Texte

Texte
11
06
ISSN
1862-4804

Umweltdatenbanken und Netzwerke

Umwelt
Bundes
Amt
Für Mensch und Umwelt

Texte
11
06
ISSN
1862-4804

Umweltdatenbanken und Netzwerke

Workshop des Arbeitskreises
„Umweltdatenbanken“ der Fachgruppe
„Informatik im Umweltschutz“,
veranstaltet in Zusammenarbeit mit dem
Umweltbundesamt am
6. und 7. Juni 2005 in Hannover

Diese Publikation ist als Download
unter der Adresse www.umweltbundesamt.de
sowie als Printfassung verfügbar.

Herausgeber: Umweltbundesamt
Postfach 14 06
06813 Dessau
Tel.: 0340/2103-0
Telefax: 0340/2103 2285
Internet: http://www.umweltbundesamt.de

Redaktion: Fachgebiet IV 2.1
Gerlinde Knetsch
Fachgebiet IV 2.4
Angela Lehmann

Dessau, März 2006

Vorwort

Für die Umweltinformatik stehen Infrastrukturen und Technologien zum Aufbau von Netzwerken zunehmend im Focus vieler Anwendungsentwicklungen. Die Nutzung und der Einsatz von Komponenten der Internettechnologie für Fachanwendungen im Umweltbereich bestimmen hierbei die konzeptionelle Vorgehensweise. Im Ergebnis entwickeln sich äußerst interessante Ansätze von Softwarearchitektur, die unter dem Blickwinkel eines Service-Angebotes gesehen werden können. Dieses Vorgehen wird dadurch motiviert mit einem breiten Kreis von Nutzern und Anwendern zu kommunizieren und Daten und Informationen über Netzwerkstrukturen auf Basis der Webtechnologie auszutauschen. Schlagwörter wie Interoperabilität, offene Infrastruktur und standardisierte Schnittstellen sind dabei Bausteine einer modernen, flexiblen und service-orientierten Architektur.

In diesem fachlichen Kontext des weiteren Ausbaus von Netzwerken über Behörden- und Institutions- sowie Fachgrenzen hinweg standen die Themen des Workshops des Arbeitskreises *"Umweltdatenbanken"* des Fachausschusses *"Informatik im Umweltschutz".* Von der Koordinierungsstelle UDK/ gein® (KUG) mit Sitz im Niedersächsischen Umweltministerium Hannover organisiert, fand dieses alljährlich stattfindende Treffen am 6. und 7. Juni 2005 statt. Mehr als 50 interessierte Fachexperten aus Behörden, Instituten und Forschungseinrichtungen Deutschlands und Österreichs nahmen daran teil.

Von den 20 vorgestellten Fachanwendungen befassten sich mehr als die Hälfte der Präsentationen mit web-basierten Lösungen. Oft beinhaltete der Titel der eingereichten Vorträge bereits den Hinweis auf die Integration in nationale sowie internationale Netzwerke. Der Einführungsvortrag der Koordinierungsstelle UDK/ gein® (KUG) bildete im Sinne einer netzwerkartigen Knotenbildung einen Rahmen bezüglich flexibler und modularer Systeminfrastruktur für die Weiterentwicklung von *gein® 2.0*. Der Einsatz und die Anwendung von intelligenten *Suchmaschinen*, die Einbindung verschiedener *Thesauri* zur gezielten und strukturierten Suche sowie konzeptionelle Ideen zur fachlich übergreifenden Integration von *Fachdaten* der Umweltbeobachtung bereicherten das Thema der Metadaten- und Fachdatenrecherche in heterogenen Datenbeständen.

Neben der Vorstellung von web-basierten Fachinformationssystemen und Plattformen aus den Bereichen Gewässerschutz *(Nord- und Ostsee-Küsteninformationssystem NOKIS),* der Biodiversität *(BioCASe,),* dem Abfallbereich *(EUWAS)* und der Umweltbeobachtung *(Web-Dioxin-Datenbank)* fanden GIS-basierte Lösungen (*Naturpilot, Web-GIS*

Umweltprobenbank, GIS GRASS) ebenso ein reges Interesse bei den Teilnehmern des Workshops. Ein für die Zukunft immer wichtiger werdendes Thema umfasste die Aufbereitung und Qualitätssicherung von heterogenen Umweltdaten. So zeigten die Projekte Grundwasser-*online* und das *„Integrierte Mess- und Informationssystem zur Überwachung der Umweltradioaktivität (IMIS)"* Lösungsansätze auf Basis von XML-Technologie bzw. integrierter Workflow-Komponenten.

Der Workshop präsentierte in eindrucksvoller Weise, konzeptionelle Ansätze sowie Softwareentwicklungen mit serviceorientiertem Charakter – ausgerichtet auf Angebot, Weitergabe und Mehrfachnutzung von Umweltdaten. Diese Freiheit der Weitergabe von Daten und Informationen wird gestützt durch das am 14. Februar 2005 in Kraft getretene Umweltinformationsgesetz des Bundes sowie das am 01. Januar 2006 in Kraft getretene Informationsfreiheitsgesetz. Nicht immer ist dieses selbstverständlich – so stellte *Kofi Annan* am 16. November 2005 auf dem Weltgipfel der Informationsgesellschaft in Tunis folgendes fest:

Die Lebensader der Information ist Freiheit

Gerlinde Knetsch

Umweltbundesamt

Fachgebiet Informationssysteme Chemikaliensicherheit

Februar 2006

INHALTSVERZEICHNIS

Flexible und modulare Systemarchitektur

für InGrid 1.0 und Portal-U

Thomas Vögele

Fred Kruse

Martin Klenke

Koordinierungsstelle UDK/GEIN

im Niedersächsischen Umweltministerium

Archivstrasse 2, 30169 Hannover

mailto:kug@numis.niedersachsen.de

Einleitung

Der Umweltdatenkatalog UDK und das Umweltinformationsnetz Deutschland *gein*® durchlaufen derzeit eine grundlegende Veränderung. Aus zwei getrennten und technisch völlig unterschiedlich aufgebauten Softwaretools entsteht ein neues, integriertes System. Die neue Software (InGrid 1.0) vereint die Vorteile beider Einzelsysteme, also effizientes Metadatenmanagement (UDK) mit komfortablen Zugriffsmethoden (*gein*®). Mit InGrid 1.0 wird primär ein neues Portal für das vom Bund und den Ländern gemeinsam betriebene Umweltinformationsnetz Deutschland aufgebaut. Dieses neue Portal – Portal-U – ersetzt das bisherige *gein*® und wird von der Koordinierungsstelle UDK/GEIN (KUG) betrieben. Die Software InGrid 1.0 ist jedoch so konzipiert, dass damit im Bedarfsfall auch andere Umweltportale, z.B. auf Länder- und kommunaler Ebene aufgesetzt werden können.

Die technische Feinkonzeption des Systems wurde von der KUG in Zusammenarbeit mit den Firmen GIStec GmbH (www.gistec-online.de), wemove (www.wemove.com), MediaStyle (www.media-style.com), und dem Fraunhofer Institut Graphische Datenverarbeitung (www.igd.fhg.de) erarbeitet. Die Umsetzung der Konzeption erfolgt bis Ende 2005. Ein erster Prototyp wird Anfang 2006 ans Netz gehen.

InGrid 1.0 setzt auf eine modulare Softwarearchitektur auf Basis lizenzfreier OpenSource Software auf, die das System auf neue Anforderungen und Einsatzgebiete vorbereiten soll. So basiert das derzeitige Umweltinformationsnetz Deutschland auf einer zentral verwalteten Instanz der *gein*® Software, die als Informationsbroker das behördliche Informationsangebot beim Bund und den

Ländern zugänglich macht. Auch der Umweltdatenkatalog nutzt zur Verknüpfung der dezentral beim Bund und den Ländern betriebenen UDK Kataloge eine zentrale Broker-Instanz, den V-UDK. Die Aufgaben dieser beiden Systeme werden in Zukunft von Portal-U als dem von der KUG zentral betriebenen Umweltportal des Bundes und der Länder übernommen. Zusätzlich wird es mit InGrid 1.0 jedoch möglich sein, bei Bedarf weitere Informationsknoten auf verschiedenen Ebenen der Verwaltungshierarchie zu etablieren. Damit können z.B. landesbezogene Informationsportale aufgebaut oder aber zusätzliche Informationsanbieter (z.B. auf Grund der Einbindung der Kommunen in das Informationsnetz) verwaltet werden.

In Portal-U wird mit Hilfe einfach zu konfigurierende Datenschnittstellen die Einbindung von Fachinformationssystemen und Datenbanken, die über reguläre Suchmaschinen nicht zugänglich sind (das sog. „invisible web“), erleichtert. Außerdem kann Portal-U über standardisierte Austauschschnittstellen Daten und Informationen mit anderen Datenkatalogen und Dateninfrastrukturen austauschen. Geplant ist z.B. eine Verknüpfung mit nationalen und internationalen Geodateninfrastrukturen (GDI-DE bzw. INSPIRE). Damit kann die Bereitstellung von Umweltinformationen, wie sie in der geplanten INSPIRE Richtlinie gefordert wird, sichergestellt werden.

Das vorliegende Papier gibt einen Überblick über die wichtigsten Komponenten der für InGrid 1.0 bzw. Portal-U vorgesehenen Systemarchitektur.

1 Architektur

Die Systemarchitektur von InGrid 1.0 sieht insgesamt 4 Hauptmodule vor: das Modul „Portal“, das Modul „WMS“, das Modul „Suchmaschine“ und das Modul „iBus“ (**Abb. 1**). Im Folgenden werden die wichtigsten Merkmale dieser Module kurz skizziert.

1.1 Modul „Portal“

Zur Realisierung der Portaloberfläche setzt InGrid 1.0 die OpenSource Portal-Engine „Jetspeed“ ([ASF, 2005]) ein. Jetspeed ermöglicht die flexible Gestaltung und Verwaltung von Internetportalen. Spezifische Portalfunktionen werden über einzeln zu verwendende „Portlets“ implementiert. Dies sind kleine (in JAVA implementierte) Applikationen, die auf der Basis einer gängigen Servlet-Engine (Tomcat) serverseitig aufgesetzt werden. Je nach gewünschter Funktionalität des Gesamtportals können

einzelne Portlets aktiviert oder deaktiviert werden. Für Portale, die mit InGrid 1.0 betrieben werden (z.B. Portal-U) bedeutet das, dass zum einen die standardmäßig angebotenen Portalfunktionen vom Systemadministrator kontrolliert werden können und zum anderen jeder Nutzer über eine Personalisierung des Portals einzelne Funktionen ausblenden kann. Insgesamt kann die Funktionalität eines Portals beliebig und flexibel erweitert werden indem in dem einzelne Portlets neu erstellt bzw. direkt von Drittanbietern hinzugekauft werden.

1.2 Modul „WMS“

Die OpenSource Produkte UMN Mapserver ([UMN, 2005]) und MapBender ([CCGIS, 2005]) werden zur Umsetzung der Geoportalfunktionen in InGrid 1.0 eingesetzt. Der WMS bzw. der MapBender Map-Viewer erfüllen dabei zwei Aufgaben: Zum einen bilden sie die Grundlage für die kartenbasierte Suche in Portal-U. Durch Aufziehen eines Such-Rechtecks oder die Auswahl administrativer Einheiten per Mausklick kann dort der räumliche Aspekt einer detaillierten Suchanfrage spezifiziert werden. Zum anderen werden WMS und Map-Viewer zur Visualisierung digitaler Karten genutzt, sofern diese ebenfalls als WMS angeboten und über die Suchfunktionen des Portals bzw. alternative Zugangswege (z.B. Themenseiten) zugänglich sind. Dabei können Attributdaten abgefragt und mehrere Karten aus verschiedenen Datenquellen gemeinsam dargestellt werden.

1.3 Modul „Suchmaschine“

Die effiziente Indexierung und Suche in Webseiten und Datenbankinhalten ist eine der Schlüsselfunktionalitäten des Portal-U bzw. InGrid 1.0. Entsprechend werden hierfür moderne, leistungsfähige und gut skalierbare Softwaretools eingesetzt. Als Indexierer dient das OpenSource Produkt Lucene ([AJP,2005]), während das Crawling und Ranking der Informationsbestände mit Hilfe von Software aus dem (ebenfalls OpenSource) Projekt nutch ([AI,2005]) und Eigenentwicklungen durchgeführt wird. Für ein integriertes Ranking und Darstellung der an Portal-U angeschlossenen Datenquellen ist es notwendig, dass alle Datenquellen über Lucene indiziert werden. Um diesem Ziel möglichst nahe zu kommen wird ein sog. Data-Source Client bereitgestellt werden. Über den Data-Source Client kann auf einfache Weise ein individueller Volltextindex für jede externe Datenquelle erstellt werden.

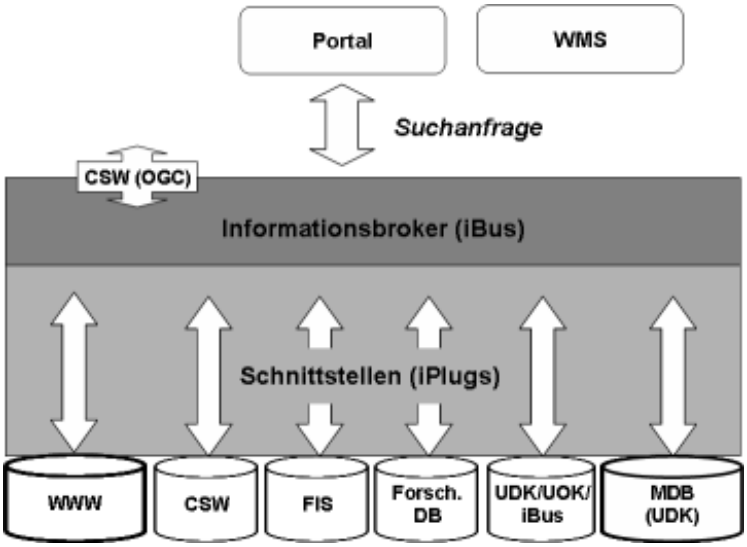


Abb. 1: Systemarchitektur InGrid 1.0

1.4 Modul „iBus“

Der iBus ist das Kernmodul eines InGrid 1.0 Portals. Hier werden Anfragen aus den Modulen „Portal“ und „WMS“, so wie aus externen Datenkatalogen, Suchmaschinen oder Portalen, angenommen und an die verschiedenen an das Portal angeschlossenen Datenquellen weitergeleitet. Der iBus bedient sich dabei einer Reihe von Schnittstellen und Schnittstellenadaptern. Als Anfrageschnittstelle werden neben einer InGrid 1.0 spezifischen Portalschnittstelle eine OGC kompatible CSW 2.0 Schnittstelle und eine UDK Schnittstelle zur Verfügung gestellt. Über die Portalschnittstelle wird die Kommunikation des Moduls „iBus“ mit dem Modul „Portal“ abgewickelt. Die CSW Schnittstelle implementiert das derzeit dem Open Geospatial Consortium (OGC) als Recommendation Paper vorliegenden *Application Profile for CSW 2.0 Catalog Services*. Über diese Schnittstelle können alle OGC CSW-2.0 kompatiblen Katalogdienste bzw. Informationsbroker (wie z.B. das GeoMIS.Bund) auf diejenigen Metadaten in Portal-U zugreifen, die in den Standardformaten ISO 19115 (Geodaten) und ISO 19119 (Dienste) abgelegt sind. Das sind z.B. alle in der Metadatenkomponente des Portal-U (also dem UDK) beschriebenen Geodaten und Dienste. Die technische Umsetzung des iBus erfolgt u.a. mit den OpenSource Werkzeugen Java Management Extensions (JMX) ([SUN,2005]) und Hibernate ([SF, 2005]).

Vom iBus aus wird der Zugang zu verschiedenen Datenquellen durch eine flexible Plug-In Architektur realisiert: Jede Schnittstelle wird als individueller Schnittstellenadapter, oder „iPlug“, implementiert. Derzeit sind iPlugs für die interne (UDK)Metadatenbank, für externe UDK-kompatible Datenkataloge bzw. den bayrischen Umweltobjektkatalog (UOK), für CSW-kompatible Geodatenkataloge, für Fachinformationssysteme, für die Semantic Network Services (SNS) des Umweltbundesamtes (zur semantischen Erweiterung der Suchanfragen), und für reguläre Web-Sites geplant. Je nach Bedarf können aber weitere Schnittstellenadapter entwickelt und in den iBus integriert werden.

1.5 Datasource Client

Als Grundbaustein eines iPlugs dient der sog. Datasource Client (DSC). Er übersetzt aus dem spezifischen Datenformat einer Datenquelle (z.B. JDBC im Falle eines FIS, oder HTML im Falle von Webseiten) in das einheitliche und performante Datenformat einer Indexdatenbank. Die Datenquellen werden zunächst über einen konfigurierbaren Quellen-Adapter an den DSC angeschlossen (**Abb. 1**). Ein auf der Open-Source Suchmaschine Lucene ([AJP,2005]) basierender Indexer erzeugt aus den so aufbereiteten Daten eine hochperformante, mit einem Ranking versehene Indexdatenbank. Das iPlugInterface schließlich macht diesen Index für InGrid Portale bzw. iBus-Instanzen zugänglich, wobei ggf. ein und derselbe Datasource Client von mehreren Portalen gemeinsam genutzt werden kann. Um die Administration der InGrid Datenquellen zu vereinfachen und eine performante Kommunikation des iBus mit den Datenquellen sicherzustellen wird auf Basis des OpenSource Produktes JXTA ([SUN,2004]) eine Peer-to-Peer Kommunikationsinfrastruktur implementiert. Dabei wird jeder Datasource Client bzw. jede über ein iPlug erschlossene Datenquelle bei einem zentralen Repository Server angemeldet. Über die auf dem Repository Server gespeicherten Verbindungsdaten kann die Datenquelle dann an jeden beliebigen iBus angeschlossen werden.

Zusammenfassend erhält also jede an ein InGrid Portal angeschlossene Datenquelle über den Datasource Client einen eigenen Lucene-Index. Unter „Datenquellen“ werden hier vorwiegend Datenbanken und Fachinformationssysteme verstanden. Die Gesamtheit der über Portal-U direkt eingebundenen Webseiten entspricht einer einzigen Datenquelle, d.h. es wird ein zentraler Index für alle „Start-URLs“ der

Informationsanbieter erstellt. Jeder Datasource Client kann die Suchanfragen des iBus hochperformant abarbeiten und eine gerankte Trefferliste zurückliefern. Da die Rankingparameter aller Datasource Clients miteinander kompatibel sind kann im Portal eine einheitliche und gerankte Gesamt-Ergebnisliste mit Treffern aus verschiedenartigen Datenquellen erzeugt werden.

Praktisch stellt ein Datasource Client ein kleines, plattformunabhängig implementiertes Softwaremodul dar, das entweder direkt beim Informationsanbieter bzw. der jeweiligen Datenquelle, oder aber zentral auf dem Rechner des InGrid Portals installiert werden kann. Da die Software nur Teile der OpenSource Suchmaschine Lucene und InGrid-Eigenentwicklungen beinhaltet, kann sie im Rahmen der VwV UDK/GEIN lizenzfrei an beteiligte Informationsanbieter weitergegeben werden. Jeder Datasouce Client verfügt über eine Administrationsoberfläche die es gestattet, das Modul vor Ort (d.h. durch den Informationsanbieter) oder von einem anderen Rechner aus (d.h. durch den Portal-Administrator) zu konfigurieren und zu warten.

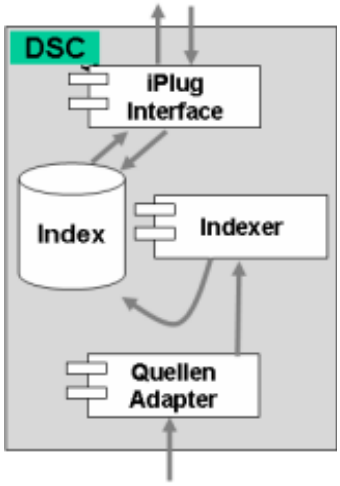


Abb. 2: Schematischer Aufbau eines Datasource Clients

Webseiten, die wie in gein auch in Portal-U eine der wichtigsten Datenquellen darstellen, werden über ein spezielles iPlug eingebunden (siehe oben). Dieses verknüpft einen Datasource Client mit einem effizienten Web-Crawler, der die Internetsites der angeschlossenen Informationsanbieter durchläuft und deren Inhalt in einem geeigneten Format zurückliefert. Der Web-Crawler wird über Software aus dem (ebenfalls OpenSource) Projekt nutch ([AI,2005]) umgesetzt. Über die automatische Verschlagwortung mit SNS erhält jede Webseite u.a. einen

Raumbezug und steht damit den räumlichen Suchfunktionen des InGrid Portals zur Verfügung.

2 Zusammenfassung

Die Software InGrid 1.0 wird zur Umsetzung von Portal-U eingesetzt, das als integriertes Werkzeug zur Metadatenhaltung und Portal für behördliche Umweltinformationen das German Environmental Information Network *gein*® und langfristig auch den Umweltdatenkatalog UDK ablösen soll. Diese Ablösung soll schrittweise und unter Einbeziehung bereits vorhandener Systeme und Datenbestände erfolgen. Technisch wird mit InGrid konsequent eine modulare und flexible Systemarchitektur auf der Basis von OpenSource Technologie umgesetzt. Insbesondere die Module „Portal“ und „iBus“ sind zusätzlich in sich für eine flexible Skalierung und Anpassung an neue bzw. sich ändernde Anforderungen ausgelegt. Damit wird es möglich bei Bedarf, neben dem zentralen Portal-U, weitere Informationsknoten auf verschiedenen Ebenen der Verwaltungshierarchie anzulegen und so ein System von vernetzten Umweltportalen zu schaffen. Diese können durch Zu- oder Abschalten von Modulen an den jeweils gewünschten Funktionsumfang angepasst werden. Durch den ausschließlichen Einsatz von OpenSource Produkten und Eigenentwicklungen ist sichergestellt, dass beim Betrieb von InGrid Informationsknoten keine Kosten für Softwarelizenzierungen entstehen. Da InGrid im Rahmen einer Bund/Länder Verwaltungsvereinbarung entwickelt wurde kann das Produkt von den Partnern der Verwaltungsvereinbarung und deren nachgeordneten Behörden kostenfrei genutzt werden. Mit InGrid und Portal-U wurden somit die technischen und organisatorischen Voraussetzungen zum weiteren Ausbau der im Bereich der deutschen Umweltverwaltung bereits bestehenden und bewährten informationstechnischen Infrastruktur geschaffen. Damit soll insbesondere die Umsetzung einschlägiger EU Richtlinien, wie der EU Richtlinie 2003/4/EG (Umweltinformationsrichtlinie) und der INSPIRE Richtlinie unterstützt werden.

3 Literatur

[ASF, 2005]

ASF: Jetspeed, an Overview, URL: http://portals.apache.org/jetspeed-1/, Apache Software Foundation, 2005.

[AI, 2005]

AI: Welcome to Nutch, http://incubator.apache.org/nutch/, Apache Incubator, 2005.

[AJP, 2005]

AJP: Lucene, an Overview. http://lucene.apache.org/java/docs/index.html, Apche Jakarta Project, 2005.

[CCGIS, 2005]

CCGIS: Mapbender – WebGIS mit freier Software, http://www.mapbender.org/, Consulting Center für geografische Informationss<ysteme, 2005

[UMN, 2005]

UMN: UMN Mapserver Version 4.4.1, http://mapserver.gis.umn.edu/, University of Minnesota, USA, 2005.

[SF, 2005]

SF: Hibernate - Relational Persistence for Idiomatic Java, http://www.hibernate.org/, SourceForge.net, 2005

[SUN, 2004]

SUN: JXTA Technologies – Creating Connected Communities, http://myjxta2.jxta.org/project/www/docs/JXTA-Exec-Brief.pdf, SUN Microsystems, January 2004.

[SUN, 2005]

SUN: Java Management Extensions (JMX), http://java.sun.com/products/JavaManagement/, SUN Microsystems, 2005.

Zur Anwendung "intelligenter" Suchmaschinen

zur Vermittlung von Umweltdaten

Erich Weihs[1]

Abstract

Das novellierte Umweltinformationsgesetz verpflichtet zur passiven (Auskunftspflicht) und aktiven Information durch Behörden über die Umwelt. Es beruht auf der Erfüllung einer EU-Richtlinie zur Informationspflicht der Öffentlichkeit. Da sich die Definition „Umwelt“ der EU-Richtlinie nicht an vorgegebene Organisationsstrukturen der Verwaltungen „hält“ und auch Bereiche aus Kultur, Gesundheit und Verbraucherschutz umfasst, ist die querschnittsbezogene Erschließung einschlägiger Daten aus dem Intranet und Internet und aus Fachinformationssystemen unterschiedlicher Verwaltungen unabdinglich geworden. Die ganzheitliche Erschließung wird mittels weiter verweisenden Katalogen (so genannten Metainformationssystemen, WEIHS 1998, 2000, 2001) und/oder durch Recherche im Internet/Intranet in ausgewählten Domänen und Auswertung der Ergebnisse erfolgen. Während die im Schwerpunkt statischen Daten der Metainformationssysteme in Datenbanken gehalten werden, haben Rechercheergebnisse aus dem Internet nur wenig gemeinsame Strukturmerkmale wie wir sie in Datenbanken vorfinden. Gleichwohl muss eine Synthese dieser dynamischen und in Struktur und Sprache sehr unterschiedlichen Quellen versucht werden.

„Intelligente“ Suchmaschinen (im Sinne transparenter Verfahren) können die ad hoc erstellten Ergebnislisten einer Recherche nach Strukturkriterien erschließen. Dabei können auch Daten von Fachinformationssystemen integriert werden. Verfolgt und bewertet werden zwei unterschiedliche Ansätze:

[1] Bayerisches Staatsministerium für Umwelt, Gesundheit und Verbraucherschutz, München
erich.weihs@stmugv.bayern.de

- spontanes Clustering, das heißt die Ergebnisliste wird erst nach der Recherche ausgewertet und nach Häufigkeitsverteilung der Begriffe gegliedert.

- Klassifizierung der durchsuchten Internetseiten nach mathematisch statistischen oder semantischen Verfahren

- oder Kombinationen davon mit Gliederung auf Grund ihrer Klassenzugehörigkeit(en).

In dem Beitrag werden zuerst die Ansätze der o.a. Methoden mit Schwerpunkt semantische Verfahren versus mathematisch-statistischer Verfahren unter Berücksichtigung der Literatur diskutiert. Dann wird über konkrete Ergebnisse und Erfahrungen berichtet, die sehr stark von der verfahrenstechnischen Umsetzung und der Art des ausgewerteten Materials abhängen: Fachdatenbanken haben einen mit semantischen Klassifizierungsregeln beschreibbaren Inhalt, während sich für die Klassifizierung von Internetangeboten wegen des inhomogenen Inhalts tendenziell multivariate Klassifizierungsverfahren mehr zu eignen scheinen.

1. Einführung

1.1. Systemüberblick

Das UIG sieht vor, dass die Daten entweder unmittelbar bereitgestellt oder zumindest nachgewiesen werden müssen. Verpflichtet zur Information sind die obersten Landesbehörden, die weiter verweisen können, sofern sie die Daten nicht halten oder Zugriff darauf haben. Da die meisten Daten räumlich und organisatorisch dezentral vorhanden sind, kommt ihrem Nachweis in Umweltinformationssystemen (UIS) besondere Bedeutung zu. Dabei sind zwei „Nachweissysteme“ eingeführt:

- Kontrolliert und in gewissem Sinne statisch durch in Datenbanken oder Filesystemen gespeicherte Katalogdaten, teilweise mit Link auf die Datenquelle/Datenbank

- Dynamisch als Ergebnisliste einer ad hoc Recherche mit Abstracts, wie diese aus Intranet/Internetsuchmaschinen bekannt sind, immer mit link auf die Datenquelles

Das UIG sieht diese Nachweissysteme direkt und indirekt vor. Die daher aus dem novellierten UIG abzuleitenden Informationsverpflichtungen gegenüber der Öffentlichkeit ergeben sich folgende wesentliche fachliche Anforderungen an die Informationstechnologie der UIS:

- Querschnittbezogene Erschließung einschlägiger Daten aus dem Intranet und Internet und aus Fachinformationssystemen
- Erschließung und Erstellung von Katalogen (so genannten Metainformationssystemen) zum Nachweis von Fach- und Verwaltungsdaten
- Für die Öffentlichkeit einen benutzerfreundlichen Internet-Zugang zu den Umweltinformationen

Natürlich bestehen weiterhin die Anforderungen der Verwaltung zum reibungslosen und zuständigen Informationsaustausch.

Bei der Konzeption des in der Konsequenz umfassenden bayerischen Umweltinformationssystem das den Anforderungen des UIG und des „inneren“ Verwaltungsbetriebes genügt, ist davon auszugehen, dass die fachliche und technische Selbstständigkeit der Betreiber von Fachinformationssystemen erhalten bleibt, selbst wenn der technische Datenbankbetrieb innerhalb eines Rechenzentrums erfolgt. Erforderlich ist daher eine Zwischenschicht als Koppelungsplattform im Sinne einer Service Orientierten Architektur (SOA) nach Abbildung 1, die zum Recherchezeitpunkt eine spontane Integration der Fach- und Verwaltungsdaten ermöglicht. Die Zwischenschicht, deren Qualität letztendlich das UIS bestimmt, ist daher der eigentliche Kern des Systems. Sie verbindet - neben den für die Fachanwendung direkten Netzzugang – die Fach- und Verwaltungsdaten mit den einzelnen Anwendungen entweder über Schnittstellenprotokolle nach dem http-Protokoll oder über Verweise aus „festen“ oder „spontanen“ Katalogeinträgen (als Ergebnis-Liste einer Recherche).

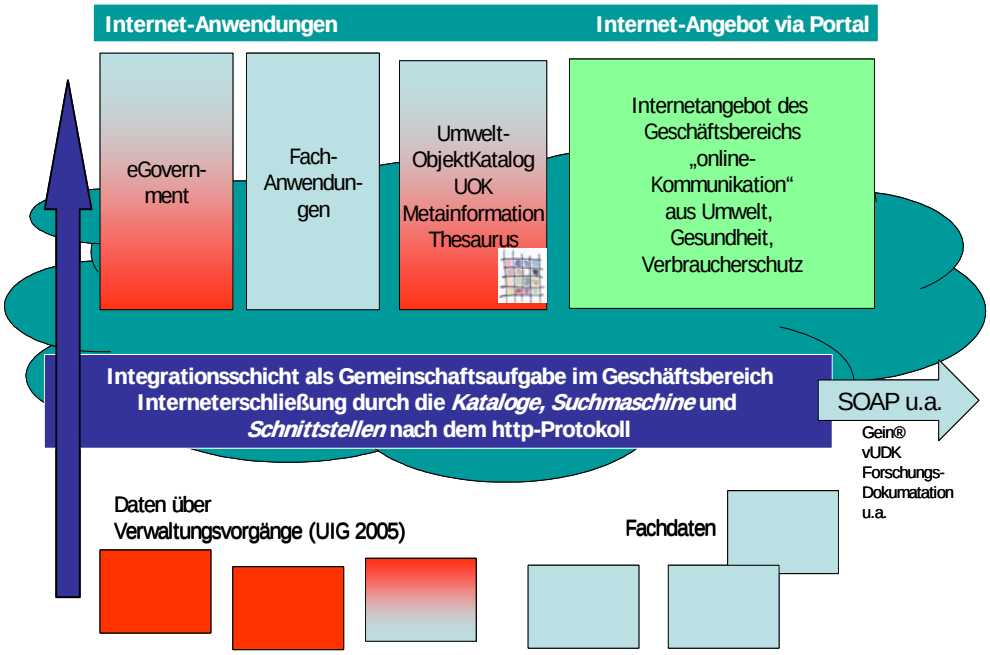


Abb. 1: Integration der Fach- und Verwaltungsdaten des Geschäftsbereichs.

Dargestellt ist in Abbildung 1 vereinfacht die Verbindung der Zwischenschicht zu Fach- und Vollzugsdaten, über die nach der EU-Richtlinie zur Umweltinformation Auskunft gegeben werden muss. Dazu gehören raumplanerisch bedeutsame Maßnahmen, wie die Umweltverträglichkeitsprüfung, deren Verfahrensstand auskunftspflichtig sein wird.

Der Einsatz einer „eigenen“ Suchmaschine (im Gegensatz zur Einbindung der am Markt befindlichen Suchmaschinen wie google etc.) wird nur berechtigt sein, wenn ein Mehrwert zu erwarten ist. Aus diesem Grunde beobachten wir das Verhalten von Suchmaschinen auf Grundlage eines einfachen Recherchesets von etwa 100 Fragen seit 01.02.2005 in dem in unregelmäßigen Zeitabständen die kommerziellen Suchmaschinen wie google, Altavista oder auch gein®, der UDK, ULIDAT etc. abgefragt werden. Den Abbildungen 2 – 4 kann entnommen werden, dass die Ergebnismenge im Zeitablauf erheblichen – und auch kurzfristigen – Schwankungen unterworfen ist. Da nie die gesamte Ergebnismenge untersucht wird sondern in der Regel nur die ersten 10 – 50 Treffer, fällt dieser Mangel nicht auf. Die Qualität der Ergebnislisten aus Recherchen ist, natürlich von der Fragestellung abhängig, deshalb oft als gering einzuschätzen.

-

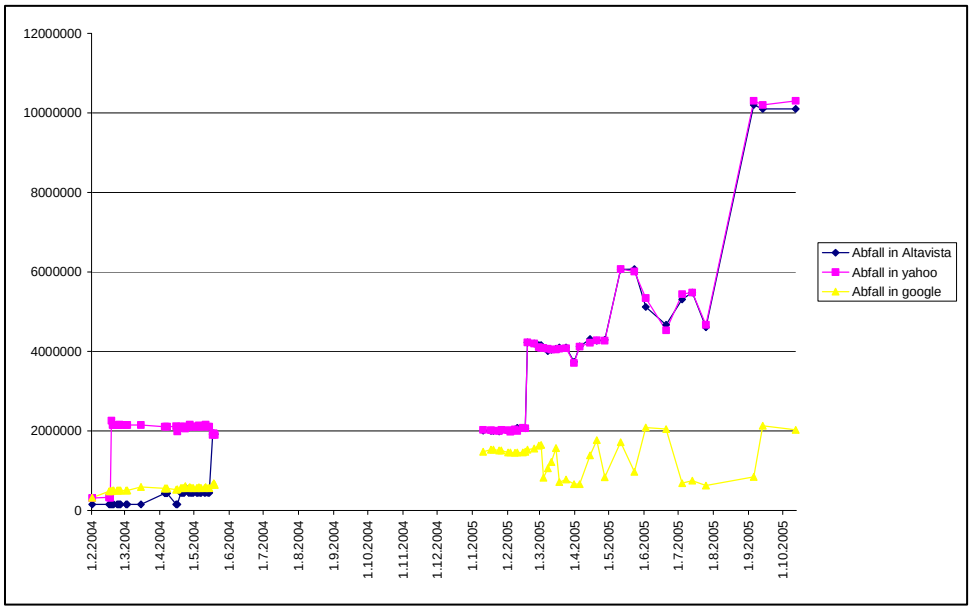


Abb. 1: Trefferstatistik für den Begriff „Abfall“ in öffentlichen Suchmaschinen recherchiert[2]

Die Anforderungen, die den Einsatz/Betrieb einer eigenen Suchmaschine rechtfertigen sind:

Effiziente Befriedigung von Suchanfragen

- aus dem Geschäftsbereich
- der Öffentlichkeit.

In den Daten des Geschäftsbereichs sind folgende Kriterien zu berücksichtigen:

- Internetangebot in den Domänen des Geschäftsbereichs
- Berücksichtigung von Datenbanken (CMS, UOK) mit festem URL-Servletaufruf
- Datenbankanbindungen (Call Schnittstelle, SOAP)
- Performance, Benutzerfreundlichkeit, Zugangsrechte (Intra-, Internet)
- Erstellung kurzer relevante Suchlisten

[2] Im Zeitraum zwischen 15.05.2004 und 15.01.2005 fand keine Erhebung statt.

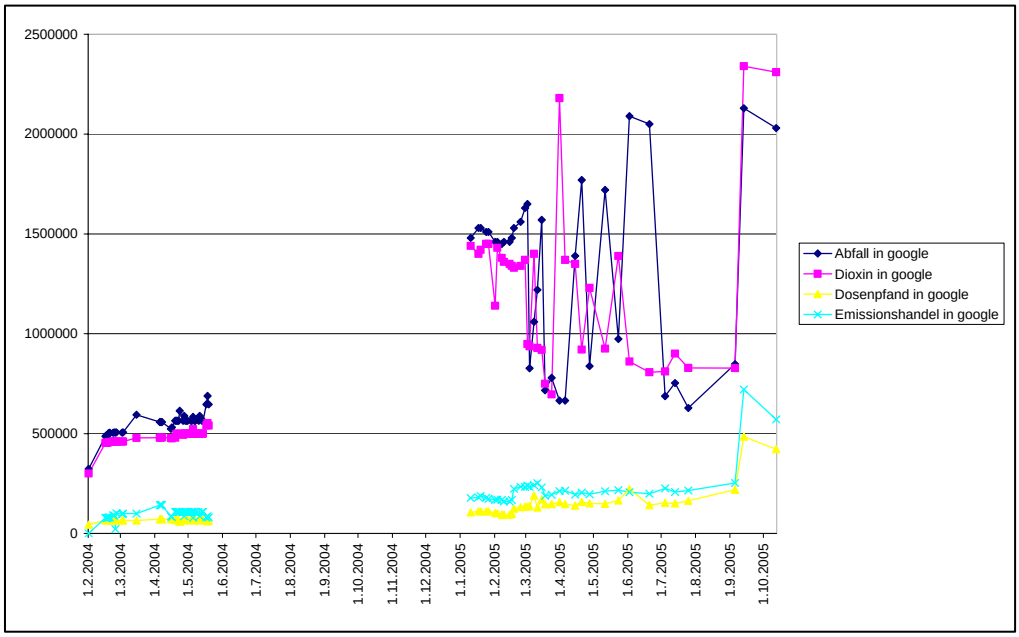


Abb. 2: Trefferlisten in google recherchiert für Abfall, Dioxin, Dosenpfand, und Emissionshandel

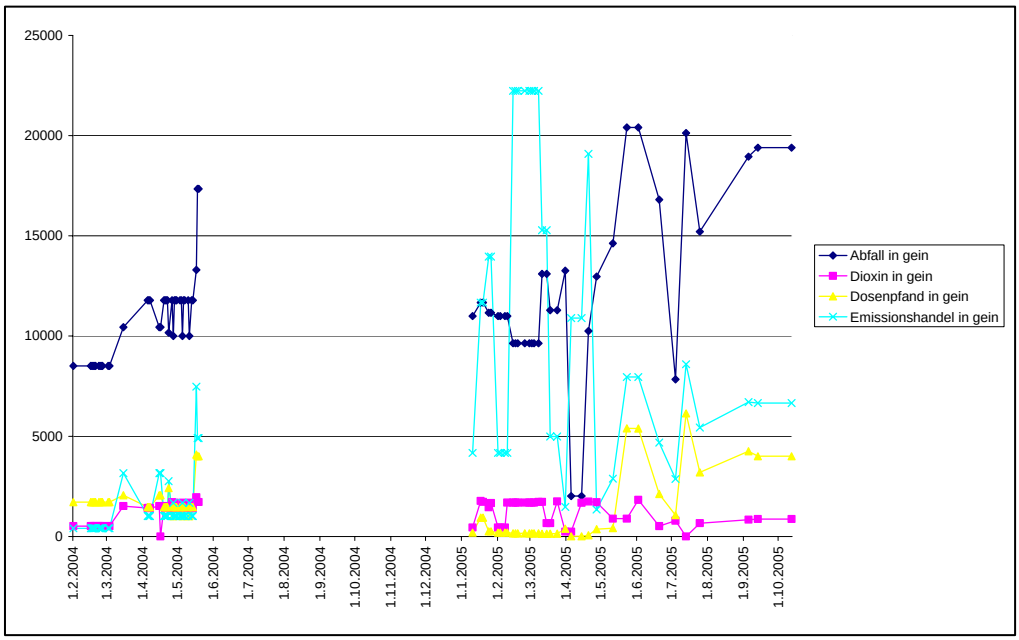


Abb. 3: Trefferliste für Abfall, Dioxin, Dosenpfand und Emissionshandel in gein®, der Suchmaschine des Bundes und der Länder für Umweltfragen

Im Folgenden werden die Methodik der Suchmaschine betreffenden Anforderungen beschrieben und über erste Ergebnisse einer Machbarkeitsstudie für eine kategorisierende Suchmaschine berichtet. Als apostrophiert „intelligent“ haben wir die untersuchte Suchmaschine bezeichnet, da der zu Grunde liegende Algorithmus im Sinne unserer Themenstellung hohe Anforderungen an die Verfahrenstechnik stellt und weitgehend manuelle Eingriffe erübrigen soll.

Zum Vergleich haben wir ein Verfahren der spontanen Klassifizierung mit den selben Daten getestet. Spontanes Klassifizieren erstellt aus der Verteilung des Wortgutes die Trefferliste, also nach der Recherche, eine Klassifizierung. Da das Wortgut abhängig vom Rechercheergebnis verteilt ist, entstehen abhängig von der Recherche und den Texten der Treffer unterschiedliche Verteilungen.

2. Die Machbarkeitsstudie

2.1. Zur Methode der automatischen Kategorisierung/Klassifikation als Ordnungselement der Ergebnislisten

Die klassischen Suchmaschinen – so notwendig sie heute noch sind – haben sich für viele Zwecke als unzureichend erwiesen, da das eigentliche Problem weniger daran liegt, Informationen zu finden und nach Relevanz zu gewichten, als darin, Informationen zu erschließen und mit bereits vorhandenen Informationen zu harmonisieren.

