultz@bmu.gv.at, <http://www.bmu.gv.at>

⁸ Technisches Büro Hermann Stallbaumer, Favoritenstr. 182/5, A-1100 Wien, hermann@tbhs.co.at

sowie einer erleichterten Bürgerpartizipation in § 10 die Einrichtung eines Umweltdatenkataloges als Zugangssystem zu Umweltdaten vor. Das UIG verleiht dem Einzelnen durch die Verpflichtung der Behörden und Ämter, ihre Umweltdaten transparent zu halten, einen neuen Informationsanspruch im Sinne demokratischer Mitgestaltung.

Im August 1993 schlossen Deutschland und Österreich eine „Vereinbarung über die Zusammenarbeit beim Aufbau, bei der Entwicklung und bei der Pflege eines gemeinsamen Umweltdatenkataloges“ ab. Im Rahmen dieser Kooperation übernahm Österreich die Aufgabe der Koordinierung, Entwicklung und Pflege eines Thesaurus (UDK-Thesaurus) sowie aller dazu erforderlichen Softwarewerkzeuge.

Mit Hilfe des UDK und des UDK-Thesaurus lassen sich mehr Informationen von höherer Qualität auf effektivere Art und Weise beschaffen und verwalten als vorher. Dies führt zu einer spürbaren Verbesserung der Informationsversorgung sowohl der interessierten Bevölkerung als auch der Umweltverwaltungen und damit zu einem effizienteren Umweltschutz [1].

Im Zuge der Validierung der gemeinsam angestrebten Ziele der deutsch-österreichischen Kooperation zum Aufbau des Metainformationssystems „Umweltdatenkatalog“ hat das Niedersächsische Umweltministerium 1995 eine Studie zur Untersuchung der konzeptiven und EDV-technischen Ausprägung des internationalen Entwicklungsfeldes UDK in Auftrag gegeben. Der Auftragnehmer der Studie, Prof. Dr. Oliver Günther, Direktor des Instituts für Wirtschaftsinformatik der Humboldt-Universität zu Berlin, hält in seinem „Gutachten zur Entwicklung des Umweltdatenkataloges (UDK)“ fest:

„Zusammenfassend ist festzustellen, daß es sich bei dem UDK um einen nach Kenntnis des Gutachters weltweit einmaligen Ansatz handelt, den Umweltdatenbestand eines Landes durchgehend und systematisch zu dokumentieren. ... Hier besteht eine besondere Chance, den wissenschaftlich-technischen und organisatorischen Vorsprung, den die Bundesrepublik und Österreich auf dem Gebiet der Umweltinformatik international genießen, praktisch umzusetzen. ... Wie bereits erwähnt, ist Österreich hinsichtlich der praktischen Einführung des UDK am weitesten fortgeschritten. Österreich ist bisher das einzige Land, in dem die Einführung des UDK gesetzlich festgeschrieben ist.“

Seit dieser Studie (1995) wurde in vielen Staaten und fachübergreifend die Notwendigkeit des Aufbaus von Metainformationssystemen erkannt. Einige dieser Projekte sind zwischenzeitlich weit fortgeschritten. Exemplarisch seien einige davon angeführt:

Im europäischen Raum ist der „Catalogue of Data Sources“ (CDS) der Europäischen Umweltagentur EEA der bedeutendste Ansatz, Metainformationen staatenübergreifend zu sammeln und zu verwalten. (<http://www.mu.niedersachsen.de/cds>)

Das Projekt „The Australia New Zealand Land Information Council“ (ANZLIC) stellt geographische Informationen auf Regierungsebene zur Verfügung (<http://www.anzlic.org.au>).

Das "Global Change Master Directory" GCMD, (<http://gcmd.gsfc.nasa.gov>) stellt Metadaten zur Unterstützung der Forschung in den naturwissenschaftlichen Fächern bereit.

2 Win-UDK 4.0

Basierend auf den Praxiserfahrungen der UDK-Versionen 1, 2 und 3 wurde 1997/98 die Konzeption des UDK grundlegend neu überdacht und an die Bedürfnisse der Nutzer angepaßt.

Die bisher verfügbare Software (UDK 3.0) wurde größtenteils zur zentralen Erfassung der Metadaten verwendet. Dies war vor allem in Österreich der Fall, wo auf diese Weise ca. 12.000 UDK-Objekte in den vergangenen Jahren erhoben wurden. Der UDK 3.0 war aufgrund mangelnder Benutzerfreundlichkeit nicht hinreichend dafür geeignet, die geplante Dezentralisierung der Pflege und Erweiterung des UDK-Datenbestandes sicherzustellen. Daher hatte die Erhöhung der Benutzerfreundlichkeit oberste Priorität bei der geplanten Neukonzeption der Software.

Ebenso wurde großer Wert auf die breite technische Einsetzbarkeit gelegt. Wichtigste Kriterien waren hierbei Lauffähigkeit, Zuverlässigkeit (Stabilität), einfache Installationsroutinen und akzeptable Performancewerte.

In der Konzeptphase wurde gründlich abgewogen, ob die Eingabe der Metadaten WWW-basiert erfolgen sollte. Dies böte folgende Vorteile:

- Die Metadaten würden direkt in die Masterdatenbank eingetragen und wären sofort im WWW verfügbar. Aufwendige Datenbankabgleiche würden entfallen.
- Die Verteilung der Software und die Versionenkontrolle würde entfallen. Ein Update der Software würde nur auf dem WWW-Server erfolgen müssen.
- Die Bedienung der Software dürfte für WWW-geübte Nutzer unproblematisch sein.

In der Konzeptphase (1997) überwogen jedoch die Nachteile, die mit dem WWW damals verbunden waren. Diese bestehen auch heute noch zum Teil, so dass die wichtigsten in Kürze genannt seien:

- Der Zugang in der öffentlichen Verwaltung zum WWW ist noch sehr heterogen. Eine Beschränkung der Eingabemöglichkeiten auf das WWW würde viele Benutzer daher ausschließen.
- Dort, wo der Zugang vorhanden ist, schützen oft Firewalls die behördeninternen Netze vor fremden Zugriffen. Diese sind häufig so konfiguriert, dass nur HTML-basierte Informationen durchgelassen werden. ActiveX-Controls, Java-Applets, ja selbst JavaScript, werden somit nicht akzeptiert. Dies reduziert die Möglichkeiten zur Gestaltung einer benutzerfreundlichen Oberfläche erheblich.
- Die Performance von WWW-Anwendungen ist in den meisten Fällen noch nicht zufriedenstellend. Insbesondere während der Kernarbeitszeiten der Bediensteten in den öffentlichen Verwaltungen ist das Internet oft sehr überlastet.

Interessant könnte ein WWW-Eingabewerkzeug daher höchstens für den Einsatz im Intranet sein, wenngleich auch hierbei zu konstatieren ist, dass die Performancewerte meistens immer noch erheblich niedriger als bei konventionell programmierten Windows-Programmen sind.

- Viele Behörden bevorzugen eine offline-Eingabe, um hausinterne Qualitätssicherungen der eingegebenen Daten vor ihrer Veröffentlichung im Internet durchführen zu können.
- Die Nutzungsdauer von WWW-Anwendungen ist aufgrund der sich ständig wandelnden technischen Rahmenbedingungen sehr kurz. Die im Vergleich zu konventionellen Programmen erforderlichen Versionswechsel erhöhen die Kosten.

Die Abwägung der angegebenen Vor- und Nachteile hat dazu geführt, daß eine MS-Windows-basierte Software zur Eingabe der Metadaten in Auftrag gegeben wurde (Version UDK 4.0). Diese sollte auch komfortable Recherchemöglichkeiten bieten. Die Software UDK 4.0 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). Es werden alle ODBC-fähigen Datenbanken unterstützt.

Der UDK 4.0 ist seit Februar 1999 verfügbar. Unter der WWW-Adresse <http://www.mu.niedersachsen.de/udk/> kann er zur Ansicht heruntergeladen werden. Weitere Informationen zur Konzeption des UDK sind in [2] enthalten.

Die wesentlichen Neuerungen des UDK 4.0 sind:

- Überarbeitung des Datenmodells:
Die Erfassung der Metadaten erfolgt in sogenannten UDK-Objekten, welche hierarchisch zueinander in Beziehung gesetzt werden können („Parent/Child-Struktur“). In den bisherigen Versionen des UDK wurde die Hierarchisierung über eine dekadische Notation realisiert, die vom Nutzer vergeben werden mußte. Dies führte zu erheblichen Akzeptanzproblemen bei den Nutzern. Im UDK 4.0 erhält jedes UDK-Objekt beim Neuanlegen eine eindeutige Objektidentifikation (OID), welche vom System automatisch generiert wird. Die Strukturierung wird über Verweise zwischen den OID's geregelt. Zwei Verbesserungen konnten hierdurch erzielt werden: der Nutzer wird nicht mehr mit der „unhandlichen“ dekadischen Notation konfrontiert und erhebliche Performancesteigerungen konnten erzielt werden.

Die Anzahl und Art der Beschreibungsfelder ist abhängig von der zu beschreibenden Umweltinformation. Zur Beschreibung von Karten benötigt man beispielsweise andere Beschreibungsfelder als zur Beschreibung von Dokumenten oder Datensammlungen. Eine Reihe von Feldern können hingegen gleich sein (z. B. Ansprechpartner). Um dieser Problematik gerecht zu werden, wurden bereits in der Version UDK 3.0 die Objekte in Klassen eingeteilt. Diese Objektklassen

wurden im UDK 4.0 entsprechend der Nutzerwünsche neu eingeteilt und inhaltlich überarbeitet. Es existieren jetzt folgende Objektklassen:

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

Die bestehenden und für alle Kooperationspartner einheitlichen Objektklassen können um maximal drei Objektklassen erweitert werden.

Die Feldbezeichnungen werden über einen Schlüssel gesteuert. Hierdurch wird die Konfigurierbarkeit der Bezeichnungen ermöglicht (Beispiel: Im Raumbezug entsprechen die österreichischen „Bezirke“ den deutschen „Kreisen“). Es können beliebig viele Felder hinzugefügt werden.

- Jedem UDK-Objekt können mehrere UDK-Adressen zugeordnet werden. Die Adressen werden in einer eigenen Adreßverwaltung eingegeben und gepflegt. Aufgrund österreichischer Anforderungen ermöglicht das Datenmodell des UDK 4.0 die Abbildung der Organisationshierarchie, wobei bei der Eingabe einer neuen Adresse die übergeordneten Adressdaten „vererbt“ werden (aber änderbar sind). Beispiel: Bei Eingabe der Adreßdaten einer Person müssen die Adreßdaten der entsprechenden Institution bzw. Abteilung nicht noch einmal eingegeben werden. Durch den Ausbau der Adreßverwaltung wird die Datenstruktur des früheren österreichischen UDK-Moduls „Amtsverzeichnis“ integriert.
- Erleichterung der Eingabefunktionalitäten: Die Beschreibungsfelder wurden in übersichtlichen Karteikarten angeordnet. Zu jedem Beschreibungsfeld wird ein Hilfetext angezeigt, der dem Nutzer Hinweise zum Ausfüllen gibt. Eine Eingabemaske wird beispielhaft in Abbildung 1 gezeigt.

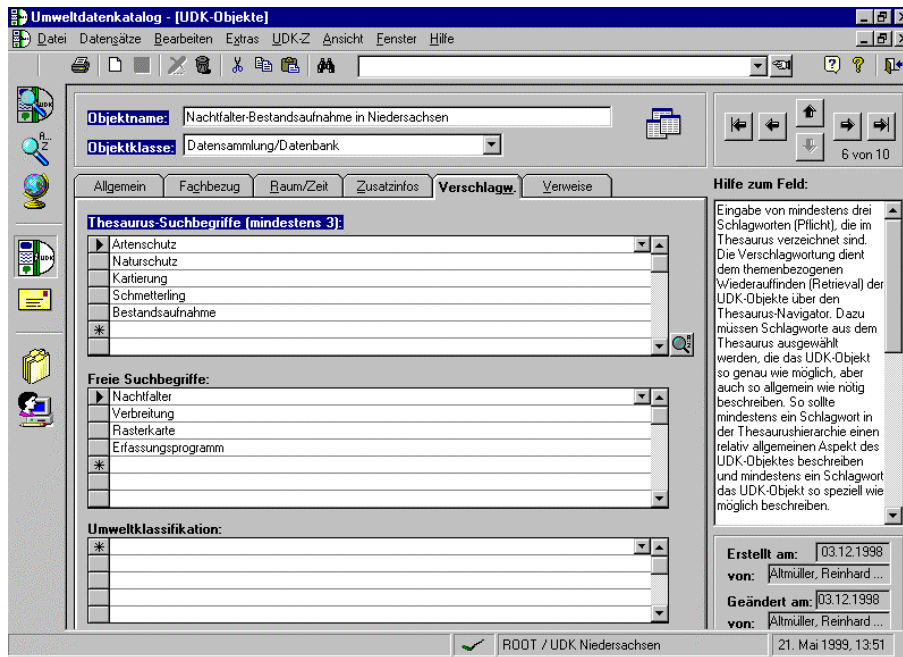


Abbildung 1
Eingabemaske für ein UDK-Objekt

- Recherchefunktionalitäten:

Der UDK 4.0 verfügt über komfortable Recherchefunktionalitäten (siehe auch[2]):

Der „UDK-Explorer“, der dem Date Explorer von Windows 95 bzw. NT nachempfunden wurde, ermöglicht die Navigation in der Baumstruktur der UDK-Objekte. Eine komfortable Suchmaschine wurde integriert. Über eine „Experten-suche“ sind auch freie SQL-Abfragen möglich.

Erstmals integraler Bestandteil des UDK ist der Thesaurus der Umweltbundesämter Berlin und Wien. Mit Hilfe des „Thesaurus-Navigators“ kann der Nutzer durch ein strukturiertes Wortgut navigieren und sich dynamisch die UDK-Objekte anzeigen lassen, die mit dem aktuell eingestellten Begriff verschlagwortet sind.

Eine geographische Suche ermöglicht mit Hilfe von zoombaren Karten das Auffinden der UDK-Objekte über ihren Raumbezug.
- Automatischer Aktualisierungszyklus

Der UDK ist für die dezentrale Eingabe und Pflege der Metadaten konzipiert. Technisch gesehen handelt es sich um eine verteilte Datenbank oder genauer: eine Datenbank wird an verschiedenen Orten mehrfach installiert und geändert.

Durch geeignete Replikationsmechanismen – beim UDK spricht man von einem Aktualisierungszyklus – muss in gewissen Abständen der Datenbestand abgeglichen werden, so daß alle Datenbanken wieder den gleichen Stand haben. Jede beteiligte Datenbank wird als UDK-Instanz bezeichnet. Der Abgleich zwischen den UDK-Instanzen wird dadurch ermöglicht, daß beim Neuanlegen, Ändern oder Löschen eines Datensatzes ein spezifischer Flag gesetzt wird. Es werden dann diejenigen Datensätze ausgetauscht, die einen Flag besitzen. Nachdem ein Austausch abgeschlossen ist, werden die Flags wieder zurückgesetzt. Der Aktualisierungszyklus wird über den Austausch von ASCII-Dateien realisiert (pro Datenbanktabelle eine Datei), deren Inhalt über entsprechende Funktionen ein- bzw. ausgelesen werden. Der UDK 4.0 verfügt über ein optional zu installierendes Modul, das automatisch die Änderungsdateien erzeugt und per Email an eine zentrale Stelle (Zentralkatalogadministrator) schickt. Dort werden die Dateien von allen beteiligten UDK-Instanzen automatisch eingelesen und wieder an alle verteilt. Dieser Prozeß läuft vollkommen im Hintergrund ab. Er kann vom Benutzer auf Wunsch auch manuell angestoßen werden, falls z. B. eine Qualitätskontrolle vor der Datenweitergabe erfolgen soll. In der näheren Zukunft wird der automatische Aktualisierungszyklus in der Praxis getestet. Es bleibt abzuwarten, ob damit eine Alternative zur Eingabe über das WWW – ohne die damit verbundenen Nachteile (siehe oben) – geschaffen wurde.

- Zur technischen Unterstützung des Zentralkatalogadministrators dient das optionale Modul UDK-Z. Es ermöglicht die zentrale Pflege des Datenbestandes eines Kataloges.

Der Problematik der Altdatenübernahme von der Version 3.0 in die 4.0 ist angesichts des großen Datenbestandes im österreichischen Katalog von 12.000 UDK-Objekten und 1.100 UDK-Adressen sowie des geänderten Datenmodells speziell im Klassenkonzept große Aufmerksamkeit entgegenzubringen. Diese Aufgabe soll im zweiten Quartal 1999 abgeschlossen werden.

3 Umweltinformationssystem WWW-UDK

Das Umweltinformationsgesetz (§ 10 UIG 1993) in Österreich verpflichtet das Umweltressort, den öffentlichen Zugang zum Umweltdatenkatalog UDK zu gewährleisten. Dem Stand der Technik entsprechend veröffentlicht das Bundesministerium für Umwelt, Jugend und Familie bzw. Umweltbundesamt Wien die UDK-Daten bereits seit März 1996 im Internet mittels der Web-Applikation WWW-UDK [3]. Der Einstieg erfolgt über <http://udk.bmu.gv.at> bzw. <http://udk.ubavie.gv.at>.

Allgemeine Information	
Objekt-Name:	Abwasser-Kontroll-Datenbank
UDK-Klasse:	Datensammlung/Gutachten/Berichte
Dekadische Notation:	02.01.13.03.04.03.02
Freie Suchbegriffe:	UK: Wasser
Thesaurusbegriffe:	Datenbank, Abwasser (benutze Unterbegriffe oder Klasse WA), Abwasserbehandlung, Abwasserbehandlungsanlage, Kläranlage, Abwasserreinigung, Anlagenbetrieb, Anlagenüberwachung, Industrieanlage, Eigenüberwachung, Fremdüberwachung, Emissionsüberwachung, Indirekteinleiterüberwachung, Biochemischer Sauerstoffbedarf, Chemischer Sauerstoffbedarf, Phosphat, Chlorid, Sulfat, PH-Wert, Leitfähigkeit, Alkalität, Nitrat, DOC, Kontrolle (benutze Unterbegriffe), Probenahme, Stichprobe, Qualitative Analyse, Ammonium
Datenhaltende Stelle:	Amt der Kärntner Landesregierung, Unterabteilung 15 W: Gewässerökologie, Polzer
Datenauskunft:	Amt der Kärntner Landesregierung, Unterabteilung 15 W: Gewässerökologie, Polzer
Beschreibung:	Probenahmen im Ablauf durch die Gewässeraufsicht in Kläranlagen (direkt) und in Betrieben (indirekt); Maßnahmen bei Grenzwertüberschreitungen; früher Stichproben, Industrieanlagen und Kommunalanlagen haben meist Mischprobenentnahme; Parameter bei kommunalen Kläranlagen: pH, Leitfähigkeit, Alkalität, BSB, CSB, NH4, NO3, Gesamtphosphor (Orthophosphat), Chlorid, Sulfat, DOC; Eigenüberwachung und Fremdüberwachung der Kläranlagen.
Letzte Änderung:	20.10.1998
Zusatzinformation	
Datenart:	digital
Freigabe:	j
gesetzliche Grundlage:	Wasserrechtsgesetz 1959
Datenzugänglichkeit:	verwaltungintern zugänglich
Fachbezug	
Titel/Thema:	Abwasser
Basisdaten:	Gewässeraufsicht
Raumbezug	
Land:	Kärnten
Bezirk:	alle Bezirke

Abbildung 2
UDK-Objekt „Abwasser-Kontroll-Datenbank“ im WWW-UDK mit Verweis auf das im Rechtsinformationssystem des Bundeskanzleramtes im Volltext verfügbaren Wasserrechtsgesetz.

3.1 Integrationscharakter des österreichischen WWW-UDKs

Ein Katalog bzw. Metainformationssystem hat einen inhärenten integrativen Charakter, da verteilte Informationsbestände so beschrieben werden, daß sie auffindbar und zugänglich gemacht werden. Der österreichische Ansatz geht allerdings über die derzeitigen deutschen Ansätze im Sinne der Integration verschiedener Datenbestände in verschiedenen Punkten deutlich hinaus:

- In der Einführungsphase des UDK in Österreich wurden sämtliche UDK-Daten zentral erfaßt. Das hat den Vorteil, daß in *allen* österreichischen UDK-Daten *aller Bundesländer und des Bundes gleichzeitig* gesucht werden kann und nicht – wie bisher in Deutschland – ausschließlich in denen eines Bundeslandes bzw. des Bundes. Falls erwünscht, können Suchfilter aber die Suche auch auf einzelne Bundesländer bzw. Behördenbereiche einschränken.
- Die Präsentation des UDK im Internet ermöglicht technologisch erstmals auch die Bereitstellung von Informationen, die über die Datenbankinhalte der PC-Version weit hinausgehen. Der UDK in Österreich ist ein „Portal“ zu den Umweltinformationen des Bundes und der Länder. Der Begriff „Portal“ stammt dabei aus dem Bereich der Web-Suchmaschinen und bezeichnet den Eingangsbereich zu verschiedenen, auch aggregierten, Inhalten und Diensten rund um ein Thema. So ist es möglich, die Vernetzung der Umweltinformationen mit anderen Informationen auszudrücken. Als Beispiel seien die Verweise von den UDK-Objekten über deren gesetzliche Grundlage direkt auf die entsprechenden Gesetzestexte, die von einem Web-Server des österreichischen Bundeskanzleramtes bereitgestellt werden. Desweiteren gibt es umfangreiche aktuelle Informationen (Publikationen zum download, eine Linksammlung zu umweltrelevanten Angeboten im Internet und Hintergrundinformationen zum Projektumfeld) rund um den UDK. URL-Verweise von den Metadaten auf die eigentlichen Daten, so daß auf diese einfach und schnell zugegriffen werden kann, sind für die Zukunft vorgesehen. Ausführliche Aufgabenbeschreibungen von Behördenmitarbeitern ermöglichen das Auffinden von kompetenten Ansprechpartnern, die über Email weitergehende Auskünfte erteilen.
- Umweltdaten sind nicht nur von nationalem sondern auch von grenzüberschreitendem und internationalem Interesse (vgl. Abschnitt 5.). Österreich hat daher seit Einführung des WWW-UDK auch eine englischsprachige Benutzungsoberfläche bereitgestellt.

Der UDK in Österreich wird durch diesen stark integrativen Charakter zu einer Kernkomponente eines Umweltinformationssystems aufgewertet und gewinnt dadurch eine völlig neue Dimension und Qualität. Der WWW-UDK soll in Zukunft über seine Funktion als nationales Metainformationssystem hinaus Teil eines weltweit vernetzten „Environmental Information Locator Service“ sein, gleichermaßen für Behörden sowie für die Öffentlichkeit zugänglich.

Die Entwicklung des Angebotes ist noch keinesfalls abgeschlossen, sondern ist vielmehr als Beginn zu sehen. Die Integration eines elektronischen Bestellwesens für Publikationen des UBA Wien sowie des BMUJF sei hier beispielhaft für nächste Projektschritte genannt.