Daher sind in letzten 10 Jahren vor allem von universitärer Seite zahlreiche Anstrengungen unternommen worden, automatische Verfahren zur Informationserschließung aus unstrukturierten Dokumenten zu entwickeln. (Alle heute kommerziell angebotenen Kategorisierungsverfahren stammen aus der universitären Forschung).

1. Dokumenten-Repräsentation

Jedes System operiert nicht mit den Original-Dokumenten, sondern mit einer Repräsentation dieser Dokumente. Die Dokumente werden mit automatisierten Verfahren vorbehandelt, um den benötigten Speicherplatz zu reduzieren, die Response-Zeiten des Systems zu erhöhen und die Vergleichbarkeit der Dokument-Daten zu gewährleisten.

2. Kategorisierung

Dokumenten-Kategorisierung meint zunächst nur die Verbindung von bestimmten Dokumenten mit einem vordefinierten Set von Kategorien. Sie umfasst allerdings auch den Lernprozess, in dem ein System Kategorisierungsmuster lernt, um neue, unbekannte Dokumente einordnen zu können. Insofern ist Kategorisierung definiert als eine überwachte Lernaufgabe, in der es darauf ankommt, vordefinierte Kategorien neuen Dokumenten zuzuordnen, basierend auf der Ähnlichkeit zwischen diesen neuen Dokumenten und einem Trainingsset von Beispieldokumenten.

Es lassen sich dabei 7 fundamentale Methoden der automatischen Kategorisierung unterscheiden. In der Praxis treten diese Methoden häufig in Kombination auf.

Schwerpunkt Semantik

- Entscheidungsbaum (semantisch, etwa nach einem mehrstufig gegliederten Themenkatalog)
- Entscheidungsregeln (semantisch wenn Begriff ‚X' vorkommt und nicht ... usw.)
- Zugriffe auf sql-Datenbanken

Schwerpunkt Statistik

- K-nearest Neighbor
- Bayes'sche Verfahren[3]
- Neuronale Netzwerke
- Linear Least Squares Fit (LLSF)
- Support Vector Maschine (SVM)

Für die Machbarkeitsstudie wurde ein Verfahren auf Grundlage der Support Vector Maschine (SVM) Methode in Verbindung mit verschiedenen semantischen und strukturellen Regeln gewählt:

[3] Vgl. Volker (2005) zur Methode

Grundidee dieses Verfahrens ist die Konstruktion eines prototypischen Vektors für eine Kategorie durch die Verwendung eines Trainingssets von Dokumenten. Bei gegebener Kategorie wird den Vektoren der Dokumente, die zu dieser Kategorie gehören, ein positives Gewicht gegeben, und den Vektoren der übrigen Dokumente ein negatives Gewicht. Durch Aufsummierung der positiv und negativ gewichteten Vektoren erhält man den prototypischen Vektor der Kategorie.

Realisiert wird diese Idee durch die Trennung von positiven und negativen Beispielen während der Trainingsphase. Positive Beispiele sind Dokumente, denen eine bestimmte Klasse/Kategorie zugeordnet werden soll, negative Beispiele sollen nicht klassifiziert werden. Um diese Trennung zu vollziehen, erfolgt die Betrachtung eines mehrdimensionalen Vektorraums, in dem die Trainingsdokumente als Punkte dargestellt sind. Das Ziel ist, die optimale Trennungsmöglichkeit zwischen den positiven und negativen Beispielen zu finden. Es werden die möglichen Linien berechnet, mit denen die Trainingsdokumente repräsentierenden Punkte getrennt werden können, ohne dass sich Fehlklassifikationen ergeben. Diese Linien bilden einen Grenzraum, der ausschließlich durch die negativen und positiven Beispiele/Punkte mit dem geringsten Abstand definiert ist. Diese Beispiele repräsentieren die sog. Supportvektoren. Alle Supportvektoren sind gleichweit von der optimalen Trennebene entfernt, und nur sie sind für die Kategorisierung wirksam.

Das bedeutet, dass man genau die gleiche Entscheidungsfunktion erhält, auch wenn man alle anderen Punkte nicht betrachtet. Darin unterscheidet sich SVM von allen anderen Methoden, die mit dem vollen Trainingsset arbeiten. Das System lernt die Gewichtung der Supportvektoren entsprechend der Trainingsbeispiele. Sobald die Gewichtung gelernt ist, sind neue Dokumente, dargestellt als binäre Vektoren (ein Term kommt vor oder nicht), klassifizierbar.

Der Vorteil dieses Verfahrens liegt darin, dass das System sowohl mit großen Dokumentenmengen (da nur die Supportvektoren berücksichtig werden müssen) als auch mit hohen Kategorienzahlen umgehen kann. Allerdings wächst die benötigte Trainings- und Rechenzeit proportional zur Zahl der Klassen, da jede Klasse mit jeder verglichen wird. Dieses sog. Multi-Class-Problem wird durch den Einsatz spezieller Optimierungs-Algorithmen gelöst.

2.2. Die Ergebnisse

Insgesamt wurden 1482 Domains mit 314812 Seiten gescannt, davon:

- Domänen des Geschäftsbereichs (Ministerium und nachgeordnete Dienststellen): 43 Domains und 32932 Seiten vertreten
- Bayerische Gemeinden 1439 Domains und 281880 Seiten
- Domänen anderer Anbieter (z.B. UBA, BMU etc.) 19 Domains mit 2344 Seiten
- UOK als Metainformationssystem
- Bodeninformationssystem (Teilbereiche)

Die Klassifizierung wurde aus den Webseiten des Geschäftsbereichs und des Umweltobjektkatalogs abgeleitet. Das Angebot der Gemeinden ist breit gefächert und in seinen die Inhalte betreffenden Zielvorgaben heterogen. Es eignet sich nur bedingt für das „Lernen“ der Klassifikation. Eine Einordnung der Webseiten der Gemeinden erfolgte auf Grundlage der Klassifizierung der Geschäftsbereichsdomänen.

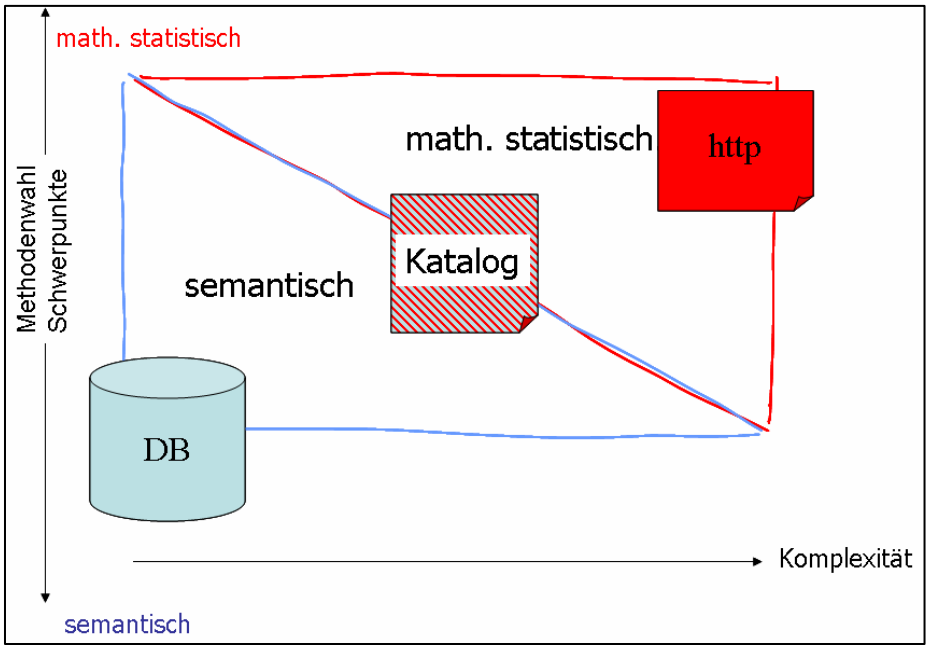


Abb. 4: Verfahrensschwerpunkte zur „intelligenten“ Klassifizierung

Die Praxis der Klassifizierung bestätigte die Annahme, dass fest formatierte Inhalte wie die des Bodeninformationssystems besser nach semantischen Regeln erschlossen werden können. Der Umweltobjektkatalog hat eine Zwitterstellung, die sowohl semantischer wie mathematisch statistischer Regeln bedarf (vgl. Abbildung 4), da verschiedene Merkmale in ihren Alternativen fest vorgegeben und formatiert sind, andere aber als Freitext vorkommen.

Eine spontane Clusterung oder Klassifizierung erbrachte keine konsistente Gliederung, da diese von der Verteilung des Wortgutes der (oft eingeschränkten) Ergebnisliste abhängt.

Abbildung 5 illustriert ein Rechercheergebnis nach „Aschaffenburg“ und vergleichend nach „Nürnberg“. Die Kategorisierung ist identisch, die Zahl der Treffer in ihrer Verteilung über die Kategorien aber unter Umweltgesichtspunkten recht unterschiedlich.

Die Kategorien „Umwelt“, „Mensch und Gesundheit“, „Verwaltung“ usw. sind zwar plausibel nach dem Lebenslagenprinzip gewählt, in erster Linie aber, um zu prüfen, inwieweit eine automatische Kategorisierung möglich ist. So finden sich unter der Subkategorie Formulare, um ein einfaches Recherchieren nach einer Fragestellung z.B. zur Erledigung „Anmeldung zur Hundesteuer“ unkompliziert zu finden ist.

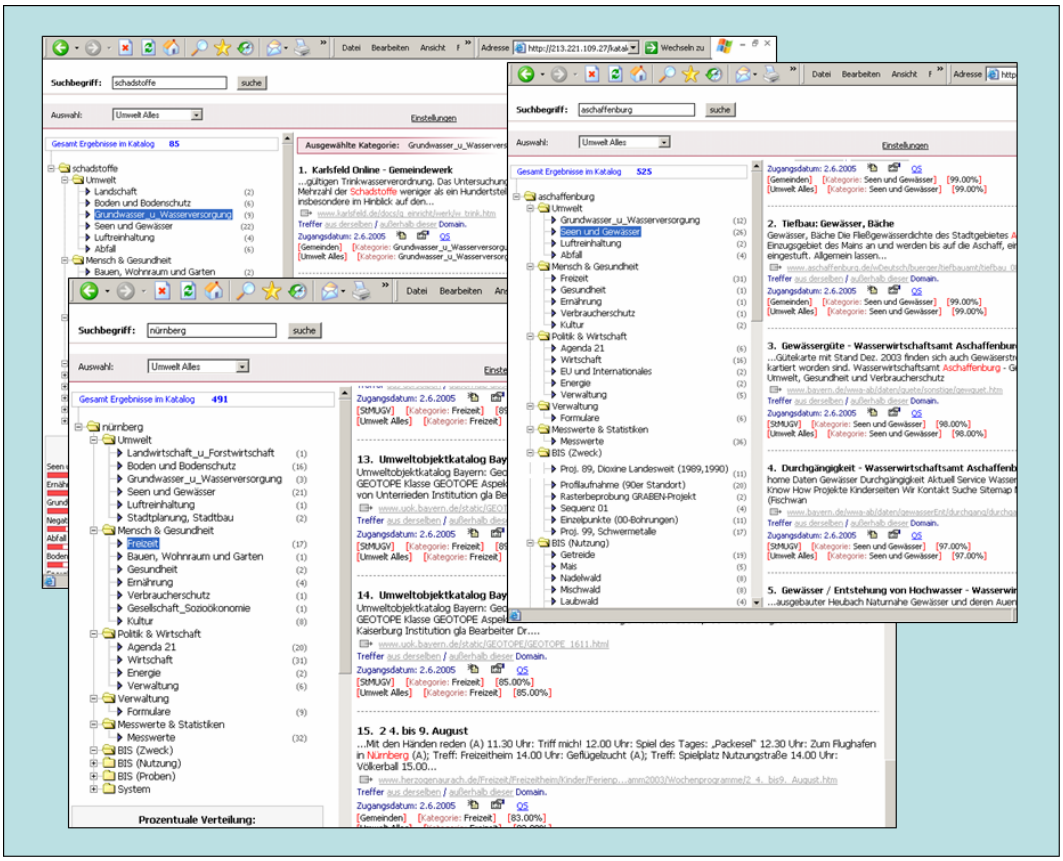


Abb. 5: Vergleich verschiedener Ergebnislisten bezüglich der Kategorisierung

2.3. Das Résumé

In Tabelle 1 sind Bewertungen auf Grundlage der Ergebnisse der Studie angeführt. Im Vergleich zu einer spontanen Clusterung (ex post) zeigt die Übersicht, dass Verfahren der ex ante-Methode zu weit aus positiveren Ergebnissen führt. Der ex ante Methode wird daher der Vorzug gegeben. „Intelligente“ ex ante Methoden, wie hier diskutiert, kategorisieren mittels statistischer und semantischer Verfahren vor der Recherche und gewähren eine stabile und vergleichbare Kategorisierung der Ergebnisse der Suchlisten.

Eine ‚intelligente’ Suchmaschine jenseits von Google zielt also nicht darauf ab, möglichst viele Dokumente zu indexieren, um im Nachgang die Ergebnismengen durch Segmentierung zu reduzieren. Sie verlagert den Focus vielmehr auf die jeder konkreten Suche vorgelagerte Aufbereitung und inhaltliche Erschließung der Dokumente (ex ante), um sodann im Moment einer Recherche mit hoher Präzision

die relevantesten zurückgeben zu können. Die Machbarkeitsstudie hat erwiesen, dass mit vertretbarem fachlichem Administrationsaufwand (Redaktion) eine konsistente Gliederung der recherchierten Ergebnisliste nach thematischen Prinzipien (z.B. Lebenslagen) möglich ist.

Bereich	Spontane Cluster	„intelligente“ Methoden
Kurze Ergebnislisten	+++	+++
Arbeitsaufwand Erfassung	++	--
Konsistenz der Themen der Ergebnislisten	--	+++
Plausibilität der Gliederung	+	+++
Zuschalten von Nachbardaten	--	++
Pflegeaufwand	++	o
Gezielte Informationsvermittlung	--	++
Auskunftssystem, Call-Center	+	o
Internet Anwendung Öffentlichkeit	--	++
Relevanzabschätzung der Recherche	--	++

Tabelle 1. Bewertung spontaner versus „intelligente“ Verfahren

Literaturverzeichnis

Volker, D (2005): Die Baye'sche Variante in: Physik Journal 4(2005) Nr 8/9, p.67 - 72

Weihs, E. (1998): On the classificatiom of environmental data in the Bavarian Environmental Information System using an object-oriented approach in: Studies in Classification, data Analysis, and Knowledge Organization: Data Science, Classification, and Related Methods; Tokyo p. 728 - 735

Weihs, E. (2001): Zum Rechercheerfolg und Typologisierung thesaurusbasierter Suche nach Umweltdaten bei: Explorative Datenanalyse in der empirischen Forschung, 25. Jahrestagung der Gesellschaft für Klassifikation, München

Weihs, E (2005) : Das Bayerische Umweltinformationssystem, in: Umweltinformationssysteme, Hrsg. Peter Fischer-StabelWichmann, Heidelberg 2005, p. 233 – 241

Keywords:

DATA MINING, META INFORMATION, CLASSIFICATION, INTERNET, SUCHMASCHINE, CRAWLER, SEMANTIC WEB

Erfahrungen mit der Anbindung externer Thesauri

Dominik Ernst, Bayerisches Geologisches Landesamt, Dominik.Ernst@gla.bayern.de
Josef Scheichenzuber, Bayerisches Geologisches Landesamt, Josef.Scheichenzuber@gla.bayern.de

Abstract

Von einem Informationssystem wird erwartet, dass man schnell und zielsicher die gesuchten Daten findet. Geowissenschaftliche Informationssysteme mit sehr umfangreichen, differenzierten Datenbeständen kommen in der Praxis daher ohne einen Thesaurus nicht aus.

Im Rahmen des Bodeninformationssystem Bayern (BIS) [BIS1] wird derzeit erprobt, wie an ein proprietäres, komplexes Schlüssellistensystem erfolgreich externe Thesauri aus dem Geo- und Umweltbereich (GEMET [GEMET], UOK-Thesaurus [UOK], UMTHES® [UMTHES], SNS [SNS], etc.) angebunden werden können. Mittels der von dem Open Source Framework Apache Lucene [LUC] angebotenen Analysemöglichkeiten werden Crosskonkordanzen[4] zwischen dem Fachvokabular des BIS (Schlüssellisten, Metadaten, Inhalte alphanumerischer Felder[5]) und den Thesauribegriffen ermittelt.

Unter Zuhilfenahme eines geeigneten Datenmodells kann das Thesaurussystem über eine XML Schnittstelle mittels der modifizierten Open Source Software Touchgraph [TG] visualisiert werden.

1. Einleitung

Die Frage nach der Verwendung von Thesauri stellte sich vor allem, weil wir glauben, dass mithilfe des Wortschatzes und der enthaltenen Verweise neue Zugänge zu den

[4] Crosskonkordanzen sind Verknüpfungen zwischen Termen verschiedener Thesauri.
5 Datenfeld in dem Buchstaben, Zahlen und Sonderzeichen eingegeben werden können.

im Bodeninformationssystem enthaltenen Daten erarbeitet werden und die Informationen des BIS in der Konsequenz einem vergrösserten bzw. einem anderen Nutzerkreis zugänglich gemacht werden können.

Um diese Vorteile zu erreichen, musste ein gangbarer Weg gefunden werden, um die Wortschätze – auf der einen Seite den Wortschatz unseres Bodeninformationssystemes, auf der anderen Seite den Wortschatz eines oder mehrerer Thesauri – miteinander zu verbinden, ohne dass hierbei zuviel Aufwände in Erstellung und Pflege dieser Verbindungen resultieren.

Da sich der Versuch einer Anbindung eines Thesaurus stark mit der Fachlichkeit, der Qualität sowie der Quantität der enthaltenen Daten sowie der Technik des BIS beschäftigt, sollen hier einleitend kurz das BIS und seine Daten dargestellt werden, von denen ausgehend ja die Verknüpfungen zu den Thesauri erarbeitet werden sollen.

Schliesslich wird erläutert, was – vor allem am Datenmodell – getan werden musste, um neben dem bisherigen Fachwortschatz auch mehrere Thesauri verfügbar zu haben und es wird skizziert, wie und welche Verbindungen zwischen BIS sowie Thesauri erstellt wurden und wie diese Verbindungen visualisiert werden.

Die dabei gewonnenen Erfahrungen im Hinblick auf Datenmengen, Handling, Darstellung und Verfügbarmachung der Verbindungen sowie positive und negative Effekte sollen abschliessend dargestellt werden, wobei – beruhend auf den bisherigen Erkenntnissen – das weitere Vorgehen skizziert werden soll.

2. Das Bodeninformationssystem Bayern (BIS)

Das BIS findet seine rechtliche Grundlage im Bayerischen Bodenschutzgesetz BayBodSchG, in dem in Art. 7 der Zweck des Bodeninformationssystems erklärt und dessen Betrieb verbindlich vorgeschrieben wird.

Das BIS ging im wesentlichen aus den Daten der Vorgängeranwendung ZDB (Zentrale Datenbank) hervor und wurde im September 2003 in Betrieb genommen.

Das am Bayerischen Geologischen Landesamt (BayGLA) betriebene Bodeninformationssystem ([BIS1], [BIS2]) enthält punktförmige wie flächenhafte Informationen, wobei die Flächendaten aus Raster- und Vektordaten bestehen. Sämtliche

Daten werden in einem gemeinsamen, metadatengestützten Datenmodell verwaltet. Die Anwendung basiert auf dem Zusammenspiel einer Java Applikation mit kommerzieller Software aus dem GIS- (ArcSDE, ArcIMS) bzw. Geodatenbereich (GeODin). Zusätzlich angebunden sind mehrere Verwaltungstools (Benutzerverwaltung, Schlüssellisten- bzw. Metadatenpflege, Systemsteuerung, etc.).

Für den Betrieb des BIS existieren zwei Arten von Clients:

- **der Behördennetz Client**

ermöglicht insbesondere den Mitarbeitern des Geschäftsbereiches die Recherche, Pflege und den Export der Daten. Potentiell wird hier – gestaffelt nach den einzelnen Berechtigungen – der komplette Datenumfang zur Verfügung gestellt.

- **der Internet Client (GeoFachdatenAtlas)**

ist ein HTML basierter Client, der der breiten Öffentlichkeit den Zugang über das Internet ermöglicht. Hierbei gibt es keine Zugriffsbeschränkung, es sind alle Objekte des Behördennetz Clients verfügbar, jedoch mit weniger Objektdetails und einer unscharfen Lageinformation.

Die Daten des BIS gliedern sich in vier grosse Bereiche:

Fachdaten

Hier werden alle Fach und Labordaten zur Verfügung gestellt, dabei wird innerhalb der Fachdaten nach Punkt- und Flächendaten unterschieden.

Metadaten

Dieser Datenbereich beinhaltet diejenigen Metadaten die zur Steuerung der Anwendung benötigt werden. Solche Metadaten sind z.B. die Fach- und Recherchemodelle der enthaltenen Objekte, deren interner Aufbau sowie die physikalischen Mappings in die jeweiligen Speicher- und Klassenstrukturen.

Daneben werden in den Metadaten auch Beschreibungen der fachlichen Dateninhalte nach ISO 19115[6] verwaltet.

[6] Die ISO-Norm 19115 definiert einen im Jahr 2003 verabschiedeten Standard für Metadaten und etabliert damit ein gemeinsames Verständnis zu Metainformationen. Die Norm beinhaltet Angaben zu eindeutigen Identifikation, zur Ausdehnung, zur Qualität, zum räumlichen und zeitlichen Schema, zum Referenzsystem und zur Nutzbarmachung.

Schlüssellisten

In den Schlüssellisten wird der zur fachlichen Beschreibung benötigte Wortschatz in hierarchisierter Form gepflegt und der Anwendung zur Verfügung gestellt.

Benutzerdaten

Zur Steuerung der Zugriffsberechtigung einzelner Benutzer bzw. Benutzergruppen werden Daten aller zugelassenen Benutzer des BIS erfasst und gepflegt.

2.1. Architektur des BIS

Die Zweiteilung des BIS in Behördennetz Client und Internet Client wird auch bei einem Blick auf die stark vereinfachte Systemarchitektur deutlich (Abb. 1). Beim Behördennetz Client handelt es sich um Java Swing basierte Clients, die über Webstart verteilt werden. Diese Clients greifen über CORBA auf einen Applikations-Server zu, der wiederum mithilfe von JDBC seine Daten aus einer Oracle Datenbank bezieht und diese Daten in entsprechend aufbereiteter Form dem Client zur Verfügung stellt. Physikalisch sind der Behördennetz Client sowie der Internet Client aus Sicherheitsgründen getrennt, weshalb der Datenbestand des Internet Client in regelmässigen Abständen auf den neuesten Stand gebracht wird.

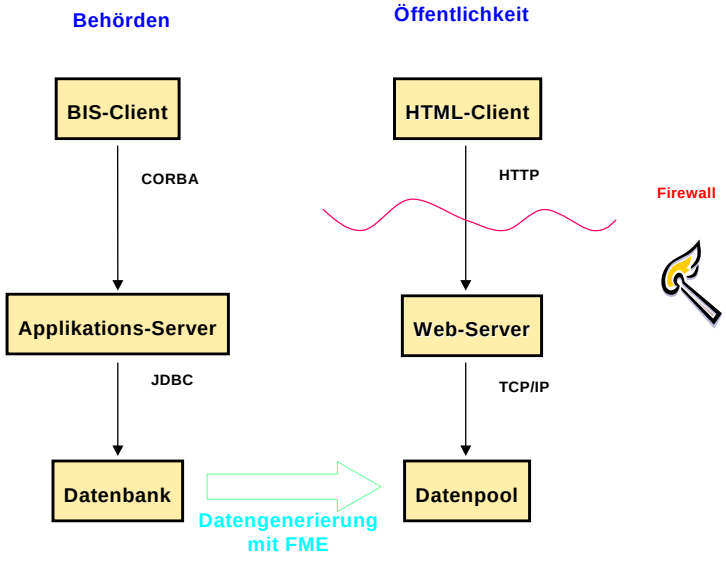


Abb. 1: Schichtenarchitektur des BIS

Der Internet Client ist ein HTTP basierter Client, der seine Daten von einem Web-Server erhält, der sich hinter einer Firewall befindet. Der Web-Server generiert seine Daten unter Zuhilfenahme eines Datenpools, auf den er mit TCP/IP zugreift.

2.2. Metadatenbasierter Ansatz

Natürlich verwendet auch das Bodeninformationssystem Metadaten, wobei zwischen strukturellen und semantischen Metadaten unterschieden wird. Die semantischen Metadaten werden, wie es auch nach ISO 19115 definiert ist, verwendet, um Angaben zu Ausdehnung, Qualität, dem räumlichen und zeitlichen Schema sowie zum Referenzsystem zu machen.

Die strukturellen Metadaten beschreiben dagegen das Fach- sowie das Recherchemodell, die Fachattribute, die verwendeten Plausibilitäten u.a.m. Aufgrund dieser Verwendung der strukturellen Metadaten ist es auch möglich, dass es als Systembesonderheit im BIS nur ein abstraktes Datenmodell gibt. Es werden keinerlei fachliche Klassen im Java Code implementiert, sondern die benötigten Klassen werden zur Laufzeit aus den strukturellen Metadaten dynamisch erzeugt und mit Werten befüllt.

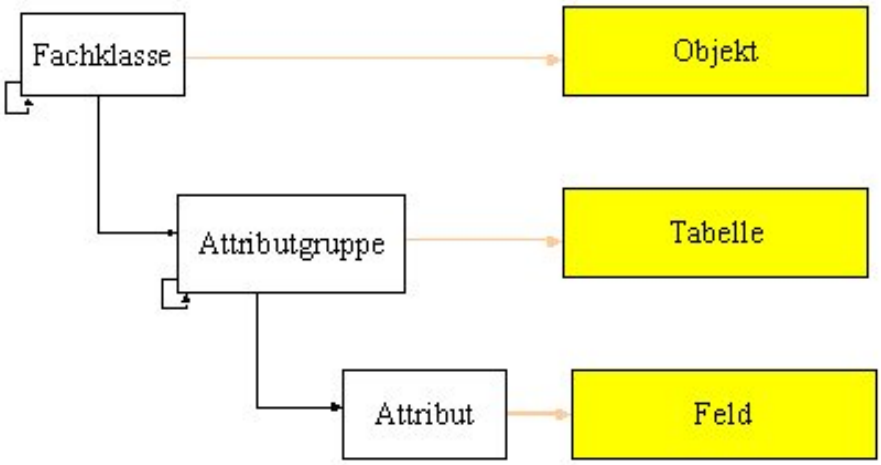


Abb.2: vereinfachtes Modell der strukturellen Metadaten des BIS

Abb. 2 stellt exemplarisch dar, in welcher Art und Weise Metadaten verwendet werden, um unsere fachliche Anforderungen in Java Klassen zu mappen. Alle Fachklassen des BIS sind als Fachklassen in den Metadaten beschrieben worden.

Fachklassen enthalten wiederum n “Attributgruppen”, also Gruppen von thematisch zusammengehörigen Informationen, die immer 1:1 bzw. 1:n zur jeweiligen Vater-Information zugeordnet werden. Eine Attributgruppe selbst setzt sich ihrerseits aus

Attributen zusammen. In Abb. 2 sind in gelb die entsprechen Datenbankobjekte dargestellt. Das metadatengestützte Modell des BIS ermöglicht es jedoch zu jeder Zeit, hier auch ein anderes Produkt bzw. sogar eine andere Technik, z.B. objekt-orientierte Datenbanken, zu verwenden.

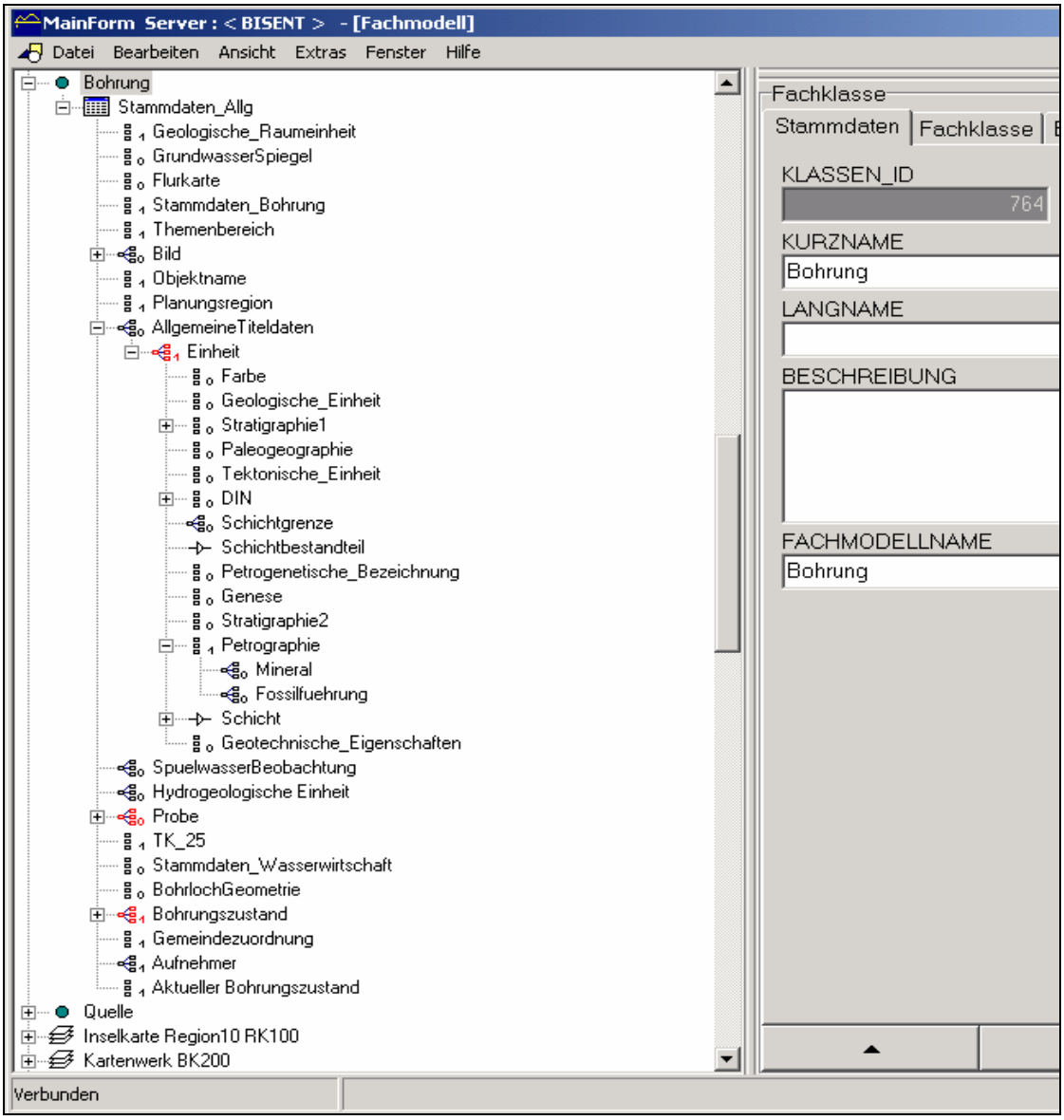


Abb.3: Metadatenpflegetool mit exemplarischer Darstellung der Fachklasse “Bohrung”

Zum Startzeitpunkt des BIS Servers werden alle Informationen aus den Metadaten ausgelesen und zu jeder in den Metadaten beschriebenen Fachklasse wird eine Java Klasse erstellt. Die Attributgruppenbeschreibungen der Metadaten werden verwendet, um das objektrelationale Mapping zwischen den Attributen des Java Objektes sowie den Feldern einzelner Oracle Tabellen zu definieren.

2.3. Verwendung von Schlüssellisten im BIS

Zur Beschlagwortung derjenigen Felder, für die im BIS nur ein kontrolliertes Vokabular[7] zugelassen ist (am BayGLA als “Schlüssellisten-Felder” bezeichnet), werden im BIS nur Begriffe zugelassen, die in Form von sogenannten Schlüssellisten vorliegen. Hierbei handelt es sich um hierarchisch organisierte Glossare, deren Begriffe in einer eigenen Anwendung, dem Schlüssellisten-Pflegetool gepflegt werden.

Im Unterschied zu einem echten Thesaurus hat dieser - auch als Fachthesaurus bezeichnete - Wortschatz nur die Beziehungen

- BT broader Term (Spezialisierung)
- NT narrower Term (Generalisierung)
- Liste Listenzugehörigkeit

Weitere thesaurus-typische Beziehungen wie “RT – related Terms” und die Verwendung von “Nicht-Deskriptoren” und “Synonymen” erfolgt in den BIS Schlüssellisten und deren Begriffen derzeit nicht.

Im BIS werden z.Z. ca. 300 solcher Schlüssellisten verwendet, in denen zusammen ca. 50.000 Begriffe enthalten sind. Ein weiterer Unterschied zu eigentlichen Thesauri ist, dass die Begriffe des BIS immer aus den Attributen

- ***ID*** eine abstrakte, über alle Begriffe aller Listen eindeutige Nummer, die vom System generiert wird
- ***Kurzname*** ein innerhalb der Liste eindeutiger Kurzname, oft auch ein nur Fachleuten zugänglicher Code, z.B. die Kurzbezeichnungen für Auflagehorizonte gemäß Bodenkundlicher Kartieranleitung

[7] Ein kontrolliertes Vokabular ist eine Sammlung von Bezeichnungen (Wortschatz), die eindeutig Begriffen zugeordnet sind, so dass keine Homonyme auftreten. In vielen Fällen gilt auch die umgekehrte Richtung (jeder Begriff hat nur eine oder eine bevorzugte Benennung, d.h. es gibt keine Synonyme). Vorteil eines kontrollierten Vokabulars bzw. einer terminologischen Kontrolle ist es, dass sowohl beim Indexieren bzw. Beschlagworten als auch bei der Recherche ein einheitliches Vokabular verwendet wird, das von vorne herein bekannt ist.

- ***Langname*** *ein sprechender Langname*
- ***Beschreibung*** eine fachliche Beschreibung des Begriffes

bestehen, aus denen die Begriffe und damit der Wortschatz zur Indexierung sowie zur späteren Recherche gebildet werden und nicht, wie bei Thesauri oft üblich, nur aus einem einzigen Begriff und einer Beschreibung.

3. Herkömmlicher Zugang zu den Daten im BIS

Wie in Abb. 4 ersichtlich ist, werden im BIS über 50% aller Fachattribute über eine kontrolliertes Vokabular bzw. Schlüssellisten verschlagwortet, was sich auch bei der Recherche nach diesen Fachattributen bewährt. Durch Verwendung eines Thesaurus könnten allerdings weitere 27 % der BIS-Felder, und zwar die alphanumerischen und CLOB-Felder, inhaltlich erschlossen werden.

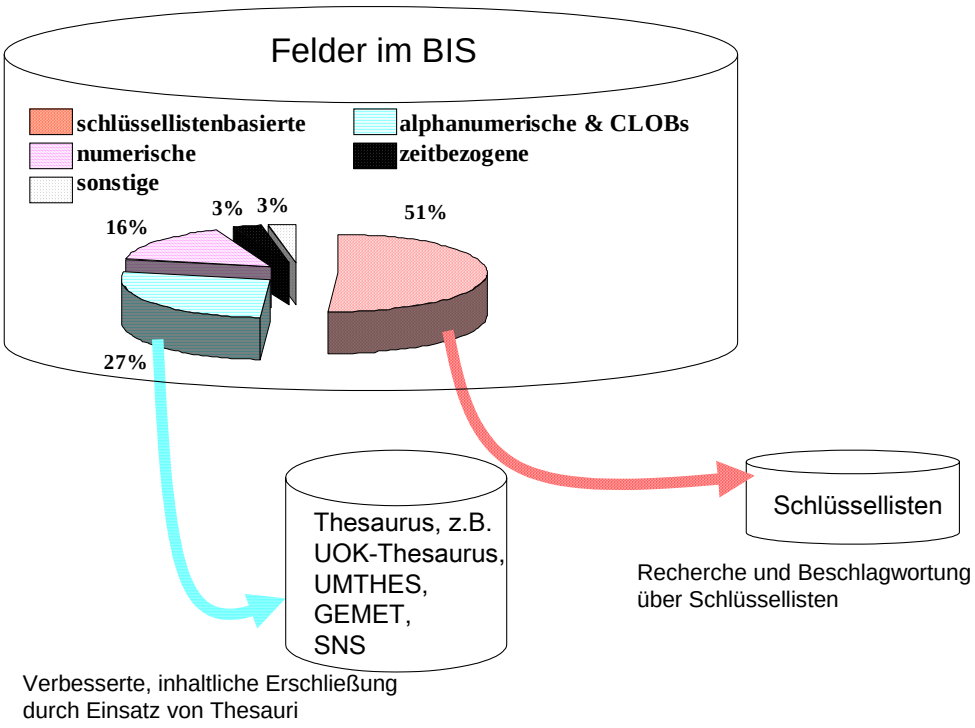


Abb. 4: Zugang zu den Daten im BIS

3.1. Recherchemodell und Fachthesaurus

Der aktuelle Zugang zu den Informationen im BIS erfolgt – neben dem räumlichen Zugang - über eine gewisse Kenntnis des Fachmodells. Erst dieses Fachmodell beinhaltet die Kenntnis darüber, wo sich welches Attribut verbirgt, in welcher Kardinalität ein Attribut bzw. eine Attributgruppe zu anderen steht und um welchen Typ von Attribut es sich handelt. Diese Informationen verbergen sich in den strukturellen Metadaten des BIS und können auch bei der Recherche eingesetzt werden.

Dadurch ist es, wie in Abb. 5 dargestellt, möglich, alle Attributgruppen sowie die Attribute einer ausgewählten Attributgruppe darzustellen und – abhängig vom Typ des zu recherchierenden Attributes – eine Hilfe bei der Eingabe der Recherchewerte zu geben.

Im Beispiel aus der Abb. 5 handelt es sich um ein numerisches Feld (den Jahresniederschlag in mm, wie er verwendet wird, um einen bodenkundlichen Standort zu beschreiben).

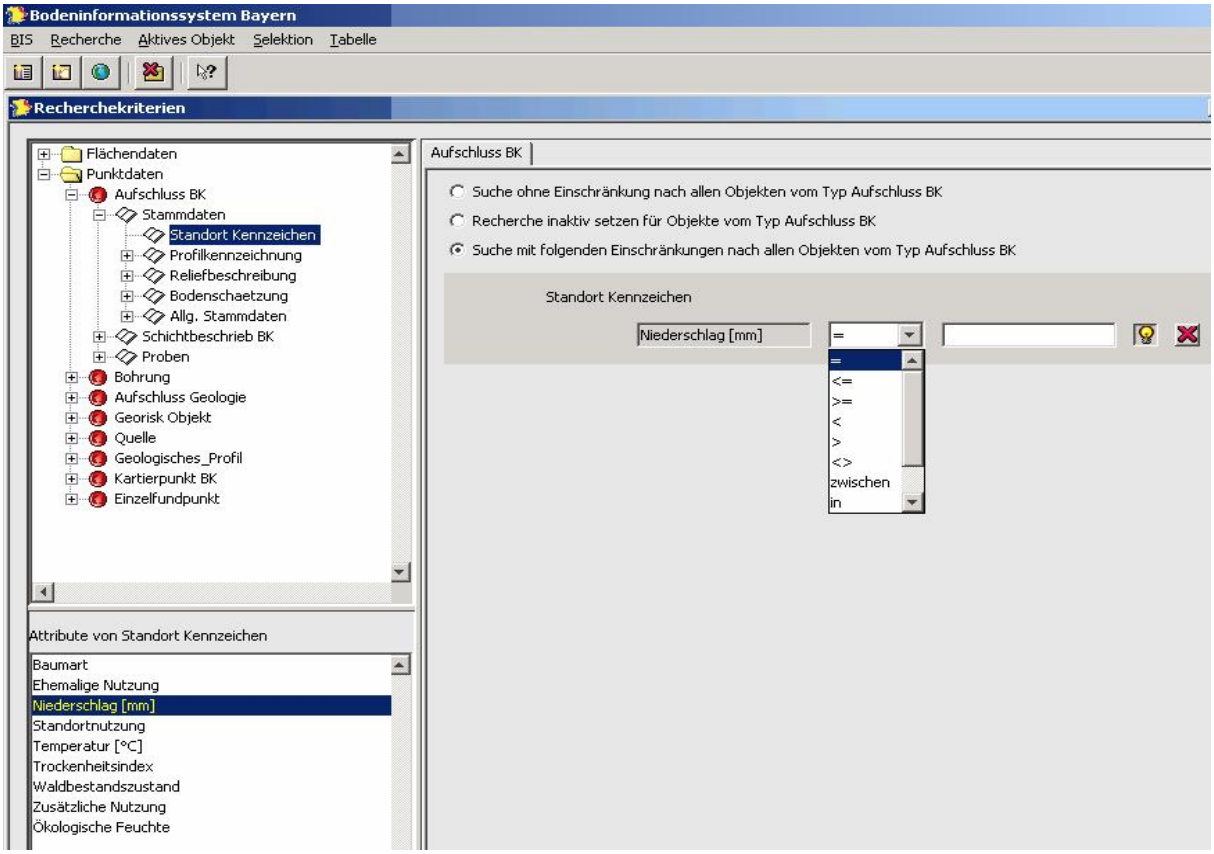


Abb. 5 : Beispiel für eine metadatenbasierte Unterstützung bei der Recherche im BIS

Sofern hier ein Schlüssellistenfeld verwendet wird, kann der Recherchedialog gleich alle erlaubten und recherchierbaren Begriffe zur Übernahme zur Verfügung stellen.

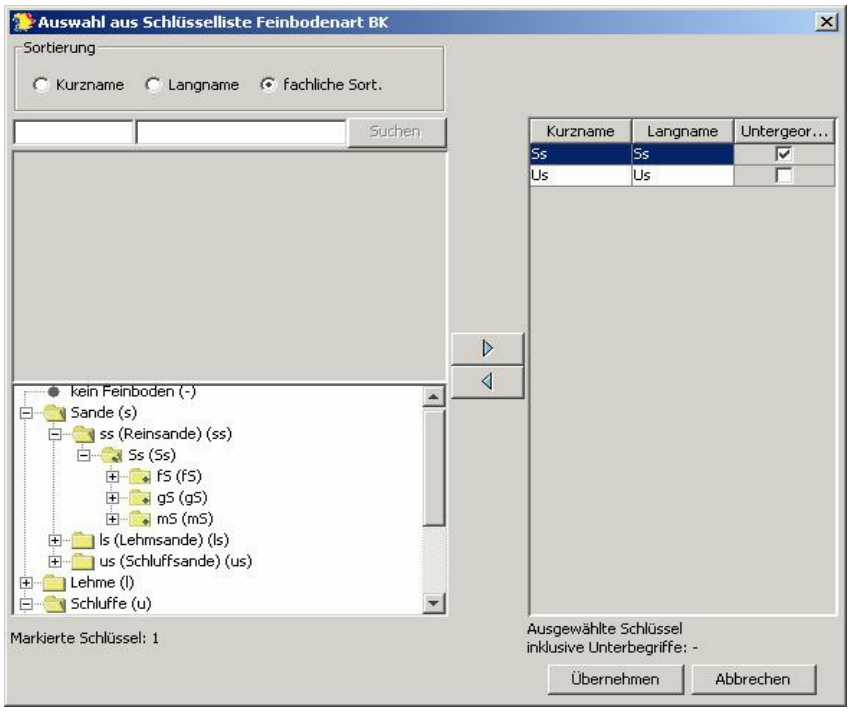


Abb. 6: Beispiel für Rechercheunterstützung eines Schlüssellistenfeldes durch kontrolliertes Vokabular

Eine solche Hilfestellung bei der Auswahl der Recherchevorgaben ist in Abb. 6 für das Beispiel „Feinbodenart in der Bodenkunde“ zu sehen. Hier wird die über die Metadaten mit dem aktuell zu recherchierenden Feld verbundene Schlüsselliste dynamisch ermittelt. Die erlaubten Werte der Liste werden – sofern es sich um eine hierarchisch organisierte Liste handelt – hierarchisch dargestellt und können in die Recherchekriterien übernommen werden.

Problematisch gestaltet sich jedoch die Recherche nach den alphanumerischen Feldern, da der Suchende hier nie sicher sein kann, ob zu seinen Suchvorgaben überhaupt ein einziger Treffer gefunden wird.

Ein weiteres Problem bei der Aufbereitung der bisherigen Informationen für die Recherche ist es, dass die zahlreichen Daten, mit denen die strukturellen Metadaten des Systems beschrieben sind, zwar dem System selbst intern zur Verfügung stehen, dass diese Informationen aber zur Unterstützung einer Recherche nur zu einem kleinen Teil verwendet werden. So enthält beispielsweise die Beschreibung

einer Fachklasse, z.B. des “Aufschluss Bodenkunde” mehrere Zeilen textueller Beschreibung über Art, Struktur und Inhalte dieser Fachklasse, aber diese Information wird z.Z. nicht aktiv bei der Recherche verwendet.

Desgleichen gilt für alle Beschreibungen aller Attribute des Systemes: Zu fast jedem Fachattribut gibt es eine von den Fachabteilungen erarbeitete Beschreibung über Entstehung, Verwendung und Inhalte des jeweiligen Attributes, ohne dass diese Informationen – ausser in unserer Systemhilfe – verwendet werden.

4. Untersuchung neuer Zugangswege zu den BIS-Daten

An den in Kap. 3.1 aufgezeigten Problemen und Schwachstellen der aktuellen Recherche im BIS setzt auch der aktuelle Versuch an, Thesauri zu verwenden, um Zugänge zu unseren Daten gerade im Bereich der alphanumerischen Daten sowie der Metadaten zu erstellen.

Es stand fest, dass zwischen den Begriffen eines Theaurus und den Begriffen, wie sie in alphanumerischen Feldern des BIS verwendet werden, Verbindungen erstellt werden müssen. Um dies zu tun, gibt es klassischerweise zwei Ansätze:

1. Eine “intellektuelle Indexierung” der einzelnen Begriffe – sowohl aus dem Thesaurus als auch aus dem System, um dann, bei erkannten Konkordanzen eine Relation zwischen dem jeweiligen Begriff des Thesaurus sowie dem jeweiligen Begriff aus dem BIS zu erstellen.
2. Eine “automatische Indexierung”, bei der die Begriffe der beiden Quellen auf deren Wortstämme reduziert, auf Gleichheiten geprüft, gewichtet und anschliessend als gewichtete Relation zwischen einem Thesaurusbegriff und einem BIS-Begriff erfasst werden.

Obwohl rein technisch nichts gegen die, im Einzelfall meist überlegene, Möglichkeit der intellektuellen Indexierung spricht, wurde dieses Verfahren zwar zugelassen, aber aufgrund der immens hohen Aufwände bei deren Erarbeitung und vor allem bei der späteren Pflege (im Falle von Updates und/oder Änderungen im einen oder anderen Datenbestand) nicht durchgeführt.

Anstattdessen wurde für das BIS der Weg einer automatischen, computergestützten Indexierung gewählt. Hierzu mussten zunächst die geeigneten Thesauri gefunden werden.

4.1. Einsatz unterschiedlicher Thesauri

Bald war klar, dass es sich um einen Fachthesaurus handeln musste, der das gleiche bzw. ein ähnliches Vokabular verwendet, wie dasjenige, das im BIS verwendet wird, wodurch sich die Auswahl auf folgende Thesauri beschränkte:

- GEMET (GEneral Multilingual Environmental Thesaurus)
- UOK-Thesaurus (Umwelt Objekt Katalogs-Thesaurus)
- UMTHES (Thesaurus des Umweltdatenkataloges UDK)
- SNS (Semantic Network Services)

Da, wie in Kap. 4.2 noch deutlich wird, alle Begriffe aller verwendeten Thesauri einmal komplett durchiteriert werden mussten, um die jeweiligen Wortstämme zu bilden und zu speichern, wurde aus Belastungsgründen der jeweiligen Services darauf verzichtet, dies online über Web-Services zu machen. Dadurch beschränkten sich die Arbeiten in der Folge auf den GEMET, den UOK-Thesaurus sowie den UMTHES, weil in der Kürze der Zeit diese Thesauri freundllicherweise von den jeweiligen Institutionen auch offline zur Verfügung gestellt wurden.

4.2. Die Ermittlung der Crosskonkordanzen

Die Ermittlung von Crosskonkordanzen, manchmal auch als Crosswalk bezeichnet, ist ein Verfahren, bei dem es um Herstellung von Links zwischen den äquivalenten, gleiche Begriffe repräsentierenden Benennungen zweier Thesauri bzw. wie in unserem Fall zwischen je einem Thesaurus und unserem gesamten Fachwortschatz geht. [KUNZ]

Diese Crosskonkodanzen ermittelten wurden durch ein eigens dazu entwickeltes Java Programm, das in Teilen auch Module des Open Source Produktes Apache Lucene [LUC] verwendet, ermittelt. Lucene entstand als Projekt der Apache Jakarta Gruppe und wurde 1997/98 von Doug Cutting gegründet und wird seitdem auch als Open Source Anwendung weiterentwickelt. Lucene ist vollkommen Java basiert, obwohl mittlerweile auch Implementierungen in anderen Sprachen existieren.

Lucene selbst enthält verschiedene Funktionalitäten, die hier kurz exemplarisch und in der Reihenfolge des typischen Ablaufes skizziert werden sollen:

- **Indexierung**
    - Aufbereitung des Suchraums
    - Aufbereitung der Texte
    - Eigentliche Indexierung
    - Wortstammbildung „Stemming“
- **Abspeichern des gebildeten Suchindex**
- **Suche**
    - Interpretation einer Anfrage
    - Durchführen der Suche
    - Darstellung der Ergebnisse

Für die Erstellung der Crosskonkordanzen wird nur ein Teil der in Lucene verfügbaren Funktionen verwendet und zwar:

- Die Aufbereitung der Texte unter Berücksichtigung einer Stoppwortliste, das Parsen und die Wortfindung
- Die Reduzierung der gefundenen Worte auf deren Stämme

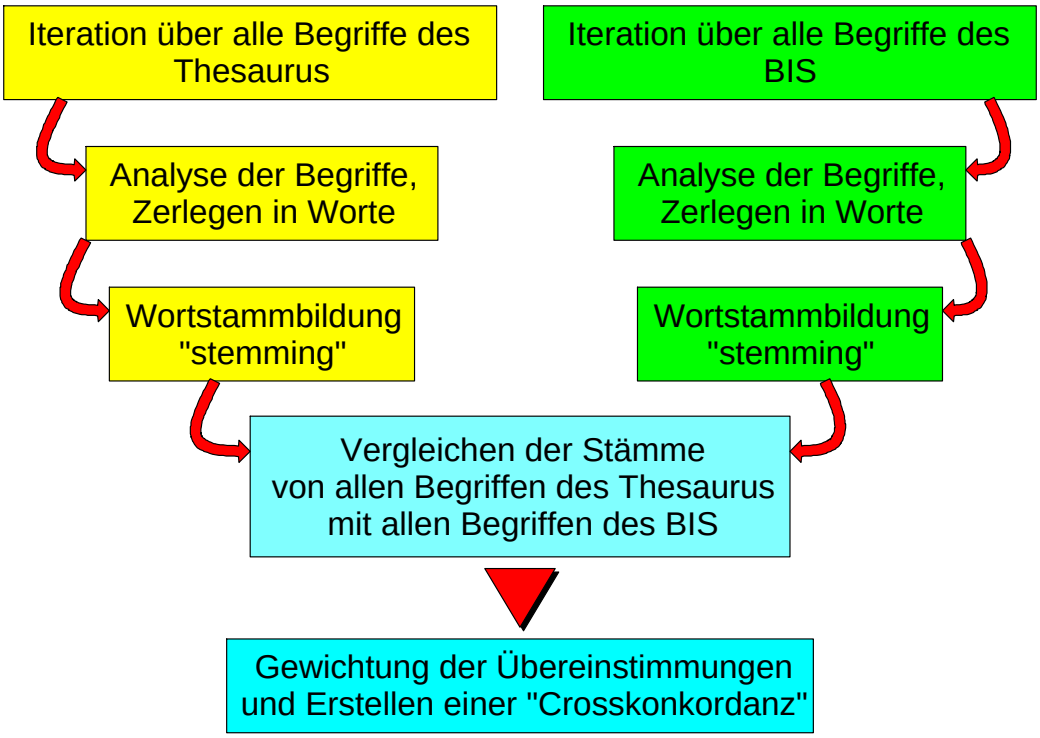


Abb. 7 : Prozess zur Bildung der Crosskonkordanzen zwischen einem Thesaurus und dem BIS

Zunächst wird, wie in Abb. 7 dargestellt, in einem Arbeitsschritt der verwendete Thesaurus mit all seinen Begriffen gelesen. Die gefundenen Begriffe, die sich zu einem guten Teil auch aus mehreren Worten pro Begriff zusammensetzen, werden in einem zweiten Schritt analysiert. Diese Analysefunktion wird von Lucene zur Verfügung gestellt und ermöglicht es auch, eine Liste von Stoppworten, also Worten, die bei der Indexierung keine Bedeutung haben, zu verwenden, so dass diese Stoppworte zur Konkordanzbildung nicht verwendet werden.

Außerdem werden alle Worte bereinigt, was für Lucene bedeutet, dass sie in Kleinbuchstaben umgewandelt werden und bestimmte Sonderzeichen innerhalb von Worten, wie z.B. Punkte oder Unterstriche entfernt werden. Die deutschen Umlaute werden in deutscher Analysefunktion von Lucene, dem „GermanAnalyzer“ in ihre Entsprechungen, also „ä“ zu „ae“, „ö“ zu „oe“, etc., umgewandelt.

Aus den bis dahin verbliebenen Worten bildet der GermanAnalyzer durch den Prozess des „stemming“ die Stammformen. Nach der Bildung der Stammformen für alle Begriffe des Thesaurus sowie für alle Begriffe des BIS erfolgt der eigentliche Vergleich: Alle Stammformen aller Begriffe des Thesaurus werden mit allen Stammformen aller Begriffe des BIS verglichen und zwar jeweils pro Begriff. Finden sich zwischen zwei Begriffen (einem Thesaurusbegriff und einem BIS-Begriff) keine gleichen Stammformen, so wird für diese beiden Begriffe keine Crosskonkordanz erstellt.

Finden sich dagegen 1..n übereinstimmende Stammformen zwischen den beiden Begriffen, so erfolgt die Ermittlung der Stärke des Zusammenhanges, des Gewichts (*weight*), nach folgendem Muster:

$$weight = \frac{2 * nEqualStems}{nStemsThesaurus + nStemsDataSource}$$

nEqualStems = *Anzahl gleicher Wortstämme zwischen Thesaurus und Datenquelle*
NStemsThesaurus = *Anzahl aller Wortstämme im Thesaurusbegriff*
NStemsDataSource = *Anzahl aller Stämme im BIS-Begriff*

Eine Crosskonkordanz wird zusammen mit dem gemäß obiger Formel ermittelten Gewicht als „Inter-Thesaurus-Relation“ gespeichert.

4.3. Verwendetes Datenmodell

Das Einbringen verschiedenartiger, fremder Thesauri in das BIS Schlüssellistensystem erforderte eine abstraktere Betrachtungsweise und resultierte in einem verallgemeinerten Datenmodell (vereinfacht dargestellt in Abb. 8), welches nachfolgend erläutert wird.

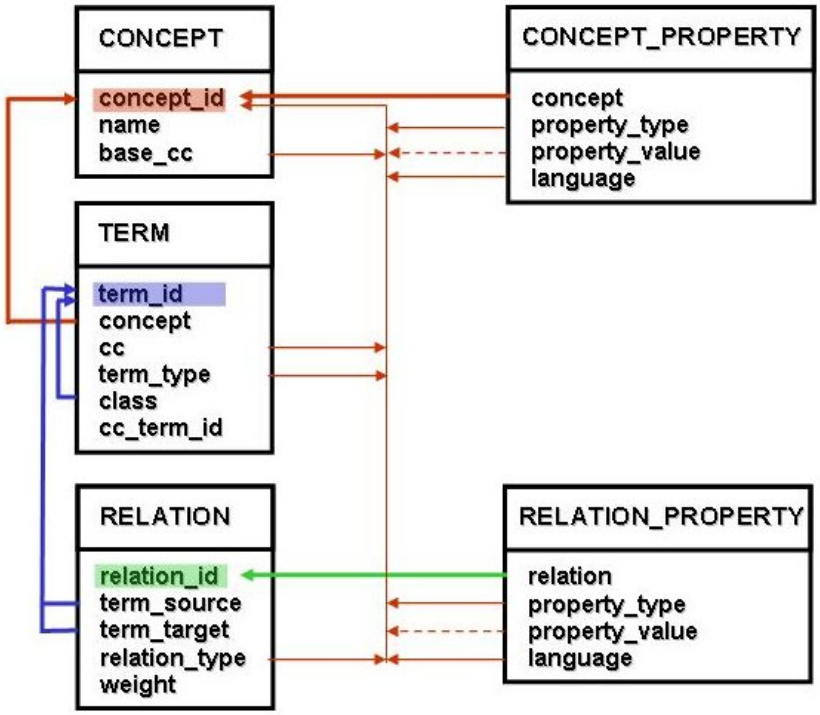


Abb 8 : verwendetes Datenmodell (vereinfacht)

Eine Schlüsselliste oder auch ein Thesaurus wird jeweils als eigene *Begriffssammlung* (*Concept Collection*, CC) angesehen. Jede Begriffssammlung besteht aus einer Menge von *Termen* (Tabelle TERM), wobei jeder Term genau einen *Begriff* (*Concept*, Tabelle CONCEPT) repäsentiert. Es kann verschiedene Begriffe mit gleichem Namen (z.B. "Bank" als Sitzgelegenheit oder "Bank" als Kreditinstitut) geben. Häufig gibt es in unterschiedlichen Begriffssammlungen Terme, welche sich auf den gleichen Begriff beziehen (z.B. "Boden" im GEMET und "Boden" in BIS-Schlüsselliste „Aufschlusstyp“ meinen jeweils den gleichen Begriff "Boden"). Somit brauchen Begriffe, welche in verschiedenen Begriffssammlungen benutzt werden, nur ein einziges Mal abgespeichert werden, wobei festgehalten wird (Feld `base_cc`), welche Begriffssammlung die Aufnahme des Begriffs erforderlich machte. Ein Begriff

hat immer einen Namen (Feld `name`) und kann durch weitere *Eigenschaften* (*Properties*, Tabelle CONCEPT_PROPERTY) gekennzeichnet sein. Properties werden über Paare von Typ (Feld `property_type`) und Wert (Feld `property_value`), ggf. in verschiedenen Sprachen (Feld language), spezifiziert.[8]

Jeder Term gehört genau zu einer Begriffssammlung (Feld cc) und ist von einem bestimmten Typ (Feld `term_type`, z.B. Deskriptor, Gruppe, Thema, etc.). Da in einigen Thesauri die Terme einer Systematik unterliegen (z.B im GEMET: Supergruppen - Gruppen - Deskriptoren), wird diese in einem eigenem Feld `class` festgehalten.[9] Es ist vorteilhaft, wenn bei der Weitergabe eines Thesaurus die diesem Thesaurus zugrunde gelegten Termidentifier (Feld `cc_term_id`[10]) enthalten sind; ein späteres Update wird damit vereinfacht.

Sämtliche Beziehungen werden in der Tabelle RELATION klassifiziert nach Beziehungstyp (Feld `relation_type`) geführt. Dies sind grob betrachtet zum einen Beziehungen zwischen Termen innerhalb einer Begriffssammlung (z.B. (poly)hierachischer Aufbau einer Schlüsselliste, bekannte Thesaurusbeziehungen und Verweise, Themenzugehörigkeit, etc.) und zum anderen Inter-Thesaurusbeziehungen; letztere werden insbesondere durch Lucene ermittelt und mit einem Ähnlichkeitsmaß (Feld `weight`) versehen. In der Regel handelt es sich um binäre Relationen (Felder term_source und term_target); sind bei einer Beziehung mehr als zwei Terme beteiligt, so werden diese Beziehungen derzeit unter Einführung von "imaginären Zwischentermen" in Paarbeziehungen aufgelöst. Weitere Informationen zu einer bestimmten Beziehung (z.B. Erfassungsdatum, Bearbeiter, etc.) werden analog zu Begriifen in einer zusätzlichen Tabelle RELATION_PROPERTY gehalten.

Das dargestellte Datenmodell nimmt die eigentlichen Daten der Begriffssammlungen (BIS-Schlüssellisten, Thesauri, etc.) und zusätzlich die Begriffssammlungen des Metamodells auf (in Abb. 8 durch dünne rote Pfeile angedeutet), dies sind: Termtypen, Beziehungstypen, Propertytypen, Sprachen und schließlich die Liste aller

[8] Aus Performanzgründen wird der Name eines Begriffs nicht als Property ausgelagert, sondern ist in der Default-Sprache der primär zugeordneten Begriffssammlung (base_cc) direkt in Tabelle CONCEPT aufgeführt.
[9] Eine mögliche Abbildung der Systematik über die Tabelle RELATION würde diese unnötig aufblähen.
[10] Da über einen Term einer Begriffssammlung derzeit nur die ggf. bekannte Original-Term-ID (cc_term_id) festgehalten wird, wurde von der Einführung einer zusätzlichen Tabelle TERM_PROPERTY abgesehen.

Begriffssammlungen.[11] In den Metabegriffssammlungen können somit auch Beziehungen (z.B.: innerhalb der Beziehungstypen ist "Hierarchie" der "Abstraktion" übergeordnet) hinterlegt oder Properties (z.B. gewünschte graphische Darstellung von Termtypen und Beziehungstypen) annotiert werden. Diese identische Grundlage für Daten und Metadaten spiegelt sich auch im daraus generierten TheVi XML Format (Kap. 5.1) wieder und ermöglicht schließlich ohne zusätzlichen Programmieraufwand auch eine Visualisierung des Metamodells (Kap. 5.2).

5. XML Austauschformat und Visualisierung

Wie in Abb. 9 dargestellt, kann die TheVi Visualisierung sowohl als Applet als auch als eigenständige Applikation am PC betrieben werden. Die vom Anwender gewünschten Begriffssammlungen werden vom GIRG (GLA Interactive Report Generator [SCH]) rechtegesteuert in einem speziellen XML Format (TheVi XML) aus der BIS-DB gelesen. TheVi XML stellt die Eingabeschnittstelle für die Visualisierung mittels TheVi dar.

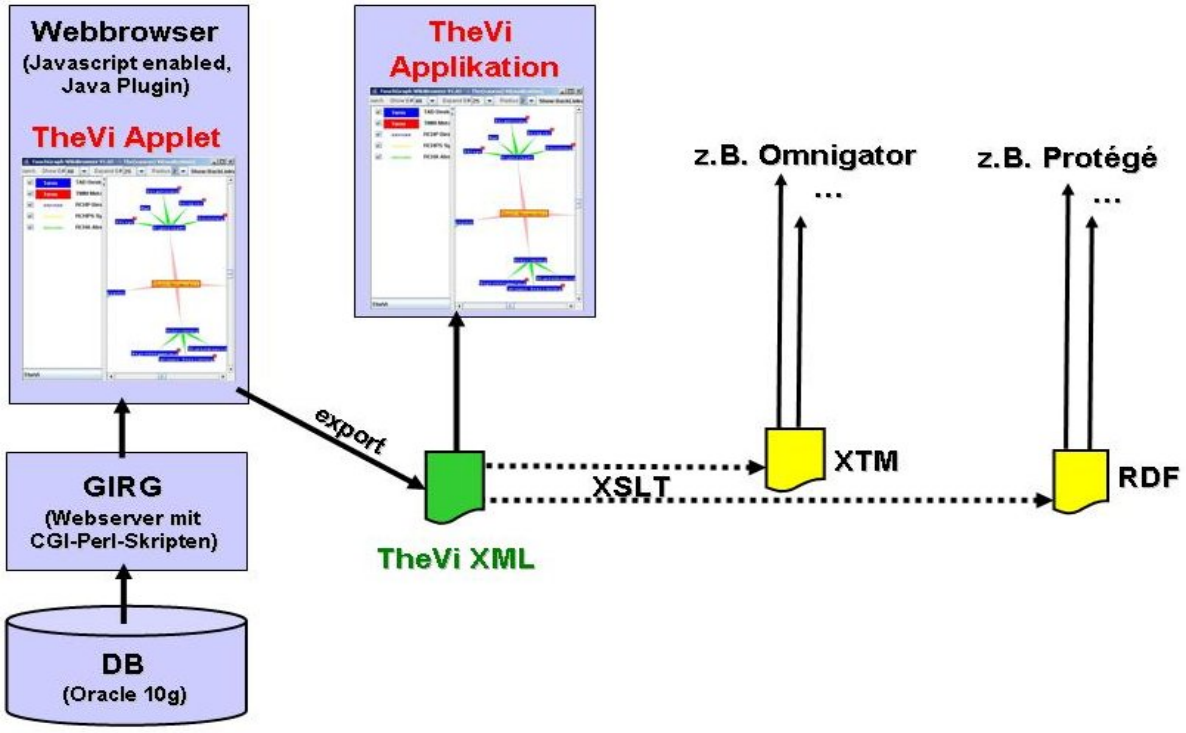


Abb 9 : Systemarchitektur

[11] Aus Performanzgründen werden die Metabegriffssammlungen derzeit zusätzlich redundant in jeweils eigenen Tabellen gehalten.

5.1. XML Austauschformat

Ein XML Format als Schnittstelle zum Visualisierungstool bringt nicht nur den Vorteil dort vorhandene Programmbibliotheken (DOM, SAX) nutzen zu können, zusätzlich kann eine XML Datei auch von Externen bequem gelesen werden und somit als Austauschformat dienen. Mit XML gewinnt man ein höheres Maß an Produktunabhängigkeit und Flexibilität, und eine Vielzahl von XML Werkzeugen stehen zur Verfügung. So können aus einer TheVi XML Datei über XSLT Skripte korrespondierende Dateien in den Standards XML Topic Map [XTM] und RDF [RDF], für welche u.a. etliche Visualisierungstools am Markt vorhanden sind, erzeugt werden (Abb. 9).

Für die Definition des TheVi XML Formats wird auf die DTD im Anhang A und die Illustrationen im Anhang B verwiesen.

5.2. Visualisierung

Bei der Arbeit mit Thesauri und den Daten des Bodeninformationssystems sowie den gebildeten Crosskonkordanzen zeigte sich immer wieder, dass die Komplexität eines solchen Thesaurisystems sich mittels rein textueller Darstellungsweise nur mehr schwer überblicken lässt. Der Thesaurusverwalter einerseits muss sofort alle Relationen, in welche ein Term verwickelt ist, sehen können; der Benutzer andererseits will nur die ihn interessierenden Typen von Relationen oder Termen sehen und über diese Relationen einfach navigieren können. Daher wird am BayGLA auch an einem Werkzeug zur Thesaurusvisualisierung (TheVi) gearbeitet. Ein auf Touchgraph [TG] basierender Prototyp stellt Terme als Knoten und Relationen als Kanten eines Graphen dar.

Um die ermittelten Crosskonkordanzen darzustellen, wurden zwei Wege erarbeitet, die jeweils auf dem in Kap. 5.1 geschilderten XML Austauschformat beruhen.

1. eine Visualisierung mithilfe des Open Source Tools Touchgraph WikiBrowser
2. weitere Visualisierungen über Tool, welche das Topic Map Format unterstützen

Der Touchgraph WikiBrowser ist ein Open Source Tool, das in Java entwickelt wurde und ursprünglich dazu dient, um die auf einem Wiki enthaltenen Informationen in grafischer Form darzustellen. Der Java Sourcecode wurde dazu so modifiziert, dass jetzt die Daten unserer Begriffssammlungen über das TheVi XML Format unter

Verwendung von JDOM [JDOM] eingelesen und dargestellt werden. JDOM ist ebenfalls ein Open Source Java Projekt, das es ermöglicht, ein XML Dokument in Java darzustellen, indem es solche Dokumente einlesen, verändern und schreiben kann.

Im Touchgraph WikiBrowser gab es initial nur zwei Arten von graphischen Objekten:

- ***Knoten***, die verwendet werden, um die Terme und Termtypen des Thesaurus darzustellen
- ***Kanten***, durch die die Beziehungen (Relationen) und Beziehungstypen zwischen zwei Termen visualisiert werden.

In Abhängigkeit von den Typen der Terme und denen der Relationen werden die Knoten und Kanten entsprechend ihrer jeweiligen Darstellungsvorschrift, welche im XML Element <representations> der Metabegriffssammlung enthalten ist, visualisiert.

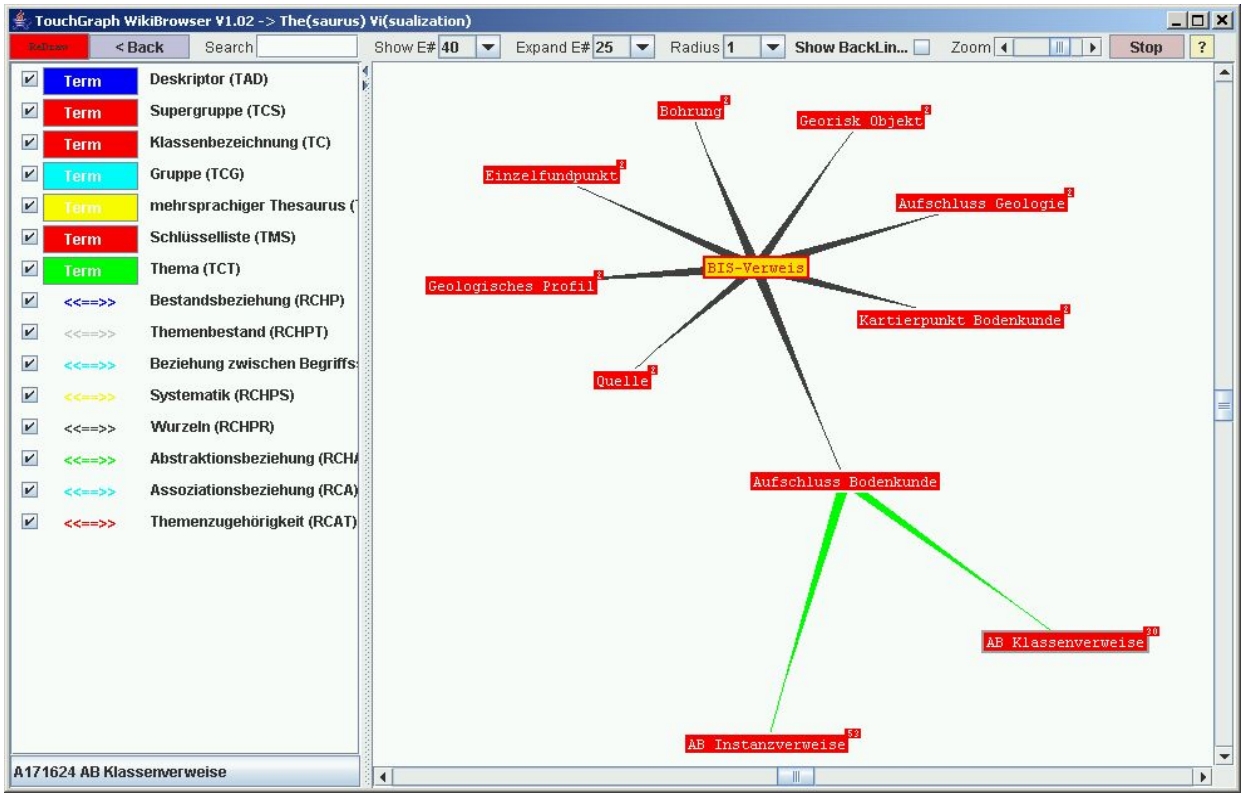


Abb. 10: Visualisierung einer Begriffssammlung mit dem Touchgraph Wikibrowser (Beispiel)

Zu der im Touchgraph WikiBrowser enthaltenen Information wurde, wie in Abb. 10 gezeigt, auch eine interaktive Legende entwickelt, über die die Typen von Kanten und Knoten identifiziert und ausserdem die gewünschten Kanten- und Knotentypen ein- oder ausgeblendet werden können.

Der Touchgraph Wikibrowser verfügt von Hause aus über die Möglichkeit nach interessierenden Knoten zu suchen, einzelne ausgewählte Knoten zu expandieren oder zu kollabieren oder diese Knoten vollständig auszublenden. Außerdem kann über den Radius die Reichweite der vom aktuell fokussierten Knoten ausgehenden, weiter anzuzeigenden Knoten festgelegt werden.

6. Ergebnisse der Methoden und Ausblick

Die Arbeiten mit den Thesauri sowie mit der gewählten Methode zur Ermittlung der Crosskonkordanzen erbrachte bisher folgende Ergebnisse:

Am Beispiel des GEMET Thesaurus zeigte sich, dass durch die Crosskonkordanzermittlung schnell sehr viele Übereinstimmungen zwischen den GEMET Begriffen und dem Fachvokabular des BayGLA gefunden wurden. Insgesamt wurden alleine zwischen dem GEMET und dem BIS Fachwortschatz 16.368.122 Konkordanzen ermittelt. Inwieweit diese Trefferquote bei anderen Thesauri sich erhöht oder erniedrigt, konnte noch nicht geprüft werden.

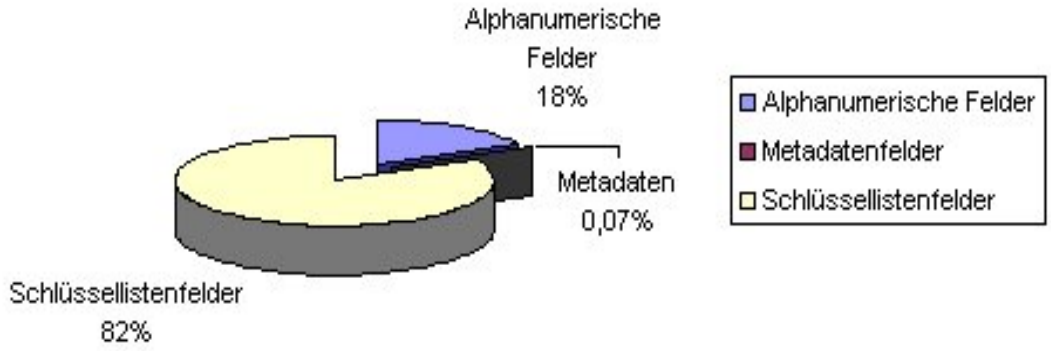


Abb. 11: Verteilung der gefundenen Crosskonkordanzen zwischen GEMET und BIS

Abb. 11 zeigt, dass sich ca. 80% aller Konkordanzen zu Inhalten der Schlüssellisten-felder, 18% zu Inhalten der alphanumerischen und nur 0,07% zu Inhalten der Metadatenfelder ergeben.

Mit dem bisher erarbeiteten Datenmodell als Grundlage zur Speicherung verschiedenster Thesauri sowie der entwickelten und erprobten Methode zur Konkordanzbildung sollen nun auch weitere Thesauri, z.B. der UOK Thesaurus, SNS sowie der UMTHES bearbeitet werden.

Zukünftig ist geplant, die einzelnen Thesauri nicht – wie es jetzt geschehen ist – redundant vorzuhalten und somit Änderungen und Updates gar nicht oder nur zeitverzögert zu realisieren, sondern direkt und online auf die verfügbaren Thesaurus Services zuzugreifen.

7. Literaturverzeichnis

[BIS1] Handbuch zum Bodeninformationssystem Bayern, 2005, Bayerisches Geologisches Landesamt

[BIS2] BIS Internet Client: http://www.bis.bayern.de (23.5.2005)

[GAUS] Gaus, W.: Dokumentations- und Ordnungslehre. Theorie und Praxis des Information Retrieval. Berlin Heidelberg New York, 2000.

[GEMET] General Multilingual Environmental Thesaurus: http://www.eionet.eu.int/GEMET (23.5.2005)

[HARDT] Hardt, M., Theis, F.: Suchmaschinen entwickeln mit Apache Lucene. Frankfurt a. Main, 2004.

[JDOM] JDOM: http://www.jdom.org/docs/faq.html (23.5.2005)

[KUNZ] Kunz, M.: Sachliche Suche in verteilten Ressourcen: ein kurzer Überblick über neuere Entwicklungen. Frankfurt / M., 2002.

[LUC] Apache Lucene: http://lucene.apache.org/java/docs/ (23.5.2005)

[NIK] Nikolai R.: Thesaurusföderationen: Ein Rahmenwerk für die flexible Integration von heterogenen, autonomen Thesauri, Diss., Karlsruhe, 2002.

[RDF] RDF (W3C Standard): http://www.w3.org/TR/rdf-primer (23.5.2005)

[SCH] Scheichenzuber J., Schinhärl J.: Flexibles Reporting und Controlling, Vortragsband zur 17. Deutschen ORACLE-Anwenderkonferenz, Mannheim, 2004, S. 150-160, ISBN-Nr. 3-928490-15X

[SCHN] Schneider, T.: Effizientes Suchen mit Jakarta Lucene, 2004.