3.2 Aktuelle statistische Daten zum WWW-UDK

Derzeit enthält die österreichische UDK-Datenbank 12.000 Datensätze (UDK-Objekte) und 1.100 Adreßdatensätze (UDK-Adressen). Im Auswertungszeitraum 22. September 1996 bis 6. Juni 1999 (987 Tage) wurden nachstehende statistische Daten betreffend die Zugriffe auf den WWW-UDK ermittelt (Zahlen in Klammern beziehen sich auf die letzten 7 Tage):

Erfolgreich bearbeitete Anfragen:	681 094 (5 825)
Durchschnittlich bearbeitete Anfragen pro Tag:	690 (832)
Anzahl unterschiedlicher anfragender Hosts:	16 540 (442)
Anzahl neu anfragender Hosts (letzte 7 Tage):	262
Menge verschickter Daten:	2 911 Mbytes (42 958 kbytes)
Durchschnittliche Tagesmenge verschickter Daten:	3 020 kbytes (6 137 kbytes)

Der Domain-Report verzeichnet Zugriffe aus etwa 60 verschiedenen Ländern. Davon stammen 20 % der Zugriffe aus österreichischen Universitäten, 27 % von österreichischen Behörden und 40 % der Zugriffe von Rechnern im Ausland.

3.3 Technische Aspekte

Die derzeit verwendete Applikation WWW-UDK 3.1 stellt bereits die zweite Generation dar, aufbauend auf einer Beauftragung an das FZI Karlsruhe durch das Land Baden-Württemberg im Jahre 1994, weiters durch Österreich im Jahre 1995/96. Diese Version basiert auf CGI-Skripten die mittels SQL auf die UDK-Datenbank zugreifen. Als Client ist ein Standard-Web-Browser wie Netscape oder Internet Explorer ausreichend, da eine dynamische HTML-Oberfläche erzeugt wird. Eine Zusatzkomponente für die komfortable Benutzung des Thesaurus, GenThes [5], ist als Java-Applet realisiert (Java 1.1-fähiger Browser erforderlich).

WWW-UDK 4.0 [4], welcher auf dem geänderten Datenmodell UDK 4.0 aufbaut, wird die derzeitige Version voraussichtlich im dritten Quartal 1999 ablösen. U. a. aufgrund der gewonnenen Erfahrungen mit WWW-UDK 3.1 und GenThes – für die schnelle Recherche wird eine HTML-Oberfläche benötigt, die minimale technische Voraussetzungen beim Client erfordert – wird es mit HTML-UDK 4.0 eine Basisanwendung, die eine vollständige Retrieval-Engine ähnlich der Vorgängerversion anbietet, geben. Verschiedene graphische Werkzeuge, u. a. GenThes in einer „leichteren“ Version, werden in Form eines integrierten Java-Applets bereitgestellt.

Mit WWW-UDK 4.0 wird durch die Implementierung als Java-Servlet erstmals auch die Plattformunabhängigkeit des Servers erreicht. Dies ist die Konsequenz einer zunehmenden Verbreitung von Windows NT als Web-Server. WWW-UDK 4.0 kann dann sowohl auf NT-basierten als auch auf Unix-basierten Web-Servern betrieben werden [6].

Ein weiterer wesentlicher Vorteil des WWW-UDK 4.0 ist dessen Möglichkeit der verteilten Suche (virtueller UDK; sucht über *einen* Einstiegs-Server in *verteilten* UDK-Datenbanken). Dies ermöglicht in Österreich die zukünftige Verteilung der UDK-Datenbestände, ohne daß auf die übergreifende Suche (vgl. Abschnitt 3.1) verzichtet werden muß.

4 Die Befüllung des österreichischen UDK

4.1 Einsatz- & Erfassungskonzept

Wesentliche Grundlagen für die erfolgreiche Einführung des UDK in Österreich waren die Erstellung eines Einsatz- sowie eines Erfassungskonzeptes, die die organisatorischen, logistischen, DV-technischen und inhaltlichen Grundlagen und Rahmenbedingungen festlegten, sowie deren Umsetzung durch ein zentrales Expertenteam, das in Folge alle relevanten Behörden Österreichs aufsuchte.

Als Besonderheit mußte dabei die Tatsache berücksichtigt werden, daß der UDK bei den Organen der Verwaltung dezentral geführt werden wird, wobei die Daten im Zuge von Aktualisierungszyklen ausgetauscht werden. Die Recherchen sollen jedoch für den gesamten Datenbestand ein bundesweit einheitliches, homogenes Detaillierungsniveau der Ergebnisse erbringen.

Voraussetzung dafür war somit eine homogene Ersterfassung als Grundlage aller weiteren Aktualisierungszyklen [7], um zu gewährleisten, daß die beabsichtigten Datensammlungen tatsächlich bundesweit harmonisiert und vergleichbar erhoben werden, ohne die Bediensteten der betroffenen Stellen allzusehr zu belasten.

4.2 Strukturbaum-Prototyp

Am Beginn der Erstellung des Einsatzkonzeptes stand die Durchführung einer Ist-Analyse, bei der all jene Organe der Verwaltung bestimmt wurden, die über Umweltdaten im Sinne des UIG verfügen, sowie jene Auskunftsstellen, bei denen umweltrelevante Daten nachgefragt werden können.

Als Grundlage zur Ermittlung der betroffenen Organe der Verwaltung auf Bundes- und Landesebene wurden der Amtskalender 1993/94 und die im „Rundschreiben zur Durchführung des Umweltinformationsgesetzes (UIG)“ des Bundesministeriums für Umwelt, Jugend und Familie angeführten Gesetze herangezogen.

Die im Amtskalender 1993/94 beschriebenen umweltrelevanten Zuständigkeiten und Fachaufgaben der so ermittelten Behörden wurden entsprechend ihrer organisatorischen Zugehörigkeit in den Prototyp des UDK-Strukturbaums eingeordnet, der in späterer Folge durch Interviews in den jeweiligen Präsidien und EDV-Abteilungen validiert und ergänzt werden sollte.

In den Bundesministerien, dem Magistrat der Stadt Wien und den Ämtern der Landesregierungen waren rund 80 Sektionen, 150 Gruppen, 1300 Abteilungen, mehrere hundert Referate sowie zu- oder nachgeordnete Dienststellen, Gesellschaften und Kommissionen zu untersuchen. Betroffene Organisationseinheiten und deren entsprechende Zuständigkeiten und Fachaufgaben wurden in den Strukturbaum-Prototyp aufgenommen, der im Zuge der Vor-Ort-Interviews anhand der jeweils gültigen Geschäftseinteilungen validiert und durch die dort erhobenen Datenbestände verbreitert und vertieft werden sollte.

4.3 Strukturbaum – 1. Phase (Einsatzkonzept-Erstellung)

Dieses Mengengerüst wurde im Rahmen der Erstellung der Einsatz- und Erfassungskonzepte im Zuge von Vor-Ort-Interviews bei den betroffenen Verwaltungsbehörden anhand der zum jeweiligen Zeitpunkt gültigen Geschäftseinteilungen validiert und durch zusätzlich erhobene Zuständigkeiten, Fachaufgaben und insbesondere zugehörige Datenbestände verbreitert resp. vertieft.

Die für die UDK-Adreßverwaltung erforderlichen Daten wurden ebenfalls im Zuge von Vor-Ort-Interviews als „Stammdaten“ der Interviewpartner erfaßt.

Außerdem wurden die EDV-mäßigen Rahmenbedingungen zur UDK-Einführung und die von den betroffenen Verwaltungsbehörden (resp. dem UDK-Administrator) geplante UDK-Realisierung (Client-Server, Stand-Alone, Userverwaltung etc.) erhoben. Nach Durchführung zweier Pilotprojekte zur Erstatenerfassung (Steiermark, BMUJF) war die Struktur der Datenlandschaft in Österreich soweit zumindest prototypisch erfaßt, daß die detaillierte Befüllung und Einführung des Umweltdatenkataloges in Angriff genommen werden konnte.

Der im Rahmen der UDK-Einsatzkonzepterstellung („1.Phase“) erstellte Strukturbaum enthielt insgesamt 4445 [8], nach Durchführung der Pilotprojekte insgesamt 7106 UDK-Objekte [9].

4.4 Basisbefüllung

Als Basis für die Vor-Ort-Erfassung bei den betroffenen Verwaltungsorganen wurde für die bereits im Zuge des Einsatzkonzeptes erhobenen Teilstrukturbaume der Bundesministerien und Ämter der Landesregierungen eine harmonisierte Beschreibung und Beschlagwortung der im Zuge der Einsatzkonzepterstellung bereits erhobenen Datenbestände durchgeführt („Basisbefüllung“).

Zur Schaffung der Grundlagen für die Harmonisierung der vorhandenen Datenbestände erfolgte eine tabellarische Zusammenstellung aller „Nicht-Basisobjekte“ (d. h. aller Datenbestände) mit zugehöriger Institution, Abteilung, Dekadischer Notation, Objektname, UDK-Klasse und Gesetzlicher Grundlage, die sich an vergleichbaren, im Zuge des Pilotprojektes bei den verschiedenen Abteilungen des Amtes der Landesregierung Steiermark erhobenen, Datenbeständen orientierte.

Auf Basis dieser Tabelle wurde die Basisbefüllung der bereits erhobenen Datenbestände mit der Eingabe der Beschreibung sowie der Befüllung der UDK-Masken „*Fach-/Raum-/Zeitbezug*“ und „*Objekt-Information*“, die Eingabe der Suchbegriffe und die Verbreiterung und Vertiefung des Strukturbaums (auf Basis der zuvor durchgeführten Erhebungen sowie des vorliegenden Sekundärmaterials) abgeschlossen.

Ebenso wurde die Ergänzung des Umweltamtsverzeichnisses um die entsprechenden Adreßinformationen jener Organisationseinheiten, die zwar im Zuge der Einsatzkonzept-Erstellung und der österreichweiten Erstdatenerfassung keine Datenbestände genannt haben, aufgrund ihrer Aufgaben jedoch mit sogenannten Basisobjekten im UDK aufscheinen, abgeschlossen.

4.5 Vor-Ort-Erfassung

Die im vorigen Abschnitt beschriebene tabellarische Zusammenstellung bereits erhobener Datenbestände wurde nach entsprechender Umsortierung nicht nur als Basis zur Harmonisierung, sondern auch als Auflistung potentieller Datenbestände bei den Vor-Ort-Interviews herangezogen. Als weitere wichtige Unterlage dienten die Ergebnisse der Basisbefüllung, die bei den Interviews kontrolliert, ergänzt und vertieft wurden.

Die Ergebnisse der stattgefundenen Vor-Ort-Interviews sowie zusätzliches Sekundärmaterial wurden ebenfalls vom zentralen Expertenteam sukzessive in den UDK eingegeben und mit den bereits vorhandenen UDK-Objekten harmonisiert.

Der im Rahmen der Ist-Analyse erstellte Strukturbaum-Prototyp wurde im Zuge der Vor-Ort-Interviews anhand der zum jeweiligen Zeitpunkt gültigen Geschäftseinteilungen validiert und durch zusätzlich erhobene Zuständigkeiten, Fachaufgaben und zugehörige Datenbestände, sowie zur Verfügung stehendes Sekundärmaterial verbreitert resp. vertieft. Ermittelte Zuständigkeiten und Fachaufgaben, denen keine konkreten Datenbestände zugeordnet werden konnten, wurden im Strukturbaum belassen.

Der im Rahmen der UDK-Einsatzkonzepterstellung vom zentralen Expertenteam erstellte Strukturbaum („1. Phase“) enthielt insgesamt 7106 UDK-Objekte, während nach Abschluß der Vor-Ort-Erhebungen der UDK **11980 UDK-Objekte** sowie **1121 Adreßkennzeichen** aufweist.

4.6 UDK-Beschlagwortung

Richtlinien zur Beschlagwortung („*Indexierung*“) des österreichischen UDK wurden frühzeitig im UDK-Erfassungskonzept [10] festgelegt und entsprechend umgesetzt, wobei folgenden Punkten besondere Bedeutung beigemessen wurde:

- Die Indexierung hat in erster Linie im Hinblick auf die recherchierenden UDK-Nutzer zu erfolgen, um sie in die Lage zu versetzen, das in Frage kommende UDK-Objekt mit Hilfe der gewählten Deskriptoren finden zu können.
- Die Beschlagwortung hat bereits bei der Dateneingabe die unterschiedlichsten Sichtweisen potentieller UDK-Nutzer auf potentielle Datenbestände vorwegzunehmen.
- Die Beschlagwortung muß so detailliert sein, daß sie auch für Experten von Nutzen ist, aber gleichzeitig so allgemein, daß sie auch von Laien des entsprechenden Fachgebietes verwendet werden kann.

Insgesamt wurden im österreichischen UDK **2261** verschiedene Thesaurus-Suchbegriffe (von „*Abbau (Bergbau)*“ bis „*Zwischenlagerung*“) und **416** verschiedene Freie Suchbegriffe (von „*Abbaufeld*“ bis „*Zuwachsminderung*“) verwendet.

4.7 Umweltklassifikation des UDK

Um der interessierten Öffentlichkeit einen intuitiven Zugang zu den UDK-Beständen mit möglichst wenig „Einstiegspunkten“ zu erleichtern, wurde rasch die Bedeutung einer geeigneten „Umweltklassifikation“ erkannt. In Ermangelung eines eigenen Eingabefeldes „UDK-Klassifikation“ in der UDK-Version 3.0 wurden zutreffende Begriffe der Umweltklassifikation des UBA-Berlin in die erste(n) Zeile(n) der „Freien Suchbegriffe“ eingegeben, wobei jeweils das Kürzel „UK“: vorangestellt wurde, um dadurch eine bessere Unterscheidung von den übrigen „Freien Suchbegriffen“ sowie später eine automatische Zuordnung bei der Altdatenübernahme in die UDK-Version 4.0 zu ermöglichen.

5 Umweltthesaurus

5.1 Der UDK-Thesaurus

Wesentlichen Einfluß auf ein zufriedenstellendes Rechercheergebnis im Umweltdatenkatalog hat eine einheitliche Beschreibung und Beschlagwortung der Datenbestände, die bereits bei der Dateneingabe die unterschiedlichsten Sichtweisen potentieller UDK-Nutzer auf Datenbestände vorwegnehmen muß. Sie muß daher einerseits so detailliert sein, daß sie auch für Experten von Nutzen ist, aber gleichzeitig so allgemein, daß sie auch von Laien des betreffenden Fachgebietes verwendet werden kann.

Allgemein formuliert ist ein Thesaurus ein hierarchisch strukturierter, begrenzter Wortschatz, welcher der natürlichen Sprache entnommen ist und der ein Hilfsmittel für das vereinheitlichte Beschreiben („*Indexieren*“) und Auffinden von Informationen eines bestimmten Fachgebietes mittels normierter Begriffe („*Deskriptoren*“) darstellt.

Der UDK-Thesaurus ist daher ein wesentlicher Bestandteil des österreichischen Umweltdatenkataloges. Er soll es ermöglichen, die Inhalte eines Fachgebietes mit wenigen, einheitlich benutzten Bezeichnungen wiederzugeben oder zu umschreiben und so ein adäquates Instrument für die Datenverwaltung, -pflege und -recherche darstellen.

Das UBA Berlin und das UBA Wien haben im Rahmen der deutsch-österreichischen Kooperationsvereinbarung zum Aufbau des Umweltdatenkataloges UDK gemeinsam einen erweiterten Umweltthesaurus (UDK-Thesaurus) entwickelt. Er baut auf dem Umweltthesaurus des Berliner Umweltbundesamtes auf und umfaßt rund 8.500 Deskriptoren sowie rund 16.000 Synonyme.

Basis der Thesaurusarbeit am UBA Berlin war und ist der anwendungsgerechte Aufbau und die entsprechende Weiterentwicklung des o. a. Thesaurus für seine bibliographischen Datenbanken (Umweltliteraturdatenbank ULIDAT, Umweltforschungsdatenbank UFORDAT und die Umweltrechtsdatenbanken URDB). Diese Datenbanken können (in einem weiteren Sinn) auch als Metainformationssysteme verstanden werden. Alle Datenbanken enthalten Informationen zu den Aspekten „wer hat welche Daten mit welchem Inhalt und ggf. geographischen Bezug zu welchen Zugriffsbedingungen und eventuellem Zeitbezug“, wobei diese Informationen nicht immer gleichzeitig in jedem Datensatz vorhanden sein müssen. Deshalb war es nur konsequent, den bereits bestehenden und jahrelang erprobten Umweltthesaurus auch für das Projekt UDK, als reinem Metainformationssystem anzuwenden. Dabei wurde Wert darauf gelegt, die Nutzer (Datenhalter und -abfrager) in den Entwicklungsprozeß des Thesaurus einzubeziehen. Zu diesem Zweck wurde von den beteiligten Kooperationspartnern eine Koordinierungsstelle Thesaurus am UBA Wien eingerichtet, die sich u. a. um die zielgerichtete Anpassung des Thesaurus an die Bedürfnisse der Partner kümmert. Als Instrument der Thesaurusarbeit wurde das Arbeitsgremium Wortgutredaktion (WGR) eingerichtet. Diese ist im WWW über die URL: http://www.cedar.univie.ac.at/wgr_home/ erreichbar, die technische Administration erfolgt durch die Internationale Gesellschaft für Umweltschutz, IGU.

Der Umweltthesaurus gibt mit seinem Datenbestand, den Schlagwörtern, einen umfassenden Überblick über die Schwerpunkte der aktuellen wissenschaftlichen Diskussion im deutschsprachigen Raum, die sich in entsprechenden Fachpublikationen bzw. aktuellen Forschungs- und Entwicklungsvorhaben sowie der gültigen Rechtsetzung niederschlägt. Dabei zeigt sich, daß durch den am aktuellen Inputmaterial orientierten Ausbau des Thesaurus manche Bereiche des Umweltschutzes stärker im Wortgut vorhanden sind, andere dagegen unterrepräsentiert erscheinen. Da sie aber direkt an den Input gekoppelt sind, spiegeln sie die Schwerpunktsetzung der Umweltpolitik und ihrer Förderziele sowie die „in der Branche“ für notwendig erachtete Forschungsausrichtung wieder.

Mit Beginn der Arbeiten am Thesaurus im Umfeld des UDK zeigte sich, daß die beteiligten Institutionen bzw. Personen in den Verwaltungen höchst unterschiedliche Vorstellungen zum Thema inhaltliche Erschließung und Wiederauffindbarkeit von gespeicherten Informationen hatten. Die Vorzüge eines normierten Wortgutes für Indexierung und Retrieval sowie die Unterstützung der Arbeit durch ein polyhierarchisches Gebilde wie den UDK-Thesaurus mußten erst einmal vermittelt werden, da die Nutzer in der Regel aus Bereichen stammen, die bisher mit den genannten Arbeitsmitteln nicht so vertraut waren bzw. sind wie die herkömmlichen Anwender in Bibliotheken und Dokumentationsstellen.

Es hat sich gezeigt, daß das vorhandene Wortgut im UDK-Thesaurus für die inhaltliche Erschließung der UDK-Objekte in der Regel gut ausreicht. Bei der in Österreich von einer zentralen Stelle vorgenommenen Erfassung von UDK-Objekten war festzustellen, daß rund 60 Vorschläge für neue Deskriptoren bzw. 80 für zusätzliche Nondesriptoren (Synonyme) für den UDK-Thesaurus erfolgten. Diese werden von der Wortgutredaktion geprüft und dann entsprechend in den Thesaurus eingearbeitet, indem die notwendige Verknüpfung zu den Ober- und Unterbegriffen sowie verwandten Deskriptoren hergestellt wird. Dies erfolgt unter Nutzung der Software-Werkzeuge, die für die Pflege des UDK-Thesaurus entwickelt wurden. Aus der Arbeit an den Umweltdatenkatalogen der Deutschen Bundesländer ist bisher noch wenig Rücklauf zum Thesaurus erfolgt, da erst mit Version 4.0 des UDK eine leistungsstarke Unterstützung der Thesaurusnutzung vorliegt und die Anwender jetzt leichter in der Lage sind, alle Funktionen des UDK-Thesaurus zu nutzen. Um Unsicherheiten bei der Anwendung des Thesaurus zu beheben, wird in Kürze, insbesondere für die mit dem Thesaurusumgang noch nicht so vertrauten Nutzer / Anwender, ein Workshop abgehalten.

Die Anwendung des UDK-Thesaurus in Österreich hat eine Reihe von Austria-zismen zu Tage gefördert, die in den Gesamtkontext des Thesaurus eingearbeitet werden müssen, um eine komfortable Nutzung sicherzustellen.

Der UDK-Thesaurus wird nunmehr in drei Publikationsformen angeboten:

- **CD-ROM "THESshow"**
- **Online-Thesaurus auf der Homepage des UDK:** <http://udk.ubavie.gv.at>
- Druckausgabe in drei Bänden

Bestellmöglichkeit für die Thesaurusprodukte besteht über die Homepage der Wortgutredaktion: http://www.cedar.univie.ac.at/wgr_home/

Es ist vorgesehen, eine erweiterte, aktuelle Fassung des UDK-Thesaurus als CD-ROM unter Verwendung von THESshow zum Herbst 1999 herauszugeben, die dann auch den im Rahmen des ETC/CDS entwickelten Multilingualen Umweltthesaurus der Europäischen Umweltagentur GEMET enthalten soll, sowie ein multilinguales Abfallglossar mit rund 500 Begriffen, das in den UDK-Thesaurus integriert wird.

5.2 Software für die Erstellung und Pflege von Thesauri

Im Rahmen der bestehenden Kooperationsvereinbarung vom 22. August 1993 über die Zusammenarbeit beim Aufbau, bei der Entwicklung und bei der Pflege eines gemeinsamen Umweltdatenkataloges mit der Bundesrepublik Deutschland hat Österreich die Aufgabe übernommen, eine Koordinierungsstelle Thesaurusentwicklung einzurichten. Diese hat als wesentliche Aufgaben die Bereitstellung und Pflege eines Thesaurus für den UDK (UDK-Thesaurus) sowie die Entwicklung geeigneter Softwarewerkzeuge zur Pflege, Übersetzung, Aktualisierung und Visualisierung multilingualer Thesauri sicherzustellen. Die Entwicklung der Programme erfolgte in Wien durch die Firma TBHS. Um internationale Synergieeffekte zu erzielen wurde die Entwicklung von THES*main* im Zusammenwirken mit dem "European Topic Centre for Catalogue of Data Sources" (ETC/CDS) der Europäischen Umweltagentur EEA vorgenommen. Das ETC/CDS verwendet THES*main* zur Verwaltung und Pflege des zwölfsprachigen Europäischen Umweltthesaurus GEMET (G*eneral* M*ultilingual* E*nvironment* T*hesaurus*). Geplant ist weiters, künftig die Verwaltung des ENVOC-Thesaurus der UNEP mittels THES*main* vorzunehmen [11].

Die Produkte "THES*show*" und "THES*main*" können auch für kundenspezifische Anforderungen adaptiert und zur Verfügung gestellt werden. Lizenzierungsmodelle für den Bereich Forschung und Lehre sowie für die kommerzielle Nutzung können unter der e-mail-Adresse udk@ubavie.gv.at angefordert werden.

Das Softwarepaket besteht, neben einer Reihe von Utilities, aus den Programmen "THES*main*" und „THES*show*“. Die Erstellung eines Thesaurus wird üblicherweise von einem kleinen Team zentral durchgeführt. Der fertige Thesaurus wird von einer großen Anzahl von Benutzern verwendet. Dies führt zu unterschiedlichen Anforderungen an die Software für die Erstellung und für die Benutzung eines Thesaurus, wobei noch festzuhalten ist, daß die meisten kommerziell erhältlichen Thesaurusverwaltungsprogramme den Nutzern den Thesaurus ohnedies nur auf Papier zur Verfügung stellen.

Erstellungs- und Wartungsprogramm THES*main*

- Definition von Thesauri
- Vielseitige Editiermöglichkeiten mit Sicherstellung der Datenintegrität
- Flexible Gestaltung von Zugriffsrechten
- Export- und Importfunktionen
- Kompatibilität zu gängigen Programmen (MS-Access, MS-Excel)

Visualisierungsprogramm THES*show*

- Leicht verständliche Darstellung des Thesaurusinhalts
- Einfache Bedienung
- Gute Online Hilfe
- Einfache und robuste Installation
- Unempfindlichkeit gegen Eigenheiten von Rechnern
- Keine Möglichkeiten zum Editieren des Datenbestands

5.2.1 THESmain: eine Anwendung zur Thesaurusverwaltung

Wesentliche Funktionen

- THESmain dient der Erstellung und Wartung von Thesauri gemäß DIN 1462/1, DIN 1462/2 sowie ISO5964.
- Das Programm wurde unter Visual Basic 4.0 entwickelt und ist unter WfW 3.11, Windows 95/98 und Windows NT lauffähig. Die Daten werden in einem zu MS-Access kompatiblen Datenbankkern abgelegt.
- Das Programm besitzt eine grafische Benutzeroberfläche mit hierarchisch gegliederter Funktionalität. Zu jedem Fenster steht kontext sensitive Hilfe zur Verfügung. Die aktuellen Fenstereinstellungen können auf Wunsch zur Wiederverwendung gespeichert werden. Alle Programmfunktionen können sowohl mit der Maus als auch über Tastaturkürzeln aufgerufen werden.
- Die maximale Anzahl von Begriffen und Relationen ist nur durch den verfügbaren Massenspeicher begrenzt.
- Mehrere (theoretisch unbegrenzt viele) Thesauri können gleichzeitig bearbeitet werden.
- Jeder Thesaurus kann bis zu 30 Sprachen beinhalten.
- Das Programm ist passwortgeschützt und verfügt über mehrere Klassen von Zugriffsrechten. Ein separates Hilfsprogramm zum Erzeugen und Verwalten von Benutzerberechtigungen steht für den Systemadministrator zur Verfügung.
- Es ist möglich Verbindungen zu externen Datenbeständen herzustellen. Mit diesem Verfahren ist es leicht möglich, Mikrothesauri einzubinden.

Descriptors C:\THESMAIN\Gemet-07.mdb Cnfg																				
Window Languages Descriptors Extras Reports ?																				
Exit			Navigation		Pg		Grafic		Pg		Termmanager		Details		Report					
3197:5064			End First		Select		-X-		Splits		Sort Order:		English							
1 English			2 US-English			3 Italian			4 German			5 Dutch			6 Spanish			7 French		
3189		organic waste	3189		Organischer Abfall	3189		rifiuto or												
3190		organisation	3190		Gesellschaft	3190		società (1												
3191		organisation of teachi	3191		Organisation des Unterrichts	3191		organizza												
3192	Toptern	organisation of the leg	3192	Toptern	Gerichtsverfassung	3192	Toptern	organizza												
3193		organisms (not system	3193		Organismen (nicht systemati	3193		organism												
3194		organisms (taxonomy)	3194		Organismen	3194		organism												
3195		organochlorine compo	3195		Organische Chlorverbindung	3195		composti												
3196		organohalogen compo	3196		Organische Halogenverbindu	3196		composti												
3197		organoleptic property	3197			3197		proprietà												
3198		organometallic compo	3198		Metallorganische Verbindun	3198		composti												
3199		organonitrogen compc	3199		Stickstoffverbindung (organi	3199		composti												
3200		organooxygen compo	3200		Organische Sauerstoffverbin	3200		composti												
3201		organophosphorous c	3201		Organische Phosphorverbin	3201		composti												

Abbildung 3
Tabellarische Darstellung von Deskriptoren in drei Sprachen
mit englischer Sortierung

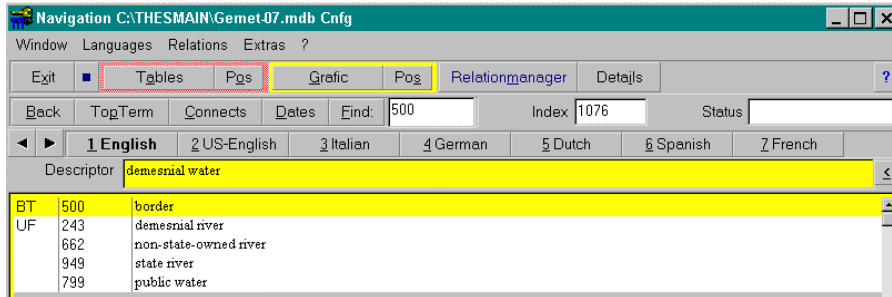


Abbildung 4
Darstellung eines Terms im Navigationsfenster.

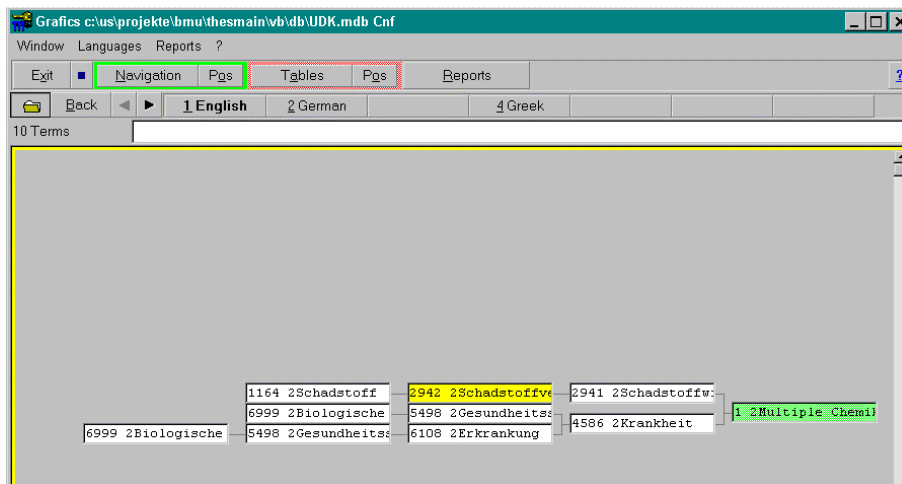


Abbildung 5
Display von Hierarchieebenen in graphischer Darstellung

Sprachen

- Bis zu 30 Sprachen können definiert werden
- Jede Sprache kann ihren eigenen Font und Zeichensatz verwenden
- Die Reihenfolge der Sprachen in den Tabellen und im Navigationsfenster kann vom Benutzer eingestellt werden.

Eine wesentliche Eigenschaft der Spracheinstellung ist die Möglichkeit Sprachen mit unterschiedlichen Zeichensätzen gleichzeitig darzustellen. Dazu muß aber auch das Betriebssystem des Rechners vorbereitet werden.

Export / Import

Alle Daten eines Thesaurus können exportiert und importiert werden. Als Format steht ein SGML Austauschformat zur Verfügung, das auch in anderen Anwendungen wie etwa dem Umweltdatenkatalog sowie den Programmen der europäischen Umweltagentur zur Anwendung kommt.

Benutzerverwaltung

Die Thesaurusdatenbank ist immer mit Nutzernamen und Passwörtern geschützt. Mittels eines mitgelieferten Tools können neue Nutzer mit entsprechenden Zugriffsrechten erzeugt werden.

5.2.2 THESshow: eine Anwendung zur Thesaurusvisualisierung

THESshow ist das Visualisierungswerkzeug für THESmain basierende Thesauri. Es wird derzeit für den Thesaurus des Umweltdatenkatalogs, UDK-T 4.0 sowie für den Thesaurus der europäischen Umweltagentur, GEMET 2.0 verwendet. Es gestattet dem Nutzer in einfacher Weise den Datenbestand zu durchsuchen. Zum Einstieg in die Daten eignet sich entweder die systematische Darstellung, wo von wenigen Top-terms aus die Menge der Deskriptoren durch Durchwandern der Hierarchien erschlossen werden kann, oder die alphabetische Darstellung, wo durch Eingabe eines Wortes ein Einstiegspunkt gefunden werden kann. Es ist dabei möglich, per Knopfdruck von einer Darstellung in die andere zu wechseln, wobei auf den gleichen Term positioniert wird.

Wesentliche Funktionen

- Verfügbar auf CD-ROM und als installierbare Datei via Internet
- Einfache, automatische Installation
- Der Datenbestand kann wahlweise von CD-ROM gelesen werden. In diesem Fall ist der Platzbedarf nur wenige Megabyte
- Vielsprachig.
- Einfache und flexible Konfiguration durch den Benutzer
- Möglichkeit der Erstellung von Teildatenbeständen.
- Möglichkeit zur Darstellung polyhierarchischer Daten.
- Darstellung des Thesaurus in systematischer, alphabetischer und thematischer Form
- Sehr schnelle Suchfunktion in alphabetischer Darstellung
- Automatisches Auffinden eines Begriffs in anderen Darstellungen
- Möglichkeit des Einbindens von Mikrothesauri
- Möglichkeit der Verwendung mehrerer Thesauri
- Darstellung der Details eines Terms inklusive aller Übersetzungen und deren Synonyme unabhängig von Sprache und Zeichensatz (griechisch!, kyrillisch!)
- Datenbestände in allen europäischen Sprachen können visualisiert werden sofern das Betriebssystem diese Sprachen unterstützt.
- Auf Knopfdruck können durch Relationen verknüpfte Begriffe lokalisiert werden.

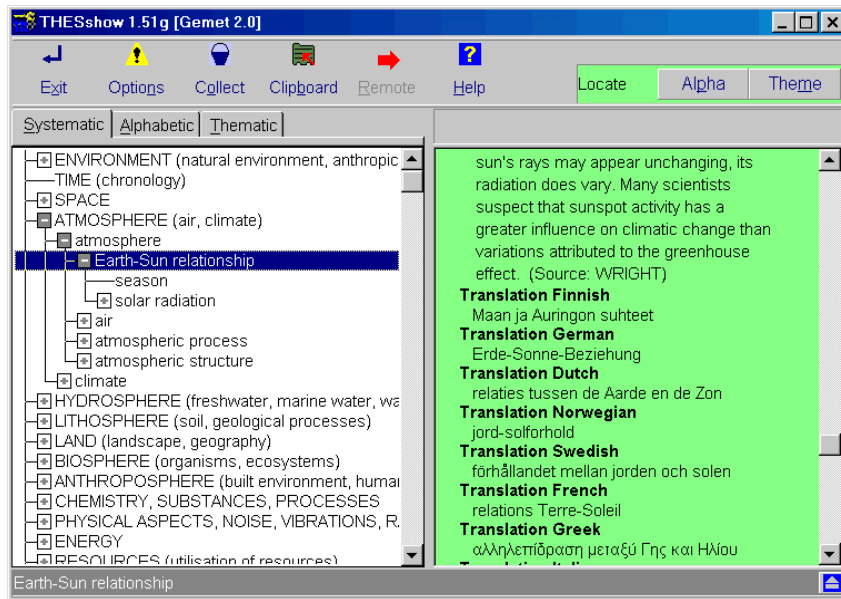


Abbildung 6
zeigt eine typische Ansicht in systematischer Darstellung.
Beachten Sie bitte auch die gleichzeitige Darstellung verschiedener
Zeichensätze im grünen Detailfenster.

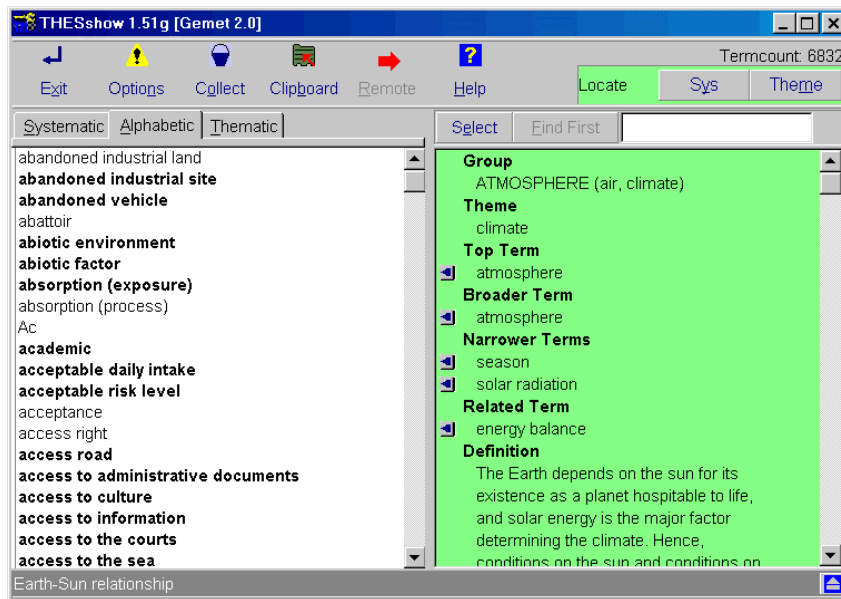


Abbildung 7
zeigt eine typische Ansicht in alphabetischer Darstellung.
Deskriptoren sind dabei fett dargestellt. Die Schaltflächen mit den
Pfeilen dienen zur Lokalisierung von Ober- und Unterbegriffen.

6 Internationale Zusammenarbeit mit dem ETC/CDS

Seit dem Beginn seiner Aktivitäten im Herbst 1994 hat die Europäische Umweltagentur (EEA) eine Reihe von "European Topic Centres" (ETCs) eingerichtet, um die organisatorische Basis für Schlüsselprojekte zu den klassischen Umweltmedien zu schaffen.

Um aber den Zugang zu diesen wichtigen Informationen auf der technischen und organisatorischen Ebene sicherzustellen und einen europäischen Überblick über die national bestehenden Systeme zu erhalten, ist zusätzlich die Einrichtung eines Meta-informationssystems erforderlich. Die EEA entschied daher, einen "Catalogue of Data Sources" (CDS) einzurichten. Als Organisationsform zur Umsetzung dieser Ziele wurde neuerlich das Topic Centre gewählt. Dieses "European Topic Centre on Catalogue of Data Sources & Thesaurus" (ETC/CDS&T) wurde 1995 gegründet. Aufgrund der technisch und organisatorisch weit fortgeschrittenen deutsch-österreichischen Entwicklung des UDK entschied die EEA, die fachlichen und organisatorischen Grundlagen des ETC/CDS auf diese Entwicklung zu legen.

Der CDS (<http://www.mu.niedersachsen.de/cds>) wird ausgewählte Umweltinformationen der Mitgliedstaaten enthalten, die auf der Ebene der EU von Relevanz sind. Er bietet somit Transparenz darüber an, wer in Europa welche Informationen in welcher Form besitzt und wie man sie erlangen kann.

Der CDS wird nicht die national bestehenden Systeme ersetzen, sondern funktionelle Links zu den bestehenden Metasystemen aufweisen. Das vorrangige Ziel besteht darin, eine nahtlose Suche in den verteilten europäischen Katalogen zu ermöglichen und Teil eines "global information locator system" zu sein. Der CDS wird bezüglich Datenmodell und Software auch Standards vorgeben für jene Länder, welche noch keine nationalen Metasysteme aufgebaut haben bzw. gerade dabei sind, solche zu entwickeln.

Die Voraussetzung dafür, mit dem CDS die Sprachgrenzen im EU-Raum zu überwinden, ist der Aufbau und Einsatz eines multilingualen Umweltthesaurus. Der Thesaurus, eine poly-hierarchisch strukturierte Sammlung von Begriffen, soll die Beschlagwortung der Umweltdatenbestände europaweit vereinheitlichen und einen multilingualen Zugang zu diesen ermöglichen. Das vorrangige Ziel ist es, einen Thesaurus zu schaffen, der alle Sprachen der EU-Mitgliedsstaaten beinhaltet und der von einer umfassenden Terminologie-Datenbank unterstützt wird.

Die Vergabe eines "European Topic Centre on Catalogue of Data Sources & Thesaurus" der Europäischen Umweltagentur EEA auf der Basis des UDK an ein Konsortium unter wesentlicher Beteiligung deutsch-österreichischer Institutionen zeigt in besonderer Weise den Wert und die Zukunftssicherheit der gemeinsam zwischen Deutschland und Österreich geleisteten Arbeit.

Die Bemühungen Österreichs zielen auf eine Konvergenz der Entwicklung des WWW-UDK mit dem europäischen CDS im Bereich Datenmodellierung und Web-Anwendungen ab. Aus administrativen sowie aus Kostengründen sind weitgehende Synergieeffekte der beiden Anwendungen anzustreben. Ein CDS-Workshop im November 1996 und anschließende Sitzungen verschiedener UDK-Gremien haben die Grundlagen

zu einem gemeinsamen Kern-Datenmodell (Core Data) gelegt. Der gemeinsame Auftragnehmer in diesem Bereich ist eine gute Voraussetzung dafür, Fehlentwicklungen nicht entstehen zu lassen.

Neben der Verbreiterung der internationalen Zusammenarbeit und der sich daraus ergebenden Synergieeffekte bestätigt diese Entscheidung der Europäischen Umweltagentur die Richtigkeit des von Österreich eingeschlagenen Weges bei der Entwicklung und Einführung des Umweltdatenkataloges.

7 Ausblick

7.1 Aarhus-Konvention

Anlässlich der Vierten Ministerkonferenz "Environment for Europe" vom 23. bis 25. Juni 1998 in Aarhus/Dänemark wurde das UN-ECE Übereinkommen über den Zugang zu Informationen, die Öffentlichkeitsbeteiligung an Entscheidungsverfahren und den Zugang zu Gerichten in Umweltangelegenheiten von 33 europäischen Staaten unterzeichnet. Dieses internationale Instrument wird der Öffentlichkeit mehr Möglichkeiten geben, aktiv die Umweltsituation mitzugestalten. Ein ausreichender Zugang zu Informationen ist die Voraussetzung für effektive Beteiligung der Öffentlichkeit an Umweltentscheidungsverfahren. Abgesichert wird dies durch umfassende Rechtsschutzmöglichkeiten. Die Aarhus-Konvention wird sowohl auf die Rechtslage der EU als auch auf die österreichische Situation Auswirkungen haben. Seitens der Europäischen Union ist bis Ende dieses Jahres/Anfang nächsten Jahres mit einem Kommissionsvorschlag für eine neue Umweltinformationsrichtlinie zu rechnen, der die Errungenschaften der Aarhus-Konvention sowie die bisherigen Erfahrungen mit der Richtlinie 90/313/EWG über den freien Zugang zu Informationen über die Umwelt miteinbeziehen wird. Die Aarhus-Konvention bedeutet einen ersten Schritt zu mehr Umweltdemokratie in Europa und wurde auch von den europäischen NGOs sehr positiv beurteilt. Bisher haben bereits 39 Staaten und die EU selbst die Aarhus-Konvention unterzeichnet.

7.2 Corporate Network Austria

Der Ausbau der Telekommunikations-Infrastruktur innerhalb der öffentlichen Verwaltung ist ein Anliegen der Bundesregierung. Das Bundesministerium für Finanzen hat die Entscheidung gefällt, den "Austrian Information Highway" aufzubauen. Diese Projekt wird unter dem Arbeitstitel "Corporate Network Austria" realisiert werden und dem stark steigenden Bedarf nach intensivem Datenaustausch zwischen den Ministerien, den Ländern, Gemeinden und mit internationalen Organisationen sowie der EU Rechnung tragen.

Dieses behördeninterne Hochleistungsnetz soll künftig auch die Basis für den flächen-deckenden Einsatz des Umweltdatenkataloges bilden.

Literaturverzeichnis

- [1] Günther, O. (1995): „Gutachten zur Entwicklung des Umwelt-Datenkataloges (UDK)“, Humboldt-Universität zu Berlin, Jänner 1995
- [2] W. Swoboda, F. Kruse, D. Nyhuis, H. Rousselle, Die Neukonzeption des Umweltdatenkataloges, Tagungsband des 12. Internationalen Symposiums „Informatik für den Umweltschutz, Bremen (1998).
URL: <http://www.mu.niedersachsen.de/udk/publikationen/ui98/ui98.html>
- [3] Kramer, R., Nikolai, R., Koschel, A., Rolker, C., Lockemann, P., Keitel, A., Legat, R., Zirm, K., WWW-UDK: A Web-based Environmental Metainformation System, in: ACM SIGMOD Record, Vol. 26, No 1, S. 16-25, New York 1997.
- [4] Nikolai, R., Kazakos, W., Kramer R., Behrens S., Swoboda W., Kruse, F.: WWW-UDK 4.0: Die neue Generation eines Web-Portals zu deutschen und österreichischen Umweltdaten, erscheint in: Umweltinformatik 99. 13. Internationales Symposium „Informatik für den Umweltschutz“, Magdeburg 1999.
- [5] Nikolai, R., Kramer R., Steinhaus M., Felluga B., Plini, P., Gentes: A general thesaurus browser for web-based catalogue systems. In Proc.IEEE Meta-Data'99, Bethesda, Maryland, USA, April 1999.
URL: <http://computer.org/conferen/proceed/meta/1999/papers/49/rnikolai.html>.
- [6] Kazakos, W., Kramer, R., Nikolai, R., and Rolker, C. WebCDS – A Java-based Catalogue System for European Environment Data. In Proceedings des International Workshop Issues and Applications of Database Technology (IADT'98), S. 482 – 490,
URL: <http://www.fzi.de/dbs/publications/kazakos/IADT98.ps>, 1998.
- [7] Bundesministerium für Umwelt: Schriftenreihe zum Umweltdatenkatalog, Band 3, Erfassungskonzept – Bewertung von Erfassungsvarianten. Wien 1996
- [8] Bundesministerium für Umwelt: Umweltdatenkatalog in Österreich, Einsatzkonzept – Strukturbaum. Wien 1995
- [9] Bundesministerium für Umwelt: Schriftenreihe zum Umweltdatenkatalog, Band 9, Einsatzkonzept – Strukturbaum. Wien 1996
- [10] Bundesministerium für Umwelt: Schriftenreihe zum Umweltdatenkatalog, Band 4, Erfassungskonzept – Beschreibung der Erfassung durch ein Expertenteam. Wien 1996
- [11] Legat, R., Stallbaumer, H.: Erstellung, Wartung und Visualisierung multilingualer Thesauri mittels „THESmain“ und „Thesshow“, in: Management von Umweltinformationen in vernetzten Umgebungen, S. 116-123, Nürnberg 1999.
URL: <http://bildungsforschung.bfz.de/hubuis/>

Weiterführende Literatur

- Batschi, W-D. (1994): “Environmental Thesaurus and Classification of the Umweltbundesamt (Federal Environmental Agency), Berlin“, Berlin 1994
- Batschi, W-D. (1995): “Development and State-of-the-Art of the German Environmental Thesaurus (UBA-Thesaurus) and User Experience in Germany“, Berlin 1995