[SNS] Semantic Network Service: http://www.semantic-network.de (23.5.2005)

[TG] TouchGraph: http://www.touchgraph.com/ (23.5.2005)

[UMTHES] Umweltthesaurus: http://www.umweltbundesamt.de/uba-datenbanken/thes.htm (23.5.2005)

[UOK] Umweltobjektkatalog Bayern: http://www.uok.bayern.de/ (23.5.2005)

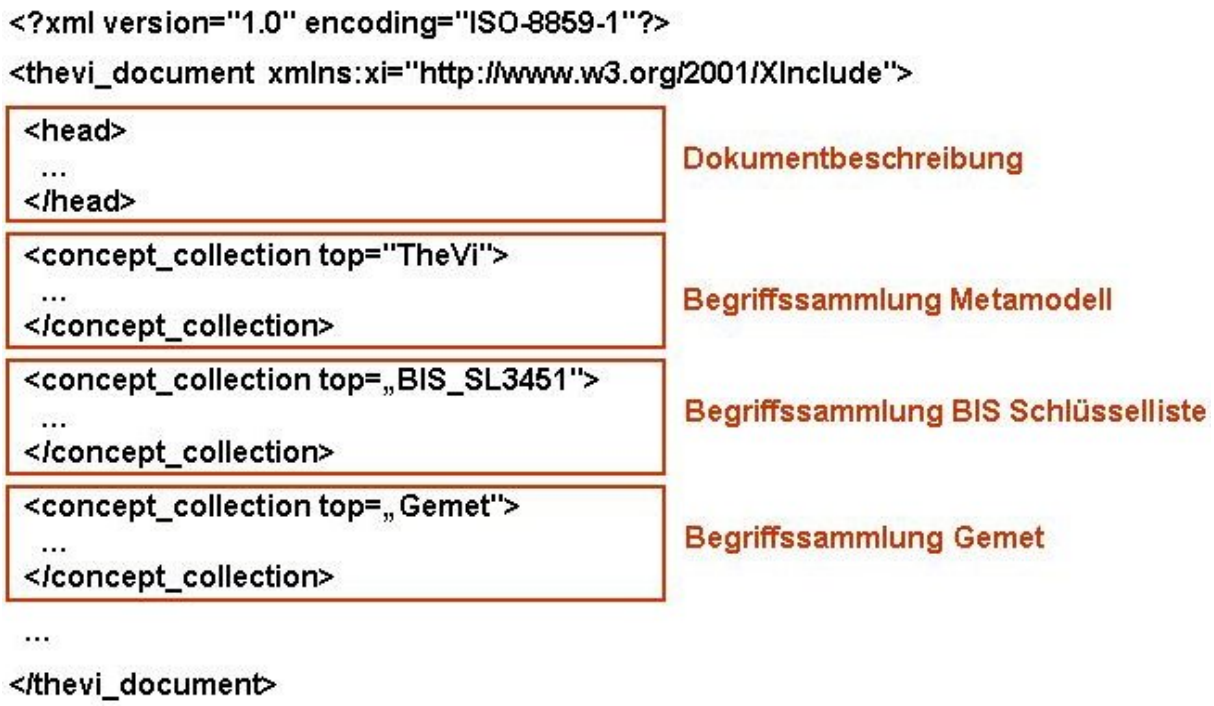
[XTM] XML TopicMaps (ISO 13250): http://www.topicmaps.org/ (23.5.2005)

Anhang A: DTD des TheVi XML Formats

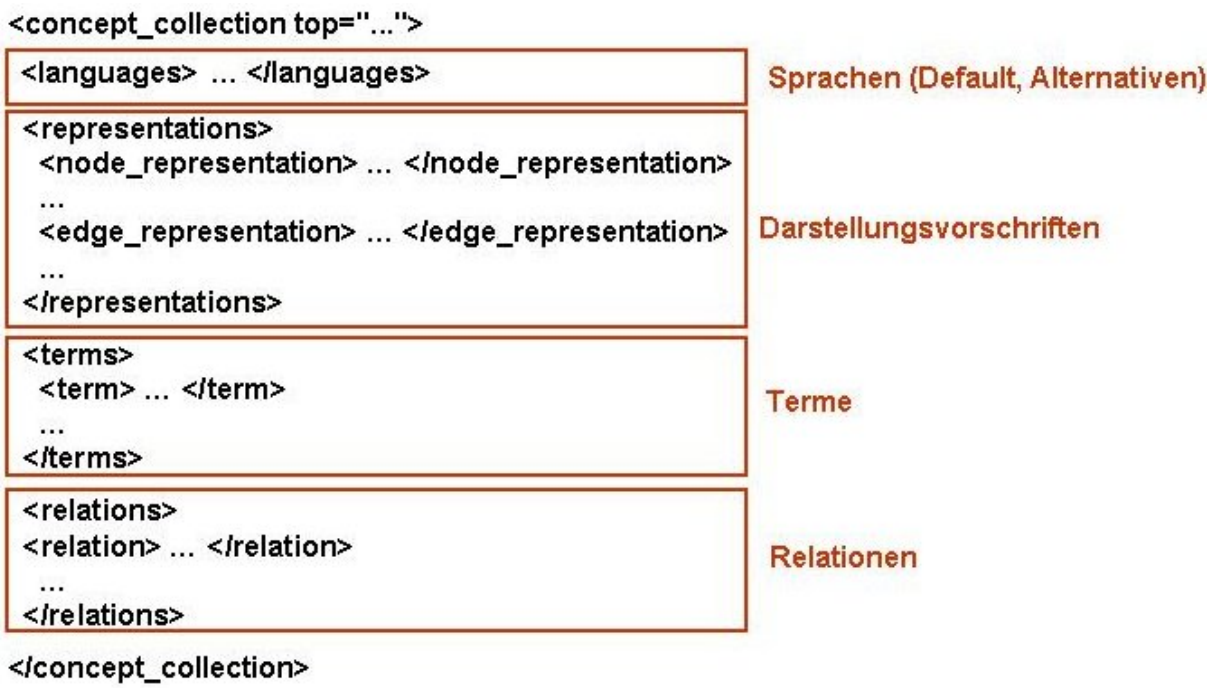
```
<!ELEMENT actuality (#PCDATA)>
<!ELEMENT alternative (#PCDATA)>
<!ELEMENT arrow (style?, direction?, color?, width?, shape?)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT border (style?, color?, width?, shape?)>
<!ELEMENT color (#PCDATA)>
<!ELEMENT concept_collection (languages, representations, terms, relations)>
<!ELEMENT date (#PCDATA)>
<!ELEMENT default (#PCDATA)>
<!ELEMENT direction (#PCDATA)>
<!ELEMENT edge (label?, tooltip?)>
<!ELEMENT edge_representation (edge?, source?, target?, arrow?)>
<!ELEMENT filling (style?, color?)>
<!ELEMENT font (#PCDATA)>
<!ELEMENT head (subject, scope, project, script, actuality)>
<!ELEMENT label (text*, style?, color?, font?, size?)>
<!ELEMENT language (#PCDATA)>
<!ELEMENT languages (default, alternative*)>
<!ELEMENT last_change (date, author)>
<!ELEMENT node_representation (label?, border?, filling?, tooltip?)>
<!ELEMENT project (#PCDATA | version)*>
<!ELEMENT property (#PCDATA | language)*>
<!ELEMENT property_type (#PCDATA)>
<!ELEMENT relation (property)*>
<!ELEMENT relations (relation*)>
<!ELEMENT representations (node_representation*, edge_representation*)>
<!ELEMENT scope (#PCDATA)>
<!ELEMENT script (#PCDATA | version | last_change)*>
<!ELEMENT shape (#PCDATA)>
<!ELEMENT size (#PCDATA)>
<!ELEMENT source (label?, tooltip?)>
<!ELEMENT style (#PCDATA)>
```

```
<!ELEMENT subject (#PCDATA)>
<!ELEMENT target (label?, tooltip?)>
<!ELEMENT term (property*)>
<!ELEMENT terms (term*)>
<!ELEMENT text (#PCDATA | property_type)*>
<!ELEMENT thevi_document (head?, concept_collection+)>
<!ELEMENT tooltip (text*)>
<!ELEMENT version (#PCDATA)>
<!ELEMENT width (#PCDATA)>
<!ATTLIST edge_representation type IDREF #REQUIRED>
<!ATTLIST language type IDREF #REQUIRED>
<!ATTLIST node_representation type IDREF #REQUIRED>
<!ATTLIST property type IDREF #REQUIRED>
<!ATTLIST relation type IDREF #REQUIRED
                   source IDREF #REQUIRED
                   target IDREF #REQUIRED
                   top CDATA #IMPLIED>
<!ATTLIST term id ID #REQUIRED
               type IDREF #REQUIRED
               top CDATA #IMPLIED>
<!ATTLIST concept_collection top IDREF #REQUIRED>
<!ATTLIST thevi_document >
```

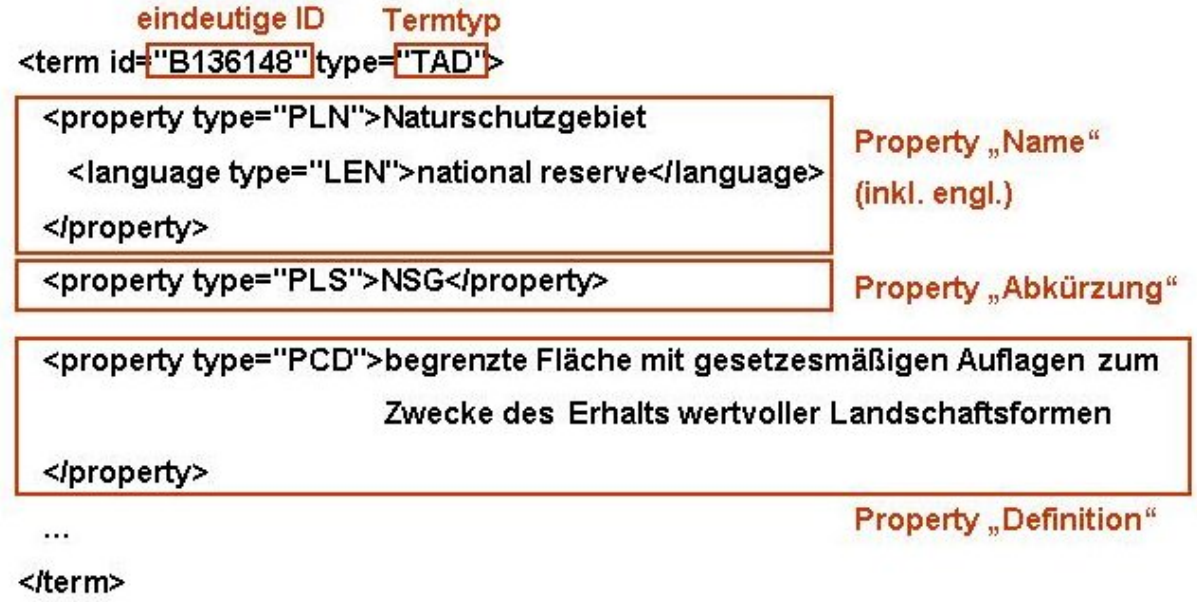
Anhang B: Illustration zum TheVi XML Format



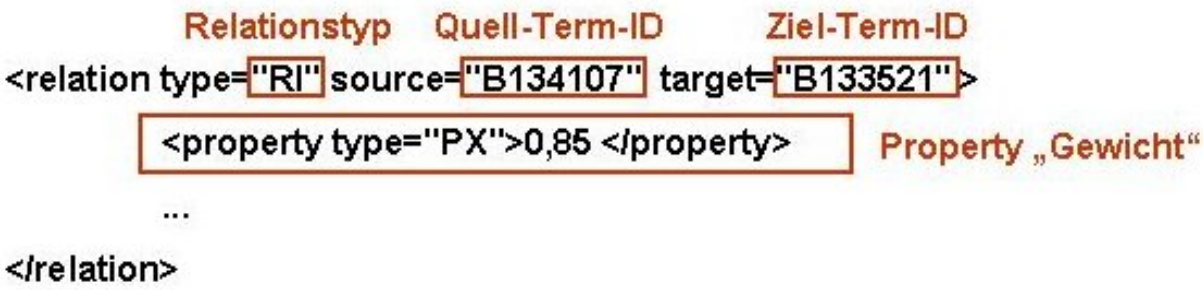
Die Grobstruktur eines TheVi XML Dokuments besteht aus einer Dokumentbeschreibung, einer Begriffssammlung für das Metamodell sowie aus beliebig vielen fachlichen Begriffssammlungen.



In jeder Begriffssammlung werden verwendete Sprachen, bei der Visualisierung anzuwendende Darstellungsvorschriften sowie alle Terme und Relationen aufgeführt.



Jeder Term ist durch eine eindeutige ID sowie einem Termtyp gekennzeichnet. Neben einem Namen in der Defaultsprache der Begriffssammlung können weitere Properties – auch in anderen Sprachen – angegeben sein.



Jede Relation bezeiht sich auf zwei Terme ist vom angegebenen Relationstyp. Weitere Informationen (z.B. Gewicht der Kante) können über Properties angehängt werden.

Ergebnisse einer Machbarkeitsstudie Integrationsschicht Umweltbeobachtung

Gerlinde Knetsch[12] Thomas Bandholtz[13]

Abstract

Der Gedanke einer integrierten Umweltbeobachtung ist nicht neu. Im engeren Wortsinne geht er auf das Sondergutachten "Allgemeine ökologische Umweltbeobachtung" [SRU 1990] zurück. Integrative Datenauswertung bedarf der Datentransparenz und der Katalyse des Datenflusses. Vorhandene technische Instrumente bieten Unterstützung, sind jedoch (noch) nicht auf eine integrative Datenextraktion der Meta-, Methoden- und Faktendaten ausgelegt. Verschiedene Sichten bezüglich des Datenzugriffs, des Datenflusses und der Datenaufbereitung für spezielle Nutzergruppen z.B. Fachexperten, allgemeine Öffentlichkeit können über geeignete innovative Technologien realisiert werden.

Innerhalb dieser Überlegungen entstand die Idee einer Integrationsschicht für Informationsbestände des UBA. Die Machbarkeitsstudie Integrationsschicht Umweltbeobachtung (IS UB) hat eine spezialisierte Sicht. Sie untersucht die rechtlichen, dokumentarischen, technischen und operativen Aspekte für die Integration bzw. Kopplung von Daten und Informationen für eine integrierte Bewertung des Zustandes der Umwelt. Durch eine übergreifende Datenauswertung aus Datenerhebungen sektoraler bzw. medienbezogener Mess- und Beobachtungsprogramme stellt sie im Kern Bewertungswissen für die Umweltpolitik zur Verfügung.

[12] Umweltbundesamt, Wörlitzer Platz 1, 16844 Dessau
email:gerlinde.knetsch@uba.de, Internet:http://www.umweltbundesamt.de;

[13] Consultant, Karl-Friedrich-Schinkelstr.2 53127 Bonn
email: www.bandholtz.info

1. Einleitung

Die Aufgabenbereiche, die durch die Integrationsschicht unterstützt werden sollen, sind keine fest definierten Verwaltungsaufgaben und lassen daher keine linearen Ableitungen von Aufgaben zu Daten und Funktionen zu. Insofern ist die Entwicklung einer Integrationsschicht ein „evolutionäres“ Vorhaben, das die kooperative Aufgabenunterstützung zwischen den Facheinheiten des UBA fördern soll. Es ist somit ein Instrument zur Verbesserung der horizontalen Kommunikation und Wertschöpfungskette einmalig erhobener Umweltzustandsdaten.

Bei der Integrationsschicht geht es darum, vorhandene und im Aufbau befindliche Informationsbestände der Umweltbeobachtung zukünftig besser zu nutzen und über eine intuitive Oberfläche einem breiten Nutzerkreis verfügbar zu machen. Die damit verbundenen Synergien fördern bidirektionale Datenflüsse und die Mehrfachnutzung von Daten der Umweltbeobachtung, ausgerichtet an den Bedürfnissen der Nutzer. Des Weiteren sind rechtliche Rahmenbedingungen Eckpfeiler zur Bereitstellung von Daten und Informationen für politische wie fachliche Entscheidungen.

2. Rechtliche Rahmenbedingungen

Die Notwendigkeit einer stärkeren Integration über alle Umweltbereiche hinweg wird aus europarechtlicher Sicht immer dringlicher. Die Mitteilung der Kommission zur Überprüfung der Umweltpolitik 2003 (KOM 2003, 274) weist ausdrücklich auf die Notwendigkeit „…der Entwicklung eines integrierten Ansatzes für die Politikgestaltung im Umweltbereich“ hin.

„Die Umweltrechtsvorschriften in Europa enthalten üblicherweise Regelungen für einzelne Schadstoffe und den Schutz einzelner Umweltbereiche, die für Umweltprobleme oft End-of-Pipe-Lösungen vorsehen und die Quelle bzw. Ursache jener Probleme oder die kombinierten Belastungen der Luftverschmutzung für

verschiedene Umweltmedien sowie die wechselseitige Abhängigkeit und Verflechtung zwischen diesen Medien außer Acht lassen.

Die Eindämmung der derzeitigen nicht-nachhaltigen Umwelttrends erfordert die Entwicklung eines integrierten Ansatzes für Politikgestaltung".

Deutlich höhere Integrationsanforderungen ergeben sich dementsprechend schon auf der Ebene der sektoralen (EU)-*Rahmenrichtlinien,* wie etwa der Wasserrahmenrichtlinie. Der Druck aus europäischem Blickwinkel lässt sich, auch im *Reportnet* innerhalb des „European Environment Information and Observation Network (EIONET)" ablesen. Dort werden über die berichtspflichtigen Zustandsdaten hinaus auch quellenbezogenen Angabe aufgeführt.

Wichtigstes Element dieser quellenbezogenen Berichterstattung ist die Entscheidung der Kommission (vom 17.Juli 2000) über den Aufbau eines *Europäischen Schadstoffregisters (EPER)* gemäß Artikel 15 der Richtlinie 96/61/EG über die integrierte Vermeidung und Verminderung der Umweltverschmutzung (IVU-Richtlinie). Sie verpflichtet einschlägige Betriebe ihre Emissionen in Luft und Wasser alle drei Jahre an die EU-Kommission zu berichten. Zu den quellenbezogenen Emissionen zählen Angaben zu 50 Schadstoffen (Klimagase, Schwermetalle, chlororganische Verbindungen), von denen 37 luftseitig und 26 gewässerseitig von Bedeutung sind.

Das EU-Weißbuch "Strategie für eine zukünftige Chemikalienpolitik - Einführung des neuen Systems zur Überwachung von Chemikalien „REACH" befasst sich in der Aktion 3H mit dem Aufbau eines Informationssystems der Schadstoffkonzentrationen in der Umwelt. Fachlicher Hintergrund dieser Aktion sind die durch Modelle abgeschätzten Expositionen von Chemikalien in der Umwelt mit Ergebnissen der Umweltbeobachtung zu verknüpfen. Durch Kombination der modellierten Expositionsdaten mit real durch Messprogramme der Umweltbeobachtung erhobenen Daten in der Umwelt (Gehalte von Chemikalien in den verschiedenen Umweltkompartimenten) können die Modellergebnisse verifiziert und die Modelle selbst validiert werden.

Auch im nationalen Recht wird der Fokus zunehmend auf eine ganzheitliche Betrachtungsweise bei der Bewertung des Zustandes der Umwelt gelegt. In § 12 des novellierten *Bundesnaturschutzgesetzes* (2002) erhielt die Umweltbeobachtung von Bund und Ländern eine gesetzliche Grundlage, zumindest was Zustand und Veränderung des Schutzgutes „Naturhaushalt“ angeht. Absatz 2 schließt ausdrücklich ein, *„… den Zustand des Naturhaushaltes und seine Veränderungen, die Folgen solcher Veränderungen, die Wirkungen von Umweltschutzmaßnahmen auf den Zustand zu ermitteln, auszuwerten und zu bewerten.“*

Die Novellierung des Umweltinformationsgesetzes (UIG 2004), welches am 14. Februar 2004 in Kraft getreten ist und die EU-Richtlinie über den Zugang der Öffentlichkeit zu Umweltinformationen (KOM 2003/4/EG) in deutsches Recht umsetzt, unterstreicht die *„… Notwendigkeit einer schutzgutübergreifenden, die Belastungsquellen mit einschließenden Informationserhebung.“* Für die Integrationsschicht von besonderem Wert ist die Festlegung auf elektronische Kommunikation. *„Die informationspflichtigen Stellen … wirken darauf hin, dass Umweltinformationen... zunehmend in elektronischen Datenbanken oder in sonstigen Formaten gespeichert werden, die über Mittel der elektronischen Kommunikation abrufbar sind.“*

Das Umweltbundesamt hat in einer Hausanordnung zur Umsetzung der Anforderungen aus dem Umweltinformationsgesetz (UIG) konkrete Vorgehensweisen für das Verwaltungshandeln beschrieben und festgelegt.

Diese nationalen sowohl internationalen Gesetzgebungen fördern die integrative Sicht und Bereitstellung von in vielfältigen Kontexten erhobenen Daten und Informationen. In diesem Sinne stellen diese gesetzlichen Grundlagen Herausforderungen auf verschiedenen Ebenen dar. Administrative, technologische und dokumentarische „Wege“ sind zu begehen, um Daten und Informationen in dem Sinne zu erschließen, die eine ganzheitliche Herangehensweise an eine Fragestellung unterstützen.

3. Nutzeranforderungen

3.1. Fachliche Ebene

Bedürfnisse definieren sich aus Anforderungen, die in umfangreichen Interviews mit den Facheinheiten des UBA im Verlaufe der Erarbeitung der Machbarkeitsstudie durchgeführt wurden. Anhand von strukturierten Fragebögen erfolgte eine Aufnahme der Nutzeranforderungen. Folgende Fachaufgaben können u.a. mit einer integrierten Sicht auf die Daten unterstützt werden:

- Darstellung und Absicherung beobachteter zeitlicher und räumlicher Trends des Umweltzustandes durch Verknüpfung mit Daten anderer Beobachtungsprogramme
- Pfadbetrachtungen für Stoffgruppen über die verschiedenen Umweltkompartimente hinweg bis zum Schutzgut menschliche Gesundheit
- Medienübergreifende Betrachtung des Verhaltens von Stoffen in der Umwelt (Bioakkumulation, Abbau, Freisetzung)
- Bereitstellung von Umweltzustandsdaten als Basis für die Entwicklung von Modellen und Szenarien (modellgeleiteter Zugriff auf verschiedene Zustandsdaten)
- Prüfung der Maßnahmen zur Minderung des Eintrags von Schadstoffen in die Umwelt (siehe Beispiel Dioxin-Bericht „Daten für Deutschland“)
- Prüfung von Vorsorgemaßnahmen – z.B. Verbotsverordnung von Tributylzinn (TBT)
- Ableitung von Qualitätszielen und ökotoxikologischen Werten für verschiedene Umweltkompartimente und Stoffe
- Verifikation von Ausbreitungsmodellen zur Abschätzung von Schadstoffexpositionen in der Umwelt (Fachbereich IV des UBA)

- Bereitstellung von Zustandsdaten für Umweltindikatoren – State-Indicators (Fachbereich I des UBA)
- Anreicherung von themenbezogenen Umweltberichten mit Ergebnissen aus der Umweltbeobachtung (Fachbereich I des UBA)

Die fachlichen Zuständigkeiten der Facheinheiten bleiben bei der Einführung einer Integrationsschicht unberührt, da die Anwender der Zustandsdatenbanken konkrete Aufgabenstellungen unter Anwendung ihrer eigenen Instrumente zu erfüllen haben z.B. für Berichtspflichten und Indikatorenableitungen.

Ziel dieser Interviews war es auch Erwartungen an die Integrationsschicht heraus zu arbeiten und „Integrationswillige“ Systeme zu identifizieren. Die Chance der Realisierung einer Integrationsschicht wächst mit der so weit als möglichen Verwendung bestehender Komponenten und Fachinformationssystemen, die unter drei Aspekten betrachtet wurden:

- Nutzung ihrer bestehenden Funktionalitäten
- Einbindung ihrer Datenbestände
- Einfluss dieser Komponenten auf die Integrationsschicht Umweltbeobachtung

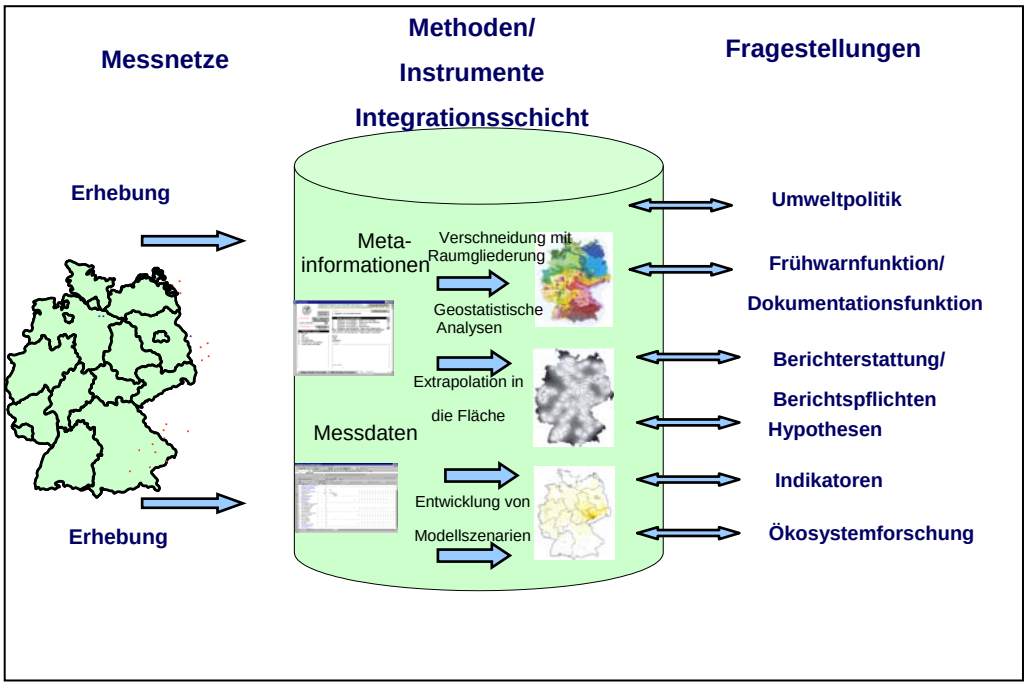


Abb.1: Methodische Schritte beim Vorgehen zur Generierung von Bewertungswissen unter Nutzung von Metadaten- und Zustandsdatenbanken

Einige dieser möglichen Komponenten können im Bereich der Zustandsdatenbanken sein:

- Informationssystem der Umweltprobenbank, IS UPB und GIS UPB

  www.umweltprobenbank.de (siehe auch Beitrag im Tagungsband)
- Web-Service Dioxin-Datenbank des Bundes und der Länder www.pop-dioxindb.de (siehe auch Beitrag im Tagungsband)
- Auswerte- und Auskunftssystem für Immissionsdaten (AAI)
- Auswertesystem wasserrechtlicher Vollzugsdaten (UDIS)
- Bundesweites Bodeninformationssystem (bBIS)

Für Metadatensysteme bieten sich folgende Komponenten zur Einbindung und Nutzung in die Integrationsschicht Umweltbeobachtung an:

- Umweltdatenkatalog des Bundes und der Länder sowie das German Environmental Information Network **(UDK/ gein®)**
  www.umweltdatenkatalog.de

  www.gein.de
  - Allgemeine Metadaten
  - technische Integrationsplattform

- Semantic Network Services (SNS) www.sns.de
  - Terminologie der Umweltbeobachtung
  - Generierung von Metadaten („Automatische Verschlagwortung“)

- Geografisches Informationssystem Umwelt (GISU) gisu.uba.de
  - Von Metadaten zu Daten und Kartenmaterial

- **Forschungsdokumentation**
  - Umweltliteratur- und Umweltforschungsdatenbank des UBA (ULIDAT/UFORDAT) www.ulidat.de, UBABSWEB
  - digital vorliegende Berichte

- **Stoffreferenzdaten**
  - Gemeinsamer Stoffdatenpool Bund-Länder www.gsbl.de
  - Meta-Informationssystem zu Chemikalien FINDEX
  - Stoffdatenbank für bodenschutz- /umweltrelevante Stoffe
  - STARS http://www.stoffdaten-stars.de/

- Umweltbeobachtungsprogrammkatalog **(UBPK)**
    - hoch aufgelöste Metadaten zu Umweltbeobachtungsprogrammen anderer Bundesressorts (Stand: 2002)

3.2. Dokumentarische Ebene

Bestehenden Metadatensysteme kommt hierbei eine besondere Rolle. Metadatenpflege bezüglich Qualität und Aktualität ist in der Vergangenheit wie heute immer noch unbeliebt. Metadatengenierung erfolgt meistens noch von Hand und Maßstäbe bzw. Kriterien für die inhaltliche Qualitätssicherung sind oft unscharf oder fehlen ganz.

Doch zunehmend entwickeln sich Technologien, die eine (teil)-automatisierte Metadatengenerierung und -extraktion unterstützen. Diese Technologien sind nicht nur auf Dokumente anwendbar sondern zunehmend auch auf strukturierte Daten.

Ein vom UBA gefördertes Forschungsprojekt Semantic Network Services [SNS 2004] hat Web-Services etabliert, über welches der UBA-Thesaurus (UmThes®), der sog. Geo-Thesaurus und der gein®-„Umweltkalender“ in Anwendungen eingebunden werden kann.

Um einen Zugang zu Umweltinformationen zu schaffen hat das Projekt Integrationsschicht Umweltbeobachtung bestehende Systeme auf ihre Nutzbarkeit hin untersucht. So konnten im Rahmen der Planungen zur Weiterentwicklung der Systeme UDK / gein® fachliche Anforderungen zur Einbindung von Fachanwendungen des UBA in UDK / gein® auf der 7. Lenkungsausschusssitzung der Bund/Länderkooperation UDK / gein® eingebracht werden. Die Potentiale dieser beiden Systeme sind für die Entwicklung einer Integrationsschicht ausbaufähig und im Zuge der Weiterentwicklung zum Portal U (siehe erster Beitrag dieses Tagungsbandes) zu beeinflussen.

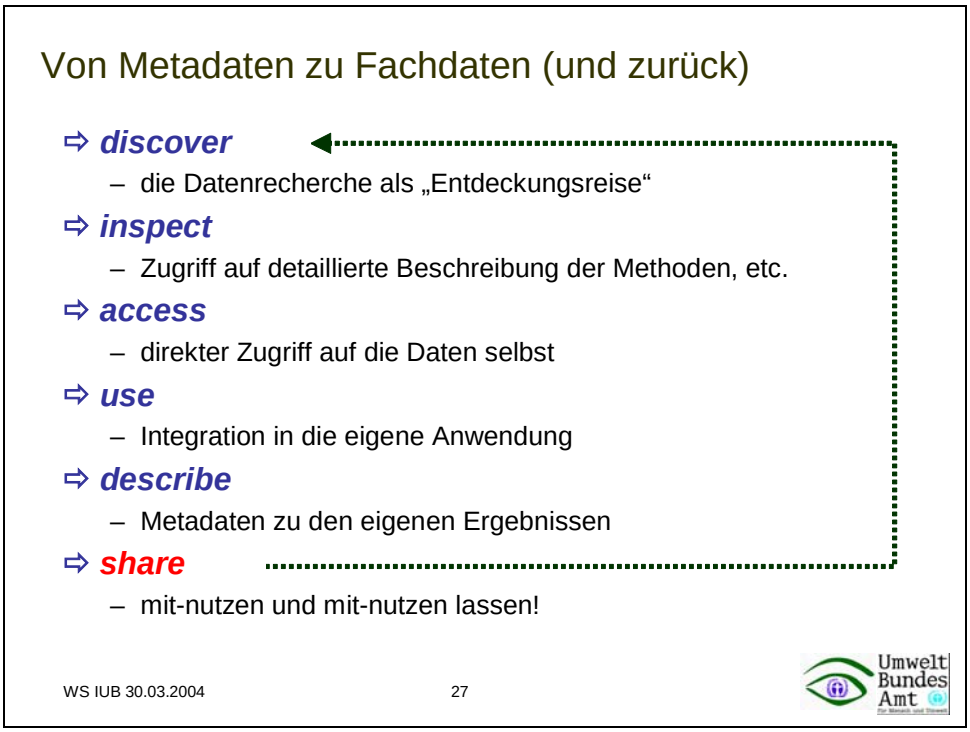


Abb.2: Erschließung von Fachdaten über Metadaten, Quelle: Projektworkshop im UBA am 30.3.2004

3.3. Technische Ebene

Ein Ergebnis der Machbarkeitstudie war die Strategie, keine Eigenentwicklung einer technischen Integrationsschicht vorzunehmen, sondern die zur selben Zeit (2003/2004) geplante Weiterentwicklung von gein® 2.0 (heute: PortalU) entsprechend zu beeinflussen und mit zu nutzen. Eine derartige Nutzung durch die Partner der Verwaltungsvereinbarung UDK/gein® war und ist grundsätzlich vorgesehen, und die Anforderungen der Integrationsschicht sind nach wie vor mit den erklärten Zielen dieser Kooperation konform. Parallel dazu sollten die knappen Projektmittel der Integrationsschicht Umweltbeobachtung im UBA dazu eingesetzt werden, um entsprechende Schnittstellen bei den einzelnen beteiligten Datenbanken bzw. Anwendungen zu koordinieren.

Auf dieser Grundlage wurde die Frage nach der Machbarkeit Anfang 2004 aus technischer Sicht positiv beantwortet.

Die in der Machbarkeitsstudie vorfolgte Technologie würde man heute als "Service Orientierte Architektur (SOA)"[14] bezeichnen. Man muss in diesem Zusammenhang einmal mehr betonen, dass unter "Web Service" nicht jede Darstellung von

[14] vgl. den gleichnamigen GI-Arbeitskreis www.soai.org

Informationen ("Service") im Internet ("Web") verstanden wird, sondern eine spezielle Form von maschinenlesbarer Schnittstelle, die spezifische Anfragen (requests) in wohl definierter Weise beantwortet (response), üblicherweise in Schema-basierten XML Strukturen. Die SOA erweitert dies zu einer Gesamtarchitektur, in der die einzelnen Komponenten (Datenbanken und Anwendungssysteme) ihre Aufgabe darin sehen, derartige Services bereit zu stellen. SOA Anwendungen können dann in dieser "Landschaft" wohl definierter Services ihre spezifische Auswahl und Zusammenstellung verfolgen, was auch als gerne bildhaft als "Choreographie" bezeichnet wird. Zur SOA gehören entsprechende Verzeichnisse der Services und ihrer Besonderheiten sowie einige grundlegende Integrationswerkzeuge. Zunehmend wird auch die Notwendigkeit einer gemeinsamen Ontologie innerhalb einer Service-Community gesehen.

Diese Strategie ist aber im Anschluss an die Studie nicht verfolgt worden.

Die konkrete Konzeption und gegenwärtige Umsetzung von gein® 2.0 (jetzt unter dem Namen PortalU) steht einem derartigen Vorhaben nicht prinzipiell entgegen, die im Kontext der Studie vorgeschlagenen Schwerpunkte sind in die aktuelle Konzeption jedoch nicht eingegangen. PortalU wird zwar bis zu einem gewissen Grade Elemente enthalten, die auch der Konzeption einer Integrationsschicht ähnlich sind (wie dies auch bereits für das bisherige gein® gilt), die Schwerpunkte liegen jedoch auf anderen Aspekten, vornehmlich auf der öffentlichen Präsentation von Umweltinformationen mit Hilfe einer gegenüber gein® wesentlich verbesserten Volltextsuche. In welchem Ausmaß die Weiterentwicklungen dennoch der Machbarkeitsstudie entgegen kommen, kann erst nach dem für Mitte des Jahres (2006) geplanten Relounch bewertet werden.

Der "integrative" Aspekt von PortalU oder gein® beschränkt sich in jedem Fall darauf, dass die Recherche alle Quellen gemeinsam durchsuchen kann. Was gegenüber den Konzeption der Studie vollkommen fehlt, sind Werkzeuge, die eine Integration der Ergebnisse ermöglichen oder wenigstens erleichtern. Es ist nicht auszuschließen, dass PortalU künftig in dieser Richtung ausgebaut wird, Planungen hierzu liegen jedoch noch nicht vor.

Aber auch die Koordination der zu integrierenden Informationssysteme im UBA wurde nicht wie vorgeschlagen verfolgt. Umstrukturierungen und Umzug der Behörde führten zunächst dazu, dass das Projekt Integrationsschicht Umweltbeobachtung nach der Machbarkeitsstudie gar keine Fortsetzung fand. Insofern sind einzelne Weiterentwicklungen der Systeme vielleicht von der Studie beeinflusst oder auch "zufällig" konform, die beabsichtigte Synchronisierung von Schnittstellen und Service-Paketen durch einen entsprechend autorisierten Arbeitskreis "Fachübergreifende Integration" ist aber vollständig ausgeblieben.

Von daher kann der Erfolg der vorgeschlagenen technologischen Strategie heute nicht nach seinen praktischen Ergebnissen bewertet werden. Die kritische Betrachtung der Ende 2003 entstandenen Vorschläge aus der heutigen Perspektive führt nicht zu grundsätzlichen Revisionen. Lediglich einige angesprochene Standards sind inzwischen weiter entwickelt worden, was die Konzeption aber gerade bestätigt und ihre Umsetzung vereinfachen könnte. Derartige Aktualisierungen vorausgesetzt, können die technologischen Empfehlungen der Machbarkeitsstudie auch heute gelten und umgesetzt werden.

4. Zusammenfassung und Ausblick

Die Ergebnisse der Interviews mit MitarbeiterInnen des UBA und der Analyse der Workflow-Prozesse spezieller Aufgabengebieten ermöglichten es, 35 Indikatoren der Machbarkeit für eine Integration von Daten der Umweltbeobachtung herauszuarbeiten. Diese Indikatoren benutzen die Klassifikation der Europäischen Umweltagentur – ungünstiger, neutraler, positiver Trend – und symbolisieren den gegenwärtigen Stand. Aufbauend darauf kristallisierten sich 32 Empfehlungen für die Umsetzung der Ergebnisse der Machbarkeitsstudie heraus. Diese Empfehlungen beziehen sich auf vier Themenbereiche:

1. **kulturell/Leitbildorientiert**

    a) Image von fachübergreifender Arbeit

2. **fachlich/organisatorisch**

    b) interdisziplinäre Kommunikation

    c) zentrale Archivierung digitaler F&E Dokumentation

3. **technisch**

    a) Besser nutzbare Metainformation

    b) Zugang von Metadaten zu Fachdaten

    c) digitale Verfügbarkeit & Volltextsuche in Forschungsberichten

4. **wirtschaftlich**

    a) Kosten der Informationsbeschaffung heute

    b) entgangener Nutzen durch nicht entdeckte Daten

    c) Nutzung/Ausbau bestehender Werkzeuge

Letztendlich bewies eine Wirtschaftlichkeitsbetrachtung, dass der Nutzen einer Umsetzung Integrationsschicht Umweltbeobachtung positiv zu bewerten ist. Beispiele für entgangenen Nutzen wurden in allen Fachgesprächen im Rahmen der Studie berichtet. Diese Erkenntnisse trägt das Umweltbundesamt in verschiedene Gremien, Kolloquien, Workshops und IT-Projekte hinein, um Synergien zwischen den verschiedenen Akteuren der Umweltbeobachtung zu fördern.

Quellen:

SRU (1990): Der Rat von Sachverständigen für Umweltfragen (SRU, www.umweltrat.de), Allgemeine Ökologische Umweltbeobachtung, Sondergutachten Oktober 1990. Stuttgart: Metzler-Poeschel, 1991. (Bundestags-Drucksache 11/8123)

Bandholtz, Thomas (2004): Machbarkeitsstudie Integrationsschicht Umweltbeobachtung im Auftrag des Umweltbundesamtes, UMPLIS-Projekt it087, unveröffentlicht

EU-KOM (2000): Richtlinie 2000/60/EG des Europäischen Parlaments und des Rates zur Schaffung eines Ordnungsrahmen für Maßnahmen der Gemeinschaft im Berreich Wasserpolitik,
http://www.umweltbundesamt.de/wasser/themen/wrrl_chronologie 3.htm
EU-KOM (2000): Entscheidung der Kommission 2000/479/EC vom 17.Juli 2000 zur Einrichtung eines Europäischen Schadstoffregisters EPER http://www.eper.cec.eu.int

EU-KOM (2001): EU-Weissbuch – Strategie für zukünftige Chemikalienpolitik http://europa.eu.int/eur-lex/de/com/wpr/2001/com2001_0088de01.pdf

EU-KOM 2003/4/EG: EU-Richtlinie über den Zugang der Öffentlichkeit zu Umweltinformationen (UIG)

http://www.bmu.de/files/pdfs/allgemein/application/pdf/ui_richtlinie.pdf

Bandholtz, Thomas (2004): Erstellung eines semantischen Netzwerkservices (SNS) für das Umweltinformationsnetz Deutschlands - German Environmental Information Network (gein®), UFOPLAN 201 11612, Abschlussbericht

Knetsch, Gerlinde; Rosenkranz, Dietrich (2003): Umweltbeobachtung – Konzepte und Programme des Budnes. In [LfUG-SH 2003]

Harmonisierung heterogener Grundwasser-Informationsbestände auf Basis eines dynamischen Datenbank-Mappings

Prof. Dr.-Ing. Uwe Rüppel, Dipl.-Ing. Thomas Gutzke, Dipl.-Ing. Peter Göbel

Institut für Numerische Methoden und Informatik im Bauwesen,
Technische Universität Darmstadt

Dipl.-Ing. Gerrit Seewald
CIP Ingenieurgesellschaft mbH, Darmstadt

Prof. Dr.-Ing. Michael Petersen
Fachgebiet Umweltinformatik, Fachhochschule Lippe und Höxter

Abstract

Die Verwaltung von sich über die Zeit akkumulierenden, großen Datenmengen ist substanzieller Bestandteil moderner Grundwasserbewirtschaftungssysteme und bildet die Grundlage für die analytische Beurteilung und Steuerung der wasserwirtschaftlichen Situation. Im Hinblick auf eine übergreifende Überwachung und Steuerung eines Monitoringgebietes ist die Übernahme hydrogeographisch zusammenhängender Informationsbestände unerlässlich, um auf einer einheitlichen Informationsgrundlage arbeiten zu können. Ein Abgleich der Rohdaten zwischen unterschiedlichen Bewirtschaftungssystemen ist jedoch grundsätzlich mit einem hohen Aufwand für den Export, die Datenaufbereitung und den anschließenden Import verbunden. Besondere Betrachtung erfordern dabei strukturelle Aspekte heterogener Datenhaltung (Datenhierarchien und –Beziehungen) sowie sprachlich-inhaltliche Gesichtspunkte (Synonyme).

Das in diesem Beitrag vorgestellte Mapping-System bietet eine Lösung zur Harmonisierung heterogener Grundwasserbewirtschaftungsmodelle mit dem Ziel, unterschiedliche Datenquellenarten mit der Zieldatenbank “Grundwasser-Online“ abzugleichen, um somit eine vollständige Datengrundlage für eine großflächige Grundwasserbewirtschaftung zu erhalten. Getätigte Zuordnungen inhaltlicher Daten-

konflikte (Synonyme) sowie das Mapping komplexer Datenstrukturen werden getrennt mittels XML-Schemata gespeichert und stehen – unabhängig voneinander und unabhängig von der Art der Datenquelle – für eine Wiederverwendung bzw. für Erweiterungen zur Verfügung. Durch dieses Verfahren kann die Übernahme und Aktualisierung verteilter Datenbestände gewährleistet, Informationsverluste vermieden und der Aufwand für die Datenaufbereitung minimiert werden. Darüber hinaus wurden Methoden entwickelt, die eine Veredlung der Daten ermöglichen. Über parametriesierbare Rechenoperationen wird dabei die Plausibilität der zu importierenden Daten überprüft und dem Fachanwender zur inhaltichen Korrektur angeboten.

1. Einleitung

In Deutschland wird Trinkwasser zu mehr als 70 Prozent aus Grundwasser gewonnen – in Hessen sogar zu mehr als 95 Prozent. Grundwasser steht in komplexen hydrogeologischen, klimatischen und wasserwirtschaftlichen Wechselbeziehungen zu Flora, Fauna und Habitat. Niedrige Grundwasserstände können beispielsweise zu Setzungsrissen an Gebäuden, Waldsterben und Erntevernichtungen durch Austrocknung führen, während zu hohe Grundwasserstände überflutete Keller und überschwemmte Ackerflächen mit Ernteausfällen nach sich ziehen können. Um allerdings die Grundwassersituation in großen hydrogeologischen Einzugsgebieten aktiv steuern zu können, ist die genaue Kenntnis aller räumlichen und zeitlichen Einflussparameter erforderlich.

1.1.Motivation

Die feingliedrigen Strukturen der Wasserwirtschaft in Deutschland mit ca. 6.500 Wasserversorgungsunternehmen, einem föderalen Behördenapparat und zahlreichen externen Umweltingenieurbüros, die häufig für Erfassung, Datenbestandszusammenführung, Aus- und Bewertung beauftragt werden, führen dazu, dass die Datenbestände räumlich verteilt vorliegen und in unterschiedlichen Systemen erfasst und abgespeichert werden.

Rechtliche und vor allem wirtschaftliche Aspekte sind aktuell die treibenden Faktoren, die Anlass zu strukturellen Änderungen geben. In vielen Regionen können Firmenfusionen und intensive Kooperation zwischen den verschiedenen Beteiligten

beobachtet werden, wobei ein z.T. kontinuierlicher Abgleich von Datenbeständen erforderlich ist. Die Einführung der EG-Wasserrahmenrichtlinie kann als weiterer initiativer Aspekt angesehen werden. Die hierbei angestrebte, großflächige Bewirtschaftung einzelner Wassereinzugsgebiete/Flussregionen setzt einen intensiven Datenaustausch bzw. Datenabgleich zwischen den verschiedenen, räumlich verteilt agierenden Akteuren voraus.

1.2. Ausgangssituation

Das zur vollständigen Überwachung (und Steuerung) erforderliche Grundwasser-Monitoring umfasst neben einer Fülle an objektspezifischen Stammdaten eine Vielzahl unterschiedlicher Objekte und Gerätschaften mit entsprechenden Messgrößen und Erfassungsformaten:

- An den Entnahmeobjekten „Brunnen“ und „Quelle“ sowie bei den zur Grundwasseranreicherung einsetzbaren „Infiltrationsorganen“ werden Wassermengen über Durchlaufzähler mit angeschlossenen Datenloggern erfasst und in herstellerspezifischen ASCII-Dateien abgespeichert.
- Aussagen über die hydrogeologische Situation des Grundwasserkörpers und der Oberflächengewässer sowie deren Einflussparameter werden über „Grundwasser- und „Gewässermessstellen“ sowie an „Klimamessstationen“ aufgenommen. Dabei herrschen z.Z. noch manuelle Ablesungen vor Ort vor, die i.d.R. händisch erfasst und später in ein digitales Format übertragen werden. Die manuelle Erfassung wird allerdings auch in diesem Bereich sukzessive durch Datenlogger (ohne und mit DFÜ) abgelöst.
- Wasserproben, die eine Aussage über den qualitativen Zustand des Grund- bzw. Trinkwassers erlauben, werden an verschiedenen Stellen entnommen, in Laboratorien untersucht, protokolliert und den Wasserversorgern bzw. den zuständigen Behörden i.d.R. tabellarisch zur Verfügung gestellt.

Die heterogenen Formate der verteilt vorliegenden Datenbestände reicht dabei von Listen in Papierform über ASCII-Dateien und Excel-Tabellen bis hin zu komplexen Grundwasserbewirtschaftungssystemen auf Basis relationaler Datenbankschemata. Die Datenvolumina erreichen je nach Messintervall sehr große Mengen, die eine manuelle Übernahme aus Effizienzgründen und Aspekten der Fehleranfälligkeit i.d.R.

ausschließen. Einheitliche Formate bzw. Schnittstellendefinitionen oder gar eine gemeinsame Ontologie im Bereich der Wasserwirtschaft existieren derzeit nicht. Stattdessen jedoch eine Vielzahl von Insellösungen, die über ungenügende Schnittstellen verfügen.

2. Anforderungen an ein dynamisches Mapping

Ziel der entwickelten und in diesem Beitrag vorgestellten Applikation ist eine möglichst vollständige Übernahme heterogen vorliegender Datenbestände in eine Zieldatenbank am Beispiel von „Grundwasser-Online“ (www.grundwasser-online.de). Die Basis der Zieldatenbank bildet ein uniformes, relationales Datenbankschema, welches auf der Schnittstellendefinition des Hessischen Landesamtes für Umwelt und Geologie (HLUG, v1.13 vom 14.11.2000) aufsetzt und sukzessive aufgrund zusätzlicher Erfordernisse der verschiedenen Beteiligten unter Beachtung der ersten drei Normalformen erweitert wurde. Die Zieldatenbank umfasst derzeit ca. 130 Tabellen.

Um einen an die verschiedenen fachlichen Bedürfnisse angepasstes Mapping-Tool zu entwickeln, wurden in einem ersten Schritt die folgenden Anforderungen definiert:

- Unterstützung von ASCII-Dateien aus Datenloggern und Durchlaufzählern, beliebiger Excel-Tabellen und Datenbank-Schemata ab der 1. Normalform
- Zuweisungen von beliebig komplexen Strukturen und Synonymen
- Effiziente Erstellung und Speicherung der Zuweisungsschemata
- Prüfung, Nachbearbeitung und Wiederverwendung bestehender Schemata
- Aktualisierung (Update) von Datenbeständen der Zieldatenbank
- Konsistenzhaltungsmechanismen im Konfliktfall
- Veredlung der Mess- und Mengendaten durch Plausibilitätsprüfungen
- Protokollierung der Importprozeduren
- Bereitstellung einer UnDo-Funktion zur vollständigen Wiederherstellung (Rollback)

3. Technische Realisierung

3.1. Grundlagen der Konzeption

Ein grundlegender Ansatz ist die getrennte Betrachtung strukturellen und inhaltlichen Mappings (Abb. 1). Zuerst wird die Zuweisung der Quelldatenstruktur an die der vorgegebenen Zieldatenbank vorgenommen. In einem zweiten Schritt erfolgt die Erstellung der Importdatei, wobei die Quelldaten objektweise ausgelesen und inhaltlich von der Applikation in Bezug auf typenverträglichkeit bzw. Synonyme und Wertebereiche überprüft werden. Typenunverträglichkeiten bzw. nicht definierte Begrifflichkeiten werden von einem Fachanwender individuell behandelt und in der Synonymdatei abgelegt. Beide Mapping-Arten werden in separaten Schemata verwaltet und in Form von XML-Dateien gespeichert. Dabei sind die strukturellen Mapping-Einstellungen (Struktur-Schemata) an die Datenquelle gebunden. Die inhaltichen Mapping-Einstellungen (Synonym-Schemata) können hingegen ungebunden verwaltet und genutzt werden, so dass die gleiche Synonymdatei beispielsweise als Basis für die Erstellung einer Importdatei aus einer Excel-Quelle und einem RDBMS des gleichen Datenlieferanten verwendet werden kann.

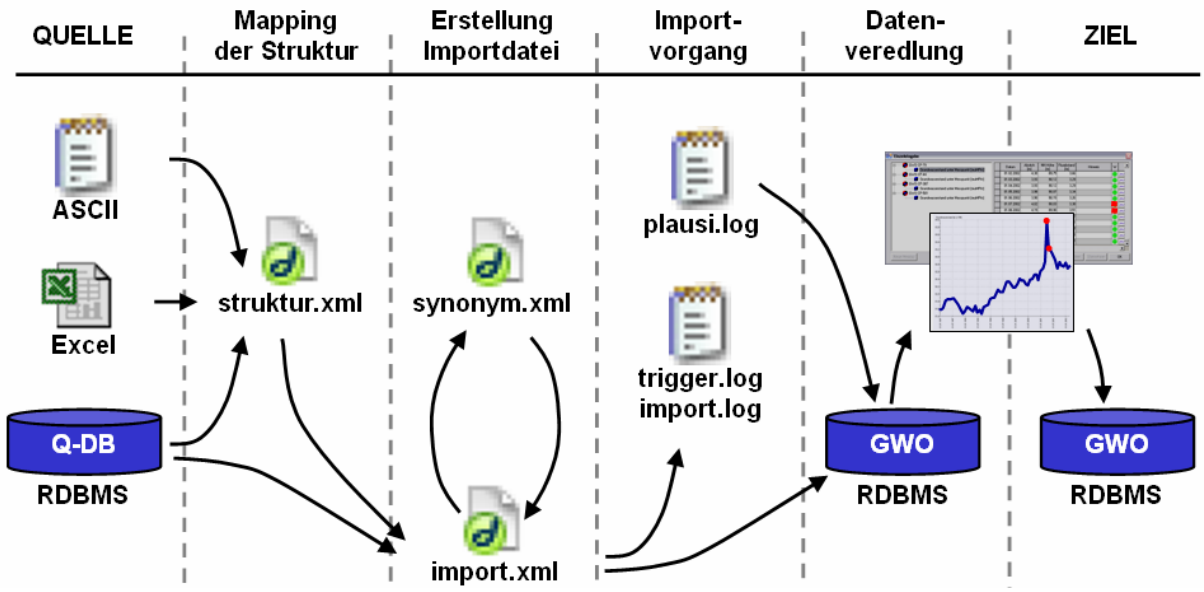


Abb. 1: Skizze des Gesamtkonzepts

Nachdem alle inhaltlichen Prüfungen abgeschlossen, die Synonymdatei aktualisiert und die Importdatei vollständig erstellt wurde, kann der eigentliche Importvorgang stattfinden. Während des Importierens der Dateien aus der Importdatei in die Zieldatenbank werden alle Ereignisse (insert, update, delete) mit weiteren Zusatzinformationen im Importprotokoll zur einfachen Nachvollziehbarkeit abgelegt. Um den Importvorgang im Bedarfsfall vollständig rückgängig machen zu können, werden die Triggerstände der Zieldatenbank in einer separaten Datei gespeichert. Mess- und Mengenwerte, die nicht den definierten algorithmischen Plausibilitäts-kriterien entsprechen, werden ebenfalls in einem separaten Protokoll vorgehalten. In einem letzten Schritt wird der Fachanwender über die nicht eingehaltenen Kriterien benachrichtigt und kann ggf. Korrekturen am Datenbestand vornehmen (siehe 3.6).

3.1.1. Struktur-Schema

Das Struktur-Schema beinhaltet Informationen darüber, welche Attribute in der Datenquelle vorhanden sind und wo genau diese zu finden sind (Tabellenname, Excel-Datenblatt etc.).

Der Aufbau dieses Struktur-Schemas (siehe Abb. 2) ist in drei Teile gegliedert:

1. Haupttabelle

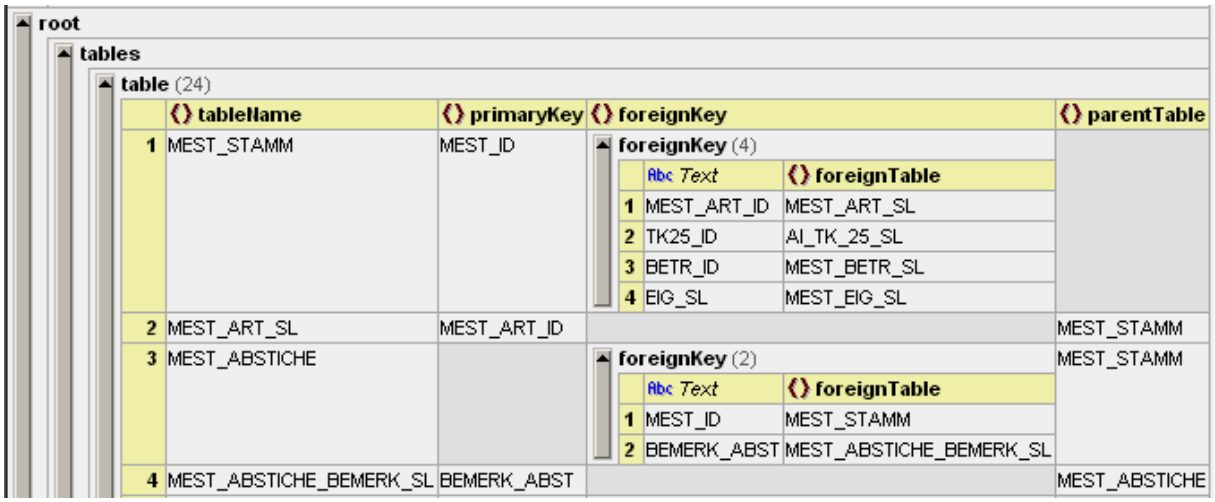
Dies ist die Tabelle, bei der jedes Tupel ein wasserwirtschaftliches Objekt (z.B. einen Brunnen) repräsentiert und in der Regel an zentraler Stelle innerhalb der untersuchten relationalen Datenbanken vorliegt. Diese Tabelle wird als Ausgangs-punkt bei der Navigation über die Quelldaten benutzt. Sollte die Quelldatenbank mehrere Haupttabellen besitzen, so wären mehrere Mapping- und Importvorgänge durchzuführen.

2. Definition der Beziehungen (Relationen)

Dieser Teil enthält Angaben zu Primär- und Fremdschlüsseln (mit dem Namen der Tabelle auf die der Fremdschlüssel zeigt = “foreignTable“) und dem Namen der nächsten Tabelle “in Richtung“ der Objekttabelle.

3. Attributzuweisungen

Ein letzter Teil schließlich speichert die Zuordnung einzelner Attribute der Quell-datenbank zu Spalten und Tabellen.



root
tables
table (24)

	tableName	primaryKey	foreignKey	parentTable
1	MEST_STAMM	MEST_ID	foreignKey (4)	
2	MEST_ART_SL	MEST_ART_ID		MEST_STAMM
3	MEST_ABSTICHE		foreignKey (2)	MEST_STAMM
4	MEST_ABSTICHE_BEMERK_SL	BEMERK_ABST		MEST_ABSTICHE

foreignKey (4)

	Text	foreignTable
1	MEST_ART_ID	MEST_ART_SL
2	TK25_ID	AI_TK_25_SL
3	BETR_ID	MEST_BETR_SL
4	EIG_SL	MEST_EIG_SL

foreignKey (2)

	Text	foreignTable
1	MEST_ID	MEST_STAMM
2	BEMERK_ABST	MEST_ABSTICHE_BEMERK_SL

Abb. 2: Beispiel der Strukturdatei

3.1.2. Synonym-Datei

Dieses Schema speichert die inhaltlichen Zuweisungen und Einstellungen jeder unterstützten Art von Datenquelle. Synonyme (ggf. in hierarchischer Abhängigkeit, wie bei Betreiber, Wasserwerk, Anlage) und Filter (z.B. Objekte vom Betreiber XY oder vom Typ YZ nicht importieren) können hier abgelegt werden.

Beim Auslesevorgang wird aus Performancegründen zunächst die Synonymdatei auf inhaltliche Übereinstimmungen geprüft und ggf. direkt zugewiesen. Werden keine Übereinstimmungen gefunden wird in einem zweiten Schritt die entsprechende Schlüsselliste durchsucht. Bei neuen Begrifflichkeiten erfolgt die Interaktion mit dem Benutzer, so dass eine gezielte Erweiterung der Synonymdatei möglich ist.

Objekttyp
Zuordnung (5)

	GWM	User
1	Grundwassermessstelle	Grundwassermessstelle
2	Grundwassermessstelle	Grundwassermeßstelle
3	Brunnen	Brunnen
4	Brunnen	SPR

Abb. 3: Beispiel der Synonymdatei

3.2. Zugriff auf die Quelldaten

Für den Zugriff auf die Quelldaten wird bei Excel sowie bei den relationalen Datenbanken Microsoft ActiveX Data Objects (ADO) bzw. die ADO-Jet Engine benutzt.

Im Folgenden werden weitere, wesentliche Ansätze zur Generierung der Abfragen auf relationale Datenbanken stichpunktartig zusammengefasst:

- Datenbankspezifische Dialekte, die vom SQL-Standard abweichen, werden vermieden.
- Das Auslesen der Quelldatenbank erfolgt zeilenweise, um auch relationale Datenbanken als Datenquelle verwenden zu können, welche der zweiten (und dritten) Normalform nicht genügen.
- Die Abfragen auf Objektebene werden vor dem Auslesen geordnet. Dabei werden die zwingend erforderlichen Attribute (Prerequisites) vor den optionalen Attributen plaziert. So können frühzeitig unzureichende Angaben zu Objekten erkannt und diese u.U. übersprungen werden. Die Prerequisites werden für den Benutzer innerhalb des Assistenten visuell hervorgehoben.
- Bei Attributen, welche Historien bilden können (wie z.B. der zeitliche Verlauf von Geländehöhen) und solche, welche in der Quelldatenbank auf mehreren Tabellen verteilt sind, in der Zieldatenbank aber nur in einer Tabelle (Spalte) vorliegen, können mehrere dieser Quellspalten ein und demselben Attribut zugeordnet werden. Dabei erhält das Attribut jeweils einen anderen Index. Z.B. können Messwerte in der Quelldatenbank in zwei verschiedenen Tabellen vorliegen und müssen in der Zieldatenbank in einer zusammengeführt werden (Abb. 4).

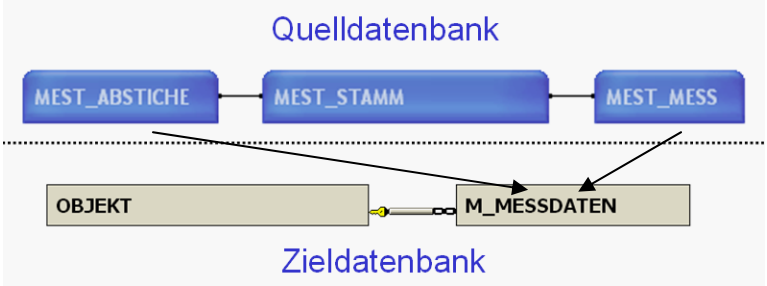


Abb. 4: Beispiel für die Verwendung zusätzlicher Indizes

3.3. Mapping-Assistent

Dem Anspruch auf intuitive Bedienbarkeit der dynamischen Mapping-Einstellungen wird durch die Verwendung eines Assistenten im “Wizard-Stil“ Rechnung getragen. Dieser ist zugleich geeignet, die (Wieder-)Verwendung gleicher Module für die verschiedenen Anwendungsfälle (ASCII, Excel, RDBMS) durch gleiche Benutzer-führungselemente zu gewährleisten.

3.4. Anwendungsbeispiele des Datenmappings

3.4.1. Mapping von ASCII/ Excel-Datenquellen

Aufbauend auf einer Analyse von bestehenden Strukturen zur Speicherung wasserwirtschaftlicher Informationen in Tabellenkalkulationen, wie sie häufig in Wasserwerken zum Einsatz kommen, wurde ein Assistent zum Import von ASCII- oder Excel-basierten Datenquellen vorgenommen. In Abbildung 5 ist die grafische Benutzungsoberfläche zur Mapping-Einstellung einer tabellarischen Datenquelle zu sehen. In diesem Beispiel wird das Excel-Arbeitsblatt “1068“ (1) eingebunden. Dort befinden sich Informationen zu einem Blatt (2), dessen Stammdaten spaltenweise vorliegen (3), (vorwiegend) in Spalte “A“ (bei der Einstellung “Ein Objekt je Blatt“ kann theoretisch jedem Attribut eine Zelle in unterschiedlichen Spalten und Zeilen zugeordnet werden). Die Mengendaten liegen zeilenweise (6)+(8) von der 3. bis zur 37. Zeile (7) vor. Der Messtyp ist bei allen Messwerten gleich und kann statisch festgelegt werden (9). Die Messwerte werden beim Auslesen mit dem Faktor 10 multipliziert, um die Maßeinheit anzupassen (z.B. Umrechnung: m³/h in l/s). Das zugehörige Messdatum befindet sich in Spalte “A“ und liegen im Format dd.mm.yyyy vor. Die getätigten Einstellungen können in einem dynamischen Vorschaufenster (10) überprüft werden.

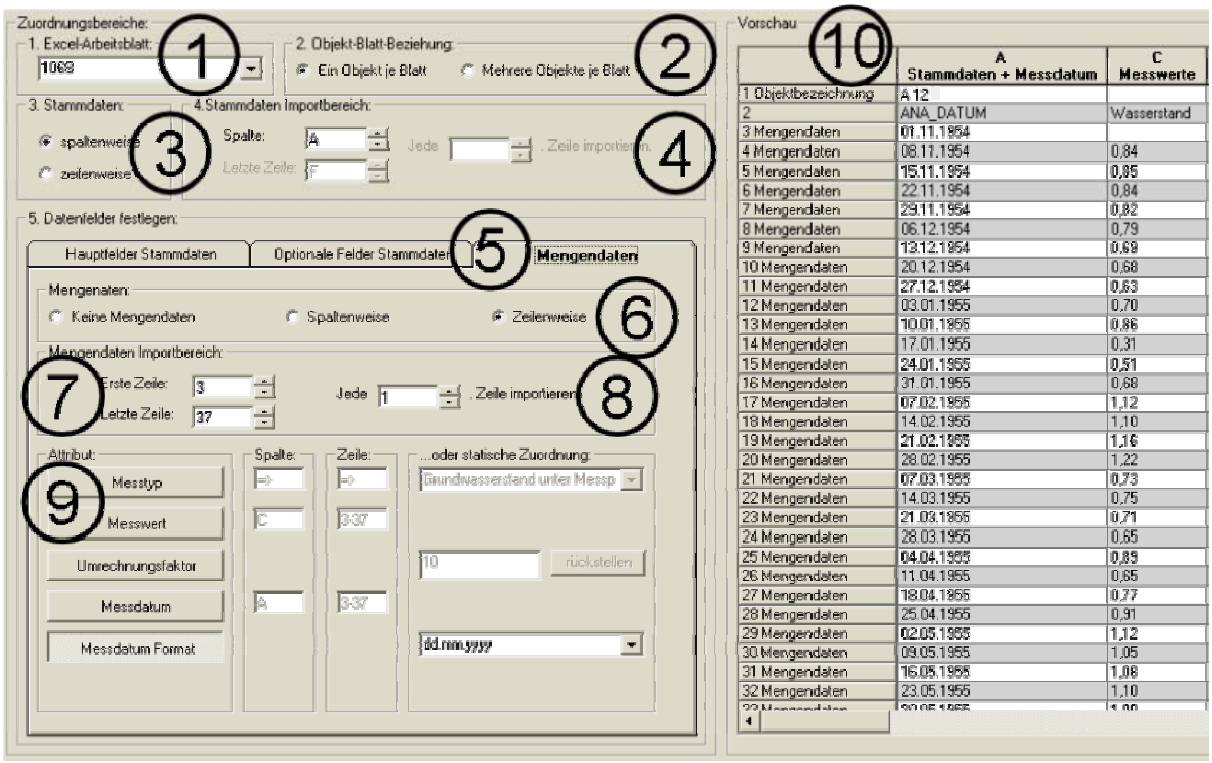


Abb. 5: Excel-Zuordnungsdialog

3.4.2. Mapping von RDBMS

Die Zuordnung für Relationale Datenbanken (Abb. 6) stellt eine erweiterte Form des Mappings von Tabellen (Kap. 3.4.1) dar. Dazu ist es zunächst notwendig, die Inhalte der Datenquelle (Tabellen, Attribute, Datentypen, Beschreibungen, Tabelleninhalte etc.) entsprechend aufbereitet darstellen, um eine Zuordnung und Überprüfung zu ermöglichen. In einem weiteren Schritt werden die Beziehung zwischen den einzelnen Tabellen definiert, wobei sich generell in „Richtung“ der Haupttabelle (1) orientiert wird.

Die Tabellenspalten selbst können sowohl Primär- bzw. Fremdschlüssel als auch bestimmten Attributen der Zieldatenbank (ggf. mit statischen Zuweisungen oder unter Angabe bestimmter Bedingungen) zugewiesen werden (2)+(3).

Die so getätigten Zuweisungen werden, getrennt nach den Tabellen- und Spaltenbeschreibungen, in separaten Übersichten dargestellt (4)+(5). Die Inhalte der Quelltabelle können zur besseren Orientierung in einer Vorschau angezeigt werden (6).

Abb. 6: RDBMS-Zuordnungsdialog

Diese Übersichten sowie die gesamte Abbildung der Quelldatenbank in Tabellen und Spalten, werden mit Symbolen dargestellt, welche die aktuellen Einstellungen übersichtlich visualisieren und auf etwaige Zuweisungsfehler aufmerksam machen sollen.

	Primärschlüssel		Nicht bekannt
	Fremdschlüssel		Notwendig / unvollständig
	Attribut		Bekannt / vollständig

Abb.7: Legende der verwendeten Symbole

3.5. Erweiterte Mapping-Funktionalitäten

3.5.1. Kontrollierte Fortschreibung von Informationsbeständen

Vor dem Schreiben der Daten in die Zieldatenbank wird bei jedem Attribut, sofern das Objekt nicht neu angelegt wurde, überprüft, inwieweit es in der Datenbank schon vorliegt und nur beim Nicht-Vorhandensein angelegt. Diese Logik wird eingesetzt, um eine Aktualisierung der Zieldatenbank zu ermöglichen. Dadurch wird den verschiedenen Institutionen ermöglicht, ihre vorhandenen Erfassungssysteme beizubehalten.

Auf diese Weise können beispielsweise Überwachungsbehörden in regelmäßigen Abständen die Datenbestände aller Wasserversorger in ihrem Monitoringgebiet einfordern, mit dem vorgestellten Werkzeug in ihre Datenbank überführen und somit ihre Überwachungs- und Steuerungsaufgaben zeitnah auf aktuellen Grundwasser-informationen durchführen.

3.5.2. Protokollierung

Alle Ergebnisse des Imports, z.B. auch der Plausibilitätskontrollen (siehe Kapitel 3.6), werden in XML-Import-Protokollen gespeichert. Hierdurch werden alle vorgenom-menen Inserts und Updates persistent archiviert und stehen für eine Nachverfolgung dauerhaft zur Verfügung.

3.6. Plausibilitätsprüfung und interaktive Datenveredelung

Um ggf. vorliegende, inhaltliche Fehler im Datenbestand der Quelldaten nicht in das Zielsystem zu übernehmen, wurden Prozesse zur Datenveredelung entwickelt. Bei den objektspezifischen Stammdaten werden Vollständigkeitsprüfungen und logische Plausibilitätskontrollen eingesetzt. So werden Gültigkeitsbereiche überprüft und Abweichungen dem Benutzer während des Auslesevorgangs (also vor der Erstellung der Import-Datei) zur optionalen Anpassung angezeigt.

Bei der Prüfung von Rohdaten werden algorithmische Plausibilitätskriterien (Maximum, Minimum, Abstand zum Vorwert bzw. Abstand zur linearen Regressionsgrade der letzten drei Messwerte und Korrelationskennwert) erst während des Importvorgangs geprüft. Importierte Werte, die den definierten Plausibilitätskriterien widersprechen werden in einem separaten XML-Protokoll mit der entsprechenden Konfliktbeschreibung gespeichert. Auf Basis dieser Liste erhält der Benutzer die Konflikte visuell und in Diagrammform aufbereitet, kann sie fachlich prüfen und ggf. Korrekturen vornehmen.

4. Zusammenfassung und Ausblick

4.1. Gesammelten Erfahrungen

Neben den unterschiedlichen Anwendungsszenarien (ASCII-/Excel- bzw. RDBMS-Zuweisung und -Import) muss grundsätzlich zwischen der erstmaligen Nutzung des Werkzeugs und einer Wiederverwendung der Mapping-Schemata im Routinebetrieb differenziert werden. So erfordert das strukturelle Zuweisen eines Datenbankschemas die größten Fachkenntnisse, während die Erstzuweisung von Excel-Tabellen bzw. ASCII-Dateien eines Datenloggers durchaus von dem entsprechenden Fachanwender seitens der Behörden bzw. der Wasserversorger und Ingenieurbüros selbst vorgenommen werden kann. In jedem Fall ist es ratsam, dass bei der inhaltlichen Zuordnung (Definition der Synonyme) sowohl Fachanwender der Quell- als auch der Zieldatenhaltung kooperieren, um Begrifflichkeiten verlustfrei abzustimmen. Bei einer Wiederverwendung eines definierten Synonymschemas kann es u.U. zu neuen Zuweisungen von Begrifflichkeiten kommen, die vom Fachanwender vorgenommen und im Synonymschema aktualisiert werden, um für zukünftige Mapping-Vorgänge zur Verfügung zu stehen.

4.2. Projektstand und aktuelle Erweiterungsgedanken

Das vorgestellte Werkzeug befindet sich mit allen Komponenten bereits im praktischen Einsatz. Zahlreiche Excel-Formate sowie drei RDBMS wurden erfolgreich eingebunden und importiert.

Das entwickelte Mapping-Tool stößt insbesondere bei schlecht modellierten RDBMS auf noch nicht behandelte Ausnahmefälle, die sukzessive mit dem Anspruch der Allgemeingültigkeit in die Applikation eingearbeitet werden. Neben einigen inhaltlichen und strukturellen Erweiterungen werden z.Z. verstärkt die besonderen Anforderungen der ASCII-Daten der Datenlogger untersucht. Hier ist es zusätzlich erforderlich, dass die dichten Datenbestände gefiltert, d.h. mit fachlich sinnvoll reduzierten Datenmenge in die Zieldatenbank überführt werden. Auch an die Korrektur von fehlerhaften Langzeittrends der Drucksonden bzw. von temporären Fehlmessungen und Messlücken werden erhöhte Anforderungen gestellt.

NOKIS++

Informationsinfrastrukturen für Nord-und Ostseeküste

Wassilios Kazakos[1], Marcus Briesen[2], Rainer Lehfeldt[3], Hans-Christian Reimers[4]

Abstract

Mit dem NOKIS Web-Portal http://nokis.baw.de steht ein von der Praxis getragenes Instrumentarium zum Informationsaustausch innerhalb der deutschen Küstenforschung zur Verfügung, das notwendige Informationsflüsse unterstützt. Dies gilt horizontal auf Bearbeiter-Ebene zwischen Dienststellen und vertikal zwischen übergeordneten Informationssystemen wie etwa in Zusammenarbeit mit dem GEIN/UDK. Voraussetzung für gezielte Informationsverarbeitung im Sinne von Recherche-Mechanismen, mit deren Hilfe effizient gesucht werden kann, sind aussagekräftige Metadaten. Diese stellen vor allem einen räumlichen und zeitlichen Bezug zu den inhaltlichen Fragestellungen her, charakterisieren Qualität und Geltungsbereich der Daten und zeigen einen Zugang zu den jeweiligen Quellen auf.

[1] FZI Forschungszentrum Informatik, Database Systems Department, Haid-und-Neu-Str. 10-14, D-76137 Karlsruhe, Germany, mailto: kazakos@fzi.de http://www.fzi.de/dbs.html

[2] disy Informationssysteme GmbH, Stephanienstraße 30, 76133 Karlsruhe. briesen@disy.net, http://www.disy.net

[3] Bundesanstalt für Wasserbau DS Hamburg, Wedeler Landstrasse 157, D-22559 Hamburg, mailto: rainer.lehfeldt@baw.de; http://www.baw.de

[4] Landesamt für Natur und Umwelt des Landes Schleswig-Holstein, Hamburger Chaussee 25, D-24220 Flintbek, email: hreimers@lanu.landsh.de

Der Aufbau von fachbezogenen Informationssystemen ist Gegenstand vieler auch von der EU geförderter Programme. Bekannte Produkte wie etwa CoastBase verwenden jedoch vorwiegend proprietäre und sektoral geprägte Metadaten. Eine konsequente Anwendung des international verwendeten Metadaten-Standards ISO19115 in NOKIS garantiert die Kompatibilität zu globalen Informationssystemen, die derzeit in Deutschland, Europa und weltweit im Aufbau sind. Damit wird sowohl der vertikale Informationsfluss, für den eine Informationsverdichtung kennzeichnend ist, unterstützt, wie auch die horizontale Informationsverbreitung erleichtert.

Für die notwendigen Arbeiten in der Praxis sind standardisierte Metainformationen alleine nicht ausreichend. Sie bilden lediglich die Grundlage für die Auswahl relevanter Daten und Informationen, aus denen die maßgebenden Informationen für sachgerechte Entscheidungen abgeleitet und bewertet werden.

In dem F&E-Projekt NOKIS++ sollen daher Web-Services gemäß ISO19119 aufgebaut werden, die den Zugriff auf Daten ermöglichen und Methoden zur Visualisierung und Analyse anbieten. Durch Erfüllung zweier ISO-Standards können unterschiedliche Nutzergruppen mit Daten und Informationen aus verschiedenen Fachdisziplinen versorgt werden.

Das Projekt besteht aus zwei Projektphasen.

- In der ersten Projektphase sollen die notwendigen Anpassungen des NOKIS Metadaten-Profils an die neu hinzukommenden Datenbestände aus Modellierung, Planung und Gewässerkunde mit den betreffenden Dienststellen festgelegt werden. Eine der Hauptaufgaben wird der Praxistest der erweiterten Werkzeuge und Bewertung des neuen Profils sein. Die erforderliche Systematisierung von Sprache und raumbezogenen Begriffen wird durch Arbeiten an einem mehrsprachigen Thesaurus und an einem Gazetter durch die Projektgruppe und in Zusammenarbeit mit externen Fachleuten bearbeitet. Die aus vorhandenen Prototypen vorhandenen Methoden zur Visualisierung und Analyse werden praxisgerecht weiterentwickelt und für den online Einsatz optimiert, und bilden die technologischen Grundlagen für Planungstools der Integrierten Küstenhydrographie (AG-Synopse) und des Küstengewässerschutzes (WRRL/BLMP) sowie für Digitale Atlanten (BAW).

- In der zweiten Projektphase werden die geplanten Web-Services implementiert. Die dann vorhandene Toolbox wird um weitere Module nach Anforderung aus der Praxis in Zusammenarbeit mit externen Software-Entwicklern erweitert. Die Fortschreibung der Arbeiten an Thesaurus und Gazetteer werden mit GEIN/UDK koordiniert. Der Austausch von Metadaten mit GeoMIS.Bund und UDK wird optimiert. Insbesondere wird die Umsetzung von Schnittstellen entsprechen OpenGIS und ISO 19119 bearbeitet.

Ein wesentlicher Schwerpunkt des Projekts ist der Zugriff auf Daten über Metadaten. Zurzeit werden neue Konzepte erarbeitet, die eine mögliche Lösung für die immer wiederkehrenden Fragen des Übergangs von Daten zu Metadaten adressieren. Hierbei gehen wir speziell auf das Problem der Granularität der Metadaten sowie der Verbesserung der Suche bei unterschiedlicher Detaillierung der Metadaten ein. Mit diesem Beitrag möchten wir den aktuellen Stand unserer Forschungsarbeit und erste Ergebnisse präsentieren.

Literatur

Kazakos, W., Kramer, R., Schmidt, A. (2000): Coastbase - The Virtual European Coastal and Marine Data Warehouse. In Armin Cremers and Klaus Greve, editors, Computer Science for Environmental Protection '00. Environmental Information for Planning, Politics and the Public, volume II, pages 646-654.

Swoboda, W., Kruse, F., Nikolai, R., Kazakos, W., Nyhuis, D., Rousselle, H. (1999): The UDK Approach: the 4th Generation of the Environmental Data Catalogue for Austrian and German Public Authorities, Proc. IEEE Meta-Data'99, http://computer.org/proceedings/meta/1999/papers/45/wswoboda.html

Swoboda, W., Kruse, F., Legat, R., Nikolai, R., Behrens S. (2000): Harmonisierter Zugang zu Umweltinformationen für Öffentlichkeit, Politik und Planung: Der Umweltdatenkatalog im Einsatz. In Armin Cremers and Klaus Greve, editors, Computer Science for Environmental Protection '00. Environmental Information for Planning, Politics and the Public, volume II Pp 595-607

Ein wissensbasiertes System zum Risikomanagement für komplexe räumlich und zeitlich orientierte Umweltdaten

Dipl. Geol. MSc Tilmann Steinmetz
Martin-Luther-Universität Halle
Umweltgeologie
Von-Seckendorff-Platz 3, 06120 Halle (Saale)
tilmann.steinmetz@geo.uni-halle.de

Abstract / Einleitung

Für die Speicherung und Beurteilung der Wissensbasis im Projektraum einer regionalen Chemie-Altlast sind umfangreiche, räumlich und zeitlich referenzierte Messdaten sowie toxikologische Referenzwerte zu speichern. Vor dem Hintergrund der Landnutzung dargestellt, dienen sie als Werkzeug zur Risikoabschätzung und als Planungsunterstützungssystem. Das projektierte System vereinigt im Ontologieeditor Protégé [Protégé, 2005] unterschiedliche Open Source Projekte als wissensbasiertes System mit Raum- und Zeitbezug. Die Fülle der bisher isoliert gehaltenen vorhandenen Daten soll auf diese Weise übersichtlicher repräsentiert und dem auswertenden Bearbeiter ein Teil der Last der Beurteilung abgenommen werden. Dazu sollen mit dem „Wissen“ des Systems (welches auf einfachen Regeln basiert, die auf die in Ontologien gespeicherten Fakten angewendet werden) eine Voraus-‚Beurteilung’ verschiedener Fakten teilautomatisiert übernommen werden. Bezieht man die Zeitdimension mit ein, kann man Szenarien planen und unterstützt so Planer durch die Berechnung und Darstellung verschiedener Alternativen. Durch die Verwendung strikter Ontologien wird außerdem sichergestellt, dass die erforderlichen Daten vollständig und korrekt eingegeben werden, wenn man die Datenbanken ergänzt oder erweitert.

Das System ist momentan noch im Planungsstadium, es findet eine Evaluierung der bestehenden Module statt und es wird geprüft, wie groß die Änderungen sein

müssten, um Vorhandenes in ein kohärentes System zu integrieren. Die Implementation lehnt sich aber in Teilen eng an vorhandene gebrauchsfertige Lösungen an (OpenMapTab [OpenMapTab, 2004], [Popovich, 2005] FLUMAGIS [ifgi, 2005]).

1. Datenlage und Problemstruktur

Die strukturierte Speicherung und Auswertung der Messdaten aus einem über 15 Jahre andauernden Grundwasser-Monitoringprogramm von über 300 Stoffen stellt schon für sich allein kein triviales Problem dar. Sind diese Daten im Hinblick auf eine Expositions- bzw. Risikoabschätzung zudem vor dem Hintergrund der Landnutzung zu betrachten, sowie den zugrunde liegenden gesetzlichen normativen Vorgaben zur Altlastenfreiheit bzw. Unbedenklichkeit, geologisch-sedimentologischem Aufbau des Untergrundes und der Grundwasserströmung, ergibt sich aus der Fülle der Daten eine gewissermaßen immanente Unübersichtlichkeit, die bei der Nutzung und Auswertung zwangsweise hinderlich sein wird.

Durch die Integration bestehender Modelle der Untergrundgeologie, Grundwasserströmung und Stoffverteilung mit den vorhandenen Daten zur Landnutzung soll schließlich die quantitative Beurteilung der Exposition von Schutzgütern ermöglicht werden. Mit Hilfe der so berechneten Expositionsintensität und –länge sowie der jeweiligen Gefährlichkeit (Toxizität, Kanzerogenität) der Einzelstoffe wird schliesslich eine Risikoabschätzung durchgeführt, deren Ergebnis in die Planung eingehen muss.

Traditionell liegt der Fokus der Darstellung in GeoInformationsSystemen (GIS) auf den räumlichen Bezügen der enthaltenen Objekte zueinander (Topologie), während z.B. Kausalzusammenhänge wegen der statischen Darstellung, also dem Fehlen der zeitlichen Dimension, praktisch nicht gezeigt werden können.

Als Planungswerkzeug eignen sich solche Systeme jedoch insbesondere dann, wenn nicht nur der status quo, sondern auch die Zustände in der Vergangenheit und – gegebenenfalls – Prognosen zukünftiger Zustände aufgezeigt werden. Gerade im Bereich der Landschaftsplanung und sozioökonomischer Entwicklung lässt sich häufig erst durch die Integration aufeinanderfolgender Zeitschritte in einer Entwicklung der verantwortliche Prozeß verständlich machen. Im vorliegenden Fall

stellt sich die Landnutzung vor dem Hintergrund der regionalen Kontamination des Grundwassers als die planerisch zu bewältigende Größe dar.

Durch gezielte Variation der möglichen Variablenwerte lassen sich im Hinblick auf die Zukunftsplanung Szenarien „durchspielen“, die bei der Bewältigung von Planungsunsicherheit, z.B. für katastrophale Ereignisse, wie die Hochflut im Projektgebiet Bitterfeld 2002, hilfreich sein können.

Das vorliegende Projekt strebt an, im Rahmen einer Doktorarbeit vorhandene Modelle und Software-(Module) auf eine Weise zu kombinieren, dass wichtige Entscheidungen im Projektraum korrekt, schnell und reproduzierbar getroffen werden können, indem aus der Fülle der Daten eine im räumlichen und zeitlichen Kontext relevante Auswahl visualisiert wird.

1.1. Projektgebiet und Daten

Die verwendeten Grundwasser-Kontaminations-Daten entstammen dem seit 1989 durchgeführten Grundwassermonitoringprogramm im Gebiet der ehemaligen und aktuellen Industrie- und Stadtgebiete aus dem Ökologischen Großprojekt (ÖGP) Bitterfeld-Wolfen, und ihrer näheren Umgebung. Zahlreiche Luftbilder, ATKIS (– **A**mtliches **T**opographisch-**K**artographisches **I**nformationssystem) - und CORINE (**C**o**or**dination of **I**nformation on the **E**nvironment) LandCover-Daten beinhalten Informationen zur Landnutzung. Die Untergrundgeologie liegt als detailliertes Modell diskreter Volumenkörper, wie auch in Bohrdaten und in interpolierten Profilschnitten als Ergebnis mehrerer vorangegangener Diplomarbeiten vor. Grundwasser-Strömungs-Vektoren wurden – insbesondere für den Zeitraum vor dem Hochwasser August 2002 – in der Arbeitsgruppe modelliert und werden zur Zeit für den Zustand nach der Flut erstellt. Datenbanken der US-EPA (**U**nited **S**tates **E**nvironmental **P**rotection **A**gency) zu Human- und Ökotoxikologie und deutsche ‚Grenz’ -(oder: Prüf-/Hintergrund-/Maßnahmen-) werte sowie Daten zu Löslichkeit, Persistenz und Bioakkumulierbarkeit liegen vor.

1.2. Verwendete Software

- Protégé Ontologieeditor (Uni Stanford)
- Ontologien für
    - ISO19115-Metadaten (Uni Drexel) und
    - CEDEX für Umweltdaten (Österreichisches Umweltbundesamt)
- können teilweise Anwendung finden
- regelbasiertes System mit
    - JESS (JAVA Expert System Shell) und
    - CLIPS (C Language Integrated Production System)
- OpenMap GIS als OpenMapTab.
- PostgreSQL [PostgreSql, 2005] mit räumlichem Postgis-Aufsatz [POSTGIS, 2005]
- Möglicherweise Modflow [Modflow, 2005] zur Grundwassermodellierung, sonst ggf. fertige Grundwasserströmungsvektoren (Umweltgeologie Uni Halle)

2. Software Überblick

Anstatt ein System von Grund auf neu zu entwickeln, wird der Plan verfolgt, die vorhandenen spezialisierten Softwarelösungen so zu kombinieren, dass auf einfache, schnelle und kostengünstige Weise ein vielseitiges Werkzeug entsteht, das dennoch auf seinen Aufgabenbereich spezialisiert ist und durch die Koppelung mehr als die Summe seiner Einzelteile darstellt – also Auswertungen ermöglicht, die den jeweils einzelnen Systemen vorenthalten wären. Aus den genannten Gründen bietet sich Open Source Software an, die veränderbar und preiswert ist.

2.1. Agentenbasierte Systeme

Neben dem Objektorientierten Paradigma neuerer Software stellen Agentenbasierte Systeme eine Entwicklung dar, die der zunehmenden Verteilung von untereinander

unabhängigen Softwareteilen, die miteinander über Netze in Verbindung stehen, Rechnung tragen. Eine einfache Definition von Agenten könnte wie folgt lauten: Agenten sind autonom, verfolgen ihre eigenen Ziele, nehmen ihre Umgebung wahr und reagieren auf sie und interagieren mit ihrer Umgebung (und gegebenenfalls anderen Agenten). Wie bei Zambonelli dargestellt, verschwimmen die Grenzen zwischen objektorientiertem und agentenbasiertem Paradigma inzwischen [Zambonelli, 2003].

Im vorliegenden Projekt wären die handelnden Agenten beispielsweise die Grundwassermessstellen, klar abgrenzbare Grundwasserkörper (Quartär, Tertiär), Flächenschläge oder z.B. Betriebsflächen, Deponien, Kippen.