- Bericht des Bundesministers für Umwelt über die Erfahrungen mit der Vollziehung des Umweltinformationsgesetzes (UIG). Wien 1995
- Bundesgesetz über den Zugang zu Informationen über die Umwelt (Umweltinformationsgesetz – UIG 1993), BGBl 495/93, Wien 1993
- Bundesministerium für Umwelt, Jugend und Familie (1993): „Das Recht auf Umweltinformation“, Informationsbroschüre, Wien 1993
- Bundesministerium für Umwelt: Schriftenreihe zum Umweltdatenkatalog, Band 1, Grundlagen und Methodik des Umweltdatenkataloges (UDK). Wien 1995
- Bundesministerium für Umwelt: Schriftenreihe zum Umweltdatenkatalog, Band 5, Einsatzkonzept – Adressen-Instanzen-Aktualisierung. Wien 1996
- Bundesministerium für Umwelt: Schriftenreihe zum Umweltdatenkatalog, Band 6, Einsatzkonzept – Stufenplan. Wien 1996
- Bundesministerium für Umwelt: Schriftenreihe zum Umweltdatenkatalog, Band 7, Einsatzkonzept – Aufwandsabschätzung. Wien 1996
- Bundesministerium für Umwelt: Schriftenreihe zum Umweltdatenkatalog, Band 8, First International Workshop CDS and Thesaurus – Strukturbaum. Wien 1996
- Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (1998): „UDK Version 4.0, Benutzerhandbuch“, Bonn, Hannover, 1998
- European Environment Agency (1996): Newsletter issue 8, Copenhagen, June 1996
- Hashemi-Kepp, H., Legat, R. (1996): „Der Umweltdatenkatalog, ein Anwendungsbeispiel für Metainformationssysteme“, Informatikforum Band 10, Wien September 1996
- Hashemi-Kepp, H., Legat R.(1994): „Der Umweltdatenkatalog – Ein bundesweites Metainformationssystem über umweltrelevante Datenbestände in Österreich“. Salzburger Geographische Materialien Heft 21, Salzburg 1994
- Legat, R., Hashemi-Kepp, H. (1994): „Der Umweltdatenkatalog – Ein bundesweites Metainformationssystem über umweltrelevante Datenbestände“, VGI – Österreichische Zeitschrift für Vermessung & Geoinformation, Heft 1+2/94, Wien 1994
- Niedersächsisches Umweltministerium & Bundesministerium für Umwelt, Jugend und Familie der Republik Österreich: Umwelt-Datenkatalog Version 3.0 – Benutzerhandbuch. Hannover, Wien 1996
- Schober, W., Lopatta, H. (1994): Umweltinformationsgesetz. Verlag Österreich, Wien 1994
- Umweltbundesamt Berlin, Umweltbundesamt Wien (1997): „Thesaurus des Umweltdatenkatalogs (UDK-Thesaurus 3.0)“, Band I bis III, Berlin, Wien 1997
- Umweltbundesamt Berlin, Umweltbundesamt Wien (1998): „THESshow“, Thesaurus des Umweltdatenkatalogs (UDK-Thesaurus 3.0)“, CD-ROM, Berlin, Wien 1998

Erfassungsverfahren für Umweltmetadaten im Umweltdatenkatalog in Baden-Württemberg

Thomas Sattler¹, Ralf Nikolai²

1 Einleitung

1.1 Historie und Motivation

Der Umweltdatenkatalog (UDK) (Günther et al. 1996), (Swoboda et al. 1999) Baden-Württemberg steht seit 1995 im Intranet der Umweltverwaltung von Baden-Württemberg und seit 1996 im Internet zur Verfügung. Seither war und ist der UDK eine wichtige Komponente bei den Entwicklungen des baden-württembergischen web-basierten Umweltinformationssystems WWW-UIS. Für die Erfassung der Metadaten, die im UDK verwaltet werden, wird in Baden-Württemberg bisher ein zentraler Ansatz verfolgt. Dabei werden im Rahmen des seit 1987 betriebenen Umweltinformationssystems Baden-Württemberg (UIS) vorhandene Daten und Metadaten genutzt. Ferner wird konsequent nach Möglichkeiten gesucht, Metadaten, die im Rahmen von Fachanwendungen anfallen, in den Umweltdatenkatalog einzubringen. Hier bot es sich an verschiedener Stelle an, die Erfassung der Metadaten softwaretechnisch zu unterstützen, um den Erfassungsaufwand möglichst gering zu halten, soweit als möglich Fehler bei der Metadatenerfassung zu vermeiden. Die aus diesen Anforderungen resultierenden Softwareverfahren sollen in diesem Beitrag vorgestellt werden.

1.2 Überblick

Zunächst wird in *Abschnitt 2* ein kurzer Überblick über die Einbindung des Umweltdatenkatalogs in das Web-Angebot der Landesanstalt für Umweltschutz (LfU) und über die Metadatenerfassung in Baden-Württemberg gegeben. Daraufhin werden die verschiedenen in Baden-Württemberg eingesetzten Verfahren zur Unterstützung bei der Erfassung und Pflege der vom Umweltdatenkatalog verwalteten Umweltmetadaten

¹ Landesanstalt für Umweltschutz Baden-Württemberg, Griesbachstraße 1, 76185 Karlsruhe, email: Thomas.Sattler@lfuka.lfu.bwl.de, udk@lfuka.lfu.bwl.de

² Forschungszentrum Informatik (FZI), Haid-und-Neu-Straße 10-14, 76131 Karlsruhe, email: nikolai@fzi.de

in Abschnitt 3 erläutert. Der Einsatz dieser Verfahren wird in Abschnitt 4 diskutiert. Abgeschlossen wird der Beitrag mit einem Ausblick auf neue denkbare Einsatzbereiche automatisierter Verfahren zur Metadatenerfassung und -pflege in Abschnitt 5.

2 Einbindung des UDK ins WWW-Angebot der Landesanstalt für Umweltschutz Baden-Württemberg

Das Informationsangebot des Ministeriums für Umwelt und Verkehr ist im WWW unter der Adresse

<http://www.uvm.baden-wuerttemberg.de>

abrufbar. Aus der folgenden, stark vereinfachten Darstellung lässt sich die Einbindung des Umweltdatenkatalogs in das Informationsnetz (Web) der LfU im WWW erkennen.

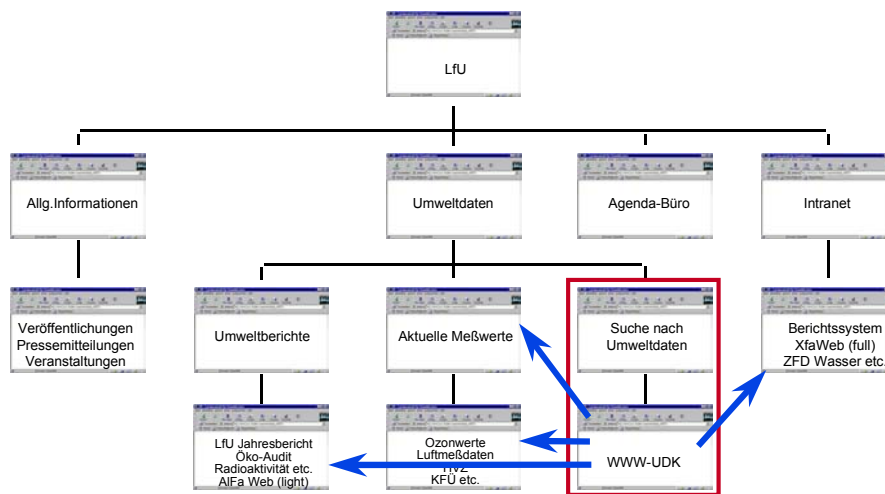


Abbildung 1

Übersicht über das Intranet/Internetangebot der LfU

Der UDK kann hier zur Suche nach Daten inner- und außerhalb des WWW-Angebots genutzt werden. Zu den zugrundeliegenden Umweltdaten, die im WWW verfügbar sind, können im UDK Verweise auf die entsprechende URL eingetragen werden.

3 Metadatenerfassung

Da es in Baden-Württemberg keinen Beschluss über die flächendeckende Einführung des UDK in der Umweltverwaltung gibt, wurden bisher nur zentral von der LfU aus UDK-Objekte erfasst; eine Verteilung des UDK an andere Dienststellen zur dezentralen Erfassung wurde nicht vorgenommen.

3.1 Erfassungsverfahren

In den folgenden Abschnitten 3.1.1 – 3.1.5 werden die bisher realisierten Erfassungsverfahren vorgestellt.

3.1.1 Manuelle Erfassung dezentraler Datenbestände (Interviews)

Die Möglichkeit der manuellen Erfassung durch Interview der Datenverantwortlichen wurde insbesondere in der ersten Pilotphase genutzt, wobei Ende 1994 Daten aus den Bereichen Wasser und Boden bei der LfU und bei einem damaligen Wasserwirtschaftsamt (heute Gewässerdirektion) erhoben wurden. Die so erfassten Metadaten wurden anschließend mit dem PC-basierten UDK-Erfassungsprogramm in den UDK eingetragen.

3.1.2 Teilweise Word-gestützte Erfassung zentraler Datenbestände

Die Geodatenbestände des Räumlichen Informations- und Planungssystems (RIPS) wurden im UDK beschrieben, indem die beim ITZ der LfU vorhandenen Informationen und die Daten selbst ausgewertet wurden. Für neue Datenbestände wurde hier ein Word-Formular entwickelt, in dem direkt mit Word oder auch von Hand die entsprechenden Metadaten eingetragen werden können. Um das Erfassungsformular nach elektronischer Bearbeitung in den UDK zu übertragen, wurde ein Programm entwickelt, das die erzeugten ASCII-Dateien (im SGML-Format) in die UDK-Datenbank einlesen kann.

3.1.3 Automatische Übernahme zentraler Messstelleninformationen

Ein anderer Weg der Metadatengewinnung ist die Analyse von Datenbeständen in der zentralen Messstellendatenbank der LfU. Wenn die Anzahl der Messstellen nicht zu groß ist, ist es sinnvoll, Metadaten über einzelne Messstellen zur Verfügung zu stellen. Hierzu wurde ein Programm entwickelt, das (Meta-)Informationen aus dem Messstellenverzeichnis oder auch aus den Messdaten selbst ermittelt und für jede Messstation eines Messnetzes in den UDK einträgt (Nikolai et al. 1997), (Nikolai et al. 1998).

3.1.4 Automatische Übernahme aus Fachinformationssystemen

Für die Erfassung von Informationen über Umweltberichte wurde im Rahmen des Altlasten-Fachinformationssystems (AlFaWeb) die Möglichkeit entwickelt, Metadaten über Berichte in den UDK zu übernehmen. Dabei wurde zunächst eine automatische Schlagwortvergabe (Generierung von Thesaurusbegriffen) eingesetzt, die allerdings wegen mangelnder Benutzerakzeptanz wieder verworfen wurde. Eine große Anzahl von Berichten aus dem Altlastenbereich wurde auf diese Weise in den UDK übernommen. Dieser Ansatz wird mit der Entwicklung gleichartiger Systeme für andere Fachbereiche wie Natur- und Bodenschutz weiterverfolgt.

Beim Technosphäre- und Luftinformationssystem (TULIS) wurde der Weg beschritten, vorhandene Hilfetexte zu den inhaltlichen Ergebnissen der Datenabfragen zu übernehmen. Die Texte werden von verschiedenen Fachleuten aus dem Bereich der Gewerbeaufsicht gepflegt und nur um noch fehlende Informationen z. B. Schlagworte für den UDK ergänzt. Auch hierbei konnte das oben genannte SGML-Einleseprogramm eingesetzt werden, da die Texte nach einem Schema aufgebaut sind, das leicht in das notwendige Format umgesetzt werden konnte.

Das SGML-Einleseprogramm ist eine Erweiterung des im Rahmen von AlFaWeb genutzten Programms, das dem Einlesen der Textdateien mit den automatisch generierten Schlagworten und weiteren Informationen in die UDK-Klasse Berichte diente, auf andere UDK-Klassen. Bei der Entscheidung für ein SGML-Datei war von großer Bedeutung, dass im europäischen Umweltmetadaten-System CDS beim Datenaustausch ein SGML-Format Verwendung findet.

3.1.5 Erfassung bei der Berichtserstellung

Der konsequenteste Ansatz, Informationen über Umweltberichte in den UDK zu übernehmen, ist mit dem Projekt Hypermediatechniken für Umweltdaten (HUDA) realisiert worden. Der HUDA-Arbeitsplatz, an dem umfassende Umweltberichte wie z. B. „Umweltdaten Baden-Württemberg“ erstellt werden, bietet die Möglichkeit, bereits bei der Berichtserstellung die notwendigen Metadaten zur Erstellung von UDK-Objekten zu sammeln, so dass schon mit der Erstellung eines Berichts dessen Erschließung für den UDK möglich wird (Tochtermann et al. 1998).

3.2 Technische Realisierung

In den folgenden Abschnitten wird die technische Realisierung ausgewählter Beispiele der oben vorgestellten Erfassungsverfahren vorgestellt.

3.2.1 Word-Formulare und SGML-Dateien zur nachträglichen Erfassung

Bei der Word-gestützten Erfassung zentraler Datenbestände sind Word-Formulare das zentrale Hilfsmittel zur Metadatenerfassung. Es existieren für die verschiedenen UDK-Klassen jeweils eigene Word-Formulare, die entsprechende Felder enthalten. Diese Felder müssen vom Benutzer entweder durch Eingabe von Freitext ausgefüllt werden oder aber der Benutzer wählt Einträge aus entsprechenden Listen aus. Diese Auswahllisten können ausschließlich für statische und relativ kleine UDK-Auswahllisten angeboten werden (vgl. Abb.2).

The image displays two side-by-side screenshots of a Microsoft Word form titled 'Schallminderungsplan'. The form is divided into several sections:

- Header:** Contains fields for 'Name', 'Titel', 'Objekt', 'Datum', and 'Seite'. The 'Objekt' field is filled with 'Schallminderungsplan'.
- Objekt-Identifikation:** A section for identifying the object, including fields for 'Objektname', 'Adresse', 'Institution', 'Name', 'Vorname', and 'Telefon'. Below these fields is a list of 'Objektklasse' options with checkboxes.
- Suchbegriffe:** A section for search terms, including a table for 'Suchbegriffe' and a list of 'Umweltklassifikation' options.
- Umweltklassifikation:** A list of environmental classification options, including 'LE Lärm/Erschütterung', 'LF Umweltaspekte der Land-, Forstwirtschaft, usw.', 'LU Luft', 'NL Natur und Landschaft/Räumliche Entwicklung', 'SR Strahlung', 'UA Allgemeine und übergreifende Umweltfragen', 'UR Umweltrecht', 'UW Umweltökonomie', and 'WA Wasser'. A dropdown menu is open over this list, showing the selected option 'LE Lärm/Erschütterung'.

Abbildung 2

Word-Formular mit aufgeklapptem Auswahlfeld zur Umweltklassifikation

Das ausgefüllte Word-Formular kann als SGML-Datei abgespeichert werden, die anhand der eingetragenen Daten von Word-Makros generiert werden. Da UDK 3 keine SGML-Definition besaß, wurde hier eine eigene DTD definiert. Mit der Umstellung auf UDK 4 wurde auch die SGML-Generierung an die SGML-DTD des UDK 4 angepasst. Allerdings sind hier einige Besonderheiten zu beachten. Diese sind zum einen unabhängig von der DTD (z. B. wird das Pflichtfeld Objekt-ID nicht automatisch mit einer global eindeutigen ID ausgefüllt sondern mit einer ID, die nur innerhalb des SGML-Dokuments eindeutig ist) zum anderen musste die DTD leicht modifiziert werden (Felder, die nicht ausgefüllt werden können, bekommen einen definierten „LEER“-Wert, der bei der Nachbearbeitung gefüllt werden muss (vgl. Abb. 3).


```

<!DOCTYPE UDK SYSTEM "UDK.DTD">
<udk>
<data-source>
<control-identifier>
<!-- hier muss eine id stehen, die fuer diese Sgml-Datei eindeutig ist -->
Grundwasser-Aufschluß (AU)
</control-identifier>
<catalogue-identifier>
UDK Baden-Württemberg
</catalogue-identifier>
<creation-identifier>
<!-- Wird beim Eintrag erzeugt -->
</creation-identifier>
<modification-identifier>
<!-- Wird beim Eintrag erzeugt -->
</modification-identifier>
<type-of-information-resource class-id="5">
Datensammlung / Datenbank
</type-of-information-resource>
<title>
Grundwasser-Aufschluß (AU)
</title>
<abstract>
Aufschluß: Als Aufschluß gilt jeder Zugang zum Grundwasser. Dies kann z. B. eine Bohrung, ein Brunnen oder eine Quelle sein. Je nach Art eines Aufschlusses werden unterschiedliche Aufschlußtypen unterschieden, die auch eine unterschiedliche technische Beschreibung erfordern.
</abstract>

```

Abbildung 3
Ausschnitt aus der SGML-Datei (UDK 4.0)

Die generierten SGML-Dateien können z. B. per E-Mail an die zentrale Metadaten-erfassung geschickt werden. Dort werden sie mit einem Werkzeug zur Nachbearbeitung der Metadaten (s. Abschnitt 3.2.4) in die qualitätsgesichert, ergänzt und in die UDK-Datenbank eingespielt.

3.2.2 Automatische Auswertung von Datenbeständen

Verschiedene Datenbestände (z. B. die Messstellendatenbank der LfU) beinhalten bereits Metadaten. Um diese Metadaten aus den Datenbeständen zu extrahieren wurde ein Programm entwickelt, das diese Extraktion anhand von Regeln vornimmt. Dieses Programm zur Extraktion greift auf verschiedene Datenquellen zu, z. B. Dateien, die manuell erstellte Hinweise geben und auf verschiedene Datenbanken an unterschiedlichen Orten innerhalb des Intranets der LfU. Da das Programm in C entwickelt und die Regeln „hartcodiert“ wurden, besteht eine große Flexibilität in der Umsetzung der Regeln. Allerdings ist es recht aufwendig, neue Regeln zur Extraktion und Ableitung der Metadaten aus den Datenbeständen zu ergänzen.

Um die so gewonnen Metadaten aktuell zu halten, wurde ein Ansatz mit flexibleren Regeln entwickelt (Nikolai et al. 1998). Änderungsereignisse an den Datenbeständen werden automatisch erkannt und von einer Ereignisverarbeitung werden entsprechende Reaktionen ausgelöst. Die Regeln der Ereignisverarbeitung können innerhalb der Expertensystemshell CLIPS formuliert werden.

3.2.3 Werkzeugintegrierte Erfassung von Metadaten

Ziel ist es, bereits bei der Erfassung der Daten die notwendigen Metadaten bereits mitzuerfassen. Dies ist sowohl mit der HUDA-Toolbox möglich als auch bei der Erstellung von Dokumenten für ALFaWeb. Im ersten Fall werden aus Metadaten, die für eine fein-granularere Beschreibung der Dokument(teile) sowieso benötigt werden für den UDK entsprechend aggregiert. Dazu werden Administrations-Komponenten verwendet, die mittels COM (Component Object Model) und OLE (Open Linking and Embedding) realisiert sind (Tochtermann et al. 1998). Im zweiten Fall werden aus den Berichtsmetadaten direkt Metadaten für den UDK bereitgestellt, die dann mit in einem weiteren Schritt mit Werkzeugen nachbearbeitet werden (s. Abschnitt 3.2.4). In beiden Fällen sind bei den Systemen zur Datenerfassung (hier: Berichtserstellung) die Anforderungen des UDK bekannt und werden durch entsprechende Module berücksichtigt.

3.2.4 Werkzeuge zur Nachbearbeitung der Metadaten

Metadaten, die dezentral mit integrierten Werkzeugen bei der Datenerfassung oder mit Word-Formularen dezentral erfasst werden, erfordern eine Qualitätssicherung und Ergänzung. Dazu werden die Dateien an die zentrale Erfassung geschickt. Es wurden unterschiedliche Werkzeuge entwickelt (z. B. für die verschiedenen SGML-Formate), die eine Nachbearbeitung und anschließend das Einspielen in die Datenbank ermöglichen.

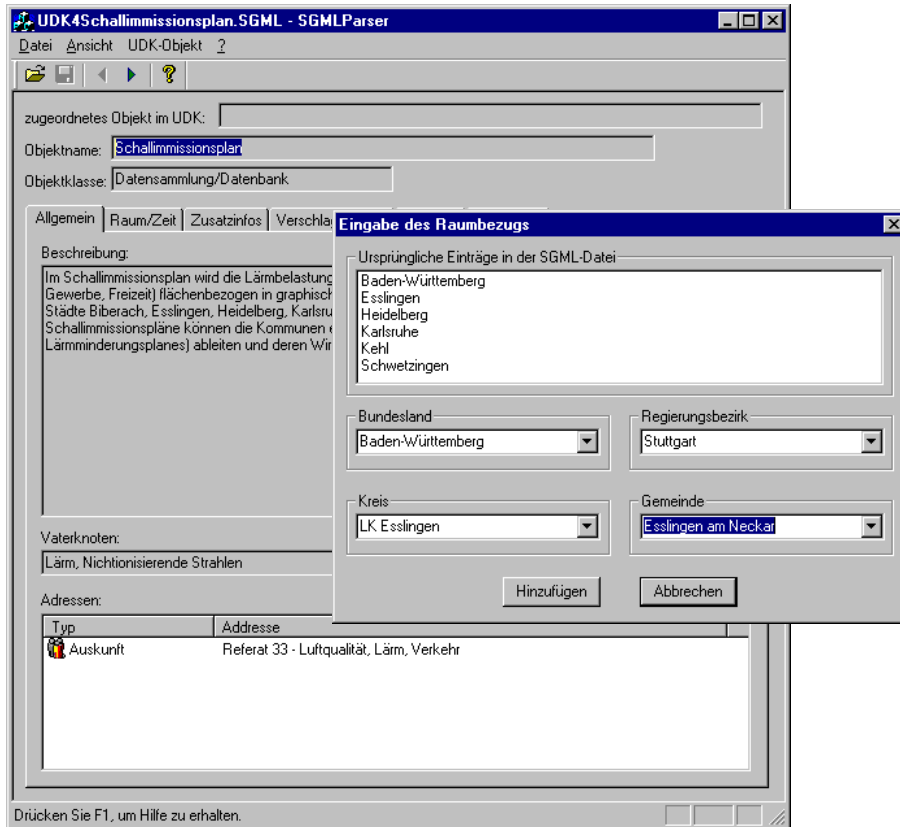


Abbildung 4
SGML-Parser mit geöffneter Dialogbox zum Eintrag der Verwaltungseinheit

Eine aktuelle Entwicklung ist ein in Visual C++ realisierter SGML-Parser für das leicht modifizierte UDK4-SGML-Format (vgl. Abschnitt 3.2.1). Dieser basiert auf dem public domain SGML-Parser SP von James Clark sowie dem darauf aufbauenden MDDL-Parser des Institut für Kernenergetik und Energiesysteme (IKE). Die SGML-Datei, die beliebig viele UDK-Objekte der unterschiedlichen UDK-Klassen enthalten kann wird in eine interne Struktur eingelesen. Dabei werden bereits Objekt-ID und weitere Daten automatisch generiert (dazu ist zu Beginn eine Authentifikation als UDK-Administrator erforderlich). Desweiteren wird bei den Schlagworten überprüft, ob es sich um Deskriptoren oder Non-Deskriptoren des UDK-Thesaurus handelt. Ist dies der Fall wird der Deskriptor bzw. die entsprechenden Deskriptoren ausgewählt. Ansonsten wird das Schlagwort als freier Suchbegriff eingetragen.

Die gesamten Daten werden in verschiedenen Reitern angezeigt, einige Felder können geändert werden, andere müssen es (z. B. Verweise auf Adressen). Wenn möglich werden für die Änderungen bereits Auswahllisten angeboten, die mit Daten aus der UDK-Datenbank gefüllt werden (z. B. Verwaltungseinheiten und Adressen). Folgende Benutzereingaben sind vorgesehen: Auswahl des Elternobjekts (Neueintrag) bzw. des zu bearbeitenden Objekts (Aktualisierung), Auswahl der Adresse, der rechtlichen Grundlage, der Umweltklassifikation, des Dokumenttyps, beim Raumbezug Auswahl der administrativen und fachlichen Raumeinheit, beim Zeitbezug Auswahl der Art des Zeitbezugs, Periodizität und Intervall (vgl. Abb. 4).

Der SGML-Parser greift zum Lesen und Schreiben der UDK-Daten mittels ODBC direkt auf die UDK-Datenbank zu.

4 Diskussion

Der UDK ist hervorragend dazu geeignet, ein „Portal zu Umweltinformationen“ bereitzustellen (Nikolai et al. 1999). Wesentliche Voraussetzungen ist allerdings eine ständige und möglichst umfangreiche Erfassung wichtiger Ressourcen (Umweltdatenbestände und Dienste), die Qualitätssicherung und Aktualisierung vorhandener Metadatenbestände und die Einrichtung von Querverweisen zu anderen relevanten Ressourcen. Denn erst ein gut befüllter und aktueller UDK-Metadatenbestand bietet die Voraussetzung, die Informationsbedürfnisse der Benutzer umfassend zu befriedigen.

Eine zentrale, rein manuelle Erfassung der Metadaten kann diesen Ansprüchen auf die Dauer nicht gerecht werden. In Baden-Württemberg wird daher ein hybrider Ansatz verfolgt: Für die manuelle Erfassung wird das Standard-PC-Erfassungsprogramm von den UDK-Experten eingesetzt. Da die Einarbeitung in diese Software den mit Metadaten weniger vertrauten Erfassern, die verhältnismäßig wenige Metadaten eintragen wollen, eine zusätzliche Hemmschwelle darstellt, wird mit Word-Formularen ein einfaches, aber allen Benutzern vertrautes Werkzeug angeboten. Der zusätzliche Aufwand (Softwareentwicklung und Nachbearbeitung der Daten) wird durch die Einbindung weiterer Mitarbeiter, die sich hervorragend mit ihren Umweltdaten auskennen, gerechtfertigt. Des weiteren ermöglicht eine automatische Übernahme von in verschiedenen Systemen bereits vorhandenen Metadaten die Vermeidung von Redundanzen und Mehrfacherfassungen. Bei neu zu entwickelnder Software, wie z. B. dem HUDA-Werkzeugkasten (Tochtermann et al. 1998) kann das Einstellen neuer Datenbestände direkt mit dem Einstellen der zugehörigen Metadaten gekoppelt werden.

Andererseits kann der UDK unmittelbar bei der Softwareentwicklung genutzt werden, indem die in Zuge der Feinkonzeption von Fachinformationssystemen benötigten Datenkataloge von vornherein mit Hilfe des UDK erstellt und gepflegt werden. Beim Einsatz so entwickelter und dokumentierter Informationssysteme sind dann UDK-konforme Metadaten mit minimalem Aufwand bereitzustellen.

5 Ausblick

Abschließend noch einige Gedanken zur Weiterentwicklung der hier beschriebenen automatisierten Verfahren, die sich allerdings noch nicht in konkreten Konzepten niedergeschlagen haben.

5.1 Aktualisierungsformulare

Ein bereits mit den **aktuellen Daten aus der UDK-Datenbank gefülltes Word-Formular** wäre für Aktualisierungen durch die Fachleute besonders günstig, da diese dann nur die notwendigen Korrekturen eintragen müssen. Hierzu ist angedacht, Möglichkeiten von Visual-Basic zu nutzen, die die gezielte Übernahme von Daten aus der Datenbank in Word-Dateien ermöglichen. Aktualisierungen von vorhandenen Metadaten werden in Zukunft eine immer größere Bedeutung haben.

5.2 Austausch von Metadaten zwischen UDK und WWW-Seiten

Es ist erklärtes Ziel der LfU, für alle im Web angebotenen Informationen im UDK Beschreibungen vorzuhalten. Auf der anderen Seite wird die Bedeutung von Metainformationen in den Webseiten selbst immer größer, um den Suchmaschinen bessere Informationen anbieten zu können (vgl. GEIN2000). Hier stellt sich die Anforderung, auch künftig die Konsistenz zwischen den Metadaten im UDK und in den Webseiten zu gewährleisten.

5.3 Integration des SGML-Import-Werkzeugs in das Standard-PC-Programm

Durch eine stärkere Integration des SGML-Einlesewerkzeugs in das UDK-PC-Programm könnten verschiedene Vorteile erreicht werden. So wäre die Nutzung vorhandener Programmfunktionen möglich, was sich besonders bei den Nutzereingaben anbieten würde. Außerdem wäre damit eine unmittelbare Ergebniskontrolle und Nachbearbeitungsmöglichkeit gegeben. Allerdings müsste vermutlich mit geringerer Flexibilität bei der Weiterentwicklung gerechnet werden.

6 Literatur

- Günther, O., Lessing, H., Swoboda, W., UDK: A European Environmental Data Catalogue, in: Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modelling, Santa Fe, New Mexico, URL: http://ncgia.ncgia.ucsb.edu:80/conf/SANTA_FE_CD-ROM/sf_papers/guenther_oliver/my_paper.html, 1996
- Nikolai, R., W. Kazakos, R. Kramer, S. Behrens, W. Swoboda, F. Kruse. WWW-UDK 4.0: Die neue Generation eines Web-Portals zu deutschen und österreichischen Umweltdaten. Papier angenommen bei 13. Internationales Symposium Informatik für den Umweltschutz (Umweltinformatik 99), Magdeburg, September 1999.
- Nikolai, R., A. Koschel, R. Kramer. Automating metadata updates exemplified by the environmental data catalogue UDK. In Proceedings 8th International Conference on Management of Data (COMAD'97), S. 263-276, Chennai (Madras), Indien, Dezember 1997. Narosa.
- Nikolai, R., A. Koschel, R. Kramer und Th. Sattler. Automatisierung der Metadatenaktualisierung am Beispiel des Umweltdatenkatalogs UDK. In: Vernetzte Umweltinformationen, Proceedings des Workshops „Vernetzte Umweltinformationen“ des GI-Arbeitskreises Umweltdatenbanken (1997), Hrsg: Jörg Hoppe, Sigrid Helle, Hansjörg Krasemann, Metropolis, 1998.
- Swoboda, W., F. Kruse, Nikolai R., Kazakos W., Nyhuis D., Rousselle, H. The UDK Approach: The 4th Generation of the Environmental Data Catalogue for Austrian and German Public Authorities. Third IEEE Meta-Data Conference, 1999.
- Tochtermann, K., W.F. Riekert, L. Kadric, R. Kramer, R. Schmidt, W. Geiger, N. Peter, M. Reissfelder, E.-E. Doberkat, H.-G. Sobottka, A. Keitel, W. Zitzmann, T. Schütz, J. Burkhardt. HUDA A Toolbox for Environmental Report Production. In H.D. Haasis and K.C. Rantze, Hrsg., Proceedings 12th International Symposium Computer Science for Environmental Protection: Networked Structures in Information Technology, the Environment and Business, Umwelt-Informatik Aktuell, S. 729-742, Bremen, September 1998. Metropolis.

Modellgetriebene Integration von Fachverfahren in Informationssysteme der öffentlichen Verwaltung

Ulrike Freitag, Hans-Peter Schreiter, Thomas Schwotzer¹

Abstract

Behörden verfügen an vielen Standorten über Fachverfahren mit Umwelt-, Infrastruktur-, Bürgerdaten etc., die heute nur lokal zugreifbar sind. Diese isolierten Datenbestände sollen künftig unter Beibehaltung der Autonomie und des Datenschutzes der Fachverfahren über Web-basierte Informationssysteme (IS) fachverfahrenübergreifend recherchierbar werden.

Die Entwicklung solcher IS erfordert häufig eine Integration unterschiedlicher, stark heterogener bestehender Softwarekomponenten in eine gemeinsame heterogene, verteilte Infrastruktur. Dies erfolgt in enger Zusammenarbeit von Fachexperten der Behörden und Spezialisten der Softwaretechnologie. Dabei sind Modelle für den Entwicklungs- und Wartungsprozeß ein geeignetes Mittel der Verständigung, das eine zuverlässige, kontrollierbare und systematische Abbildung der komplexen Struktur der real existierenden Fachverfahren auf Programm-Bausteine (Komponenten) des zu erstellenden Softwaresystems gestattet. Schlüssel für die Integration und Bereitstellung fachverfahrenübergreifender Recherchen ist ein geeignetes Meta-Modell des IS, in dem sowohl die wesentlichen Informationen über die einzelnen auf dieser Basis integrierten Komponenten (sogenannte Metainformationen) als auch deren Beziehungen untereinander beschrieben werden.

Um die Integrität sowohl zwischen Modellen und zugehörigem Softwarebaustein als auch von Modellen der Fachverfahren und Meta-Modell des IS zu gewährleisten, wird ein modellgetriebener Integrationsprozeß verfolgt. Dazu werden aus Modellen der Komponenten durch spezielle softwaretechnische Werkzeuge (Generatoren) automatisiert einzelne bausteinartig integrierbare Softwarekomponenten erzeugt. Grundlage ist ein flexibles Konzept zur Konstruktion und zum Zusammenbau solcher Komponenten.

Da diese Arbeiten zeitgleich mit zu beachtenden internationalen Arbeiten der großen Frameworks zu Komponenten-Modellen wie Enterprise JavaBeansTM (Hapner98), den CORBA Components der Object Management Group (OMGTM) (OMGCC98) für Middleware-Schichten-Architekturen entstanden, wird eine kurze Einordnung der eigenen Arbeiten in diesen Kontext vorgenommen.

¹Condat AG,
Alt Moabit 91d, D-10559 Berlin
E-Mail: {uf|hps|tom}@condat.de, Internet: <http://www.condat.de>

1 Einleitung

1.1 Gegenstand/Lösungsansatz/Mittel

Bei der Entwicklung von modernen heterogen verteilten IS ist die Integration neuer Datenbestände eine immer wiederkehrende komplizierte Aufgabe. Deshalb gilt es, die Integration von neuen oder existierenden Fachverfahren in eine einheitliche verteilte Infrastruktur durch einen flexiblen Prozeß zu unterstützen, der die Integration nachvollziehbar und technisch leichter macht. Dazu ist es erforderlich, Informationen zu erfassen, die einerseits das integrierte Fachverfahren einschließlich der dazugehörigen Objekte beschreiben. Andererseits ist es auch wichtig, Informationen zu erfassen, die Beziehungen zu bereits in der Infrastruktur integrierten Fachverfahren oder Meta-informationssystemen dokumentieren. Da diese Informationen der Beschreibung der eigentlichen Daten in den Datenbeständen der Fachverfahren dienen, werden sie als Metadaten bzw. Metainformationen ("Daten über Daten") bezeichnet.

Um die Integration besser zu unterstützen, muß das Modell des bereits existierenden IS für den Entwickler recherchierbar sein. Zugleich ist auch das Meta-Modell des IS wichtig, das beschreibt, aus welchen Arten von Bestandteilen das IS besteht und welche Informationen über diese Bestandteile zu erfassen sind, damit diese Recherchen im Kontext der Integration unterstützt werden können.

Eine wichtige Aufgabe bei der Entwicklung einer Architektur heterogener, verteilter IS die flexibel erweiterbar sein müssen, ist folglich die Festlegung der Arten von Bestandteilen bzw. Komponenten.

Sind diese Fragen geklärt, kann man den Kern des IS bauen, der dann eine bzgl. der konkret eingebundenen Datenbestände flexibel erweiterbar ist.

Lösungsansatz

Der Kern des Lösungsansatzes für die Integration neuer Bestände besteht darin für jeden einzelnen Datenbestand ein Datenmodell zu verwalten. Dieses wird mittels der Unified Modeling Language (UML) modelliert (Booch 1998, Alhir 1998, Neumann 1998). Damit werden die existierenden Fachverfahren über eine einheitliche Schnittstelle erreichbar. Darüber ist eine Integration in das neue Gesamtsystem leicht möglich.

Generatoren, die auf diesen UML-Modellen aufsetzen, erzeugen mit Hilfe eines Satzes von bestimmten Codeschablonen (templates) das zu erstellende Codepaket (Datenkapsel) für den zu integrierenden Datenbestand. Dabei wird jeweils eine Schablone für unterschiedliche zu erstellende Textbausteine (Codearten) wie beispielsweise IDL-Dateien, Java-Dateien, SQL-Scripts etc. benötigt. Jeder Datenbestand wird durch eine sogenannte *Komponente* (Paket von Dateien) realisiert. In der Literatur wird diese modellgetriebene Vorgehensweise mit "model elaboration" bezeichnet (OMGMOF98).

Bei jeder Generierung wird nicht nur das zu integrierende Codepaket selbst erzeugt, sondern gleichzeitig auch das Repository mit den Metadaten des dadurch integrierten Fachverfahrens gefüllt. Dies ist in Abbildung 1 dargestellt.

Dazu erfolgt die Beschreibung der Art der Informationen, die dem Fachverfahren zugeordnet sind und zum (Wieder-)Auffinden der Daten nötig sind.

Mittel

Zur Umsetzung dieses Ansatzes wird ein Komponentenbegriff benötigt, der folgenden Anforderungen Rechnung tragen muß:

- Verteilbarkeit bzgl. Standorten,
- Integration von vorhandenen Informationsobjekten,
- Konstruktion von neuen Informationsobjekten für verschiedene wiederverwendbare Informationen,
- Transaktionsbehandlung, Zugriffsschutz, Konfigurierbarkeit
- Integration verschiedener Datenhaltungssysteme und
- leichte Erweiterbarkeit/Änderbarkeit eines bestehenden Systems.

Um diese Anforderungen besser erfüllen zu können, werden folgende Dinge benötigt:

- *Werkzeuge zur flexiblen Erstellung und Integration neuer Recherche-Bausteine*
Um die Konstruktion neuer Recherche-Bausteine zu vereinfachen und die Integration technisch zu unterstützen, ist ein flexibles Baustein-Konzept für ggf verteilte Infrastrukturen erforderlich. Derartige Konzepte werden in der neueren Literatur (Hapner 1998, OMGCC 1998, Thomas 1998) als Komponentenmodell bezeichnet.

- *Werkzeuge zur Automatisierung der Integration der Datenbestände der Fachverfahren über technische Metadaten*

Metadaten werden softwaretechnisch häufig in sogenannten Repositories verwaltet. Ein *Repository* wird gemeinhin als zentrales rechnergestütztes Werkzeug zur Speicherung und Verwaltung von verschiedensten Unternehmensinformationen (Daten, Prozesse, Werkzeuge, Dokumente etc.) verstanden. Im Unterschied zur gewöhnlichen Datenbank liegt jedoch ein wesentlicher Schwerpunkt auf der Verwaltung der oben eingeführten Metainformationen.

Das Repository bildet beim modellgetriebenen Ansatz das Herzstück für die Integration der verschiedenen Fachverfahren in die verteilte Infrastruktur des IS. Die für die Integration benötigten Werkzeuge sind vor allem Werkzeuge zur Erzeugung der Komponenten (Generatoren), aber auch zur Werkzeuge zur Füllung und Wartung des Repository.

1.2 Aufbau des Artikels

Dieser Artikel basiert auf Arbeiten in Projekten zur Entwicklung von komplexen, verteilten IS für Behörden der Berliner Öffentlichen Verwaltung. Schwerpunkt sind die Erfahrungen im Umgang mit Modellen bei der Integration von Datenbeständen in eine heterogene, verteilte Infrastruktur. Dazu werden zunächst im zweiten Kapitel kurz die Projekte vorgestellt, in denen mit diesem Ansatz erfolgreich Implementierungen durchgeführt wurden.

Im dritten Kapitel werden die Konzepte für den oben erwähnten Generatoransatz vorgestellt. Da ein Teil der Projekte im Umfeld heterogener CORBA-basierter verteilter Systeme stattfindet, wurde neben dem Framework zur Entwicklung von Generatoren auch ein Framework zur Implementation und Integration von *lightweight-Components* entwickelt. Letzteres unterstützt ein leichtgewichtiges, pragmatisch nutzbares Komponentenmodell für Java unter CORBA 2 (bzw. 2.3) und wird im vierten Kapitel vorgestellt. Daran anschließend werden die eigenen Ansätze mit den obengenannten internationalen Arbeiten kurz vergleichend diskutiert, bevor der Artikel mit einer Zusammenfassung und einem Ausblick auf weitere interessante Arbeiten endet.

2 Beispiel-Projekte

In diesem Kapitel werden kurz zwei Projekte erläutert, die im Umweltkontext durchgeführt wurden und in deren Verlauf die obengenannten Frameworks entwickelt bzw. wiederverwendet wurden.

Der Überblick über die verschiedenen Kontexte der Projekte soll verdeutlichen, daß ein flexibler, wiederverwendungsfähiger und ggf. parametrisierbarer Generatoransatz sinnvoll ist, um den Entwicklungsaufwand projektübergreifend, zu minimieren.

2.1 VeZuDa – Vereinheitlichung bzw. Zusammenführung der verschiedenen Datenstrukturen in der Berliner Verwaltung

Kurzbeschreibung:

Die bisher voneinander getrennten und redundant gehaltenen Grunddatenbestände der Berliner Verwaltung (Bürgerdaten, geografische Informationen, Adressen etc.) sollen – soweit möglich und rechtlich zulässig – allgemein über das behördeninterne Intranet bzw. auch über das öffentliche Internet zur Verfügung gestellt werden.

Problembeschreibung:

Das Projekt VeZuDa hat deshalb die Aufgabe, ein Konzept zu entwerfen und umzusetzen, um den Mangel an Verfügbarkeit sowie vermeidbare Redundanzen in der

Erhebung und Pflege von Informationen zu beseitigen. Ein weiteres Ziel ist die Schaffung einer modernen IT-Infrastruktur, um die Entwicklung von IT-Verfahren kostengünstiger und einfacher zu gestalten. Hierbei steht insbesondere der Aspekt der Wiederverwendung im Vordergrund. Zu integrierende Datenbestände sind Verwaltungs- und Umweltdatenbestände im weitesten Sinne wie beispielsweise das Amtliche Liegenschaftsbuch- und Kataster (ALB, ALK), das Regionale Bezugssystem (RBS) des Statistischen Landesamtes, das Altlastenverdachtsflächenkataster, der Gewerbeflächenatlas, der Bodenrichtwertatlas u. v. a. m.. Die bereits existierenden Fachverfahren sollen unter Beibehaltung ihrer Autonomie gekapselt werden und i. a. nur lesend zugreifbar gemacht werden. Wichtig ist eine Unterstützung unterschiedlichster Datenhaltungssysteme.

Lösungsidee:

Recherchen über technischen Metadaten sollen entwicklerseitig die Integration neuer Fachverfahren erleichtern. Betreiberseitig sollen Recherchen über organisatorisch-technischen Metadaten Administration und Wartung dieses heterogenen, verteilten IS unterstützen. Anwenderseitig sollen Recherchen über fachliche Metadaten eine bessere Orientierung und Navigation innerhalb der komplexen fachlichen Informationsstrukturen ermöglichen.

Wesentliche Anforderungen an dieses moderne IS bzgl. der Recherchearten sind daraus resultierend

- das Bereitstellen von Recherchen über fachlichen, organisatorischen und technischen Metadaten, um Datenbestände von Fachverfahren zu finden, bzw. Administration und Integration zu unterstützen
- das Bereitstellen von häufig benötigten Recherchen über Datenobjekten der Fachverfahren durch vorkonfigurierte, bausteinartig kombinierbare Recherchen und darüber hinaus
- das Bereitstellen von fachverfahrenübergreifenden Daten-Recherchen.

Als Lösungsidee zur Bewältigung dieser Anforderungen bietet es sich an, zur Realisierung häufig verwendeter oder fachverfahrenübergreifender Recherchen, eigene, wiederauffindbare Geschäftsobjekte zu verwenden.

Solche Geschäftsobjekte können gekapselte Objekte aus den integrierten Datenbeständen der Fachverfahren (Basisobjekte) abstrahieren (vereinfachende Sichten auf komplexe Objekte der Fachverfahren), aggregieren oder um neue auswertende oder verarbeitende Funktionalität erweitern.

Es wird eine heterogene CORBA-fähige Infrastruktur aus Basis- und Geschäftsobjektenkomponenten bereitgestellt und ein Kernsystems entwickelt, das eine Laufzeitumgebung zur Verwaltung dieser Objekte bereitstellt. Diese Arbeiten basieren auf dem in Kapitel 4 beschriebenen Middleware Framework *Core*.

Die Integration der verschiedenen Datenbestände der Fachverfahren erfolgt im Rahmen von einzelnen Integrationsprojekten. Sie werden schrittweise über ihre Modelle integriert. Herzstück der Integration ist ein Repository. Dieses Repository ist Basis für die dynamisch erzeugte Web-basierte Nutzerschnittstelle des Informationssystems. Diese ist in der gegenwärtigen Form eine rein HTML-basierte Lösung, die über Servlets die CORBA-Server ansprechen und mittels Java Server Pages (JSP) dynamisch die Ergebniswebseiten erzeugen.

2.2 FIS Broker – Fachübergreifendes Informationsystem

Kurzbeschreibung:

Das Ziel von FIS Broker ist die Schaffung eines einheitlichen Informationssystems, das die existierenden räumlichen und ökologischen Datenbestände bei den Planungsämtern der Berliner Senatsverwaltung für Stadtentwicklung- und Umwelttechnik und der Bezirke (v. a. Umweltinformationssystem UIS, Bebauungspläne, Flächennutzungsplan, Wasserschutzgebiete) sowohl internen als auch externen Nutzern umfassend, schnell und aktuell verfügbar macht. Durch die verbesserte und aktuellere Bereitstellung von Basisdaten soll die Produktion von Datenbeständen (z. B. Bebauungs-, Bereichsentwicklungspläne) stärker unterstützt und beschleunigt werden. Das Projekt FIS Broker ist ein Integrationsprojekt von VeZuDa.

Problembeschreibung:

Die Datenbestände dieses Projektes enthalten umfassende Metadaten, die über verschiedene Integrationsprojekte hinweg von Bedeutung sind und die deshalb in einem speziellen Datenbestand erfaßt werden sollen. Während sonst Metadaten bzgl. eines gesamten zu integrierenden Datenbestandes erhoben werden, werden im Rahmen von FIS Broker beispielsweise für geographische Karten viele instanzbezogene Metadaten benötigt. Außerdem werden auch Metadaten erfaßt, die rein zur Anzeige in den Webseiten oder als einfache Textbausteine benutzt werden, und also nicht über Recherchen erreichbar sein müssen.

Lösungsidee:

Es wird ein vom VeZuDa-Repository abgeleitetes spezielles Repository erzeugt, das vor allem die fachlichen Semantikbezüge enthält. Für das reine Datenmodell wird der Generatoransatz eingesetzt. Die komplexe fachliche Funktionalität wird allerdings nicht über den modellgetriebenen Ansatz realisiert, sondern durch die FIS-Broker-Entwickler unter Verwendung des generierten Codes manuell implementiert (siehe Ende Abschnitt 3.2).

Die Implementierung der Komponenten erfolgte ebenfalls unter Einsatz des Middleware Frameworks *Core*.

3 Modellgetriebener Generatoransatz

Das Benutzen von UML-basierten Codegeneratoren ist an sich nichts Neues, da dies mittlerweile von vielen UML-Modellierungswerkzeugen wie beispielsweise Rational Rose98 und TogetherJ angeboten wird.

Für die obengenannten Zwecke wurden aber Erweiterungen benötigt, die beispielsweise das Codeerzeugen nach den Konventionen eines komponentenfähigen Frameworks ermöglichen. Außerdem zeigte sich, daß bei Verwendung eines eigenen Generators auch viele Vereinfachungen bei der Modellierung möglich wären, die durch Ablage von redundanten, d. h. ableitbaren Informationen im Modell verursacht werden.