Agentenbasierte Systeme sind sinnvollerweise mit GIS zu koppeln, hier stellt sich die Frage, wie der Bezug der Agenten zu anderen Objekten im Raum ist. Zambonelli et al. führen vier verschiedene Arten der Verknüpfung zwischen Agenten und Objekten im Raum auf: *Identische Abbildung, kausal, temporal* oder *topologisch*. Die Agenten stehen dabei in dem jeweiligen Wirkungszusammenhang mit der Umgebung (z.B. könnte eine abgegrenzte Fläche als Ausschlussfläche markiert werden, sobald die Konzentration in einer (oder mehreren) Grundwassermessstellen einen bestimmten Wert überschreitet. Solche Regelhaftigkeiten liessen sich in der Regelbasis über JESS implementieren.

Bei Brown et. al. [BROWN, 2005] wird ein Planungs- und Simulationsmodell zur militärischen Anwendung vorgestellt (TSUNAMI), welches mit OpenMap und dem Entwicklungswerkzeug für Agentenbasierte Systeme *RePast* [RePast, 2003] implementiert wurde. Die Integration eines agentenbasierten Systems *sensu stricto* wird in unserem Projekt zunächst nicht weiterverfolgt; zu Konzeption, Design und Bedeutung entsprechender Systeme sei auf Zambonelli et al. [Zambonelli, 2005] und [Scholl, 2001] verwiesen.

2.2. Wissensbasis, Ontologien

Die Formalisierung von Wissen unter Verwendung strikter Ontologien ist notwendig, wenn Methoden der künstlichen Intelligenz zur Generierung reproduzierbarer Entscheidungen verwendet werden sollen. Zum einen wird sichergestellt, dass die Begrifflichkeiten, die im Zusammenhang mit der Problemstellung gebraucht werden, einheitlich sind, zum anderen ist die Domäne des Wissensgebietes durch die

Onotologie vollständig und hinreichend beschrieben, so dass Eintragungen in die Wissensbasis (die eigentliche Wissens-Datenbank) eindeutig sind. Dies hat unter anderem den Vorteil, dass Regeln zur Verwendung des gespeicherten Wissens einfacher angelegt werden können. Insgesamt wird die digitale Verwaltung der Informationen und das „digitale Verständnis“ der Informationen, also die Semantik, durch Ontologien vereinfacht oder erst möglich.

Expertenwissen wird in einer Wissensbasis formalisiert abgelegt (im einfachsten Fall z.B. in einer Relationalen Datenbank). Dazu müssen die für die jeweiligen Wissensdomänen zuständigen Experten zunächst hinzugezogen werden, um interdisziplinär festzulegen, wie die Zusammenhänge zwischen den einzelnen Wissensdomänen sind. Dies ist der entscheidende Schritt bei der Modellerstellung, da schon die Auswahl der Regelhaftigkeiten und das System, unter dem das Wissen gespeichert wird, festlegen, welche Systematiken dem Modell zugrundeliegen werden. Die Auswahl des Modells spiegelt dabei die Semantik wieder, also das Verständnis oder Weltbild, das die Experten dem System zugrunde legen. Meist kann nur ein stark vereinfachtes Konzeptmodell in ein ebenso stark abstrahiertes Computermodell überführt werden. Später soll es jedoch zu einem Ergebnis führen, das durch den Vorgang der (menschlichen) Abstraktion durch den Experten erreicht werden würde. Das Ergebnis ist keine fertige Lösung, sondern durch Abstraktion ein komplexes Modell überschaubar und greifbar zu machen. Durch die Vereinfachung sind mit relativ wenig Rechenleistung auch verschiedene Ausgänge eines Vorgangs (Szenarien) zu berechnen. Allerdings kann nur ein solcher Zustand erreicht werden, der von vorneherein im Rahmen des Modells möglich ist. Durch die Abstraktion bedingt, fallen möglicherweise wichtige, in der Realität mögliche, Szenarien weg.

Bei *Protégé* handelt es sich streng genommen einen Ontologieeditor. Er ist aber zum Aufbau wissenbasierter Systeme so erweiterbar, dass die Einbindung vorhandener Daten aus relationalen Datenbanken über die ODBC-Schnittstelle möglich ist. Wie bereits an dieser Stelle vorgestellt [Schentz, 2004] sind vorläufige Ontologien zur Repräsentation der zwei Domänen Ökologie CEDEX (*C*lasses for *E*nvironmental *D*ata *Ex*change) [CEDEX, 2004] und räumliche Metadaten nach ISO19115 [Drexel, 2004] erarbeitet worden. Der Standard des OGC (Open Geospatial Consortium) - für Ontologien ist OWL (Web Ontology Language), von diesem können die genannten Ontologien aber in das Sprachmodell von Protégé umgesetzt werden können.

Durch eine übersichtlich gestaltete Oberfläche und die verhältnismässig einfache Bedienbarkeit eignet sich Protégé auch für Einsteiger im Bereich Wissensmodellierung. Experten können sich darauf konzentrieren, dass ihre Wissensdomäne richtig repräsentiert wird und müssen sich nicht zusätzlich mit der Programmierung einer Ontologiebeschreibungssprache beschäftigen. Protégé ist modular erweiterbar. Module werden als sogenannte ‚Tabs', also Karteireiter in die graphische Oberfläche mit einbezogen.

Bezieht man georeferenzierte (GIS-) Daten ein, wird die Simulation räumlicher Szenarien möglich. Wird schließlich noch der Zeitbezug integriert, kann man die Auswirkung veränderter Rahmenbedingungen im zeitlichen Verlauf simulieren und stellt damit ein effizientes Planungswerkzeug zur Verfügung.

2.3. GIS

Mit Hilfe des GIS OpenMap [OpenMap, 2003], welches als Protégé-Modul [OpenMapTab, 2004] integriert worden ist, sind auf Regeln basierende ([JESS, 2005]], [CLIPS, 2005]) räumliche und zeitliche Simulationen möglich. Diese werden als Zustandsänderungen von der Wissensbasis geliefert und dann im GIS angezeigt. Ein System aus den genannten Komponenten ist vom *OOGIS Labor St. Petersburg* implementiert worden [Popovich, 2005].

2.4. Regelbasis und Inferenzmaschine

Die in Protégé als JESSTab/CLIPSTab implementierte Regelbasis lässt sich über deren eingebaute Inferenzmaschine auswerten. Dabei übernimmt letztere die Aufgabe, durch Ziehen von Schlüssen zu Ergebnissen zu kommen. In ihr liegt also die eigentliche „künstliche Intelligenz". Dabei kommt ein sogenannter *forward-chaining*-Algorithmus zum Einsatz (WENN-> DANN-Regeln). Bei diesem Vorgang trifft ein *pattern matching*, also eine Mustererkennung, die Entscheidung, ob ein bestimmter vordefinierter Zustand eingetreten ist, oder nicht. Mit *n:m*- (mehrfach zu mehrfach)- *matching* kann man die gleichen Schlüsse aus dem Auftreten unterschiedlicher Zustände ziehen bzw. aus ein und demselben Zustand unterschiedliche Schlüsse ziehen. Dieser Vorgang ist sehr komplex und der Definition der Regelbasis kommt eine Schlüsselstellung bei der Implementation des Systems zu.

2.5. Beispiele

2.5.1. FLUMAGIS

Das Flußgebietsmanagements- und -planungssystem FLUMAGIS [ifgi,2005] stellt ein Beispiel für die Verwendung von Protégé und einer Regelbasis dar, verwendet aber als GIS MapObjects für JAVA, was der gewünschten verteilten Verwendung mit GIS-Visualisierung beim Client Rechnung trägt, während das Protégé-System auf einem Server liegt. Möltgen und May [FLUMAGIS, 2005] unterscheiden Planungs-Unterstützungs-Systeme (PSS) von räumlichen Entscheidungs-Unterstützungs-Systemen (SDSS), „wenn bei einem SDSS die eigentliche Szenarienplanung (neben der Entscheidungs-Unterstützung) in den Mittelpunkt“ gestellt wird. Nach ihrer Definition stellt ein PSS „eine Umgebung dar, die drei Komponenten integrieren sollte:

- Problemdefinition- und Planungsumgebung;
- Systemmodelle zur Analyse, Vorhersage und Simulation;
- Transformation von Daten zu Informationen als Grundlage für Modellierung und Design“

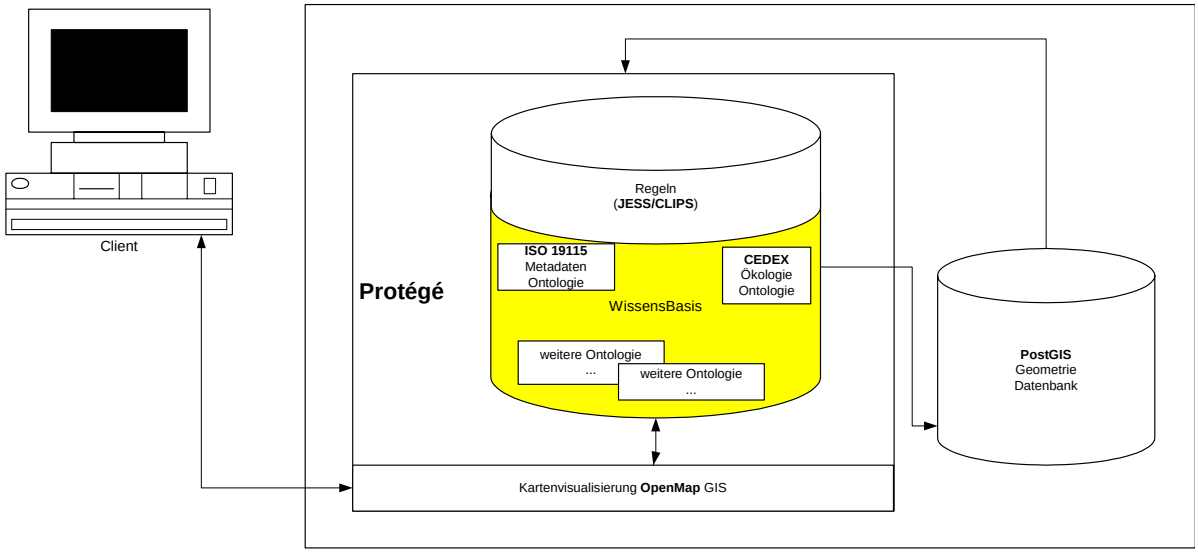
2.5.2. OpenMapTab

Im vorliegenden Projekt soll als GIS-modul das OpenSource GIS OpenMap zum Einsatz kommen, das von Ru Sorokin [OpenMapTab, 2004], [Popovich, 2005] als OpenMapTab in die Protégé-Oberfläche integriert wurde. Ausserdem ist es ihm gelungen, das GIS-Modul mit einer Regelbasis (CLIPS, JESS) zu verknüpfen und damit raumzeitliche Simulationen darzustellen.

3. Systemarchitektur

Das fertige System soll in der Lage sein auf die Anfrage eines Clients, z.B. nach dem Zustand des Grundwassers in einem bestimmten Punkt des Gebietes zu einem Zeitpunkt t_x eine Antwort zu generieren. Diese Antwort könnte z.B. darin bestehen, die Messstelle im GIS farbig codiert darzustellen, wenn die darin vorgefundenen Schadstoffgehalte eine festgelegte Konzentration überschreiten. Gleichzeitig könnte beispielsweise das davon abstromig gelegene Gebiet mit einem räumlichen Puffer belegt und farbig markiert werden sowie potenzielle Bau- oder Siedlungsgebiete mit einer Warnung versehen werden.

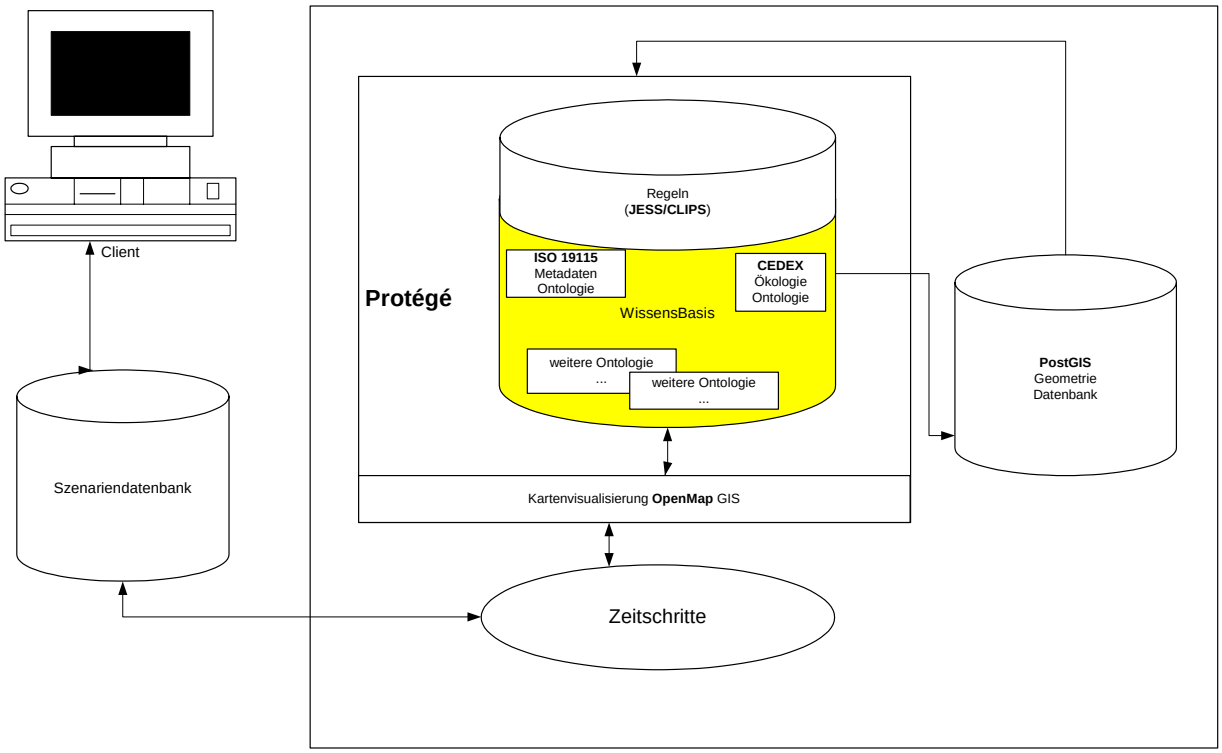
Abb. 1: Schema der Systemarchitektur. Der GIS Client muss nicht, kann aber, auf einem anderen Rechner liegen als Protegé. Dasselbe gilt für die Geometrie-Datenbank. Theoretisch können auch alle auf dem lokalen Rechner liegen.



Das Verfahren würde vereinfacht, wenn als Zeitpunkt heute (bzw. der Zeitpunkt der vorliegenden Messung) gewählt würde und als Punkt, für den eine Aussage getroffen werden soll, ein Punkt, an dem tatsächlich gemessen wurde. Dies würde für den Anfang genügen und könnte, die richtige Architektur vorausgesetzt, mit einer Interpolation der Gehalte auf die Fläche und Zeitschritten oder beliebiger Zeitwahl erweitert werden.

Alternativ zur oben gezeigten schematischen Architektur könnte der Client als Antwort eine Karte mit einem Szenario erhalten, das vorberechnet für eine bestimmte Parametrisierung und/oder einen bestimmten Zeitschritt, in einer separaten Datenbank vorliegt (s. Abbildung 2).

Abb.2: Systemarchitektur mit der Modellierung in Zeitschritten und Clientabfrage aus der Szenariendatenbank



3.1. Geographische Datenbank

Als Datenbanken bieten sich im Open-Source Bereich PostgreSQL und MySQL an, die zum einen kostenlos und quelloffen vorliegen, zum anderen auch weit verbreitet sind (und für die daher auch eine das Wissen der ‚Nutzergemeinde' zur Verfügung steht). Die objektrelationale Datenbank PostgreSQL scheint die ausgreiftere Lösung zu sein, wenn es darum geht, räumlich referenzierte Objekte zu speichern und abzufragen.

3.1.1. Räumliche Explorative Datenanalyse (Data Mining)

Bei Messungen von Umweltdaten ist ein häufig anzutreffendes (und dadurch nicht minder schwieriges) Problem, zu beurteilen, ob Extremwerte als Ausreisser auszuschliessen sind, oder ob sie mit bewertet werden sollen. Ausserdem hängt von der Repräsentativität der Messstellenverteilung deren Aussagekraft ab.

Routinemässig muß daher der Auswertung eine (räumliche) explorative Datenanalyse vorangestellt werden. Dabei kommen einfache graphische Darstellungen wie Histogramme und *quantil-quantil-plots (qq-plots)* ebenso zur Anwendung wie kompliziertere Berechnungen und deren Darstellung in Karten. Zu letzteren gehören z.B. die Lokalen Indikatoren räumlicher Assoziation (LISA – **l**ocal **i**ndicators of **s**patial **a**ssociation, [Anselin, 1995]). Die Statistiksprache *R* [R-project, 2005] kann direkt *in* der PostgreSQL-Datenbank operieren und dort solche Auswertungen vornehmen, bspw. als eingebettete Prozeduren in PL/SQL (Procedural Language/Structured Query Language)- einer Erweiterung von SQL mit prozedural abzuspeichernden Routinen. Für *R* gibt es Bibliotheken mit Routinen zur räumlichen Auswertung (splancs, sdep, maptools …) u.a. für die oben genannten Aufgaben.

Bei der Auswertung umfangreicher räumlicher Daten geht es darum, Verteilungsmuster in der räumlichen Verbreitung von Daten zu erkennen und darüber Schlüsse über die zugrundeliegenden Prozesse ziehen zu können ("Spatial data mining is the process of discovering interesting and previously unknown, but potentially useful patterns from spatial databases", [SHEKHAR, 2005]). Idealerweise könnten neben den o.g. Ausreisser-Erkennungsroutinen auch Algorithmen implementiert werden, die zur Erkennung anderer Muster dienen. (Beipielsweise könnte man eine Warnung ausgeben lassen, wenn beim Monitoring über längere Zeit

bestimmte Flächen unterrepräsentiert beprobt wurden. Was genau „unterrepräsentiert“ und „längere Zeit“ bedeuten, bliebe natürlich festzulegen!)

3.1.2. Räumliche Operationen

Mit PostGIS liegt eine räumliche Erweiterung für PostgreSQL vor. Sie dient dazu, räumliche Objekte nach Simple Features Spezifikation des OGC in der Datenbank zu speichern. Die verwendete Form ist entweder *well known binary* oder *well known text.* Die Aufgabe der räumlichen Verschneidung von Objekten (also z. B. die Abfrage, ob eine Grundwassermessstelle in einem Grundwasserschutzgebiet liegt) entfällt dabei für das GIS, da diese Information bereits in der Datenbank abgelegt ist. Weil solche Verschneidungsoperationen neben der georeferenzierten Darstellung von Daten mit zu den Hauptaufgaben des GIS gehören, resultiert daraus eine deutliche Entlastung für das GIS, bzw. für die Rechenzeit eines Clients.

Über die standardisierte Schnittstelle JDBC (JAVA DataBase Connector) lassen sich an Protégé beliebige relationale Datenbanken einbinden. Leider gehört PostGIS nicht zu den direkt unterstützten proprietären Formaten für OpenMap-GIS. Eine Möglichkeit die Datenbank zu lesen, scheint WKB4J [WKB4J, 2005] zu bieten, eine Bibliothek, die das in PostGIS verwendete *well known binary* Format nach JAVA übersetzt. Das ebenfalls als OpenSource vorliegende GIS-Projekt JUMP (JAVA Unified Mapping Platform) stellt dieses Format dagegen zum Lesezugriff bereit. Bis OpenMap den Zugriff auf das PostGIS-Format ermöglicht, wäre es daher wünschenswert, analog zum OpenMapTab ein JUMPTab zu integrieren. Es gilt abzuwägen, ob die erforderliche Entwicklung dafür nicht vergleichsweise zu arbeitsintensiv ist.

3.2. Wissensbasis

Die Erstellung der Wissenbasis soll im vorliegenden Projekt durch die Verwendung fertig vorliegender Ontologien vereinfacht werden. Geplant ist, die für Protégé angepassten Ontologien CEDEX für ökologische Daten und eine von der Uni Drexel erstellte Ontologie für Geographische Metadaten nach ISO19115 zu verwenden.

Die laut Dokumentation bereits in CEDEX umgesetzten Themen umfassen unter anderem Luft+Wasserchemie, Bodenanalytik und Bodenwasseranalytik, Vegetationskunde und Biodiversity, Depositionsanalytik und Geologie, so dass ein

grosser Teil der uns vorliegenden Daten eingearbeitet werden könnte. Kartenmaterial, das keinem der genannten Themen zugeordnet werden kann, könnte dann, sozusagen ‚verallgemeinernd' über die ISO19115 beschrieben werden.

3.3. Wissensbasis und GIS

Grundsätzlich sind 3 Arten der Kopplung wissensbasierter Systeme und GIS denkbar: eine lose Kopplung durch den Austausch von Informationen und Zwischenergebnissen über Dateien oder die entweder GIS-zentrierte Lösung mit Integration einer Wissens- und Regelbasis ins GIS, die ereignisbasiert aufgerufen wird (z.B. beim Anklicken eines mit einem Link versehen Punktes) bzw. die auf das Wissensbasierte System (WBS) zentrierte Lösung, bei der die GIS Darstellung in das WBS integriert wird.

Die letztgenannte Lösung ist in der Integration von Protégé und OpenMap gewählt worden. OpenMap ist als ‚Tab' (Karteireiter) wie auch CLIPS und JESS in die Grafische Benutzeroberfläche (GUI) von Protégé integriert und tauscht mit diesen direkt über deren jeweilige APIs (application programming interfaces) Daten aus.

Protégé von der Uni Stanford ist ein Werkzeug zur Wissensaquisition und zum „Wissensmanagement". Es liegt in Open Source und als freie Lizenz vor. Geplant als ein erweiterbarer Ontologiededitor wurde es durch eine aktive Nutzergemeinde um Plugins erweitert, die es ermöglichen fremde Ontologien zu importieren, Verbindungen zu Datenbanken herzustellen und GIS- Daten zu visualisieren. Das regel- oder objektbasierte CLIPS zur Entwicklung von Expertensystemen bzw. sein Java-Äquivalent JESS lässt sich ebenso integrieren. Diese Kombination von Fähigkeiten ermöglichen seine Nutzung als räumlich gestütztes wissensbasiertes System.

3.4. Regelbasis

Bei der Erstellung einer Regelbasis liegt die Schwierigkeit darin, ausgehend vom Konzeptmodell durch Abstraktion und Vereinfachung zu einem Prozeßmodell zu kommen und dieses formalisiert als Regelbasis zu implementieren. Ausgangspunkt muss hier wohl eine kleine Basis eindeutiger Regeln sein, bei deren Anwendung tatsächlich keine Alternativen vorliegen, über die ein Entscheider beschließen müsste (Faustregeln: wenn -> dann; das System enthält eine Inferenzmaschine, die

entscheidet, ob Fakten zutreffen und löst entsprechende Regeln aus). Diese Regelbasis muss anschliessend fortlaufend erweiterbar bleiben, um weitere (und auch komplexere) Regeln in den Entscheidungsprozess mit einzubeziehen.

Im vorliegenden Fall sind Regeln denkbar wie: Wenn Grenzwert überschritten -> Dann Gefährdung. Wenn Gefährdung -> Dann Ausschlussfläche

Hierbei sollen z.B. einzeln ausgewiesene Flächen (aus dem ATKIS), Bohrpunkte – bzw. Bohrungen, Grundwasserkompartimente, Betriebsflächen/ Deponieflächen/ Kippenflächen als Objekte in die Wissensbasis eingehen.

4. Zusammenfassung

Bestehende OpenSource Software und vorgefertigte Ontologien sollen zu einem Entscheidungsunterstützungssystem kombiniert werden. Eine Art „denkendes GIS“ soll entstehen, welches routinehaft zu bearbeitende Schritte automatisiert übernimmt, insbesondere im Bereich räumliche explorative Datenanalyse und Visualisierung.

Zusammengenommen geht es im Gegensatz zu einem herkömmlichen GIS nicht nur um die Visualisierung der räumlichen Bezüge der Messdaten, sondern auch um die Visualisierung der gegenseitigen Bezüge unterschiedlicher Wissensdomänen zueinander und damit ein verbessertes Verständnis der untersuchten Problematik bzw. der erforderlichen Lösungswege. Ausserdem sollen Regeln zum Verhalten der Fakten zueinander (insbesondere räumlicher Bezug) aber auch deren Kausalzusammenhänge und eventuell erforderliche Reaktionen oder Massnahmen eingebaut werden. Angestrebt ist räumliche Simulation mit OpenMapTab (mit Interaktion der faktisch unabhängigen „Agenten“: GW-Messstellen, Grundwasser, Stoffquellen, Grundstücke, Flächenschläge – wenn auch nicht im strengeren Sinne eines „agentenbasierten Simulationssystems“), sowie eventuell Visualisierung der Ausgabe des Modflow-Modells. Normative Vorgaben (wie einzuhaltende Grenzwerte oder toxikologisches Basis-Wissen) sollen regelbasiert Eingang finden.

Eine „strenge“ Behandlung der Eingangsdaten durch Ontologien und OGC- sowie ISO-konforme Datenhaltung sind außerdem im Sinne einer „nachhaltigen Datenbewirtschaftung“ wünschenswert.

Vereinfachend wäre die „Vorratshaltung“ einiger weniger Szenarien (s. Abbildung 2) denkbar, die sich für vorgegebene Randbedingungen ergeben und als fertige Karten in einer getrennten Datenbank zugänglich wären.

Literaturverzeichnis

[ANSELIN, 1995]

Anselin, Luc. *Local Indicators of Spatial Association: LISA*. Geographical Analysis, 27(2):93- 115, 1995. 16

[BROWN, 2005]

Daniel G. Brown, Rick Riolo, Derek T. Robinson, Michael North, and William Rand: *Spatial process and data models: Toward integration of agent-based models and GIS*, J Geograph. Syst (2005) 7:25-47

[CLIPS, 2005]

http://www.ghg.net/clips/CLIPS.html, abgerufen 23.05.05

[Drexel, 2004]

Drexel-University Online Publication: http://loki.cae.drexel.edu/~wbs/ontology/iso-19115.htm , abgerufen 23.05.05

[FLUGMAGIS,2005]

http://www.flumagis.de/veroeffentlichungen.htm, abgerufen 23.05.05

[Goovaerts, 2004]

Goovaerts, P., Avruskin, G., Meliker, J., Slotnick, M., Jacquez, G. and Nriagu, J. (2004): *Modeling Uncertainty about Pollutant Concentration and Human Exposure using Geostatistics and a Space-time Information Sytem: Application to Arsenic in Groundwater of Southeast Michigan*. Spatial Accuracy Invited Lecturer, Sixth International Symposium on Spatial Accuracy Assessment in Natural resources and Environmental Sciences, Portland, Maine, June 2004. April 2004.

[ifgi, 2005]

Brauner, J., May, M. (2005): *Einbindung einer Wissensbasis in GIS*. In: Schrenk, M. (Hrsg.) Tagungsband "Reale Modelle - irreale Welt. Der professionelle Umgang mit dem Unvorhersehbaren" Februar 2005, Corp 2005, Wien

[Jacquez et. al., 2003]

G.M. Jacquez, G.AvRuskin, E.Do, H. Durbeck, D.A. Greiling, P. Goovaerts, A. Kaufmann and B. Rommel: *Complex Systems Analysis using Space-Time Information Systems and Model Transition Sensitivity Analysis* . *In*: Accuracy 2004: Proceedings of the 6th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.

[JESS, 2005]

JESS the Rule Engine for the JavaTM Platform

http://herzberg.ca.sandia.gov/jess/, abgerufen 23.05.05

[Modflow, 2005]

http://water.usgs.gov/nrp/gwsoftware/modflow.html, abgerufen 23.05.05

[Möltgen&May, 2004]

Möltgen, J. und May, M.: *Entwicklungskriterien für ein Planungsunterstützungssystem*. *In*: Möltgen, J. und D. Petry (Hrsg.) Tagungsband: „Interdisziplinäre Methoden des Flussgebietsmanagements“ März 2004, IfGIprints 21, Institut für Geoinformatik, Universität Münster.

[OpenMap, 2003]

http://openmap.bbn.com/, abgerufen 23.05.05

[OpenMapTab, 2004]

St. Petersburg Institute for Informatics and Automation of RAS; Object oriented Geoinformatics Systems Laboratory: http://www.oogis.ru/en/projects/OpenMapTab/openmap_tab.htm, St. Petersburg, 2004. , abgerufen 23.05.05

[Popovich, 2005]

Vasily Popovich, Sergei Potapichev, Ruslan Sorokin, Andrei Pankin: *Intelligent GIS for Monitoring Systems Development*, *in:* CORP 2005 & Geomultimedia05, Proceedings/Tagungsband; Ed./Hg.: Manfred Schrenk, ISBN: 3-901673-12-1

[PostgreSql, 2005]

http://www.postgresql.org/, abgerufen 23.05.05

[POSTGIS, 2005]

http://www.postgis.org/, abgerufen 23.05.05

[Protégé, 2005]

http://protege.stanford.edu, abgerufen 23.05.05

[R-Project, 2005]

http://www.r-project.org , abgerufen 23.05.05

[RePast, 2003]

University of Chicago (2003) RePast 2.0. (Software) Chicago: Social Science Research Computing Program – http://repast.sourceforge.net , abgerufen 23.05.05

[Schentz, 2004]

Schentz, Herbert; Mirtl, Michael: *CEDEX eine erweiterbare Ontologie für ökologische Daten*. Workshop Umweltdatenbanken, Darmstadt, 2004.

[Scholl, 2001]

Hans J. Scholl: *Agent-based versus System Dynamics Modeling*. Proceedings of the 34th Hawaiian International Conference on System Sciences (HICSS-34), January 2001, Maui, Hawaii, 1-37

[SHEKHAR, 2005]

Shashi Shekhar and Ranaga Raju Vatsavai: *Spatial Data Mining Research by the Spatial Database Research Group, University of Minnesota,* http://www.cs.umn.edu/research/shashi-group/ , abgerufen 23.05.05

[UBA-AU, 2004]

http://www.umweltbundesamt.at/umweltdaten/schnittstellen/cedex/cedex_protege/, abgerufen 23.05.05

[WKB4J, 2005]

http://wkb4j.sourceforge.net/ , abgerufen 23.05.05

[Zambonelli, 2003]

Zambonelli, F., Jennings, Nicholas R., and Wooldridge, M.: *Developing Multiagent Systems: The Gaia Methodology*, ACM Transactions on Software Engineering and Methodology, Vol.12, No. 3, July 2003, 317-370

Qualitätsgesicherte Veröffentlichung von Umweltdaten am Beispiel von IMIS

Volkmar Schulz, Condat AG, vs@condat.de

Abstract / Einleitung

Das vom Bundesamt für Strahlenschutz (BfS) betriebene „Integrierte Mess- und Informationssystem zur Überwachung der Umweltradioaktivität“ (IMIS) wurde durch Condat von Grund auf neu entwickelt und am 1. April 2005 in Betrieb genommen.

Hauptaufgabe des IMIS ist die ständige Erfassung und zuverlässige Bewertung der Umweltradioaktivität. In dieser Rolle fungiert es als Frühwarnsystem und liefert den verantwortlichen Personen die notwendigen Entscheidungsgrundlagen für deren unverzügliches Handeln.

Zu den von IMIS unterstützten Routineaufgaben zählt unter anderem, die Öffentlichkeit umfassend über die aktuelle Lage zu informieren. Auf Basis der im System verankerten mehrstufigen Qualitätssicherung und den zur Verfügung stehenden Workflow-Mechanismen hat Condat eine Lösung geschaffen mit deren Hilfe die Umweltdaten über die Schritte.

Erfassung → Bewertung → Aufbereitung → Visualisierung → Veröffentlichung

der Bevölkerung automatisiert und damit ständig aktuell im Internet zugänglich gemacht werden können.

Nachfolgend werden die Kernelemente dieser flexibel einsetzbaren Lösung näher vorgestellt.

1. IMIS im Überblick

Die Umweltradioaktivität wurde in der Bundesrepublik Deutschland bereits seit 1955 als Aufgabe verschiedener Behörden großräumig gemessen. Beim Reaktorunfall in Tschernobyl 1986 zeigte sich jedoch, dass die Messergebnisse teilweise unterschiedlich dargestellt und damit auch unterschiedlich bewertet wurden.

Der zum damaligen Zeitpunkt zähe Informationsaustausch und die aufgrund unterschiedlicher Darstellung der Situation ausgelöste Verunsicherung in der Bevölkerung bewogen die Bundesregierung, noch im selben Jahr das so genannte Strahlenschutzvorsorgegesetz zu verabschieden (StrVG vom 19.12.1986). Die Umweltradioaktivität ständig nach einheitlichen Kriterien zu überwachen, ist wesentlicher Regelungsinhalt dieses Gesetzes. Deshalb wurde beschlossen, die Überwachung der Umweltradioaktivität auszubauen und zu einem einheitlichen Mess- und Informationssystem (IMIS) zusammenzuführen. Im Fall eines Ereignisses mit erheblichen radiologischen Auswirkungen soll der Schaden so gering wie möglich gehalten werden; angemessene Maßnahmen sind einheitlich und nach neuesten wissenschaftlichen Erkenntnissen zu treffen.

IMIS ermöglicht es durch permanente Messungen, bedeutsame Änderungen der Umweltradioaktivität schnell und zuverlässig zu erfassen und zu bewerten, erforderliche Maßnahmen zu identifizieren, sowie die Öffentlichkeit umfassend zu informieren.

Dem Bundesumweltminister werden mit IMIS die Entscheidungsgrundlagen für unverzügliches Handeln geliefert. Koordinierte Vorsorgemaßnahmen können getroffen werden, um Bevölkerung und Umwelt wirksam zu schützen.

Das Gesamtverfahren ist dazu in drei Ebenen aufgebaut: Datenerhebung, Aufbereitung, Entscheidung (siehe Abbildung 1). Die von Condat entwickelte IMIS-Software unterstützt alle drei Ebenen.

Im Zeitraum von Oktober 1999 bis Dezember 2004 realisierte Condat das IMIS mit einem Gesamtaufwand von 100 Personenjahren. Das System ist geprägt von einer hohen Flexibilität und einem komplexen Funktionsumfang. Allein die technische Anwenderdokumentation (ohne fachliche Beispiele) umfasst mehr als 1.100 Seiten. Die ca. 160 Fachanwender wurden vor Inbetriebnahme im Rahmen einwöchiger Schulungen in die für sie relevanten Grundfunktionen des IMIS eingeführt. Die Vorstellung des Systems muss daher auch hier auf einen kleinen Ausschnitt beschränkt werden.

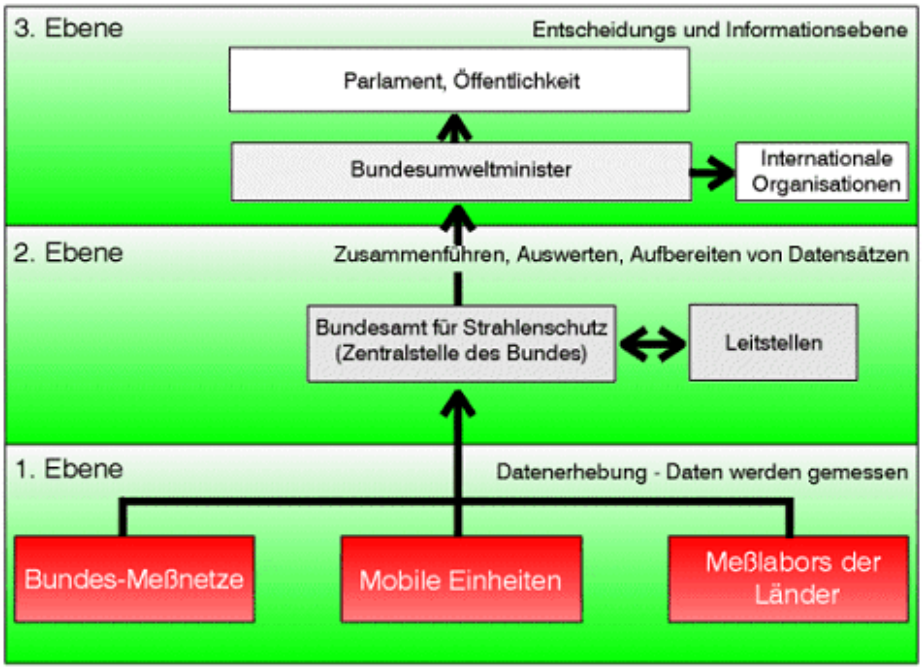


Abb.1: Ebenen des Datenflusses und Aufgabenverteilung

2. Erfassung von Daten

Das IMIS unterstützt die Erfassung von Messdaten durch zwei Komponenten. An der serverseitigen Importschnittstelle werden Datenaustauschdateien automatisiert entgegengenommen und in die Datenbank importiert. Mit Hilfe spezieller Masken in der IMIS Client-Anwendung können einzelne Messdatensätze hingegen manuell erfasst werden. Den größten Anteil der Datenzulieferung machen die Datenaustauschdateien aus den Bundesmessnetzen aus. So betreibt das BfS ein eigenes ODL[15]-Messnetz aus mehr als 2.100 Messstationen, die im Routinebetrieb zweistündlich und im Intensivbetrieb alle zehn Minuten ihre Daten übermitteln. Aber auch die über 40 Messlabors, die aufgrund ebenfalls mit IMIS erstellter Pläne Proben entnehmen (z.B. Nahrungsmittel, Futtermittel, Pflanzen, Gewässerproben, Niederschlag usw.) und untersuchen, nutzen verstärkt die Möglichkeit des automatisierten Datenimports, indem sie mit Hilfe ihres Laborinformationssystems erstellte Datendateien an die zentralen IMIS - Server übertragen. Ferner wird der Import meteorologischer Daten wie die Bahnen hypothetischer Luftpartikel (sog. Trajektorien) und Niederschlagswahrscheinlichkeiten in eigens hierfür entworfenen Austauschformaten unterstützt. Diese benötigt das integrierte Simulationssystem zur

[15] ODL: Ortsdosisleistung

Berechnung von Ausbreitungsprognosen. Schließlich verfügt das IMIS über Importschnittstellen für internationale Datenformate (IDF und EUDF), mit deren Hilfe der Austausch verschiedener EU-Ländern (Deutschland, Frankreich, Schweiz, Ostseeanrainer) erfolgt.

Insgesamt unterstützt das IMIS zehn unterschiedliche Formate für den Datenaustausch mit direkt assoziierten Anwendern und internationalen Partnern.

Sämtliche Austauschdateien werden auf den zentralen Servern des BfS in die Datenbank importiert. Die Zulieferung kann aus Sicht des IMIS passiv erfolgen (FTP, SFTP, Email-Anhang). Aber auch der aktive Download von Austauschdateien per (S)FTP oder HTTP(S) ist möglich und über die Workflowkomponente zudem automatisierbar.

2.1. Fachliche Bedeutung der Datenerfassung

Die Nutzung als Frühwarnsystem erfordert aus fachlicher Sicht neben der Sicherstellung der Datenqualität natürlich auch die Einhaltung der Anforderungen an eine höchstmögliche Verarbeitungsgeschwindigkeit. Dies gilt erster Linie für die kontinuierlich zu gelieferten Daten aus den Messstationen, auf deren Basis in kürzester Zeit Ausbreitungsprognosen berechnet werden müssen. IMIS importiert die Ergebnisse einer Ausbreitungsrechnung (ca. 500 MB) innerhalb von 15 Minuten in die Datenbank.

Neben den genannten Kriterien war eine hohe Benutzerfreundlichkeit zur Erreichung größtmöglicher Akzeptanz bei den Fachanwendern in den Laboren eine wichtige Anforderung, da ein einfacher Umgang mit dem System und eine weitgehende Unterstützung bei der Datenerfassung auch spürbare Auswirkungen auf die Datenqualität haben.

2.2. Sicherung der Datenqualität bei der Datenerfassung

Die folgenden systemseitigen Funktionen dienen der Qualitätssicherung bei der Datenerfassung:

Automatisierter Datenimport: Der Import auf Basis der in IMIS integrierten Workflowkomponente trägt zur Qualitätssicherung bei, da durch Minimierung von manuellen Arbeitsschritten nicht nur die Effizienz gesteigert, sondern auch potentielle

Fehlerquellen ausgeschlossen werden. Automatisch versendete Benachrichtigungen an die beteiligten Personen weisen diese zeitnah auf die Verfügbarkeit neuer Daten hin, die im nächsten Schritt einer manuellen Plausibilisierung zu unterziehen sind (siehe Abschnitt 0).

Konsistenzprüfung: Daten werden anhand konfigurierbarer Regeln während des Imports auf ihre Konsistenz geprüft. Dies betrifft auf der einen Seite einfache Prüfungen z.B. auf das Vorhandensein zwingender Werte oder die Einhaltung von Wertebereichen. Auf der anderen Seite werden aber auch komplexere Regeln, die die logische Konsistenz innerhalb eines Datensatzes sicherstellen, geprüft.

Beispiel: Befindet sich die in einer Ortsangabe enthaltene Koordinate innerhalb der ebenfalls spezifizierten Verwaltungseinheit?

Erkannte Probleme werden an den Datenlieferanten zurückgemeldet, wahlweise durch Systemmeldungen und/oder per Email. Hieraufhin kann der Lieferant entweder seine Datenaustauschdatei korrigieren und erneut übermitteln oder er korrigiert den inkonsistenten Datensatz direkt im System mit Hilfe der Erfassungsmasken. Dies ist möglich, da IMIS syntaktisch korrekte Datensätze selbst dann importiert, wenn sie gegen die Konsistenzregeln verstoßen, und somit dem Anwender den Weg der einfachen manuellen Korrektur im System offen hält.

Auch die Masken zur manuellen Datenerfassung bzw. -korrektur nutzen die Konsistenzregeln, um dem Anwender eine sofortige Rückmeldung zu geben.