Gleichzeitig hat die modellgetriebene Integration den enormen Vorteil, daß eine spätere Erweiterung und Pflege der „lebendigen Infrastruktur“ durch das Betreiber-Personal leichter durchgeführt werden kann. So werden zur Erweiterung vor allem Modellierungskennnisse benötigt, um die fachverfahrenstypischen Teile zu sichten und ggf. zu ändern, ohne die tiefen Details der restlichen Implementierung bis ins Einzelne zu beherrschen.

Gegenstand dieses Kapitels ist der modellgetriebene Entwicklungs- und Integrationsprozeß. Nach einer Beschreibung der wesentlichen Idee des Generatoransatzes wird anhand des Modells für das VeZuDa-Repository ein Beispiel beschrieben.

3.1 Vom Analysemodell zum Designmodell

Ausgangspunkt für den modellgetriebenen Entwicklungs- und Integrationsprozeß sind UML-Klassendiagramme des Analysemodells des Datenbestandes eines Fachverfahrens. Solche Klassen von Fachobjekten verfügen über Attribute und einfache oder komplexere Operationen zur Manipulation der Attribute.

Diesem Analysemodell werden einige Zusatzinformationen hinzugefügt,

- die entweder allein für den Generierungsprozeß benötigt werden, wie zum Beispiel Mappinginformationen, die beschreiben durch welche Tabelle eine Klasse in einer relationalen Datenbank realisiert wird.
- Außerdem könnten aber auch bestimmte Metainformationen in Form von zusätzlichen Attributen oder Stichwort-Wert-Paaren (tagged values) abgelegt werden, die dann durch die Generierung im Repository abgelegt werden.
- Aus dem erweiterten Modell sollen verschiedene Codestücken generiert werden. Dazu ist es erforderlich, die verschiedenen zu erzeugenden Codearten im Modell zu kennzeichnen.

Konventionen zur Ablage von Zusatzinformationen in Modellen

Es gibt mehrere Alternativen zur Ablage der Zusatzinformationen. So können *UML-Stereotypes* oder sogenannte *Stichwort-Wert-Paare (tagged value)* benutzt werden.

Stereotypes sind ein UML-Mechanismus zur Klassifizierung oder Markierung von Modellelementen und zur Einführung neuer Typen von Modellelementen. Leider hat die Modellierung über Stereotypes den Nachteil, daß pro Modellelement nur ein Stereotyp zugeordnet werden kann.

Deshalb werden in neueren Generatorversionen die tagged values als das allgemeinere Konzept zur Ablage solcher Zusatzinformationen benutzt, da hierbei durch UML keine Beschränkungen vorliegen.

Ein weiterer Kerngedanke des modellgetriebenen Ansatzes ist es, die UML-Design-Modellierung auf ein nötiges Maß zu reduzieren und so die Codeerzeugung dahingehend zu unterstützen, daß das langwierige Schreiben von *get()*- und *set()*-Methoden für Attribute wegfällt, und deren Erzeugen automatisch von den Generatoren übernommen wird. Dadurch braucht der Modellierer nur noch die komplexen Methoden, die nicht durch diese einfachen Zugriffsmethoden generisch erzeugt werden können, im Modell ablegen.

Das so entstandene Modell ist Ausgangspunkt für die Implementierung. Alles weitere „Wissen“ über die Abbildung von Analysemodell zu Designmodell, das sonst in einem mühseligen Designprozeß aufgebracht werden muß, steckt in Form von Schablonen (Templates) und Abbildungsregeln in den Generatoren bzw. in Form des Komponentenmodells in dem dazugehörigen Middleware Framework. Dies ermöglicht es sogar, nicht nur Methoden im Modell „wegzulassen“, sondern kann den Modellierer von den ganzen Details der technischen Implementierung befreien, wo i.a. zur effizienten Programmierung noch eine ganze Menge von neuen Implementierungsklassen hinzukommt. So wird beispielsweise die ganze Umsetzung einer Komponente auf eine kompliziert aufgebaute Menge von Klassen verborgen.

Diese erweiterten Modelle sind somit eine Art minimales Designmodell und werden im weiteren kurz Designmodelle genannt.

Zusammenfassend muß beim Übergang vom rein fachlichen Analysemodell zum Designmodell also für jede Klasse des Analysemodells festgelegt werden, welche Art von Code daraus erzeugt werden soll und welche Zusatzinformationen benötigt werden. Der Modellierer muß dazu aus dem durch den Generatorentwickler aktuell unterstützten Satz von Stereotypes und tagged values die jeweils passenden auswählen.

3.2 Der UML-basierte Generator-Baukasten

Der Generatoransatz ist zunächst allgemeingültig und damit wiederverwendbar angelegt. Die Grundidee ist dabei, ausgehend von einem UML-Klassendiagramm, das Modell durch einen Modellparser zu analysieren und in Abhängigkeit der verwendeten Stereotypes und tagged values, ein Paket von dazugehörigen Texten zu generieren. Die Struktur dieser Texte wird durch die zu den jeweiligen Stereotypes bzw. tagged values zugehörigen Schablonen (templates) oder Abbildungsregeln bestimmt. Dies ist in Abbildung 1 dargestellt.

Die durch einen Generator konkret zu erzeugenden Klassen werden erst durch die zur Generierungszeit verwendete Generatorinstanz und das jeweilig verwendete Template-Set des jeweiligen Projektes bestimmt.

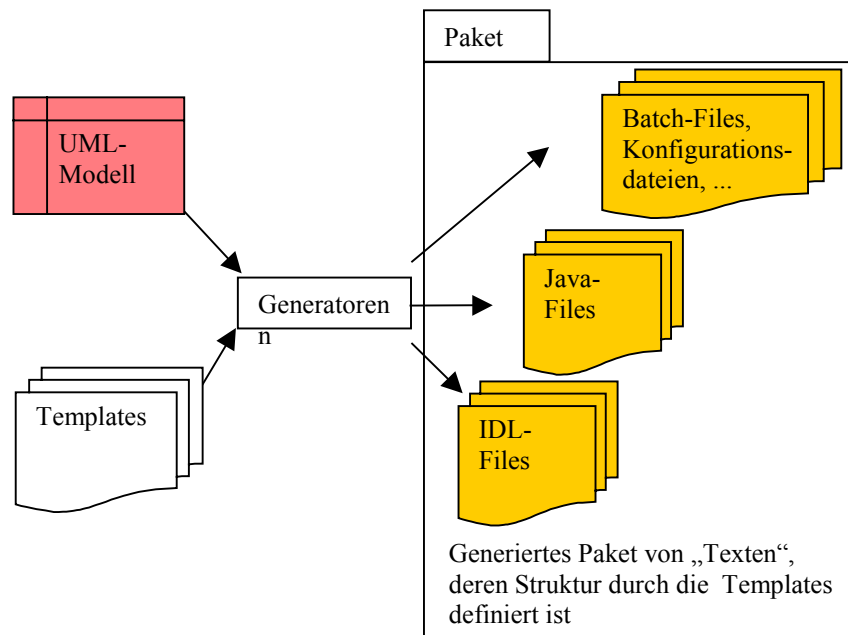


Abbildung 1
Arbeitsweise der Generatoren

Eine spezielle Form der Instanziierung des Generatoransatzes, um CORBA-fähige recherchierbare light-Components zu erzeugen, wird durch das Condat Middleware Framework *Core* spezifiziert, das die Entwicklung von bausteinartigen Komponenten unterstützt und ein entsprechendes Laufzeitsystem für solche Bausteine bereitstellt. Die Konzepte des Frameworks sind im nächsten Kapitel kurz dargestellt.

Um die Codeerzeugung sowohl für objektorientierte als auch für verschiedene Arten von relationalen Ziel-Datenbankmanagementsystemen zu unterstützen, wurden DBMS-spezifische Generatoren beispielsweise für JDBC-fähige Datenbanken und POET entwickelt. Diese benutzen DBMS-spezifische Templates. Auch hier wird ein extra Stereotyp für den DBMS-Typ (POET, JDBC, SAP, einfaches File etc.) zur Kennzeichnung benutzt. Der Generator ist selbst wiederum modular baukastenartig aufgebaut, damit er flexibel erweitert werden kann. Dieser Zusammenhang ist in Abbildung 2 dargestellt.

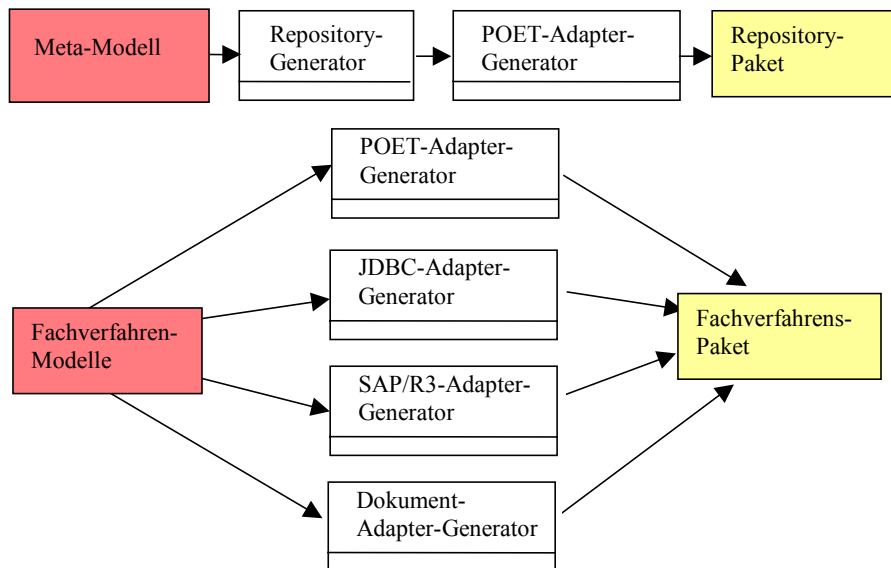


Abbildung 2
Baukasten von Generatoren

Konzept für die Separation von generiertem und manuell ergänztem Code

Die generierten Klassen mit den get()- und set()-Methoden können durch abstrakte Methoden für komplexere manuell zu implementierende Funktionalität ergänzt werden. Die komplexen Methoden werden im Rahmen einer von der generierten Klasse abgeleiteten Klasse implementiert und dadurch vom Überschreiben durch den Generator geschützt. Diese Separation von generiertem und manuell erzeugtem Code wird von kommerziellen Generatoren häufig nicht konsequent realisiert. Klassen, die solche komplexen Methoden benötigen, werden im UML-Modell ebenfalls mit einem speziellen Stereotyp markiert (XIDL_..._NEEDS_IMP).

3.3 Beispiel anhand der Erzeugung des VeZuDa-Repository

Um eine Vorstellung über ein solches Design-Modell zu vermitteln, das um generatorspezifische Informationen angereichert ist, zeigt die folgende Abbildung Ausschnitte² aus dem UML-Modell des VeZuDa-Repository.

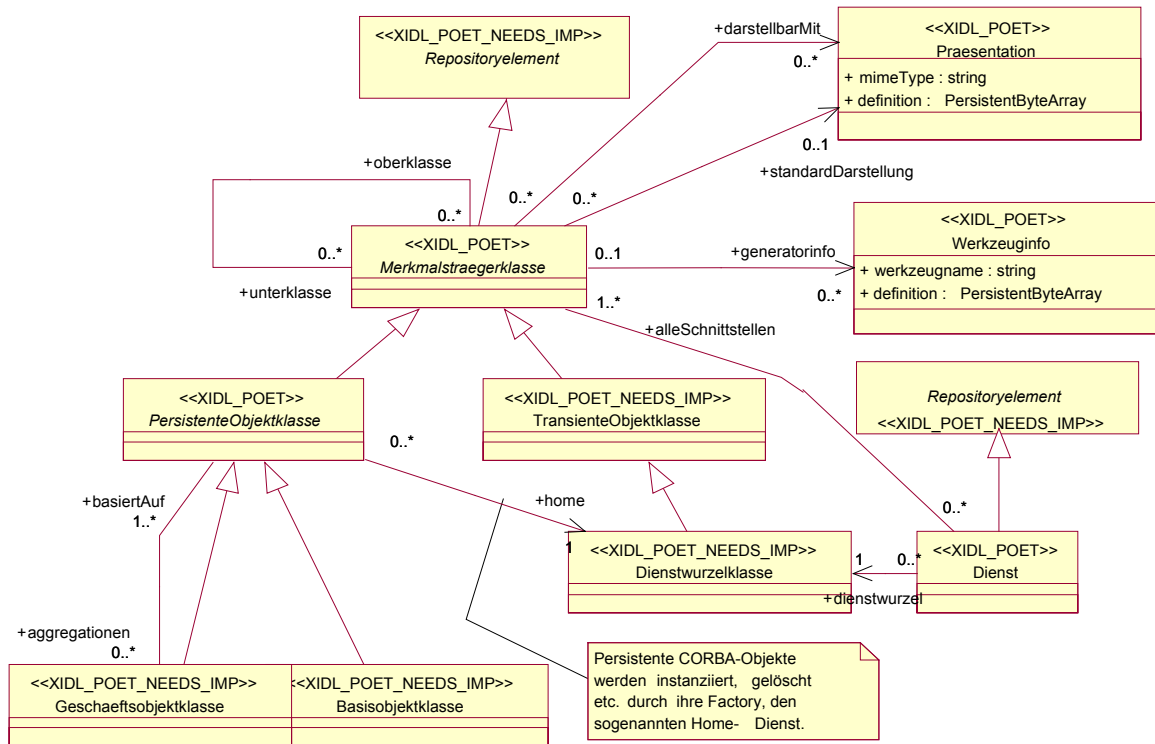


Abbildung 3
Ausschnitt aus dem VeZuDa-Repository-Modell

² Diese Ausschnitte wurden ausgewählt, da darin zugleich die wichtigsten Bestandteile des *Core-Frameworks* modelliert sind. Anhand dieses Modells sind die Unterschiede zu dem ähnlichen MOF-basierten Component Metamodel der OMG aus (OMGCC1998 S.181) ersichtlich. Die Frameworkklassen Dienstwurzelklasse, Basisobjektklasse und Geschäftsobjektklasse und Home sind die wichtigsten Metaklassen für die Konzeption des Komponentenmodells.

Der Repository-Generator nutzt Templates für

- frameworkspezifische IDL-Dateien (erster Teil im Stereotype "XIDL_POET"),
- verschiedene Templates für die verschiedenen frameworkspezifische Java-Implementationsklassen,
- ein recht komplexes Makefile für die Übersetzung der erzeugten Java-Klassen sowie
- eine Applikation zum Füllen des Repository mit den Metadaten der zu integrierenden Datenbestände des Fachverfahrens.

Für Datenbankspezifische Klassen werden spezielle Generatoren eingesetzt. Das Repository wird beispielsweise in POET persistent gehalten (zweiter Teil im Stereotype XIDL_POET). Zur Erzeugung dieser Klassen nutzt der Repository-Generator intern einen POET-Adaptergenerator, der mittels POET-spezifischer Java-Templates die Java-Klassen erzeugt, die den Datenbankzugriff realisieren. Einige Klassen benötigen über die Standardfunktionalität hinausgehende manuell implementierte Funktionen (Stereotyp XIDL_POET_NEEDS_IMP).

4 Middleware Framework für lightweight-Components Core

Im folgenden Kapitel werden kurz die wesentlichen Bestandteile des Middleware Frameworks *Core* beschrieben. Eine detaillierte Darstellung würde den Rahmen dieses Artikels sprengen. Deshalb sei an dieser Stelle auf (VeZuDa 1999) verwiesen.

4.1 Frameworkbegriff

Frameworks sind erweiterbare Mengen von objektorientierten Klassen, die so verknüpft werden, daß sie ein wohldefiniertes Verhalten zeigen. Nach (Johnson) sind Frameworks folgendermaßen definiert:

Ein Framework ist eine Anzahl von Klassen, die ein abstraktes Design für Lösungen zu einer Familie verwandter Probleme enthalten.

Man kann Frameworks auch als eine vorgefertigte Struktur oder eine Schablone (Template) eines Programms betrachten. Der größte Vorteil von Frameworks ist, daß sie Wiederverwendung von Programmwurf und -code auf einer höheren Ebene erlauben, als dies bei anderen Entwurfsmethoden möglich ist.

Marktverfügbare Frameworks werden als architektonische Bausteine eingesetzt, da sie sowohl Infrastruktur als auch flexible Schnittstellen zur Verfügung stellen und somit die Probleme und den Verwaltungsaufwand reduzieren können, die die traditionelle Programmierung den Softwareentwicklern abverlangt.

4.2 Framework *Core*

Im Laufe von mehreren Projekten im CORBA-Umfeld stellte sich zusätzlich heraus, daß für eine CORBA-basierte multi-tier Architektur ein flexibles Konzept sinnvoll ist, das ein "Plug and Play" von Bausteinen ermöglicht, das Konzepte wie Lebenszyklus, Transaktionshandling und Zugriffsschutz beinhaltet. In diesem Zusammenhang wurde das oben erwähnte eigene Middleware Framework *Core* erstellt, das die Entwicklung solcher Komponenten unterstützt und ein entsprechendes Laufzeitsystem für solche Bausteine bereitstellt.

Kombiniert man den oben beschriebenen Generator-Baukasten mit dem Komponenten-Framework *Core*, sind offene, komplexe und flexible Infrastrukturen mit verteilten Objekten wesentlich einfacher zu entwickeln, da viele ähnlich zu erstellende Codestücke durch generischen Code automatisiert erzeugt werden können und der Entwicklungsaufwand so nur noch einmalig bei der Template-Erstellung anfällt. Dies macht den Entwicklungsprozeß kürzer und die so erzeugte Software robuster, nicht zu letzt, weil so eine konsequente Trennung von generiertem Datenbankzugriffscod und manuell implementierter komplexer fachlicher Logik sichergestellt werden kann.

In Anlehnung an die in Entwicklung befindlichen Arbeiten zum Enterprise JavaBeans™ Framework (Hapner 1998) und zu den anvisierten CORBA Components (OMGCC 1998) im zukünftigen CORBA 3.0 Standard wird das Framework *Core* für die Entwicklung von Server-seitigen CORBA-2-fähigen lightweight-Components für die Realisierung von Basis-, Geschäftsobjekt- und Dienstobjekten verwendet. Dieses Framework definiert ein pragmatisches Komponentenmodell für die Entwicklung, Verteilung und Integration von CORBA-fähigen Objekten innerhalb einer Komponenten-Architektur eines verteilten Systems.

Lightweight-Components sind wiederverwendbare, konfigurierbare, CORBA-2-fähige Teile von Applikationscode, die in ein existierendes Applikationssystem integriert werden können.

Die Unterstützung des Frameworks erstreckt sich auf folgende Punkte:

- Laufzeitumgebung
 - # Lebenszyklus von Objekten
 - # Zugriffsschutz von Objekten und Methoden
 - # Sitzungsmanagement mit Zugangskontrolle
 - # Transaktionen
- Konfiguration von projektspezifischen, und applikationsspezifischen Variablen, um beispielsweise mehrsprachige Ausgabertexte oder anwenderspezifische Texte zu unterstützen
- Mittel zur Administration der Verteilung der Infrastruktur

- Teilautomatische Entwicklung von lightweight-Components durch Generatoren. Es werden drei Arten von Komponenten unterschieden:
 - # für transiente Objekte : Dienstobjekte
 - # für persistente Objekte:
 - Basisobjekte(BO) für gekapselte Objekte der Fachverfahrensdatenbestände
 - Geschäftsobjekte (abstrahierte, aggregierte oder angereicherte Basisobjekte)
- Anbindung vorhandener EDV-Datenhaltungssystemen durch Delegation an DBMS-spezifische Adapterklassen

Aufgabe des Frameworks ist es, einen Rahmen für die einheitliche Einbindung der Datenbestände zu schaffen, auf die über Dienstobjekte des zu schaffenden IS zugegriffen wird. Dabei werden nicht nur die Schnittstellen definiert, die jeder Dienst anzubieten hat, sondern es wird eine einheitliche Programmstruktur definiert, die jeder Server benutzt und in die nur noch die Teile eingefügt werden müssen, die für einen Datenbestand spezifisch sind.

4.3 Analyse alternativer Komponentenmodelle in der Literatur

Mittlerweile laufen viele Arbeiten, die sich mit der Entwicklung von Modellen für die Konstruktion und Integration Server-seitiger Komponenten für verteilte Infrastrukturen beschäftigen. Die bekanntesten sind:

- das Enterprise JavaBeans™ Framework (Hapner98) und
- das CORBA Components Model der OMG™ (OMGCC98). In der Fachpresse werden diese Components zum Ärger der OMG häufig mit CORBA Beans bezeichnet, die diesen Begriff jedoch bewußt vermieden hat, weil sie an einigen Stellen als übergeordneter Standard über die Konzepte von EJB hinaus geht.

Der Begriff der *light-Components* wird in Zusammenhang mit *Core* verwendet, weil sich die Komponenten des eigenen Frameworks gegenwärtig in einigen Punkten von dem in der Literatur verwendeten Komponentenbegriff unterscheiden.

Interessant sind allerdings zunächst die Gemeinsamkeiten in der Kernfunktionalität aller drei Ansätzen, die den eigentlichen Komponentenbegriff und die verschiedenen unterstützten Komponentenarten betreffen. Danach gibt es in allen Ansätzen eine grundsätzliche Unterscheidung zwischen transienten und persistenten Komponenten. Eine weitere wichtige Gemeinsamkeit ist die Bereitstellung von Retrievalfunktionalität über die Home-Schnittstelle persistenter Komponenten.

Außerdem wird in allen Frameworks eine Laufzeitumgebung benötigt, die zur Verwaltung der Komponenten in einer verteilten Infrastruktur dient.

Auch die Unterstützung unterschiedlicher Datenhaltungen findet sich in den anderen Frameworks. Dazu wird anstelle des Adapterbegriffs von *Core* der Container-Begriff verwendet.

Unterschiede finden sich in allen Frameworks in der Benennung von Schnittstellen und in Details der Erzeugung von Objekten. Der wohl wesentlichste Unterschied zwischen dem *Core*-Framework und den CORBA Components ist durch die Tatsache begründet, daß viele der Einschränkungen, denen das eigene Framework unterliegt, aus der ausschließlichen Verwendung der gegenwärtig verfügbaren CORBA 2.0 Version resultieren. D. h. dieser pragmatische Ansatz für Komponenten ist ein Weg, auch unter CORBA 2.0 bereits Komponenten-Konzepte zu realisieren, da noch einige Zeit ins Land gehen wird bis zuverlässige CORBA 3.0 oder vollständige EJB-Implementierungen verfügbar sind.

Da das Framework *Core* bewußt aus pragmatischen Gesichtspunkten unter diesen Einschränkungen entwickelt wurde, sind folgende Eigenschaften nicht realisiert:

- **Transaktionsbehandlung**
Das Transaktionsmodell von *Core* unterstützt zwar das Erzeugen von Transaktionen und Subtransaktionen durch die Clientseite, aber es unterstützt keine zwischen Komponenten verteilten Transaktionen wie der Java Transaction Service (JTS) von EJB oder der Object Transaction Service (OTS) von der OMG. Demzufolge werden in *Core* auch keine Notifikationsmechanismen zur Transaktionssynchronisation unterstützt, die in EJB eine wichtige Rolle spielen. Es gibt aber sehr viele Einsatzbereiche, wo zwar Komponenten durchaus benötigt werden, aber keinerlei Bedarf für verteilte Transaktionen besteht.
- **Event-Interface**
Es wird keine Eventbehandlung unterstützt. Da beispielsweise im VeZuD-Projekt bis auf das Repository vorläufig sowieso nur lesend per CORBA auf die Datenbestände zugegriffen wird, reicht ein einfacher Methodenaufruf-basierter Event Service, wie er zunächst in EJB oder CORBA 3.0 vorgesehen ist, eh nicht aus, um Events aus den Originaldatenbeständen zu fangen: Die Methodenaufrufe, die schreiben werden aufgrund der geforderten Autonomie der Fachverfahren eben gerade nicht über CORBA transportiert, sondern noch in den autonomen Fachverfahren durchgeführt. Hier wären komplexere Mechanismen wie ereignisgetriebene Monitoringdienste für heterogene Informationsquellen wie sie in (Koschel 1999) beschrieben sind vonnöten.
- **Noch keine Möglichkeit der dynamischen Konfiguration mittels Konnektoren, da das Receptacle und das Connectivity-Interface nicht angeboten werden.** Diese Schnittstellen ermöglichen das Herstellen von dynamischen Relationen. In vielen Fällen reichen bereits statische Relationsbehandlungen, so daß dann auch auf das relativ komplizierte Interface verzichtet werden kann. In den anderen Fällen muß natürlich auf die Komplexität von CORBA 3.0 zurückgegriffen werden oder es müssen eigene Mittel ergänzt werden.

Der wichtigste Vorteil von *Core* ist, daß dieses Framework eine Arbeit mit Komponenten auch ohne CORBA-IDL-Erweiterungen und CORBA Core-Änderungen ermöglicht und daß die Anforderungen an die Schnittstellen von Komponenten weniger komplex und dadurch leichter beherrschbar sind.

5 Zusammenfassung und Ausblick

Die Entwicklung des Frameworks war ein wichtiger Schritt im Zusammenhang mit dem modellgetriebenen Entwicklungs- und Integrationsprozeß für Datenbestände von Fachverfahren in ein verteiltes IS.

Der wohl größte Nutzen dieses Ansatzes liegt in einer enormen Verkürzung des Entwicklungsprozesses, da die sonst häufig sehr langwierige Designphase auf ein Minimum reduziert wird. Außerdem bieten diese erweiterten Modelle ein beiderseits noch verständliches Kommunikationsmedium, zwischen den Entwicklern der autonomen Fachverfahrenssysteme und dem VeZuDa-Entwicklungsteam einerseits, aber auch zwischen Entwicklern und externen Qualitätsgutachtern nebst Auftraggeber andererseits.

Auch die Auswirkungen auf die Softwarequalität sind nicht zu unterschätzen, da dadurch neben der einheitlichen Code-Strukturierung über die verschiedenen Integrationsprojekte hinweg auch Wege zu einer einheitlichen generierten Dokumentation verfügbar sind.

Der Ansatz verschiedenartige Datenhaltungssysteme über einen Baukasten von verschiedenen Generatoren zu berücksichtigen ermöglicht einen höheren Grad an Wiederverwendung von Code. Sicher muß für die Zukunft ein Konzept zur Migration zu standardisierten Komponentenmodellen erarbeitet werden. Da aber gegenwärtig gerade im Lager der OMG heftige Diskussionen darüber laufen, ob sich CORBA 3.0 mit seiner enormen Komplexität überhaupt durchsetzen kann, scheinen die eigenen pragmatischen Arbeiten ein Weg in richtige Richtung.

Die Integration von relationalen Datenbeständen verlief im Rahmen der Prototypentwicklung relativ problemlos. Für "exotische" ISAM-Datenbestände sind zukünftig komplizierte Datenexporte in relationale oder objektorientierte Datenmodelle vorgesehen.

Modellgetriebene Generierung vs. dynamische Datenkapselerzeugung zur Laufzeit

In diesem Papier wurde der generative Weg zum Erschließen neuer Datenbestände beschrieben: Basierend auf den technischen Zusatzinformationen des Design-Modells (technischen Metainformationen) wird der Programmcode für Datenkapseln generiert.

Eine andere Möglichkeit ist die Bereitstellung von generischen Datenkapseln. Diese werden erst im laufenden System mittels Metainformationen konfiguriert und stellen dann eine konkrete Datenkapsel zur Verfügung. Der Vorteil ist, daß keine Generierung und Übersetzung von Programmcode notwendig ist und schnell neue Datenquellen erschlossen werden können. Damit wird ein serverseitiges Rapid-Prototyping unterstützt.

Der Nachteil besteht in der schlechteren Performance. Nach der Prototypphase kann aber eine vollständige Generierung erfolgen. Die Schnittstellen ändern sich dabei nicht.

Erweiterungen für fachliche Integration und Föderation von Metadaten

Neben der eigentlichen Integration der Datenbestände, die mit diesem Ansatz im Rahmen eines ersten Prototyps mit mehreren Beständen erfolgreich getestet wurde, wird in einer neuen Projektphase die Erweiterung des modellgetriebenen Ansatzes notwendig, um künftig auch die Integration der fachlichen und organisatorischen Metadaten flexibel und modellgetrieben änderbar unterstützen zu können.

Für die fachlichen Metadaten ist ein Konzept vonnöten, daß es ermöglicht zusätzlich zu generell erhobenen Metadaten (statisch festgelegte Struktur), die Struktur von durch ein Teilprojekt bereitgestellten Metadaten dynamisch erweiterbar zu halten. Kern des anvisierten Lösungsansatzes sind hier Strukturen, die mit der Extensible Markup Language (XML) beschreiben werden. Hier gibt es mittlerweile viele Einsatzgebiete und einen aufstrebenden Markt für XML-basierte Werkzeuge (XMLReferenz 1998).

Fachliche Metadaten können sowohl einzelne Verfahren spezifisch beschreiben, als auch das Fachverfahren in Bezug zu anderen allgemeineren Metadatensystemen einordnen, die im Kontext eines IS relevant sind wie zum Beispiel geographische Raumbezugssysteme, Thesauri, Gesetzestexte etc. In diesem Kontext sind Arbeiten zur Föderation von Metainformationssystemen zu beachten (Kramer 1998, Busse 1998).

Um künftig auch mit bereits existierenden Metadatenbeständen der Fachverfahren besser kommunizieren zu können, werden Ansätze zum Import und Export von Metadaten mittels des zukünftigen von der OMG bereits standardisierten XML-basierten Metadata Interchangeformat (XMI) untersucht. Dies setzt voraus, daß das Repository dann MOF-basiert beschrieben werden kann. Da das Repository derzeit aus UML-Modellen generiert wird, sollte einer Beschreibung die konform zum Meta Object Facility (MOF) der OMGTM – einem Standard zur Metamodellierung – prinzipiell nichts im Wege stehen. Momentan gibt es nur leider wenig stabile Werkzeuge, die MOF unterstützen, und zugleich über CORBA-IDL's angesprochen werden können. Der Schwerpunkt der zukünftigen Arbeiten wird in dieser Richtung liegen.

Literatur

- Alhir, S. (1998): UML in a Nutshell, O'Reilly & Associates, Inc. September 1998; ISBN 1-56592-448-7.
- Booch, G.; Rumbaugh, J. and Jacobson, I. (1998): Unified Modelling Language User Guide, Addison-Wesley Object Technology Series 1998.
- Freitag, U.; Kutsche, R.; Busse, S. (1998): Internet-Technologie als Basis für föderierte Online- Umweltinformationsdienste am Beispiel des LUIS-Brandenburg, Tagungsband "Hypermedia im Umweltschutz" 1.Workshop, Ulm 1998, Hrsg. von Rieckert, W.-F.; Tochtermann, K.- Marburg : Metropolis, 1998.
- Freitag, U.; Schreiter, H.-P. (1999): Objekt- und Webtechnologien als Mittel zur Schaffung von modernen, verteilten individualisierten Online-Umweltinformationssystemen, In: "Heterogene, aktive Umweltdatenbanken": Workshop Vilm 1998.Hrsg. von R. Kramer und F. Hosenfeld – Marburg: Metropolis-Verl., 1999; (Praxis der Umwelt-Informatik; Bd.8); ISBN 3-89518-215-X.
- Gamma, E.; Helm, R.; Johnson R.; Vlissides, J. (1994): Design Patterns – Elements of reusable object-oriented software; Addison-Wesley Professional Computing Series 1994; ISBN: 0-201-63361-2.
- Hapner, M.; Matena, V. (1998): Enterprise JavaBeans™ – Server Component Model for Java™, Specification Version 1.1 , Sun Microsystems, Inc.
- Ralph E. Johnson (1992), Documenting Frameworks using Patterns, In Proceedings OOPSLA '92, ACM SIGPLAN Notices, pages 63-76, October 1992, Published as Proceedings OOPSLA '92, ACM SIGPLAN Notices, volume 27, number 10.
- Koschel, A.; Kramer, R. (1999): Konfigurierbare ereignisgetriebene Dienste für verteilte Umweltinformationssysteme, In: Heterogene, aktive Umweltdatenbanken: Workshop Vilm 1998.Hrsg. von R. Kramer und F. Hosenfeld – Marburg: Metropolis-Verl., 1999; (Praxis der Umwelt-Informatik; Bd.8); ISBN 3-89518-215-X.
- Neumann, H. (1998): Objektorientierte Softwareentwicklung mit der Unified Modeling Language (UML); München; Wien; Hanser Verlag 1998; ISBN: 3-446-18879-7.
- Kramer, R.; Nikolai, R. (1998): Technische und semantische Aspekte der losen Integration heterogener und autonomer Thesauri; In: Heterogene, aktive Umweltdatenbanken: Workshop Vilm 1998.Hrsg. von R. Kramer u. F. Hosenfeld – Marburg: Metropolis-Verlag, 1999; (Praxis der Umwelt-Informatik; Bd.8); ISBN 3-89518-215-X.
- OMGCC (1998): CORBA Components – Joint Revised Submission; In: OMG TC Document orbos/98-12-02, Supported by: Sun Microsystem, Inc.
- OMGMOF (1998): CORBA Meta Objects Facility – Specification Joint Revised Submission; In: OMG TC Document ad/97-08-14, Supported by: Sun Microsystem, Inc.
- Schulzki-Haddouti, C. (1999): XML – Eine Metasprache knüpft das Netz neu; <http://www.firstsurf.de/xml.htm>
- VeZuDa (1999): Spezifikation Kernsystem; interne Projektdokumentation; Condat gmbh ; siehe email: uf@condat.de;
- XML (1998): Extensible Markup Language (XML) 1.0; W3C Recommendation 10-February-1998; <http://www.w3.org/TR/1998/REC-xml-19980210#sec-bibliography>

Ein Internet-basiertes Umweltdokumentsystem für kleine Nutzergruppen

Klaus Tochtermann, Andreas Kusssmaul und Katrin Pursche¹

Zusammenfassung

Die Entwicklungen der vergangenen Jahre erlauben in Zukunft auch Nicht-Experten kleinere Umweltdokumentbestände aufzubauen, zu warten und inhaltlich zu erschließen. Vor diesem Hintergrund wird sich die Nutzung von Internettechnologien in Zukunft weg von großen, nur von Experten zu beherrschbaren Dokumentbeständen und zugehörigen Systemen, hin zu kleineren und auf spezielle Bedürfnisse der Nutzer zugeschnittene Dokumentbeständen und Systemen bewegen. Dieser Beitrag stellt ein solches System und mögliche technische Realisierungen zur Anbindung von (Umwelt-)Datenbanken über das Internet vor. Die verschiedenen Alternativen zur technischen Realisierung werden in dem Beitrag diskutiert und bewertet.

1 Einleitung

Im vergangenen Jahr endete die erste Phase der amerikanischen Initiative "Digital Libraries". Während in dieser Phase sechs Universitäten mit Arbeiten zu Grundlagenproblemen, wie Verfügbarmachung riesiger Dokumentbestände, gefördert wurden, werden in der derzeit vorbereiteten Phase 2 mehr Schwerpunkte auf Bereiche wie Erzeugung von Inhalten, Zugriff auf Inhalte sowie Nutzen und Nutzbarkeit von Inhalten gelegt.

Die im Special Issue "Digital Libraries" der Zeitschrift IEEE Computer (Computer 1999) dargelegten Erfahrungen aus der Phase 1 zeigen, daß für das 21. Jahrhundert erwartet wird, daß sich das Arbeiten mit Wissen und Informationen dramatisch ändern wird. In den vergangenen Jahren wurden riesige Datenbestände von professionellen Fachkräften katalogisiert, automatisch indiziert und anschließend der Öffentlichkeit über das Internet zugänglich gemacht (Schatz 1999). Aufgrund dieser Vorgehensweise konnte sich die Öffentlichkeit in den letzten 10 Jahren mit Internettechnologien vertraut machen. Inzwischen entsteht mehr und mehr das Bedürfnis, nicht nur recherchierend mit Dokumentbeständen zu arbeiten, sondern diese auch verändern zu

¹ K. Tochtermann, A. Kusssmaul, K. Pursche
Forschungsinstitut für anwendungsorientierte Wissensverarbeitung, Postfach 2060,
D-89081 Ulm; E-Mail: {tochterm | kusssmaul | pursche}@faw.uni-ulm.de.

dürfen. Die Möglichkeit der Veränderung spielt sich auf mindestens zwei Stufen ab. In der ersten Stufe werden von Systemen Möglichkeiten angeboten, die es erlauben, die Systeme an persönliche Anforderungen anzupassen. Ein solches System ist etwa PADDLE (Personal Adaptable Digital Library Environment) (Hicks et al., 1999). Damit nehmen Nutzer eine aktive Rolle in Form von Autoren ein. Um sie in ihrer Funktion als Autor mit der damit verbundenen Komplexität nicht zu überfordern, ist es sinnvoll, daß zunächst Systeme entwickelt werden, mit deren Hilfe von Nicht-Experten kleinere organisationsinterne Datenbestände aufgebaut, erschlossen und nutzbar gemacht werden können. Mit PADDLE ist es bspw. möglich, erhobene Metainformation zu verändern und zu personalisieren, ohne daß der Originaldatenbestand davon betroffen ist. In der zweiten Stufe werden Systeme ihren Nutzern erlauben, auch die Dokumentbestände an sich zu verändern, und zwar unabhängig davon, wer der Eigentümer dieser Dokumentbestände ist. Diskussionen über diese zweite Stufe werden derzeit unter dem Schlagwort "cross-publishing" geführt. Zukünftige Entwicklungen im Bereich Bereitstellung von Online-Informationen müssen daher diesem Trend Rechnung tragen.

Vor diesem Hintergrund wurde am FAW im Rahmen einer Diplomarbeit ein Internet-basiertes Dokumentenarchiv erstellt, das kleinere Nutzergruppen (ca. 10 Personen) in Aufbau, der Erschließung und der Nutzung von Umweltdokumenten unterstützt. Es wurde bewußt Wert darauf gelegt, zunächst nur einer überschaubaren Zahl von Nutzern Autorenrechte einzuräumen. Damit ist es möglich, die entstehenden Phänomene, aber auch Schwierigkeiten besser zu erfassen und zu kontrollieren. Auf der Basis der gesammelten Erfahrungen können zu einem späteren Zeitpunkt Anforderungen abgeleitet werden, die zu berücksichtigen sind, um einer beliebig großen Zahl von Nutzern Autorenrechte für ein Dokumentenarchiv einzuräumen. Die Bedeutung dieser Arbeit wird dadurch unterstrichen, daß das Ministerium für Umwelt und Verkehr Baden-Württemberg das FAW mit der Erstellung eines solchen Systems für die Verwaltung von Umweltdokumenten in Behörden beauftragt hat (Henning et al. 1999).

Der Beitrag stellt nun die technischen Aspekte der Realisierungsmöglichkeiten in den Vordergrund und ist wie folgt aufgebaut: Abschnitt 2 stellt das System an sich mit prinzipiellen Arbeitsabläufen. Abschnitt 3 diskutiert die technische Realisierung, bevor in Abschnitt 4 der Beitrag mit einer Bewertung und einem Ausblick beendet wird.

2 Anforderungen an das System und prinzipieller Arbeitsablauf

Mit dem System sollen Nutzer bei dem Aufbau, der Erschließung und der Nutzung von ihren täglichen Arbeitsdokumenten unterstützt werden. Darunter sind Dokumente zu verstehen, die schnell zugreifbar und recherchierbar sein müssen, ohne daß zuvor umfangreiche Metadaten (Daten über Daten) erhoben werden müssen. Versionsverwaltung sowie andere typische Eigenschaften von (Internet-basierten) Dokument-

verwaltungssystemen, wie BSCW (BSCW 1999) oder AltaVista-Forum (AltaVista 1999), wurden in dieser Phase nicht berücksichtigt. Bei der Zielgruppe der Nutzer wurde ferner davon ausgegangen, daß kein umfangreiches technisches Know-how vorliegt. Vor diesem Hintergrund ist es wichtig, die Komplexität des Systems hinter einfach zu bedienenden Benutzungsoberflächen zu verbergen.

Für das Dokumentenarchiv wurden drei Komponenten vorgesehen (vgl. Abb. 3), welche jeweils auf einem separaten Rechner in einem Intranet oder im Internet plziert werden können.

Eine Komponente bearbeitet Anfragen an eine "Accounts"-Datenbank. In dieser Datenbank werden Informationen über die Benutzer festgehalten, wie zum Beispiel der Status des Benutzers (Gast, Autor oder Administrator) oder seine Emailadresse. Die Komponente ist erforderlich, um prinzipiell nachhalten zu können, welcher Nutzer welche Änderungen am Dokumentbestand vorgenommen hat. Da nur ein kleiner Nutzerkreis angesprochen werden soll, wurde kein besonderer Wert auf Sicherheitsaspekte, etwa bei der Vergabe und Verwaltung von Paßwörtern, gelegt.

Eine weitere, sehr mächtige Komponente beinhaltet einen WWW-Server mit einem Index Server und einem FTP-Dienst sowie alle Dokumente, welche über das Dokumentenarchiv verwaltet werden und vom Index Server indiziert werden. Neben den rechenintensiven Suchanfragen an den Index Server über Netzwerkverbindungen mit dem HTTP-Protokoll werden auch FTP-Sitzungen zur Dateiübertragung abgehalten. Hierüber wird den Nutzern die Möglichkeit eingeräumt, in der Rolle eines Autors Dokumente zu verändern. Über die Dateitransfers mit FTP hinaus werden weitere Dateioperationen auf dem Dateisystem ständig zur Aktualisierung der Indizes des Index Servers ausgeführt.

Die dritte Komponente verwaltet eine Datenbank. Diese Datenbank ist der Umweltthesaurus GEMET (General European Multilingual Environmental Thesaurus) des Umweltbundesamtes und findet bei der Suche nach synonymen Begriffen, Unter- und Oberbegriffen sowie Schwesterbegriffen Anwendung. Die Komponente ist im Vergleich zu den anderen nicht erforderlich, um Nutzer in ihrer Rolle als Autor zu unterstützen. Allerdings erleichtert sie die inhaltliche Erschließung von Dokumenten, da anhand des kontrollierten Vokabulars eine systematische Verschlagwortung möglich ist. Der prinzipielle Arbeitsablauf mit dem System wird in den folgenden Abschnitten beschrieben (mehr Informationen hierzu in (Henning et. al 1999)).

2.1 Autorenkomponente

Die Oberfläche der Autorenkomponente wird in Abbildung 1 dargestellt. Die Autorenkomponente wird im Prototypen aus der Rechercheoberfläche heraus gestartet. Im linken Teil des Fensters wählt der Autor das Verzeichnis aus, in welches er das von ihm bearbeitete Dokument einstellen möchte. Es handelt sich hier um die physischen Verzeichnisse der Server, die die Dokumente des Dokumentbestandes enthalten.

Im rechten Teil des Fensters werden die lokalen Verzeichnisse des Autorenrechners angezeigt. Der Autor sucht hier nach dem Dokument, welches in den Dokumentbestand eingestellt werden soll. Dieses Dokument wird dann auf das ausgewählte Verzeichnis eines Servers kopiert. Weiterhin kann der Autor sowohl auf dem Server als auch auf seinem lokalen Rechner neue Verzeichnisse erstellen oder Dateien löschen, wenn er die notwendigen Zugriffsrechte besitzt.

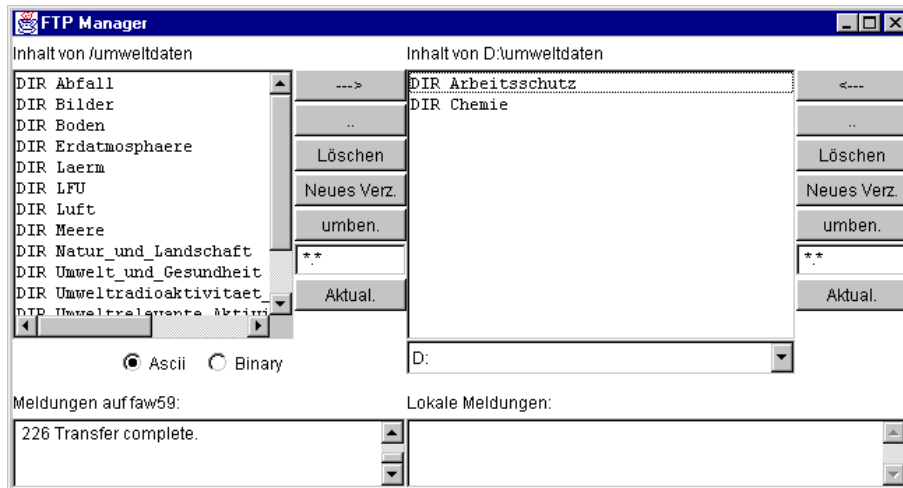


Abbildung 1
Oberfläche des Autorenkomponente

2.2 Recherchekomponente

Nach der Anmeldung erhält der Anwender die Rechercheoberfläche im WWW-Browser. Sie setzt sich aus dem Suchformular, der Einstellung für die zu recherchierenden Dokumentbestände sowie je nach Benutzerrechten den Startbuttons für die Autoren- und die Administratorenkomponente zusammen. Der Anwender legt nun fest, ob er seine Anfrage an eine Internetsuchmaschine weiterleiten will und auf welchem Dokumentbestand des Dokumentenarchivs die Suche ausgeführt werden soll. Mit der Wahl des Dokumentbestandes wird das Suchformular an die Metainformationen angepaßt, die diesem Bestand zugewiesen wurden. Der Anwender füllt nun das grau hinterlegte Suchformular aus (vgl. Abbildung 2; die Abbildung zeigt bereits die Weiterentwicklung DIWA des hier vorgestellten Systems und ist der Veröffentlichung (Henning et al. 1999) entnommen).