Intelligente Erfassungsmasken: Die Erfassungsmasken sind mit dynamischen Regeln unterlegt, die der Vermeidung von Eingabefehlern dienen. So passen sich beispielsweise die Listen der gültigen Eingaben ständig dem aktuellen Kontext an, der sich aus bereits gesetzten Werten ergibt. Je nach Feldtyp werden somit Eingaben abgewiesen (Textfeld) oder gar nicht erst angeboten (Auswahllisten). Darüber hinaus werden an allen Stellen an denen dies aufgrund der verfügbaren Regeln und Stammdaten möglich ist, Felder automatisch gefüllt. Eine weitere Funktionalität, die die Effizienz steigert und Fehler vermeiden hilft, ist die der benutzerspezifischen Vorbelegungen. Anwender können beliebige Zustände einer teilweise oder auch vollständig ausgefüllten Maske unter einem eindeutigen Namen abspeichern und jederzeit laden. Dies vereinfacht z.B. die Erfassung neuer Proben, wenn dort bestimmte Informationen immer wieder gleich einzugeben sind.

Automatische Probenzuordnung: Ein bereits erwähntes Steuerungsinstrument im IMIS sind die so genannten Probenahmepläne. Hiermit wird für einzelne Institutionen im Voraus geplant, in welchen Zyklen und aus welchen Umweltbereichen Proben zu entnehmen sind und welchen Messmethoden diese zu unterziehen sind. Mit Hilfe dieser Pläne können die betroffenen Anwender ihre Aufgaben organisieren und Tourenpläne erstellen. Darüber hinaus erzeugt IMIS anhand solcher Pläne Rohdatensätze. Beim Import von Messdaten werden diese anhand festgelegter Regeln automatisch den entsprechenden Rohdatensätzen zugeordnet, die somit vervollständigt werden. Treten hierbei Fehler auf, wird der Datenlieferant informiert. Er kann nun die Datensätze mit IMIS-Mitteln vergleichen und eine Korrektur vornehmen bzw. die Zuordnung manuell herstellen.

3. Bewertung von Daten

Nachdem die Erfassung eines Datensatzes abgeschlossen ist, wird dieser vom Erfasser für die so genannte Plausibilisierung auf der nächst höheren organisatorischen Ebene freigegeben. Beim dateibasierten Import kann diese Freigabeinformation bereits mitgeliefert werden. Bei der manuellen Datenerfassung muss sie explizit erfolgen.

3.1. Fachliche Bedeutung der Datenbewertung

Die Information, inwiefern ein Datensatz plausibel ist und damit überhaupt für eine weitere Verarbeitung und schließlich zur Lagebeurteilung herangezogen werden kann, ist im IMIS von zentraler Bedeutung.

Der Prozess der Plausibilisierung basiert auf einer vierstufigen Statusvergabe. Auf der untersten Ebene signalisiert der Datenlieferant (in der Regel die Landesmessstelle) die Vollständigkeit und Korrektheit einer Messung. In der zweiten Stufe prüft die Landesdatenzentrale alle in ihrem Bundesland durchgeführten Messungen. Die dritte Aggregationsebene stellen die Leitstellen dar, deren Zuständigkeitsbereiche – unabhängig vom geographischen Ursprung – durch Umweltbereichsgruppen definiert sind. In der letzten Stufe vergeben Vertreter des BMU den höchsten Statuswert und entscheiden final über die Freigabe jedes Datensatzes. Erst nach dieser Entscheidung ist eine Messung für alle sichtbar und für die Veröffentlichung freigegeben. In der Praxis übernehmen häufig fachlich

verantwortlich Mitarbeiter des BfS als nachgeordnete Behörde diese Freigabe im Auftrag des BMU. Lediglich in der kritischen Situation eines Ereignisses (Intensivbetrieb) erfolgt dies ausschließlich durch Mitarbeiter der BMU.

3.2. Sicherung der Datenqualität bei der Datenbewertung

Der gesamte Vorgang der Datenbewertung und Plausibilisierung dient der Sicherung der Datenqualität an sich. Das IMIS bietet dem Anwender hierbei vielfältige Unterstützung. Diese betrifft in erster Linie die Recherche und Darstellung zu plausibilisierender Daten. Da eine Prüfung auf Basis einzelner Datensätze bei dem gegebenen Mengengerüst nicht praktikabel ist, werden ganze Ansammlungen von Datensätzen im Kontext betrachtet und bewertet. So ist es z.B. möglich im Rahmen von Karten, Diagrammen und Tabellen zusammenhängende Messungen zu betrachten, um hierbei Ausreißer zu identifizieren und nur diese einer Einzelprüfung zu unterziehen. Die flexible Recherchekomponente im IMIS erlaubt es aber auch Datensätze zu selektieren die beliebigen Kriterien entsprechen. Damit ist der Anwender in der Lage gezielt Positiv- oder Negativlisten zu ermittelt.

Dem Anwender steht zudem eine faktisch unbegrenzte Sammlung von Vergleichsdarstellungen früherer Lagesituationen zur Gegenüberstellung mit aktuellen Daten zur Verfügung. Im direkten Vergleich kann er somit die aktuellen Daten synoptisch plausibilisieren.

Mit Hilfe von „drill down“ Funktionen kann der Anwender alle Detailinformationen auffälliger Datensätze aufrufen und so eingehende Analysen durchführen.

Schließlich sichert das IMIS die Plausibilisierung insofern ab, als es eine Statusvergabe ablehnt, solange entsprechend konfigurierte Konsistenzregeln verletzt sind. Bei diesen bereits oben erwähnten Konsistenzregeln handelt es sich in JavaScript formulierte Code-Fragmente, die in der zentralen Datenbank gespeichert sind und zur Laufzeit von IMIS ausgeführt werden. Auf diese Weise ist eine vergleichsweise einfache Anpassung und Erweiterung des Regelwerks möglich, da ein aufwändiges Programmieren und Verteilen neuer Anwendungsversionen entfällt.

Nicht zuletzt ist die von IMIS gesteuerte Datensichtbarkeit ein weiteres wichtiges Instrument im Kontext der Datenqualität. Hiermit wird sichergestellt, dass jede der eingangs erwähnten Organisationseinheiten in der Kette der vierstufigen

Statusvergabe nur die Datensätze sehen kann, die von der darunter liegenden Einheit freigegeben wurden. Dies verhindert Fehlinterpretationen auf Basis unvollständiger oder fehlerhafter und deshalb nicht freigegebener Daten.

4 Aufbereitung und Visualisierung von Daten

Sowohl im Rahmen der Bewertung der Datenqualität als auch zur Lagebeurteilung sind die in IMIS integrierten Visualisierungswerkzeuge wichtige Instrumente.

4.1. Fachliche Bedeutung der Datenvisualisierung

Unabhängig von der Frage, ob ein Anwender das Ziel hat, die Plausibilität neuer Daten zu prüfen oder ob er mit der Aufgabe der Lagebeurteilung auf Basis bereits plausibler Daten betraut ist, benötigt er Werkzeuge, die es ihm erlauben, Daten auf verschiedene Weise zu aggregieren und in Karten, Zeitreihen oder Tabellen anzuzeigen. Ferner besteht hier der Bedarf, diese aktuellen Messdaten mit historischen Vergleichswerten und Richtwerten in Beziehung zu setzen. Ein weiterer wichtiger Schritt ist die Möglichkeit, sich vom System entsprechend der aktuellen Lage geeignete Darstellungen anzeigen zu lassen, aus denen empfehlenswerte Maßnahmen zur Reduzierung der Strahlenexposition direkt ablesbar sind.

4.2. Sicherung der Datenqualität bei der Datenvisualisierung

Die funktionale Gestaltung des IMIS im Bereich der Datenvisualisierung war neben dem Wunsch nach hoher Flexibilität und Benutzerfreundlichkeit auch vom Ziel getrieben, Fehlbedienungen und in deren Folge eventuelle Fehlinterpretationen der Daten zu vermeiden.

Getrieben von diesem Ziel wurden die Darstellungen bei ihrer Modellierung in zwei Hauptteile untergliedert. Der eine Teil beschreibt eine Recherchedefinition. Das Ergebnis einer Recherche ist immer eine zweidimensionale Tabelle. In der Recherchedefinition werden die vom Benutzer mittels einer einfach zu bedienenden Point & Click Oberfläche zusammengestellten Recherchekriterien fixiert. Diese Kriterien beschreiben welche Daten der Anwender recherchieren will (Ergebnisspalten), ob und in welcher Form Ergebnisdaten aggregiert oder transformiert werden sollen und welchen Bedingungen die Ergebnisdatensätze entsprechen sollen. Aus diesen Angaben konstruiert IMIS eine SQL-Abfrage und

führt ggf. Berechnungen oder Formatierungen aus, bevor es das Ergebnis an die Anzeigekomponente weiterreicht.

Die Anzeigekomponente zieht für ihre Aufgabe neben dem aufbereiteten Rechercheergebnis den zweiten Informationsteil der Darstellung heran. Dieser enthält die vom Anwender in entsprechenden Konfigurationsdialogen definierten Layoutregeln für die Darstellung. Dort finden sich Angaben darüber, welcher Darstellungstyp (Karte, Diagramm, Tabelle und deren Unter- und Mischformen) zu verwenden ist. Wie die darzustellenden Messwerte zu klassifizieren sind (Grenzen der Klassifikationsstufen, zu verwendende Farben, Muster oder Symbole, einzublendende Geobasisthemen, darzustellender Kartenausschnitt und Maßstab etc.)

Die Trennung dieser Definitionselemente einer Darstellung erfolgt nicht nur im zugrunde liegenden Objektmodell, sondern wird auch an der Benutzeroberfläche vollzogen. Dies hat zur Folge, dass Recherchedefinitionen ebenso wie Layoutdefinitionen unabhängig voneinander genutzt werden können. So lassen sich die einzelnen Bestandteile beliebig verschieben, kopieren und in anderen Darstellungen nutzen (eine Darstellung kann auch die Ergebnisse mehrerer Recherchen visualisieren). Ferner werden Recherchedefinition auch als Grundlage für Datenexporte genutzt. Somit kann auf einfache Weise das Ergebnis eines Exports visuell kontrolliert werden und umgekehrt können die Inhalte einer Darstellung mit Hilfe der Exportkomponente in eine Datenaustauschdatei geschrieben werden.

Die so gespeicherten Darstellungsdefinitionen können jederzeit aufgerufen werden, um die aktuelle Datenlage zu visualisieren. Das Berechtigungskonzept erlaubt es Experten komplexe Recherchen und Visualisierungsregeln vorzugeben, die durch einfache Anwender nicht bzw. nur eingeschränkt modifiziert werden können (z.B. Änderung des Selektionszeitraums oder des gesuchten Umweltbereiches). Auf diese Weise sind fachlich kritische Rechercheparameter, deren Änderung die Aussage der Darstellung beeinflussen kann, geschützt.

Zusätzlich können notwendige Eingriffe in die Abfragebedingungen und damit potentielle Fehlerquellen durch so genannte dynamische Recherchekriterien reduziert werden. Hierbei handelt es sich um logische Platzhalter für

Abfragebedingungen, die erst zur Ausführungszeit von IMIS automatisch mit den korrekten Vergleichswerten gefüllt werden. Beispiele:

- die jüngsten Daten,
- alle Daten aus dem vorangegangenen Importvorgang (wird im Kontext von automatisierten Workflows eingesetzt, in deren Rahmen z.B. nach erfolgreichem Import eine Darstellung zu erzeugen ist),
- alle Daten aus einer bestimmten Ausbreitungsprognose (ebenfalls im Kontext automatisierter Workflows).

Ein herausragendes Element bei der Unterstützung der Datenqualität ist die konsequente Berücksichtigung von Informationen bezüglich der im Messgerät und im Messverfahren begründeten Messwertqualität (Stichwort Nachweisgrenze). IMIS kann auch dann noch Informationen bezüglich der Nachweisgrenzen ausweisen, wenn in den Lagedarstellungen nur noch aggregierte Werte visualisiert werden.

Ein weiteres Hilfsmittel zur Vermeidung von Fehlinterpretationen ist der Warnhinweis, der in Darstellungen ebenso wie in den von IMIS erzeugten PDF-Dokumenten erscheint, wenn diese Testdaten enthalten. Die Einblendung dieses Hinweises kann nicht unterdrückt werden.

Zuletzt ist die dynamische Generierung von Legendeninhalten zu erwähnen. Damit sichergestellt wird, dass Darstellungslegenden immer korrekte Auskunft über die Darstellungsinhalte geben, wurde ein Mechanismus realisiert, der dafür sorgt, dass das Legendenlayout vom Anwender zwar mittels Textbausteinen konstruiert und anschließend gespeichert wird. Das Ersetzen der Textbausteine erfolgt jedoch erst zum Zeitpunkt der Darstellungserzeugung. In diesem Moment fügt IMIS die tatsächlichen Informationen ein, die es aus den verwendeten Recherchekriterien, den Recherchebedingungen und den Rechercheergebnissen extrahiert.

5. Veröffentlichung von Daten

Die regelmäßige Information anderer Behörden sowie der Öffentlichkeit und der internationale Datenaustausch sind im Strahlenschutzvorsorgegesetz selbst verankert und spielen daher eine bedeutende Rolle im IMIS.

5.1. Fachliche Bedeutung der Datenveröffentlichung

Für das BfS als Betreiber des IMIS ist die Erzeugung und Veröffentlichung von klar gestalteten, inhaltlich korrekten und für alle Betrachter objektiv zweifelsfrei interpretierbaren Lagedarstellungen von weit reichender Bedeutung, da Fehlinformationen im Bereich der Bevölkerungsschutzes verheerende Auswirkungen in politischer, wirtschaftlicher und sozialer Hinsicht haben können. Auch verfolgte das BfS mit der Realisierung des neuen IMIS durch Condat das Ziel, die Durchlaufzeiten bis zur Veröffentlichung von Lagedarstellungen zu beschleunigen und somit aktuellere Situationsbeschreibungen anzubieten, als dies auf Basis des inzwischen abgelegten Systems der Fall war. Dieses Ziel hat das neue IMIS insofern erreicht, als dass die technischen Verarbeitungsschritte unter maximaler Ausnutzung von Automatisierung optimiert wurden. Der Zeitpunkt bis zur Veröffentlichung wird nun durch die unvermeidlichen organisatorischen Durchläufe (Plausibilisierung, Bewertung, Freigabe) bestimmt. Der in dieser Hinsicht unaufwändigste Prozess ist die Auswertung der Gamma-Ortsdosisleistung auf Basis der automatisch liefernden Messnetze des BfS. Die entsprechende Darstellung täglich aktualisiert und kann unter http://www.bfs.de/ion/imis eingesehen werden.

5.2. Sicherung der Datenqualität bei der Datenveröffentlichung

Die qualitätsgesicherte Veröffentlichung von Daten wird in erster Linie von der im IMIS integrierten Workflowkomponente getragen. Diese Komponente automatisiert alle hierfür geeigneten Schritte des Datenflusses von der Messdatenbereitstellung bis zur Veröffentlichung von Lagedarstellungen auf Basis der importierten und freigegebenen Daten.

Den Kern der Workflowkomponente bilden frei definierbare Aufträge, die in Ketten zu beliebig komplexen Workflows zusammengeschaltet werden können. Jeder Auftrag hat ein auslösendes Ereignis. Dies kann das Eintreffen einer Datei mit vorgegebenem Namen oder Namensmuster sein, in diesem Fall reden wir von dateigesteuerten Aufträgen. Zeitgesteuerte Aufträge hingegen werden nach einem Intervallmuster ähnlich zu Terminserien zyklisch ausgeführt. Die eigentlichen Arbeitsschritte eines Auftrags, die so genannte Ablaufdefinition, werden mittels JavaScript spezifiziert. Condat hat an dieser Stelle ein API zur Verfügung gestellt, über das Verarbeitungsfunktionen genutzt werden können, die entweder

Funktionalität aus dem IMIS selbst zugänglich machen oder im Rahmen der Workflowautomatisierung sinnvolle Ergänzungen darstellen. Einige Beispiele für solche Verarbeitungsfunktionen sind:

- Senden und Empfangen von Dateien über (S)FTP oder HTTP(S),
- Importieren und Exportieren von Austauschdateien,
- Erzeugen von Darstellungen und Überführung in PDF - Dokumente,
- Erzeugen und Versenden von Email-Nachrichten inklusive Anlagen.

Hiermit steht den explizit geschulten Anwendern ein mächtiges Werkzeug zur Verfügung, mit dessen Hilfe nicht nur kritische Prozesse automatisiert werden können. Das BfS nutzt die flexiblen Möglichkeiten dieser Komponente inzwischen bereits zur Unterstützung von vielfältigen Arbeitsabläufen des Tagesgeschäfts.

Im Kontext der Datenveröffentlichung ergibt sich der mit der Automatisierung verbundene Vorteil aus der Tatsache, dass einmal definierte Prozesse zuverlässig ausgeführt werden. Die fest vorgegebene Nutzung der korrekten Darstellungsdefinitionen (siehe Abschnitt 0) für die Dokumenterzeugung schafft eine verlässliche Aussagekraft und ermöglicht die objektive Vergleichbarkeit der verschiedenen Lagedarstellungen.

Die Steuerungsmittel der Workflowkomponente erlauben zudem zielgruppenspezifische Darstellungen zu erzeugen und über differenzierte Wege zu unterschiedlichen Zielen zu verteilen. Hierzu bietet die Workflowkomponente Schnittstellen, über die die Ergebnisdokumente

- in das integrierte Dokumentenmanagementsystem eingestellt (Intranet)
- per Email an fei konfigurierbare Empfänger versandt
- automatisch auf zentralen Druckern ausgegeben
- über entsprechende Gateways per Fax verteilt
- mittels FTP oder SFTP an Austauschpartner übertragen
- und schließlich auch auf dem öffentlichen Internetserver des BfS abgelegt

werden können.

6. Zusammenfassung und Ausblick

Ein großer Teil des Systementwurfs für das neue IMIS war von der Aufgabe geprägt, die Qualität der hiermit verwalteten Daten zu sichern und Instrumente zur zuverlässigen Beurteilung der Datenqualität zur Verfügung zu stellen. Vor diesem Hintergrund hat Condat eine Architektur aus lose gekoppelten Komponenten entworfen, die eine Vielzahl von unterstützenden Funktionen zur Sicherung und transparenten Darstellung der Datenqualität bereitstellen und die im Überblick in diesem Dokument vorgestellt wurden.

Mit dem Hintergrund, ein System zur Überwachung der Umweltradioaktivität zu sein, ergaben sich für IMIS zahlreiche fachspezifischer Anforderungen. Diese fachlichen Besonderheiten wurden jedoch bewusst aus den Kernmodulen für Recherche, Visualisierung, Dokumentenerzeugung und Workflowautomatisierung heraus gehalten. Diese Komponenten sind universell ausgelegt und können ihre Aufgaben unabhängig davon wahrnehmen, ob es sich bei den Fachdaten um Radioaktivität, Hochwasserpegel oder CO2-Emissionen handelt.

Diese Komponentenarchitektur bildet eine sinnvolle Ausgangsposition für zukünftige Erweiterungen und die Übertragung des Systems auf andere Bereiche der Frühwarnung und Entscheidungsunterstützung. So ist es beispielsweise einfach möglich das Fachdatenmodell zu erweitern und neue Importschnittstellen zu ergänzen (z.B. Unterstützung für Daten gemäß OGC Sensor Web Enablement).

Auch das BfS plant in Kenntnis dieser Offenheit des IMIS die Nutzung auf ursprünglich nicht vorgesehene Datenarten auszuweiten. So sollen nach einer von uns durchzuführenden Ergänzung des Datenmodells ab 2006 alle Daten nach der Richtlinie zur Emissions- und Immissionsüberwachung kerntechnischer Anlagen (REI) mit IMIS verarbeitet werden.

Leider bietet der hier gegebene Rahmen nicht genug Raum, um die mit Sicherheit beim Lesen entstandenen und offen gebliebenen Fragen zu beantworten. Ich möchte Sie daher ermutigen, mit mir Kontakt[16] aufzunehmen, um Ihre Neugier zu befriedigen und gerne auch kritische Anmerkungen zu diskutieren.

[16] Kontakt zum Autor: Volkmar Schulz, Projektleiter IMIS - Realisierung, vs@condat.de, Tel.: 030/3949-1125

Naturpilot Schleswig-Holstein - Präsentation von Natur-Highlights im interaktiven, virtuellen Ballonflug

Friedhelm Hosenfeld[17], Andreas Rinker[18], , Ernst-Walter Reiche[19], †, Dirk Bornhöft[20] und Gudrun Schultz[21]

Abstract

Der Naturpilot Schleswig-Holstein präsentiert ausgewählte Natur-Attraktionen von Schleswig-Holstein im Web. Mit dem Naturpilot entstand nicht einfach eine weitere Portalseite, sondern eine interaktive Web-Anwendung mit einer Vielzahl von Optionen und Funktionen, die den Nutzenden angeboten werden. Im „Standard-Modus“ kann jeder Interessierte in einem virtuellen Ballon über Schleswig-Holstein schweben, während gleichzeitig dazu passende Fotos und erläuternde Texte präsentiert werden. Darüber hinaus können etwa 50 „Naturhighlights“ direkt angesteuert werden. Dort ist es möglich, den Ballon abzusenken, so dass sich das Naturhighlight in Form aufbereiteter Luftbilder präsentiert. Ausführliche Informationen zu den Highlights werden ebenfalls angeboten.

Im Modus „Freies Fliegen“ kann der Ballon überall über der Schleswig-Holstein-Karte schweben und über sechs „Höhenstufen“ abgesenkt werden. Auf den beiden unteren Höhenstufen werden farbige Orthophotos des angesteuerten Bereiches angezeigt. Neben dieser Ballon-Navigation über Schleswig-Holstein wird eine Reihe

[17] DigSyLand - Institut für Digitale Systemanalyse & Landschaftsdiagnose, Zum Dorfteich 6, D-24975 Husby, mailto: {hosenfeld | rinker }@digsyland.de , Internet: http://www.digsyland.de/
[18] DigSyLand - Institut für Digitale Systemanalyse & Landschaftsdiagnose, Zum Dorfteich 6, D-24975 Husby, mailto: {hosenfeld | rinker }@digsyland.de, Internet: http://www.digsyland.de/
[19] DigSyLand - Institut für Digitale Systemanalyse & Landschaftsdiagnose, Zum Dorfteich 6, D-24975 Husby, mailto: {hosenfeld | rinker }@digsyland.de, Internet: http://www.digsyland.de/
[20] Ministerium für Landwirtschaft, Umwelt und ländliche Räume des Landes Schleswig-Holstein, Mercatorstraße 3, D-24106 Kiel, mailto: { dirk.bornhoeft | gudrun.schultz }@munl.landsh.de, Internet: http://www.umweltministerium.schleswig-holstein.de/
[21] Ministerium für Landwirtschaft, Umwelt und ländliche Räume des Landes Schleswig-Holstein, Mercatorstraße 3, D-24106 Kiel, mailto:{ dirk.bornhoeft | gudrun.schultz }@munl.landsh.de, Internet: http://www.umweltministerium.schleswig-holstein.de/

weiterer Funktionen zur Verfügung gestellt, wie etwa die interaktive Präsentation von knapp 50 Kanu-, Fahrrad- und Wandertouren, eine Suche nach allen Gemeinden und Ortschaften Schleswig-Holsteins, virtuelle 3-D-Rundflügen ausgewählter Gebiete und einiges mehr.

Eine wichtige Anforderung an den Naturpilot war die Kombination dieses breit gefächerten Funktionsumfangs mit einer einfachen, intuitiven Bedienungsphilosophie, die auch für Gelegenheitsnutzer rasch erfassbar ist.

Technisch basiert die Anwendung auf einem modularen Software-Framework unter der Verwendung von PHP, JavaScript, dem Minnesota (UMN) MapServer und dem Datenbanksystem MySQL (wahlweise Oracle).

1. Einführung

Das Land Schleswig-Holstein präsentiert seine Umwelt- und Agrarinformationen online durch den Einsatz populärer und bewährter Web-Anwendungen wie dem Umwelt- und Agrarberichtssystem[22] (Rammert/Hosenfeld 2003) innerhalb der Umwelt-Internet-Plattform InfoNet-Umwelt[23] (Bornhöft et al. 2000) und dem Umweltatlas[24] für geographische Daten (Görtzen et al. 2004).

Ergänzend zu diesen Präsentationsformen wurde eine neue Anwendung konzipiert, die eine breitere Gruppe von Nutzenden adressiert. Unter anderem sollen Schulen, Touristen und Gelegenheitsnutzer des Internets angesprochen werden. Ein stärkerer Schwerpunkt wird auf „Edutainment" und touristische Aspekte gelegt. Das damalige Ministerium für Umwelt, Naturschutz und Landwirtschaft (jetzt: Ministerium für Landwirtschaft, Umwelt und ländliche Räume) gab 2004 die Erstellung von Luftbildern (Orthophotos) von ganz Schleswig-Holstein in Auftrag, da diese im Agrarbereich benötigt werden. Dies passte sehr gut in das Konzept der neuen Web-

[22] Umwelt- und Agrarbericht Schleswig-Holstein: http://www.umweltbericht-sh.de/
[23] InfoNet-Umwelt Schleswig-Holstein: http://www.umwelt.schleswig-holstein.de/
[24] Umweltatlas des Landes Schleswig-Holstein: http://www.umweltatlas-sh.de/

Anwendung, da die Luftbilder so auch der Öffentlichkeit in einer ansprechenden Form zugänglich gemacht werden können.

Auf diese Weise entstand die Idee eines virtuellen Ballonflugs über Schleswig-Holstein. Aus der Ballon-Perspektive wird eine interaktive Exploration der Natur im Land angeboten. Für über 50 Naturhighlights wurden Hintergrundinformationen, sowie Fotos und ergänzende Daten aktuell zusammengestellt, in einer unterhaltsamen Form aufbereitet und in den Naturpilot integriert. Der Landesnaturschutzverband Schleswig-Holstein (LNV) konnte für das Projekt als Kooperationspartner gewonnen werden, der die Präsentation von Naturschutzthemen unterstützt.

Abb.1: Haupt-Anwendungsfenster des *Naturpilot Schleswig-Holstein*

2. Voraussetzungen und Anforderungen

Die zu entwickelnde Anwendung sollte unter der Berücksichtigung der gegebenen Voraussetzungen einige wichtige Anforderungen erfüllen:

- Ein gefälliges und „unterhaltsames" Erscheinungsbild,
- eine einfach zu bedienende und selbsterklärende Bedienungsoberfläche,
- keine speziellen Hard- und Software-Anforderungen für die Nutzenden,
- Unterstützung der gängigsten Web-Browser (wie Internet Explorer und Mozilla Firefox),
- Nutzbarkeit mit normalen Bildschirm-Auflösungen (ab 1024 * 768),
- Einsatz vorhandener Technik, soweit dies möglich ist,
- Präsentation der digitalen Orthophotos mit hoher Qualität aber trotzdem mit einer kleinen Dateigröße, die auch von Menschen mit langsameren Internet-Anbindungen sinnvoll genutzt werden können und
- die Möglichkeit zur späteren Erweiterbarkeit.

3. Funktionalität der Web-Anwendung

Um die einfache Bedienbarkeit, die vor allem für Internet-Nutzende mit wenig Computer-Erfahrung wichtig ist, mit dem weitreichenden Funktionsumfang, den die Mapserver-Software bietet und der vor allem für erfahrene Nutzer interessant ist, zu kombinieren, wurden zwei verschiedene Bedienungs-Modi eingerichtet:

3.1. Standard-Modus

Im *Standard-Modus* werden auf der Karte von Schleswig-Holstein die Naturhighlights durch farbige Symbole dargestellt (s. Abb. 1). Wenn eines dieser *Naturhighlights* mit der Maus ausgewählt wird, fliegt der Ballon direkt zu der entsprechenden Position auf der Karte. Dort kann der Ballon abgesenkt werden, indem ein aufbereitetes Luft- oder Kartenbild präsentiert wird, auf dem das Naturhighlight aus der Vogel-Perspektive betrachtet werden kann.

Zu jedem Naturhighlight können ausführliche Informationen abgerufen werden, die unter anderem Beschreibungen der Besonderheiten enthalten sowie Fotos, Adress- und Kontaktangaben.

Die Naturhighlights umfassen Naturschutzgebiete, Naturerlebnisräume, Natur- und Nationalparks und ähnliche Attraktionen.

3.2. Freies Fliegen

In dem Modus „*Freies Fliegen*" wird eine leicht abgewandelte Schleswig-Holstein-Karte als Grundlage präsentiert. Die Naturhighlights sind hier nicht mehr gesondert hervorgehoben. Dafür schwebt der Ballon frei über der Karte und ändert seine Flugrichtung, wenn mit der Maus in die Karte geklickt wird. Der Ballon kann an jeder beliebigen Stelle des Landes über sechs verschiedene Höhenstufen abgesenkt werden. Auf jeder Höhenstufe kann der Ballon mittels einer symbolischen Windrose in alle Himmelsrichtungen bewegt werden. Besonders reizvoll sind die beiden unteren Höhenstufen, in denen die Luftbilder angezeigt werden. Auf der zweit-untersten Ebene werden zur besseren Orientierung einige topographische Angaben wie Ortsnamen und Straßen über dem Luftbild eingeblendet.

Auf jeder Ebene (1-5) können per Mausklick zusätzliche Informationen interaktiv abgefragt werden, wie beispielsweise über die Gemeinde, Fahrrad-Routen, Naturschutzgebiete und ähnliches.

3.3. Touren

Der Naturpilot bietet eine Auswahl von knapp 50 verschiedenen Touren durch attraktive Gegenden Schleswig-Holsteins, die mit dem Kanu, mit dem Fahrrad oder auch zu Fuß unternommen werden. Wenn die Entscheidung für eine Tour gefallen ist, zeigt der Naturpilot ein Luftbild der Gegend an, auf dem die Tour gekennzeichnet ist. Um einen besseren Eindruck der Tour zu vermitteln, schwebt der Ballon anschließend über dem Luftbild entlang der markierten Route. Währenddessen werden Erklärungen zu den Sehenswürdigkeiten und wissenswerte Details vermittelt, ergänzt durch passende Fotos. Das Tempo kann dabei nach eigenem Ermessen eingestellt werden. Das Abfahren der Tour kann jederzeit gestoppt werden, um weitere Informationen zu der Tour abzurufen oder die Fotos in Vergrößerung zu betrachten. Jede Tour-Beschreibung kann zudem komplett als PDF-Datei

heruntergeladen (und ausgedruckt) werden. Für die Zukunft ist die interaktive Einbindung von Skater-Touren geplant, die bisher ausschließlich in PDF-Form zur Verfügung stehen.

Die Touren können nur im Standard-Modus interaktiv abgefahren werden. Jedoch ist das Umschalten zum Freien Fliegen jederzeit möglich.

3.4. Regionen Schleswig-Holsteins

In einer weiteren Funktion werden verschiedene Regionen Schleswig-Holsteins dargestellt. So können die *Naturräume* des Landes oder auch die *Kreise* aus einer Liste ausgewählt werden. Die jeweilige Region wird dann auf der Karte farbig markiert. Außerdem bewegt sich der Ballon in diese Region, so dass dort weitere Erkundungen gestartet werden können. Zu den Naturräumen werden umfangreiche Informationen beispielsweise zur Entstehungsgeschichte, zur Landschaftscharakteristik, zu typischen Lebensräumen und zu Schutzgebieten gegeben.

3.5. Suche

Mit der Suchfunktion des Naturpilot können alle Ortschaften und Gemeinden Schleswig-Holsteins anhand ihres Namens gesucht werden. Werden nur Teile eines Ortsnamens angegeben, wird eine Liste aller Möglichkeiten zur Auswahl präsentiert. Der Ballon bewegt sich anschließend zu dem gewählten Ort. Im Freien Fliegen kann der Ballon dann bis zur Luftbild-Ebene abgesenkt werden, um beispielsweise das eigene Haus, das aktuelle Feriendomizil oder ein Ausflugsziel in der Natur aus der Nähe zu betrachten.

Da alle Naturhighlights, Touren und Regionen in einer internen Datenbank verwaltet werden, werden diese durch die Stichwort-Suche ebenfalls berücksichtigt.

3.6. Virtuelle 3-D-Rundflüge

Für einige ausgewählte Gegenden Schleswig-Holsteins wurden auf der Basis der Luftbilder und des Digitalen Höhenmodells (DHM) 3-D-Filme berechnet und im mpeg-Format erzeugt. Diese virtuellen Rundflüge lassen sich in aktuellen Internet-Browsern direkt betrachten und vermitteln einen außergewöhnlichen Eindruck der betreffenden Gegend, wie etwa der Schlei-Mündung, dem Nord-Ostsee-Kanal oder

dem höchsten Berg Schleswig-Holsteins, dem Bungsberg. Die Ergänzung der Filme um textliche Erläuterungen und deren Vertonung ist für die Zukunft geplant.

3.7. Weitere Funktionen und Inhalte

Einige weitere Funktionen und Angebote vervollständigen die Web-Anwendung, wie beispielsweise eine Liste passender Internet-Links, ein Glossar („was bedeuten ...?“) und die integrierte Online-Hilfe.

4. Implementierung

Die Minnesota Mapserver-Software (UMN Mapserver 2005) wird eingesetzt, um alle geographischen Informationen auszuwerten und anzuzeigen. Die Sachdaten werden hingegen in einem relationalen Datenbank-Managementsystem verwaltet. Optional können MySQL oder Oracle eingesetzt werden (andere Systeme sind prinzipiell möglich).

Während der Ballon über Schleswig-Holstein schwebt, werden zugehörige Informationen und Fotos automatisch in der Informationsleiste angezeigt, wenn der Ballon ein Highlight erreicht. Befindet sich unter dem Ballon kein Naturhighlight, präsentiert der Naturpilot Informationen zu dem aktuellen Naturraum und dem Landkreis.

Ein modulares Software-Framework auf der Basis von PHP unterstützt von JavaScript-Skripten für die Nutzer-Steuerung leistet die Hauptarbeit beim dynamischen Informationsmanagement des Systems. Einige der Framework-Module basieren auf Elementen, die bereits im interaktiven Umwelt- und Agrarberichtssystem Schleswig-Holstein eingesetzt werden (Rammert/Hosenfeld 2003).

Abb.2: Präsentation der digitalen Orthofotos (links) und eine interaktive Tour (rechts)

4.1.Informationsverwaltung auf der obersten Ebene

Wenn sich der Ballon auf der obersten Ebene bewegt (über der Karte von ganz Schleswig-Holstein) wird die Ballonposition mit Hilfe eines virtuellen Gitternetzes (Kacheln mit der Fläche 10 km * 10 km) verfolgt. Beim Starten des Naturpilot werden die Kachel-Informationen dynamisch mit Hilfe von PHP in JavaScript-Arrays geladen und stehen so für alle weiteren Auswertungen zur Verfügung.

Zu jeder Kachel werden Angaben zu den Naturhighlights, Naturräumen, Kreisen sowie zu den Fotos und Texten vorgehalten. In der Datenbank werden diese Informationen so verwaltet, dass jedem Objekt (Highlight, Gemeinde, Kreis, ...) Bezeichner, Identifikationsnummern, Rechts- und Hochwerte (GK3: Gauß-Krüger-System, 3. Meridianstreifen) sowie je nach Typ weitere Angaben zugeordnet werden. Bei den flächenhaften Objekten geben die Koordinaten nur den (Mittel-)Punkt an, in dem der Ballon platziert werden soll. Die Geometrien werden vom Mapserver bereitgestellt. Den Raster-Kacheln werden in der Datenbank die räumlich zugehörigen Objekte zugeordnet. Mit Hilfe einfacher Formeln lassen sich die GK-Koordinaten in die betreffende Kachelnummer und auch in die Bildkoordinaten umrechnen. Dies geschieht zur Laufzeit durch den entsprechenden JavaScript- bzw. PHP-Code. Falls andere Bild- oder Rastergrößen eingesetzt werden sollen, sind nur wenige Umrechnungskonstanten anzupassen. Die einmalige Verschneidung des Kachelrasters mit den übrigen Geometrien wurden mit Hilfe eines GIS durchgeführt. Zusätzliche Informationsebenen, die mit dem Kachel-Raster verbunden werden und dann interaktiv zur Verfügung stehen, sind daher leicht ergänzbar.

Im Freien Fliegen erfolgt die Bewegung des Ballons in Richtung eines Mausklicks. Der Ballon behält diese Richtung schwebend bei, bis über einen weiteren Mausklick eine Richtungsänderung signalisiert oder eine andere Funktion des Naturpilot aktiviert wird. Die Geschwindigkeit des Ballons lässt sich in fünf Stufen regeln. Die automatische Bewegung („Schweben“) des Ballons wird in JavaScript mit einer geschwindigkeitsabhängigen Zeitschleife realisiert.

Wenn ein Naturhighlight (im Standard-Modus) oder ein anderer Punkt auf der Karte (im Freien Fliegen) ausgewählt wird, aktiviert sich der UMN Mapserver im Hintergrund, um bei Bedarf die aktuelle Position auf einer Referenzkarte oder die Karten-Legende zu präsentieren.

4.2. Interaktive Touren

Nach Auswahl einer Tour (s. Kap.0) werden alle relevanten Informationen mittels PHP in JavaScript-Variablen geladen. Jede Tour wird über einzelne ASCII-Dateien konfiguriert. Auf diese Weise können die Tour-Angaben leicht handhabbar gepflegt und aktualisiert werden. Auf dem Client-Rechner müssen im Browser nicht immer alle verfügbaren Touren verwaltet werden, sondern nur die aktuell ausgewählte. Die Konfigurationsdateien enthalten die Bild-Koordinaten, denen der Ballon auf seiner Route folgt, und außerdem die Angaben zu den relevanten URLs, Text- und Bilddateien, die während der Tour angezeigt werden.

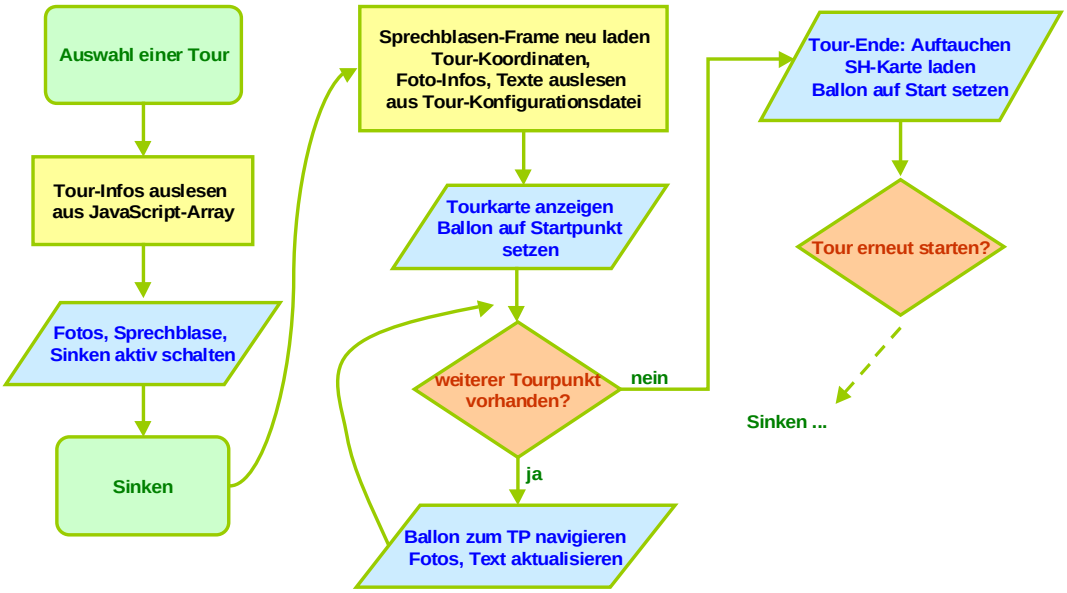


Abb.3: Symbolischer Ablauf einer interaktiven Tour

4.3. Freies Fliegen auf den unteren Ebenen

Im Modus „Freien Fliegen“ wird die Ballon-Perspektive direkt vom UMN Mapserver erzeugt (außer auf der obersten Ebene, s. Kap. 0). Die für ganz Schleswig-Holstein flächendeckend vorliegenden Orthophotos werden eingesetzt, um die beiden untersten Ebenen („Höhenstufen“) anzuzeigen (s. Abb. 2). Die drei übrigen Höhenstufen werden auf der Basis von ATKIS (Amtliches Topographisch-Kartographisches Informationssystem) und anderen Karten-Layern zusammengestellt.

5. Zusammenfassung und Ausblick

Der Minister für Umwelt, Naturschutz und Landwirtschaft des Landes Schleswig-Holstein stellte im November 2004 gemeinsam mit dem Projektpartner Landesnaturschutzverband Schleswig-Holstein den Naturpilot der Öffentlichkeit vor. Die Web-Adresse lautet: **http://www.naturpilot-sh.de/ .**

Bisher erhaltene Rückmeldungen von Interessierten waren in der Regel positiv, einige Verbesserungsvorschläge wurden eingebracht. Von zahlreichen Web-Seiten wurden Verweise auf den Naturpilot eingerichtet. Dafür soll zukünftig eine Option angeboten werden, mit der der Naturpilot z.B. direkt mit der Karte oder dem Luftbild

eines Ortes aufgerufen werden kann. Weitergehende Planungen sehen auch die Eingabe eigener Ziele vor, die per URL aufgerufen werden können und dann im Naturpilot sogar mit einer individuellen Markierung angezeigt werden können (zum Beispiel für kleinere Natur-Attraktionen, Ferienwohnungen, Rathäuser, Hofläden, ...).

Der parametrisierbare Einstieg in den Naturpilot soll zudem für eine engere Verknüpfung mit anderen Systemen wie etwa dem Internet-Angebot „Schleswig-Holstein Topographie“ genutzt werden. Bereits jetzt kann im Freien Fliegen mit dem interaktiven Abruf von Ortsinformationen auch direkt zu den entsprechenden Topographie-Seiten verzweigt werden, die ausführliche Informationen zu jeder Gemeinde Schleswig-Holsteins anbieten.