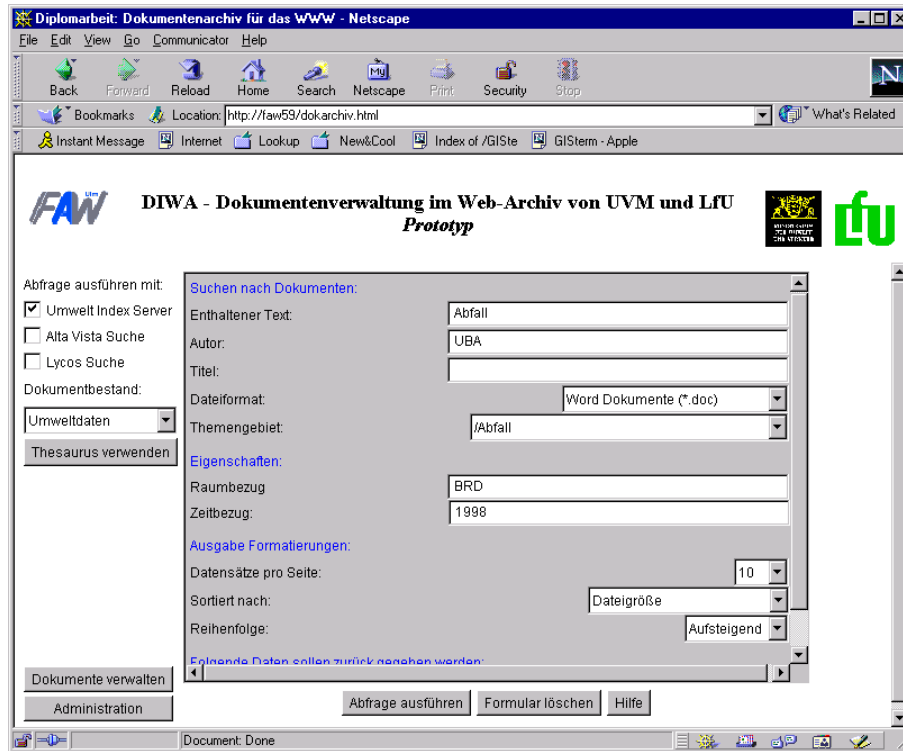


Abbildung 2
 Rechercheoberfläche des Prototypen

Dabei kann der Nutzer neben einer Volltextsuche mit logischen Verknüpfungen auch nach Dateieigenschaften (Metainformationen) suchen.

Zur Begriffsfindung für die Volltextsuche kann der Thesaurus verwendet werden. Die in ihm selektierten Begriffe werden direkt in die Volltextsuche übernommen. Es kann weiterhin das Dateiformat der zu suchenden Dokumente sowie ein Themenbereich festgelegt werden. Ebenfalls im Suchformular wird das Format der Ausgabe der Ergebnisseite eingestellt. Nach dem Start der Anfrage wird ein neues Fenster für die Ergebnisanzeige geöffnet. Über die angezeigte URL können nun die Dokumente zur Ansicht oder weiteren Verarbeitung auf den lokalen Rechner heruntergeladen werden.

Der zugrunde liegende Datenbestand, welcher von diesem System derzeit verwaltet wird und recherchiert werden kann, besteht aus über 1000 Berichtsdokumenten, die in dem vom Umweltbundesamt erstellten Umweltbericht „Daten zur Umwelt 1997“ Verwendung finden. Die Datenbank wird bis zum Jahr 2000 ein einzigartiges Archiv an Informationen über den Zustand der Umwelt in Deutschland enthalten, welche zur Umweltberichterstattung herangezogen werden.

Um das System für Nicht-Experten einfach nutzbar zu machen, wurde eine überschaubare Anzahl an Feldern für die inhaltliche Erschließung des Datenbestandes über Metadaten ausgewählt. Dieser Sachverhalt ermöglicht es, Benutzern nur die für ihre Aufgabe erforderlichen Metadaten zu erheben, so daß der gesamte Dokumentenbestand auf ihre Bedürfnisse zugeschnitten ist.

Dies ist ein wesentlicher Unterschied zu Systemen wie der Datenbank Grunddaten (DBG 1999) oder dem Umweltdatenkatalog (UDK 1999). Bei der Datenbank Grunddaten stehen bspw. etwa ca. 40 Felder für die inhaltliche Erschließung der eingestellten Umweltdokumente zur Verfügung. Viele davon sind jedoch nur für Experten in bestimmten Fachgebieten erforderlich. Erfahrungen zur Befüllung der Datenbank Grunddaten werden in diesem Tagungsband von (Dombeck et al. 1999) vorgestellt.

3 Technische Realisierung

Abbildung 3 stellt die Systemarchitektur sowie die Kommunikationswege zwischen den in Abschnitt 2 beschriebenen Systemenkomponenten dar. Um möglichst viel Komplexität nutzerfreundlich darzustellen, wurde für die Benutzungsschnittstelle ein Java Applet entwickelt. Dieses Java Applet verbirgt z. B. die Nutzung von FTP hinter intuitiv bedienbaren Funktionskomponenten. Die Kommunikation zwischen dem Java Applet und den beiden Datenbankservern erfolgt über Java Database Connectivity (JDBC) und Remote Method Invocation (RMI). RMI wird verwendet, da momentan für Microsoft Access Datenbanken noch keine JDBC-Treiber existieren, welche den Zugriff auf die Access-Datenbanken mit JDBC direkt über das Netz ermöglichen. Um mit JDBC auf Access Datenbanken zuzugreifen, benötigt man einen speziellen JDBC-Treiber (JDBC-ODBC-Bridge) von SUN (JDBC 1999), welcher JDBC-Befehle in ODBC-Befehle umwandelt und dann an den lokal installierten ODBC-Treiber für Access-Datenbanken weiterleitet. Ein RMI Server nimmt aus diesem Grund die Datenbankabfragen der Clients über das Netz entgegen, wandelt sie in JDBC-Befehle um und leitet diese über die JDBC-ODBC-Bridge an die lokale Datenbank weiter. Die Ergebnisse werden dann über RMI an die Klienten zurück gesendet.

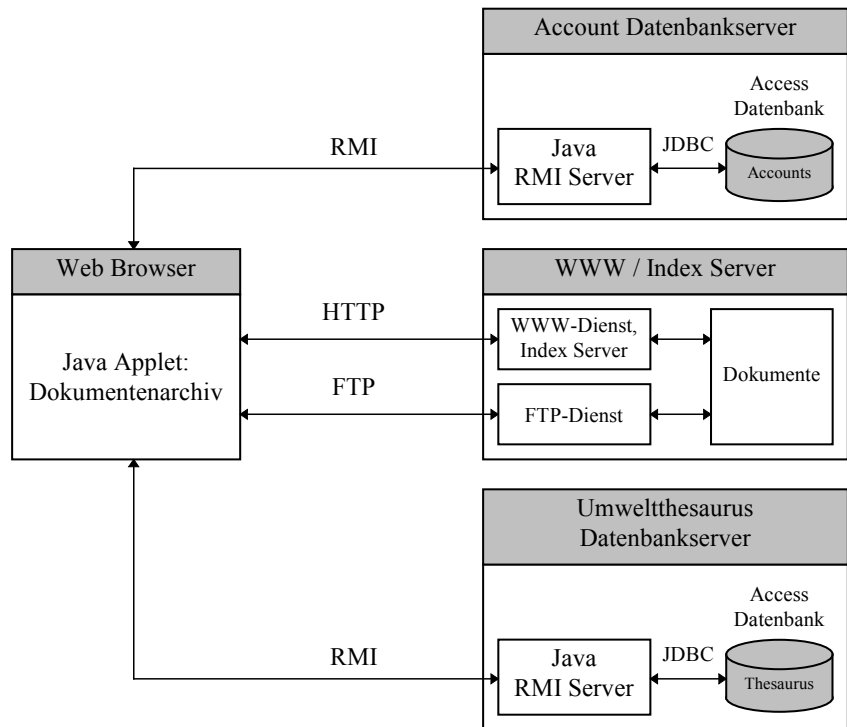


Abbildung 3
Systemarchitektur

3.1 Möglichkeiten zur technische Anbindung der Datenbanken

3.1.1 Java Database Connectivity (JDBC)

Der Begriff JDBC (Java Database Connectivity) bezeichnet ein von Sun entwickeltes Paket, das auf den Java Basisklassen aufsetzt und über die Bereitstellung von relationalen Datenbankobjekten sowie der entsprechenden Methoden den Zugriff aus Java Applets und Anwendungen auf beliebige Datenbanken ermöglicht. JDBC setzt auf dem X/OPEN SQL-Call-Level-Interface (CLI) auf und besitzt damit die gleiche Basis wie die ODBC-Schnittstelle.

Die API-Spezifikation für JDBC wurde von JavaSoft im März 1996 als vorläufige Version 0.50 vorgestellt. Sie liegt nunmehr in der Version 2.0 vor und wird im Rahmen des JDK 1.2 (Sun's Java Development Kit) ausgeliefert.

3.1.1.1 Einsatz von JDBC

Bei clientseitigem Zugriff auf eine Datenbank wird mit Java Applets unter Verwendung von JDBC eine direkte Verbindung zwischen dem Klienten und dem Datenbankmanagementsystem hergestellt. Die Verbindung erfolgt auf der Grundlage der von Sunsoft entwickelten JDBC-API (Java-Database-Connectivity), welche Java-Anwendungen und Java Applets den direkten Zugriff auf Datenbanksysteme ermöglicht.

Die Vorteile dieser Lösung liegen auf der Hand: Der Web-Server wird von intensiven Datenbankabfragen entlastet, die Programmierung über CGI oder eine spezifische API entfällt, und die Portabilität ist durch die Verwendung der Java-Schnittstelle JDBC gewährleistet.

3.1.1.2 Architektur von JDBC-Anwendungen

Ein Java Applet, das mit einem Browser vom Server auf den Klienten geladen wird, kann aufgrund der strengen Sicherheitsvorschriften für Applets nur mit dem Rechner in Verbindung treten, von dem es geladen wurde. Wenn das Applet direkt mit dem Datenbankserver kommunizieren soll, ist es daher notwendig, daß Datenbank und Web-Server auf der gleichen Maschine installiert werden. Man spricht in diesem Fall von einer "Two-Tier-Architektur" (vgl. Abbildung 4), die im wesentlichen der klassischen Client-Server-Architektur entspricht.

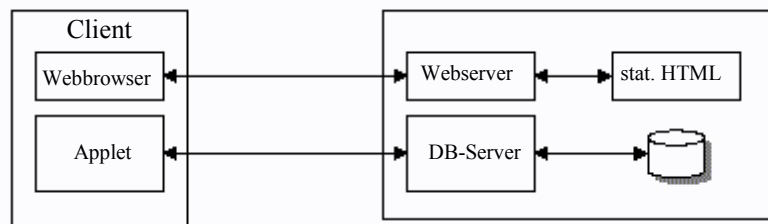


Abbildung 4
Two-Tier-Architektur

Der Klient lädt ein Java Applet vom Server, welches sich über ein Datenbankmanagementsystem (DBMS) spezifisches Protokoll (hier JDBC) mit der Datenbank verbindet, die auf der gleichen Maschine liegt wie der Web-Server.

Eine andere Möglichkeit, bei welcher der Web-Server und der Datenbankserver auf getrennten Maschinen installiert werden können, ist in Abbildung 5 dargestellt. Bei dieser sogenannten "Three-Tier-Architektur" wird ein Applet vom Web-Server geladen, das sich über ein DBMS-unabhängiges Protokoll mit einem Gateway (auf dem Web-Server) verbindet, welches dann als Datenbank-Klient die Abfrage für das

Applet übernimmt und die Ergebnisse anschließend an das Applet zurücksendet (ein Beispiel für eine solche Three-Tier-Architektur ist die Verbindung eines Java Applets mit einer Java Anwendung über die Remote Method Invocation (RMI)).

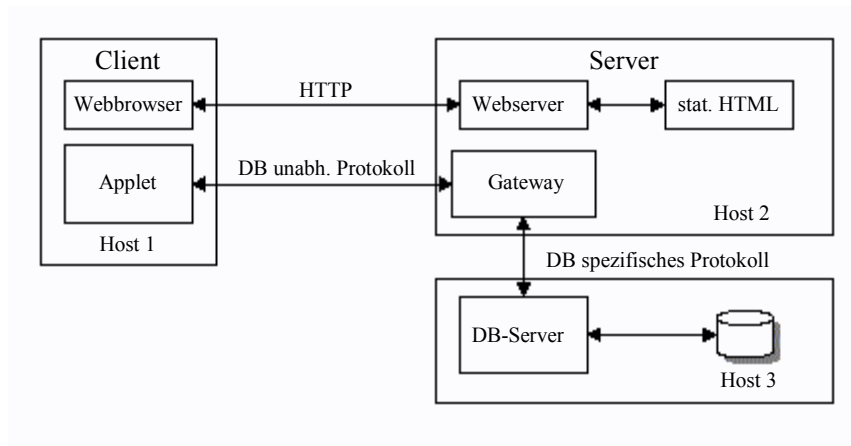


Abbildung 5
Three-Tier-Architektur

3.1.1.3 JDBC-Zugriffe über RMI

Eine Möglichkeit, die Three-Tier Architektur umzusetzen, besteht in der Verwendung der Remote Method Invocation (RMI). RMI ermöglicht beliebigen Java Applets und Anwendungen untereinander zu kommunizieren, um gegenseitig Funktionalitäten, Objekte und Methoden der einzelnen Java Anwendungen zu nutzen. Mit RMI ist es also auch möglich, ein komplexes System auf mehrere kleinere Systeme zu verteilen. Diese Möglichkeit existiert ab der Version 1.1 des JDK in Form des `java.rmi`-Packages.

Dieses Package stellt Java-Klassen mit den entsprechenden Methoden bereit, um vom Client-Applet aus mit Serverobjekten zu kommunizieren und Daten auszutauschen. In einer solchen Form kann die Three-Tier-Architektur folgendermaßen implementiert werden:

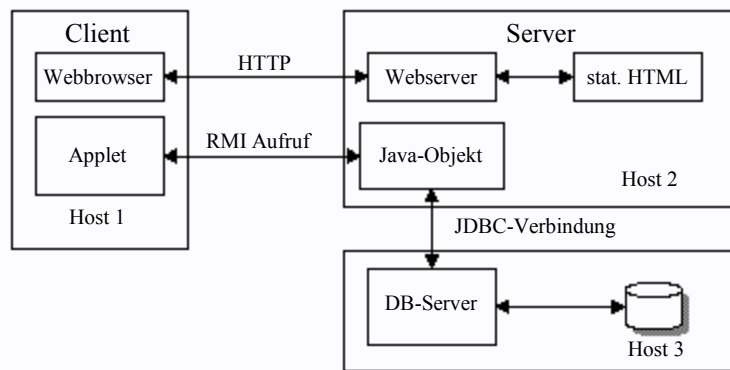


Abbildung 6
Three-Tier-Architektur mit RMI und JDBC

Der Klient lädt ein Java Applet vom Server, welches sich indirekt über RMI mit einem Serverobjekt verbindet, das seinerseits über JDBC die Verbindung mit einem ausgelagerten Datenbankserver herstellt.

3.1.2 Remote Method Invocation (RMI)

Das Package `java.rmi` stellt die Klassen und Methoden zur Realisierung einer verteilten Systemarchitektur mit Java bereit.

Allgemein geschieht das in der Form, daß ein Serverobjekt definiert wird, das die Methoden bereithält, die von den Klienten aufgerufen werden können. Zu diesem Serverobjekt muß ein Interface bereitgestellt werden, das der Klient vom Server lädt und das ihm die Schnittstelle zu den Serverkomponenten bereitstellt. Der Server selbst wird einmalig gestartet und wartet dann auf Anfragen der Klienten nach Ausführung seiner Methoden. Dies geschieht in der Form, daß der Klient das Serverobjekt ermittelt, und die Methode genauso aufruft, als wäre es ein lokales Objekt.

Die Weitergabe der Aufrufe zwischen dem Server und dem Klient geschieht über sogenannte Stubs und Skeletons, die stellvertretend für die jeweilige Gegenseite mit der aufrufenden Komponente kommunizieren.

RMI stellt somit eine geeignete Methode zur Realisierung verteilter Systemarchitekturen in einem Netzwerk zur Verfügung, sei es, um einem Applet den Zugriff auf externe Server (hier auf einen Datenbankserver) zu ermöglichen, oder aber spezielle Anwendungen (zum Beispiel umfangreiche Berechnungen) von einem eigens dafür eingerichteten Rechner durchführen zu lassen.

3.1.3 Bewertung der Architekturen

Aufwand

Die Two-Tier-Architektur verlangt weniger Aufwand bei der Implementierung und Wartung. Der gesamte Code wird auf dem Klient ausgeführt, es werden keine zusätzlichen Programme benötigt.

Performance

Ein Two-Tier Treiber kann häufig nicht die Möglichkeiten des Multithreading unter Java und dessen Vorteile ausnutzen, wenn der Klient nicht selbst multithreadfähig implementiert wurde. Threads sind für eine Steigerung der Performance sehr wichtig, da dadurch mehrere Datenbankzugriffe annähernd gleichzeitig und mit einer sicher eingehaltenen Reihenfolge ausgeführt werden können.

Die Two-Tier-Architektur erfordert eine Realisierung der JDBC Implementation explizit für jedes Applet. Dieser zusätzliche Code muß immer an den Klienten übermittelt werden und verursacht zusätzlichen Verbrauch von Systemressourcen und Overhead zu Lasten des Klienten.

Durch die Verwendung der Three-Tier-Architektur verringert sich die Komplexität des Client-Applets. Das Applet empfängt lediglich Daten und stellt diese auf der Oberfläche dar.

Realisierung

Die enge Bindung der JDBC Implementierung an die Benutzungsoberfläche bei der Two-Tier-Architektur birgt die Gefahr, daß diese beiden Komponenten oft nicht mehr sauber voneinander getrennt implementiert werden; die Wiederverwendbarkeit von einzelnen Modulen und Komponenten wird eingeschränkt.

Einen weiteren Nachteil der Two-Tier-Architektur stellen einige JDBC-Treiber dar, die nativen, plattformabhängigen Code beinhalten, welcher von den Web Browsern aus Sicherheitseinschränkungen nicht ausgeführt werden kann.

Sicherheit

Bei der Two-Tier-Architektur müssen das Applet und die Datenbank auf dem selben Rechner liegen, da das Applet aus Sicherheitsgründen nur auf die Maschine zugreifen darf, von der es geladen wurde. Alle Klienten, welche dieses Applet laden und ausführen, greifen direkt auf den Datenbankserver zu, wozu jeder Client ein verschlüsseltes Paßwort und einen verschlüsselten Benutzernamen benötigt. Dies erschwert die Konzeption eines Sicherheitskonzepts.

In einer Three-Tier-Architektur werden alle Informationen des Datenbankservers innerhalb der Mittelschicht (Gateway) von dem Applet versteckt gehalten. Dieses Gateway bildet eine Art Schutzschild für den direkten Zugriff von Klienten auf das Datenbankmanagementsystem.

4 Bewertung und Ausblick

Für kleine Nutzergruppen, die in einem überschaubaren Maß als Autoren mit einem Internet-basierten Informationssystem arbeiten, ist die Two-Tier-Architektur ausreichend. So kann eine reduzierte Performance akzeptiert werden, da nicht mit allzu vielen gleichzeitigen Datenbankzugriffen gerechnet werden muß. Auch die explizite und ressourcenaufwendige Realisierung einer JDBC-Verbindung für jedes Applet ist vor diesem Hintergrund akzeptabel. Bei kleinen Nutzergruppen kann zudem überschaut werden, in welcher Systemumgebung welche Browser auf dem Klienten installiert sind. Gegebenenfalls können diese Browser bzw. die darauf laufenden Applets so eingestellt werden, daß der z. T. erforderliche plattformabhängige Code keinen echten Nachteil der Two-Tier-Architektur darstellt. Schließlich müssen bei kleinen Nutzergruppen keine so hohen Sicherheitsanforderungen angesetzt werden, wie dies bei Systemen mit einer großen Nutzerzahl der Fall ist. Die Tatsache, daß die Two-Tier-Architektur für jeden Klienten ein Paßwort und Benutzernamen benötigt, stellt also keine echte Sicherheitslücke dar. Sollen jedoch Systeme auch von einer großen Zahl von Nutzern verwendet werden, bietet die Three-Tier-Architektur entscheidende Vorteile hinsichtlich Performance und Sicherheit.

Um Internet-basierte Informationssysteme für verschiedene Nutzergruppen effektiv nutzbar zu machen, ist es erforderlich, daß für ein und denselben Datenbestand nutzerspezifisch Metadaten verfügbar sind. Es ist geplant, diesen Aspekt in zukünftige Weiterentwicklungen des Systems zu integrieren. Ein weiterer wichtiger Aspekt ist die Möglichkeit zur Verwaltung mehrerer unabhängiger, nutzerspezifischer Dokumentbestände sowie der zugehörigen Metadaten auf einem Index Server. Das in diesem Beitrag beschriebene System kann nur für kleine Nutzergruppen effektiv eingesetzt werden. Ein Grund hierfür ist, daß der verfügbare Dokumentbestand in keiner Form strukturiert ist. Strukturierungen der Dokumente, die sich z. B. an organisatorischen Einheiten wie Abteilungen orientieren, sind jedoch notwendige Voraussetzung um große Dokumentbestände aufzubauen, insbesondere wenn diese von zahlreichen verschiedenen Nutzern erweitert werden.

Derzeit sind nur Microsoft Office Dokumente und HTML-Dokumente recherchierbar. Für die Weiterentwicklung sind sowohl XML als auch PDF als Dokumentformate zu berücksichtigen.

Eine weitere technische Herausforderung besteht in der Integration von Dokumentbeständen über technische Systemgrenzen hinaus. Diese Aspekte werden im Rahmen von dem vom Ministerium für Umwelt und Verkehr Baden-Württemberg initiierten Projekt DIWA (Dokumentenverwaltung im Web-Archiv von UVM und LfU) (Henning et al., 1999) weiter verfolgt und umgesetzt.

4.1 Literatur

- AltaVista 1999: <http://altavista.afterfive.com/forum.htm>
- BSCW 1999: <http://bscw.gmd.de/index.html>
- Computer 1999: IEEE Computer, Special Issue "Digital Libraries" Vol. 32, No. 2, 1999.
- Dombeck, T., Tochtermann, K. 1999: Praktische Erfahrungen bei der Datenerschließung für eine Umweltdatenbank des Umweltbundesamtes. Veröffentlicht in diesem Tagungsband.
- DGB, 1999: Dokumentation der Datenbank Grunddaten Version 3.1. Dr. Lippke und Dr. Wagner GmbH, Berlin.
- Hartwik, J., Einführung in JDBC: <http://home.fhtw-berlin.de/~s0109297/jdbc/jdbc.html>.
- JDBC 1999: SUN, The JDBC(tm) Universal Data Access API:
<http://www.javasoft.com/products/jdbc/index.html>.
- Hicks, D., Tochtermann, K., Kussmaul, A., Neils, S. 1999: Customization in Environmental Information Systems, *Tagungsband 13. Symposium Umweltinformatik*, Magdeburg, Metropolis-Verlag Marburg, 1999.
- Schatz, B., Hsinchun, Ch. 1999: Digital Libraries: Technological Advances and Social Impacts; in: IEEE Computer 1999, p. 45-50.
- Schmidt, A., Remote Method Invocation
<http://www.tzi.org/~ansu/papers/rmi/PowerPoint/index.htm>.
- SUN, Java Remote Method Invocation: <http://www.javasoft.com/products/jdk/rmi/index.html>.
- UDK, 1999: http://www.uvm.baden-wuerttemberg.de/www-udk/RPT_DATA/about_d.html.
- Seggelke, J. und Mohaupt-Jahr, B.: „Der Verweis- und Kommunikationsservice des Umweltbundesamts – Ein Modellfall für das Umwelt-Intranet“. In: W. Geiger, A. Jaeschke, D. Rentz, E. Simon, T. Spengler, L. Zilliox und T. Zundel (Hrsg.): *Umweltinformatik '97*. 11. Internationales Symposium, Straßburg, September 1997, Tagungsband. Metropolis-Verlag, Marburg, 1997.