Der Bildaufbau und das Laden der Objekte erfolgt aufgrund der geringen Dateigrößen (Kartenbilder: etwa 30 – 80 KB) so performant, dass auch die Nutzung per Modem möglich ist. Das relativ einfache Datenmodell stellt sicher, dass die Serverlast ebenfalls niedrig ist. Die Bedienungsführung des Systems wurde bereits optimiert und wird auch weiterhin noch stärker den Bedürfnissen angepasst. Die Integration von zusätzlichen Bedienungshilfen ist geplant.

Hinzu kommen Erweiterungen wie etwa die Integration weiterer Sachdaten (z.B. aus den Bereichen Bildung oder Tourismus), eine Web-Schnittstelle zur Administration des Systems sowie die Verwaltung und Ansteuerung von Sehenswürdigkeiten („Points of interest“). Die Stammdaten der Naturhighlights werden zur Zeit in die Naturpilot-Datenbank überführt (vormals: HTML), die zudem mit der Datenbank der Umweltbildungseinrichtungen durch die Umweltakademie Schleswig-Holstein konsolidiert wird, so dass die zukünftige Pflege über eine Import-Export-Schnittstelle erfolgen kann.

Literatur

Bornhöft, D. et al. (2000): InfoNet-Umwelt Schleswig-Holstein – Erfahrungen mit Aufbau und Betrieb eines kooperativ aufgebauten Umweltinformationssystems - In: Cremers; A. & Greve; K. (eds.): Umweltinformation für Planung, Politik und Öffentlichkeit, 14. Internationales Symposium Informatik für den Umweltschutz, Bonn 2000, Metropolis-Verlag, pp. 306-316.

Görtzen, D.; Schneberger, S. & Rammert, U. (2004): Schleswig-Holsteins Environmental Atlas for the Public and for Special Interest Groups. In: Proceedings of the 18th Con-ference Informatics for Environmental Protection. October 21-23, 2004. CERN, Geneva, pp. 634-640.

Rammert, U. & Hosenfeld, F. (2003): Dynamic and Interactive Presentation of Environmental Information. In: Gnauck, A. & Heinrich, R. (eds.): The Information Society and Enlargement of the European Union, 17th International Symposium Informatics for Environmental Protection, Cottbus 2003, Vienna, pp. 517-524.

UMN Mapserver (2005): Homepage des Minnesota Mapservers: http://mapserver.gis.umn.edu/

Mobilisierung von primären Biodiversitätsdaten: Das BioCASe Protokoll und seine Anwendung in internationalen Netzwerken

Anton Güntsch, Markus Döring & Walter Berendsohn
Botanischer Garten und Botanisches Museum Berlin-Dahlem
Abt. f. Biodiversitätsinformatik und Labors, a.guentsch@bgbm.org

Abstract

Biodiversitätsforschung fußt in vielen Bereichen auf primären Sammlungs- und Artinformationen, die weltweit verteilt in taxonomischen Institutionen vorgehalten werden, für die es aber bislang keinen standardisierten Zugang gab. Das BioCASe Protokoll spezifiziert eine XML Anfragesprache für solche heterogenen Datenquellen, mit der bislang unter anderem ca. 5 Millionen Sammlungsobjekte im Web verfügbar gemacht werden konnten. Dabei ist die entwickelte Technologie so generisch, dass sie für die verschiedenartigsten Problemstellungen auch außerhalb der Biodiversitätsforschung eingesetzt werden kann.

1. Hintergrund

Biologische Primärdaten wie zum Beispiel Beschreibungen von Sammlungsobjekten, Beobachtungen lebender Organismen aber auch Namenslisten und Artbeschreibungen werden in immer größerem Umfang elektronisch in Datenbanken erfasst und bilden eine wertvolle Informationsquelle für die biologische Systematik, Biotechnologie, Umweltstudien aber auch die Lehre in Schulen und Universitäten ([Felinks et al., 2000], [Chapman, 2005]). Ein Merkmal dieser Datenbestände ist, dass sie zwar zumeist lokal erfasst und gepflegt werden, ihr Nutzen sich aber wesentlich erhöht, wenn sie gemeinsam abfragbar gemacht werden. Soll zum Beispiel die historische geografische Verbreitung für einen Organismus auf der Grundlage von Sammlungsdaten, die in Naturhistorischen Museen vorliegen, berechnet werden, so wird diese Berechnung aussagekräftiger und

vertrauenswürdiger, je mehr Primärdaten aus Museen hierfür abgefragt werden können.

Mit der zunehmenden Verfügbarkeit von Datenbanken und Internettechnologie sind in den letzten zwei Jahrzehnten vermehrt Primärdaten elektronisch erfasst und im Internet zugänglich gemacht worden. Zusätzlich wurde daran gearbeitet, Netzwerke aufzubauen, um die Abfrage verteilter Datenquellen über gemeinsame Portale zu ermöglichen ([Berendsohn, 2003]). Beispiele hierfür sind das auf dem ursprünglich im Bibliotheksbereich verbreitete z39.50[25] aufbauende Species Analyst Netzwerk mit Beobachtungen und Sammlungsdaten mit dem Focus Nordamerika ([Vieglas, 1999]) und das European Natural History Specimen Network ENHSIN (Sammlungsdaten, Europa), das auf der Grundlage von http und XML realisiert wurde ([Güntsch, 2003]). Gemeinsam war diesen Netzwerken, dass sie thematisch oder geografisch spezialisiert waren, da keine Technologie verfügbar war, mit der sich die heterogenen Primärdatenquellen über Fach- und Ländergrenzen hinweg verknüpfen ließen.

Die Entwicklung und Implementierung eines solchen umfassenden Netzwerks für Sammlungsdaten ist eines der Ziele des im 6. Rahmenprogramms der EU angesiedelten BioCASE Projekts[26] ([Berendsohn, 2002]). Das BioCASE Netzwerk besteht gegenwärtig aus ca.150 verteilten Sammlungsdatenbanken, die mit einer gemeinsamen XML Anfragsprache, dem so genannten BioCASE Protokoll, abfragbar sind. Die im Netzwerk gültigen Begriffe werden mit einem XML Datenschema (ABCD) spezifiziert. Neben der originären Nutzung des BioCASe Protokolls und des ABCD Schemas im europäischen BioCASE Netzwerk, wird die Technologie nun auch von der Global Biodiversity Information Facility GBIF[27], des weltweiten Netzwerks für Biodiversitätsdaten, eingesetzt.

[25] http://www.loc.gov/z3950/agency/
[26] http://www.biocase.org
[27] http://www.gbif.org

2. BioCASe Protokoll

Das BioCASe Protokoll[28] ist ein XML-Schema, mit dem eine einfache generische Anfragesprache für heterogene Datenquellen spezifiziert wird. Es werden lediglich drei grundlegende Anfragetypen definiert:

- **Capabilities** ergibt die Liste der Datenschemas, für welche die Datenquelle konfiguriert wurde, und die jeweilige Menge von Datenelementen, die geliefert werden können.
- **Scan** entspricht einer „SELECT DISTINCT“ SQL Anfrage und ergibt für ein gegebenes Datenelement die Menge unterschiedlicher Werte. Die Scan-Anfrage wird zum Beispiel im BioCASE Netzwerk benutzt, um eine zentrale Indexdatenbank aufzubauen und zu aktualisieren.
- **Search** entspricht einer abstrakten SQL-Abfrage und ergibt ein XML Ergebnis-Dokument für das übergebene Suchmuster.

Eine einfache BioCASE-konforme XML Anfrage (**<request>**) könnte zum Beispiel so aussehen:

```
<request>
 <header>
  [...]
  <type>search</type>
 </header>
 <search>
  <requestFormat>http://www.tdwg.org/schemas/abcd/1.2</requestFormat>
  <responseFormat start="0" limit="50">http://www.tdwg.org/schemas/abcd/1.2</responseFormat>
  <filter>
   <and>
    <like path="/DataSets/DataSet/[...]/TaxonIdentified/NameAuthorYearString">Abies*</like>
    <or>
     <like path="/DataSets/[...]/TaxonIdentified/HigherTaxa/HigherTaxon">Pinace*</like>
     <and>
      <like path="/DataSets/DataSet/[...]/GatheringSite/Country/CountryName">*Russia*</like>
```

[28] http://www.biocase.org/dev/protocol/index.shtml#methods

```
            <greaterThan path="/DataSets/DataSet/[...]/ISODateTimeBegin">2002-04</greaterThan>
          </and>
        </or>
      </and>
    </filter>
  <count>false</count>
  </search>
</request>
```

Sie besteht also aus einem **<header>** Element, das unter anderem den Typ der Anfrage festlegt. Die eigentliche Suchanfrage wird dann mit dem folgenden Element (hier **<search>**) spezifiziert. Neben dem Suchkriterium **(<filter>**), das der WHERE-Klausel einer SQL-Anfrage entspricht, wird hier auch angegeben, in welchem XML-Format das Antwortdokument erwartet wird (**<responseFormat>**). Damit kann dieselbe Datenquelle für verschiedene Rückgabeformate konfiguriert werden und Anfragen entsprechend beantworten.

Das Protokoll ist dabei völlig entkoppelt von den damit abgefragten Datenelementen, die in einem getrennten Schema spezifiziert werden müssen. Zum einen wurde so erreicht, dass die im BioCASE Netzwerk verwendete Software bei neuen Versionen eines Datenschemas nur wenig angepasst werden muss, zum anderen ist das BioCASE Protokoll und die zugehörige Software so aber auch für den Aufbau verschiedenartigster Netzwerke auch außerhalb des Themengebietes Biodiversitätsforschung geeignet.

Eine Schwäche des BioCASe Protokolls ist die fehlende Möglichkeit der Projektion, also des Selektierens von Elementen für das Antwortdokument. Es werden demnach immer vollständige Dokumente, die sämtliche vom Datenanbieter konfigurierten Elemente enthalten, zurückgeliefert. Die gegenwärtig in der Implementierung befindliche nächste Version des Protokolls TAPIR[29] (TDWG Access Protocol for Information Retrieval) behebt diesen Mangel, indem Anfragen eine Liste von Elementen enthalten können, die im Antwortdokument enthalten sein sollen.

[29] http://ww3.bgbm.org/tapir

3. ABCD Daten-Schema

Das ABCD[30] Datenschema (Access to Biological Collection Data) ist eine umfassende Definition von ca. 700 Datenelementen, die für den Bereich von biologischen Sammlungen und Beobachtungsdatenbanken relevant sind. Neben Daten zu den Sammlungsobjekten selbst werden auch Informationen zu den Sammlungshaltern repräsentiert. ABCD wird als gemeinsame Initiative des „Committee on Data for Science and Technology – CODATA“[31] und TDWG, der „Taxonomic Databasing Working Group“[32] in einem offenen Begutachtungsprozess entwickelt.

Das Schema verfügt für jedes Element über eine strukturierte maschinenlesbare Annotation, die der Erklärung der Elemente selbst dient und Entsprechungen in anderen existierenden Standards dokumentiert. Zusätzlich können Annotationen aber auch zur automatischen Beschriftung von Elementen in Portalen verwendet werden. Das Element *FullScientificNameString*, das den vollen wissenschaftlichen Namen eines Sammlungsobjekts als Freitext repräsentiert, wird zum Beispiel so im Schema dokumentiert:

```
<xs:element name="FullScientificNameString" type="String">
  <xs:annotation>
    <xs:documentation>
      Concatenated scientific name, preferrably formed in accordance with a Code of Nomenclature,
      i. e. a monomial, bionomial, or trinomial plus author(s) or author team(s) and - where relevant -
      year, or the name of a cultivar or cultivar group, as fully as possible.
    </xs:documentation>
    <xs:appinfo>
      <sea:FullName>Full scientific name</sea:FullName>
      <sea:Audience>BioCASE</sea:Audience>
      <sea:Audience>CODATA TDWG</sea:Audience>
      <sea:Reviewer/>
      <sea:ExistingStandard>Darwin Core 2: Scientific Name.</sea:ExistingStandard>
      <sea:Content/>
```

[30] http://www.bgbm.org/tdwg/codata/Schema/default.htm
[31] http://www.bgbm.org/tdwg/codata/Schema/default.htm
[32] http://www.tdwg.org

```
            <sea:Example>Acipenser gueldenstadti Linnaeus 1758</sea:Example>
            <sea:Comment/>
            <sea:Rule/>
            <sea:EditorialNote/>
          </xs:appinfo>
        </xs:annotation>
      </xs:element>
```

ABCD ist hoch strukturiert, bietet aber häufig für den selben Datenbereich atomisierte und nicht atomisierte Freitext-Elemente an. Auf diese Weise wird Datenanbietern, deren Datenbanken gegenwärtig noch wenig atomisiert sind, der Anschluss an das BioCASE Netzwerk ermöglicht, ohne dass die Daten zuvor konvertiert werden müssen. Diese Erleichterung für Datenanbieter hat zum schnellen Wachstum des BioCASE-Netzwerkes in den letzten zwei Jahren massgeblich beigetragen, kann aber eine Hürde für Anwendungs- und Portalprogrammierer darstellen, deren Software mit dem Umstand, dass ähnliche Informationsinhalte in verschiedenen Elementen übertragen werden können, umgehen können muss. In der Praxis hat sich allerdings gezeigt, dass Kernelemente, die für gemeinsame Abfragbarkeit der Daten wichtig sind (z.B. wissenschaftliche Namen in biologischen Sammlungsdatenbanken oder Land der Aufsammlung), von nahezu allen Datenanbietern geliefert werden können.

ABCD ist seit März 2005 in der Version 2.0 als „final draft“ verfügbar, über die Verabschiedung als TDWG Standard wird im September entschieden. Implementierungen basieren gegenwärtig auf der Version 1.2.

4. Verfügbare Software

Aufbauend auf die beschriebenen Spezifikationen implementiert das BioCASE Projekt Software Module, mit denen Datenanbieter und Nutzer sich an das Netzwerk anschließen können ([Döring & Güntsch, 2003]). Ein zentrales Anliegen ist dabei, dass das BioCASE Protokoll durch leicht zu benutzende Datenbank- und Programmierer-Schnittstellen vollständig verdeckt wird.

Der *PyWrapper*[33] ist eine in *Python*[34] implementierte CGI-Schnittstelle, die BioCASE XML-Anfragen in die native Anfragesprache der angeschlossenen Datenbank übersetzt und deren Resultate wieder als ein BioCASE-konformes Antwortdokument zurücksendet (s. Abb.). Spezifika der angeschlossenen Datenbank-Management-Systeme wurden in verschiedenen Datenbank Modulen implementiert und können jederzeit durch weitere Module ergänzt werden. Gegenwärtig sind ausser einem Standard-Modul für ODBC-Verbindungen Module für Microsoft SQL-Server, PostgreSQL, MySQL, Oracle 8 und 9, Firebird/Interbase und Sybase verfügbar[35].

Die in der Datenbank vorhandenen Attribute werden in einer Konfigurationsdatei (CMF = Concept Mapping File) den zugehörigen Konzepten des verwendeten Datenschemas zugeordnet. In einer weiteren Konfigurationsdatei (PSF = Provider Setup File) werden Parameter für die Datenbankverbindung angegeben und die verwendeten Tabellen und deren Beziehungen deklariert.

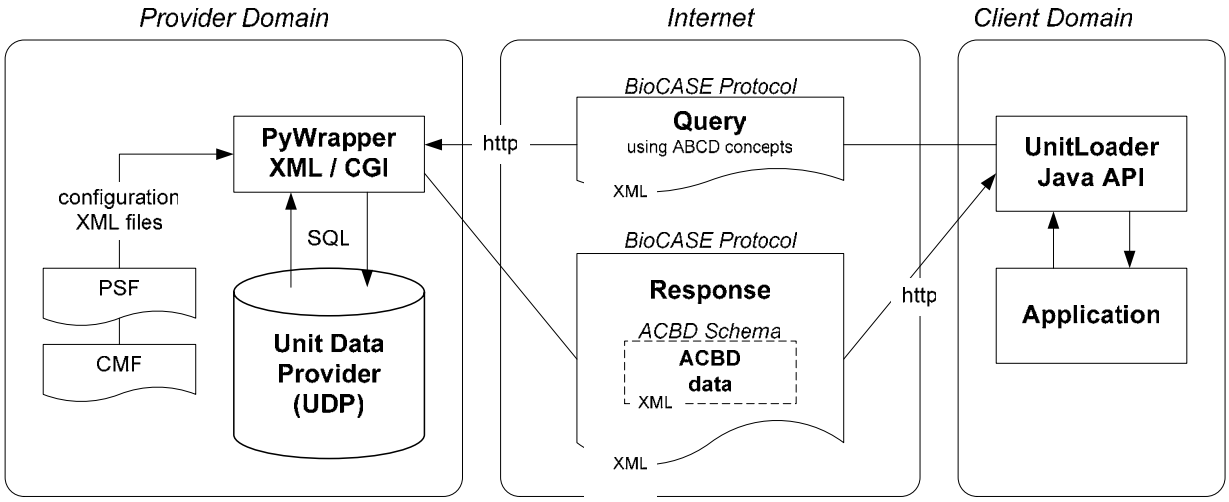


Abb.1: Provider-Client Kommunikation im BioCASE Netzwerk, Copyright © 2005 BGBM.

Um die Konfigurationsarbeit von Datenanbietern weiter zu vereinfachen, wurde ein *Provider Configuration Tool*[36] entwickelt, mit dessen Hilfe Konfigurationsdateien bequem über eine grafische Benutzeroberfläche editiert werden können. Das vollständig web-basierte Werkzeug erlaubt es, Datenbank-Verbindung zu konfigurieren, dem PyWrapper die relevanten Tabellen und Beziehungen der

33 http://www.biocase.org/dev/wrapper/index.asp
34 http://www.python.org/
35 http://ww3.bgbm.org/biocase/utilities/testlibs.cgi
36 http://www.biocase.org/dev/configtool/index.shtml

angeschlossenen Datenbank bekanntzugeben, sowie die zu veröffentlichenden Datenbankattribute den Elementen des verwendeten Datenschemas zuzuordnen. Insbesondere der letzte Schritt kann aufwändig werden, wenn grosse Datenschemas verwendet werden, in denen die für die konkrete Anwendung relevanten Elemente oft nur schwer zu finden sind. Daher wurde die Konfigurationssoftware mit einem Mechanismus versehen, der die Suche nach Elementen unterstützt und dem Anwender gefundene Elemente nach ihrer angenommenen Relevanz sortiert anzeigt.

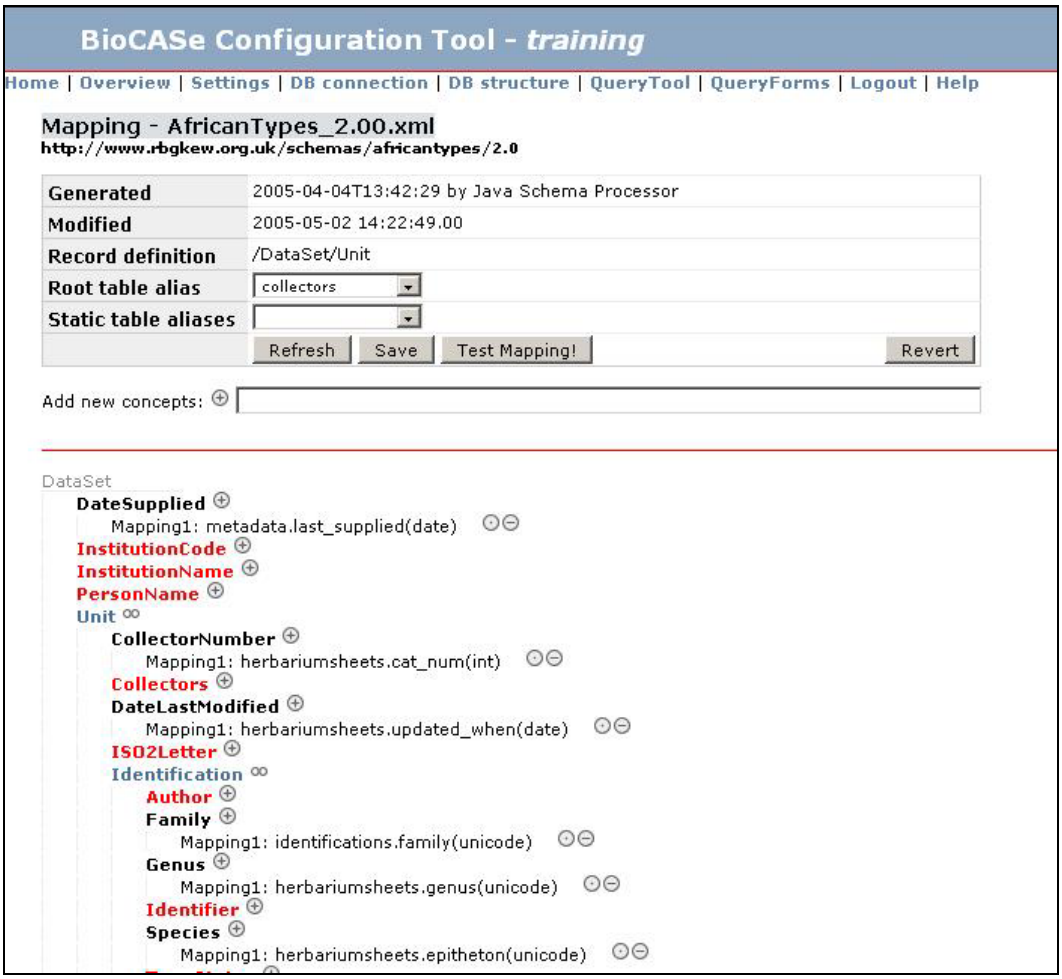


Abb.2: Verknüpfung von Datenbank-Attributen mit Schema-Elementen mit dem BioCASE Configuration Tool

Der *UnitLoader*[37] ist eine Java API, mit der Anwendungsprogrammierer und Portalentwickler syntaktisch korrekte BioCASE Anfragen generieren und an verteilte Datenanbieter parallel versenden können. Die Software wird gegenwärtig von

[37] http://www.biocase.org/dev/unitloader/index.asp

mehreren europäischen und deutschen Internet Portalen für Sammlungsdaten verwendet.

Die Registrierung geeigneter Datenquellen im Web muss dabei von der jeweiligen Anwendung selbst vorgenommen werden. So greifen zum Beispiel das europäische BioCASE Projekt und der deutsche Botanik-Knoten für GBIF jeweils auf eine eigene Registrierungsdatenbank zu, obwohl Datenanbieter zum Teil über beide Netzwerke zugänglich gemacht wurden.

Neben diesen zentralen Software-Komponenten wurde eine Reihe weiterer Software Werkzeuge entwickelt, die insbesondere für die Entwicklung des ABCD Schemas benötigt wurden. Beispiele sind der *SchemaViewer*[38], mit dem sich komplexe XML Schemas in einer für Revisoren des Schemas verständlichen Form darstellen lassen, und der *SchemaProcessor*[39], mit dem sich Datenschemas, die in einem Netzwerk verwendet werden sollen, in eine initiale Konfigurationsdatei für den *PyWrapper* umwandeln lassen. Alle Software-Komponenten, die im Rahmen des BioCASE-Projekts entwickelt werden, sind frei verfügbar, und der jeweilige Quellcode ist öffentlich zugänglich.

38 http://www.bgbm.org/scripts/ASP/TDWG/Frame.asp?config=6
39 http://www.biocase.org/dev/utilities/index.shtml

Literaturverzeichnis

[Berendsohn, 2003]

Berendsohn, W.: ENHSIN in the context of the evolving global biological collection information system. In: Scoble, M. (ed.): ENHSIN, the European Natural History Specimen Information Network. The Natural History Museum, London, 2003.

[Chapman, 2005]

Chapmann, Arthur D. (in press) : Uses of Primary Species-occurrence data. The Global Biodiversity Information Facility, Copenhagen, 2005.

[Döring & Güntsch, 2003]

Döring, M.; Güntsch, A.: Technical introduction to the BioCASE software modules. 19th annual meeting of the Taxonomic Databases Working Group (TDWG 2003), Abstract, Oeiras, Lisbon, 2003.

[Felinks et al., 2000]

Felinks, B.; Hahn, A.; Olsvig-Whittaker, L.; Los, W.: Users and uses of biological collections. In: Berendsohn, W. (ed.): Resource Identification for a Biological Collection Information Service in Europe. Botanic Garden and Botanical Museum Berlin-Dahlem, Berlin, 2000.[Güntsch, 2003]

Güntsch, A.: The ENHSIN Pilot Network. In: Scoble, M. (ed): ENHSIN - The European Natural History Specimen Information Network. The Natural History Museum, London, 2003.

[Vieglas, 1999]

Vieglas, D.: Integrating disparate biodiversity resources using the information retrieval standard Z39.50. Annual meeting of the Taxonomic Databasing Working Group (TDWG 1999), Abstract, Cambridge, USA, 1999.

Europäischer Abfallwirtschaftsassistent

Einführung eines webbasierten Wissensmanagement –systems

Ulrich Eimer, Stadt Hagen, Germany, ulrich.eimer@stadt-hagen.de
Alexandra Thannhäuser, i-world GmbH Hagen, thannhaeuser@i-world.de

Abstract/Einleitung

Die Nutzung von Online-Plattformen für den Austausch von Expertenwissen und den Aufbau von Netzwerken (*expert networks*) in der öffentlichen Verwaltung sowie in mittelständischen Betrieben insbesondere auf lokaler und regionaler Ebene, ist gegenwärtig wenig verbreitet. Das Projekt **" European Waste Sector Assistant**" – (EUWAS) ist *das* aktuelle europäische Pilotvorhaben mit dem Ziel, eine webbasierte Plattform für die Bedürfnisse von Spezialisten auf dem Gebiet der Abfallwirtschaft zu implementieren.

EUWAS zielt darauf ab, durch die Strukturierung von Informationen und vorhandenen Datenquellen den Anforderungen, die ein funktionsfähiges webbasiertes Expertensystem stellt, gerecht zu werden. Die Plattform will Mitarbeiter aus dem Bereich der Abfallwirtschaft durch die Etablierung sogenannter „Wissensgemeinschaften" innerhalb ihrer Arbeitswelt unterstützen.

Schlüsselwörter: *Wissens-Management, europäischer Abfallsektor, webbasierte Plattform, Europäisches gefördertes e-Content-Projekt, mehrsprachig, Werkzeuge, Schritt- für Schritt Anleitungen, öffentliche Datenbanken, inhaltliche Verknüpfungen, Abfallwirtschafts-Netzwerk, Informationssystem*

1. Projektgrundlagen

1.1. Projektidee, Visionen und Mehrwerte

EUWAS ist ein Projekt, das im Rahmen des EU-Programms e-Content im Sommer 2004 ausgewählt wurde. Die Laufzeit des Projektes beträgt 24 Monate von Januar 2005 bis Dezember 2006 und wird mit Partnern aus Polen, dem Baltikum und Deutschland durchgeführt (für weitere Projektinformation siehe www.euwas.org).

EUWAS zielt darauf ab, die notwendige Strukturierung, Verarbeitung und Darstellung von Daten und Informationen aus dem Abfallwirtschaftssektor im europäischen Rahmen zu fördern. Projektziel ist es dabei insbesondere, zu einer effizienten Nutzung existierender Daten und Informationen aus dem Bereich Abfallwirtschaft beizutragen.

Bis zum Ende der Projektlaufzeit wird ein mehrsprachiges Internetportal (englisch, deutsch, polnisch, litauisch, lettisch, estnisch), bestehend aus 5 Werkzeugen ("Tools") mit Spezialdienstleistungen, Hilfsmodulen und Methoden entstanden sein. Vor einer Inbetriebnahme des Portals werden evaluative Studien den praktischen Nutzen für die Abfallwirtschaft sowohl von repräsentativen Gemeinden aus Deutschland, Polen und den baltischen Staaten als auch von Unternehmen aus dem Bereich der Abfallwirtschaft getestet.

Den späteren Nutzer werden Daten, Informationen und Lehrwerkzeuge zur Verfügung stehen, welche auf der Grundlage europäischer Fortbildungsstandards erarbeitet wurden. Diese Tools sollen durch eine mehrwertbildende Strukturierung, Verarbeitung und Darstellung von Daten eine Unterstützung bei fachlich relevanten Entscheidungen bieten.

1.2. Projektpartner

Balti Keskkonnafoorum MTÜ | **Estonia** | **Latvia** | **Lithaunia** | http://www.bef.lv

Bildungszentrum für die Entsorgungs- und Wasserwirtschaft | **Germany** | http://www.bew.de

Fraunhofer Institut für Umwelt-, Sicherheits- und Energietechnik UMSICHT | **Germany** | http://www.umsicht.fraunhofer.de

i-world Ltd. | **Germany** | http://www.i-world.de

Instytut Mechanizacji Budownictwa i Górnictwa Skalnego | **Poland** | http://www.imbigs.org.pl

Kauno Technologijos Universitetas - Aplinkos Inžinerijos institutas | **Lithaunia** | http://www.ktu.lt

Umweltamt Stadt Hagen | **Germany** |http://www.hagen.de

1.3. Inhalte und Ziele

In vielen Ländern der Europäischen Union - einschließlich der neuen Beitrittsländer in Mittel- und Osteuropa - existiert umfangreiches Datenmaterial auf hohem inhaltlichem Niveau. Bedingt durch die fehlende Dokumentation im Hinblick auf die Existenz von Metadaten und bedingt durch den eingeschränkten Zugang zu Datenbasen aufgrund technischer Schwierigkeiten ist ein angemessener und sinnvoller Einsatz in Behörden und Unternehmen zurzeit nicht möglich. Dieser Umstand führt zu enormen kostenintensiven Reibungsverlusten im Umgang mit solchen Daten der europäischen Abfallwirtschaft auf allen relevanten Ebenen der Wirtschaft und der Verwaltung. Die große Anzahl heterogener, vielschichtiger und unvernetzter Datenbestände, der Mangel an transparenten Datenstrukturen komplizieren ihren effizienten Gebrauch. Aufgrund dieser Umstände ist es dringend geboten, einen vernetzten und einfachen Zugang zu den relevanten Daten/-banken zu schaffen und die große Anzahl an Datenmaterial auch in ökonomischer Hinsicht gewinnbringend zu strukturieren. Ziel des Projektes EUWAS soll in diesem Zusammenhang insbesondere sein, Daten in Hinblick auf den effektiven und

strukturierten Gebrauch im Rahmen praktischer Anwendungen und Arbeitsfelder in der Abfallwirtschaft auszuwerten und zur Verfügung zu stellen.

1.4. Zielgruppen

Erfolgreiches Abfallwirtschaftmanagement auf den unterschiedlichen administrativen Ebenen in der Europäischen Union verlangt heute zweierlei: gebündeltes, fundiertes Expertenwissen sowie zuverlässige Datenquellen und Informationen. Der Zugang zu den Daten sowie die Art und Weise ihrer Aufbereitung spielen in den betroffenen Organisationen eine Schlüsselrolle in allen Fragen der effektiven Nutzung solcher Daten. Dies wiederum beeinflusst in hohem Maße die kompetenten Entscheidungsprozesse in Planung und Überwachung.

Die zur Zeit im Aufbau befindliche Plattform EUWAS soll in diesem Zusammenhang elektronische Informationen zum Abfallwirtschaftsbereich liefern und wird sich dementsprechend an den Bedürfnissen von städtischen Behörden, Firmen oder anderen verwandten Organisationen orientieren. Marktnähe soll und muss hier das oberste Gebot sein!

EUWAS setzt sich zum Ziel, das Fachpersonal von Umweltbehörden und Unternehmen in den beteiligten Europäischen Ländern mit Daten, die eine hohe Relevanz für die tägliche Arbeit haben, zu unterstützen. Darüber hinaus soll die beiderseitige Nutzung von EUWAS öffentliche Einrichtungen und Unternehmen der Privatwirtschaft mit dem Ziel einer gegenseitigen Win-Win-Beziehung näher in Kontakt bringen.

Ein großer Vorteil von EUWAS besteht darin, unabhängige Informationen zu verschiedenen Aspekten der Abfallwirtschaft anbieten zu können.

Das Portal richtet sich hauptsächlich an:

- Öffentliche Körperschaften: Umwelteinrichtungen und -behörden in Stadtverwaltungen, regionale und nationale Behörden, die mit Abfallwirtschaftsaufgaben betraut sind, Bundesagenturen und Institute
- Firmen aus der Abfallwirtschaft: Entsorgungsfirmen (Müllbeseitigungsgesellschaften), Firmen für logistische Planung, Consultants

- Weitere Rechtskörperschaften: Universitäten, Handelskammern, regionale und nationale Abfallwirtschaftsverbände, andere private und öffentliche Interessenten

Das zukünftige Ziel des EUWAS-Projektes ist es, das internetgestützte EUWAS Portal als ein innovationsfreudiges Dienstleistungsunternehmen für private und öffentliche Kunden des Abfallwirtschaftssektors zu führen.

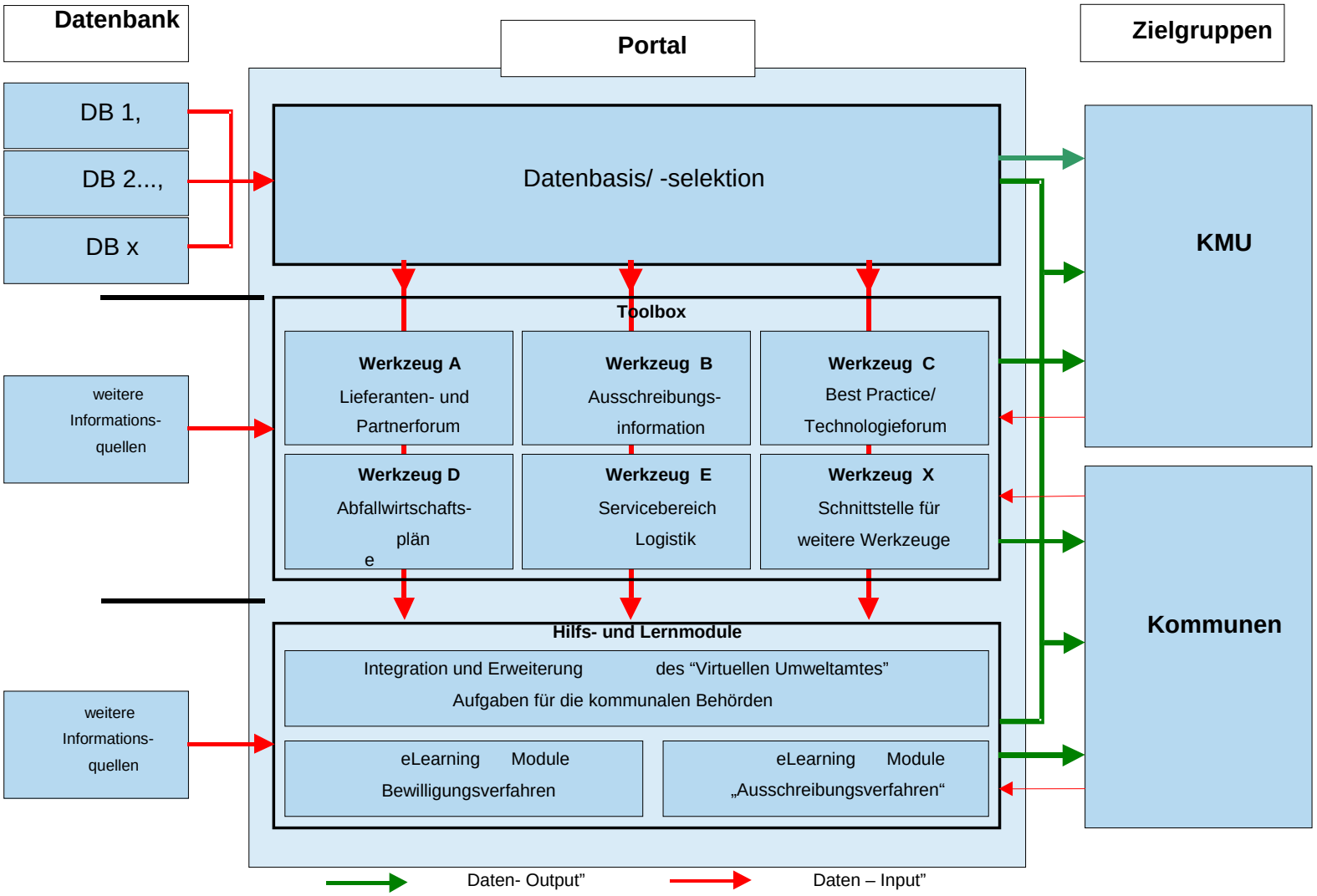
1.5. Struktur

Das Portal EUWAS stellt einen Zugangsweg zu Fachinformationen und Daten für Experten der Abfallwirtschaft dar. Im Mittelpunkt der Plattform stehen so genannte Tools, die verschiedene Funktionalitäten ziel- und passgenau anbieten. Darüber hinaus kann der Nutzer Gebrauch von Hilfsmodulen machen, welche a) das Handling der Plattform erläutern und b) strukturierte Hilfen zu Verfahren in der Abfallwirtschaft und auch beispielsweise für den Bereich des Ausschreibungswesens anbieten.

Das komplexe System umfasst eine Vielzahl von Elementen, dessen Bestandteile nach nationalen Interessenslagen, Sprachen, Inhalten, Funktionen und Benutzer (Personalisierung) strukturiert werden.

1.5.1. Gesamtplattform

Das folgende Diagramm zeigt den Inhalt und die Struktur des Portals:



- Tool A "Virtual supplier and partner forum"

Dieses Forum dient insbesondere als Informations- und Dienstleistungsplattform für Akteure der Abfallwirtschaft. Es ermöglicht den Nutzern, ihr Unternehmen und die entsprechenden Produktportfolios zu präsentieren, andere Angebote zu suchen und bietet darüber hinaus die Gelegenheit, Partner für gemeinsame Projekte und Investitionen zu finden.

- Tool B “Tender Information”

Tool B beinhaltet die Bereitstellung länderübergreifender öffentlicher Ausschreibungen und bietet Unternehmen die Möglichkeit, öffentliche und private Bewerbungen zu erstellen oder selbst Ausschreibungen anzufertigen. Durch das Angebot strukturierter Leitfäden wird der Ablauf von Bewerbungs- und Ausschreibungsverfahren insgesamt unterstützt.

- Tool C “Best Practice / Best Available Technology”

Das *Best Practice Forum* beinhaltet eine umfangreiche Auswahl an verfügbaren Technologien und Bester Praxis, die dem Nutzer helfen kann, (Investitions-) Entscheidungen auf regionaler oder lokaler Ebene zu treffen. Dieses Tool bietet Hilfestellung bei der Auswertung der verfügbaren Technologien an und vermittelt Informationen im Hinblick auf inhaltliche und politische Prozesse, die bei der Entwicklung und Verbreitung der besten verfügbaren Technik zum Tragen kommen.

- Tool D “Waste Management Services”

Mit Tool D wird ein Kerngeschäft der Europäischen Abfallwirtschaft funktional unterstützt: unter Zuhilfenahme mannigfaltiger Daten, Datenbanken und anderer Informationsquellen wird hier ein umfangreiches Portfolio an Dienstleistungen angeboten. Neben Leitfäden und praktischen Hilfestellung zur Aufstellung von Abfallwirtschaftsplänen sowie der Dokumentation von Genehmigungsverfahren und Aufgaben der Abfallwirtschaft in öffentlichen Einrichtungen und der Privatwirtschaft werden auch innovative Projekte der Abfallwirtschaft sowie Kernzahlen und -fakten zu nationalen und europäischen Abfallthemen dargestellt.

- Tool E “Logistic Services”

Im Rahmen mehrerer webbasierter Logistikbausteine sollen in Tool E praxisnahe Lösungen vorgestellt werden, die im Rahmen logistischer Fragestellungen in der

Abfallwirtschaft Einsatz finden können. Neben Routenplanungen stehen unter anderem Fragen der Distribution von Wertstoffen im Mittelpunkt der Betrachtungen.

- Weitere Hilfs- und Lernmodule

Ergänzend zur grundlegenden Struktur bietet die Plattform EUWAS so genannte "Hilfsmodule" an. Diese umfassen so genannte „Step-by-Step-Guides“, welche die oben genannten Tools unterstützen sowie zusätzliche Information in Bezug auf Bewilligungsverfahren und Überwachungsaufgaben bereitstellen. Darüber hinaus werden Projektdatenbanken, Gesetzessammlungen und eine Datensammlung zu Finanzierungsmöglichkeiten auf nationaler und europäischer Ebene angeboten.

2. Technik und Umsetzung

2.1. Einführung

Die Realisierung des Projektes erfordert im Hinblick auf Pflege der Inhalte, Verfügbarkeit der Daten und Inhalte, sowie Aktualisierung und Erweiterung des Projektes eine flexible technische Umsetzung, die den Anforderungen in vollem Umfang gewachsen ist.

2.2. Rahmenbedingungen

Folgende Anforderungen stehen bei der technischen Umsetzung und der Gesamtkonzeption im Vordergrund:

- Mehrsprachigkeit in der Benutzerführung
  Deutsch – Polnisch – Lettisch – Litauisch – Estnisch - Englisch
- Vermeidung von Datenredundanzen
  Eventuell sind z.B. englische Inhalte länderübergreifend von Interesse
- Erweiterung auf inhaltlicher Basis
  Gewährleistung der Aktualität der Daten und Informationen auf verschiedenen Gebieten

- Inhaltliche Pflege

  Die Inhalte müssen in der Form im System abgelegt werden, dass eine benutzerfreundliche inhaltliche Pflege bestehender Seiten erfolgen kann. Pflegbarkeit der Inhalte ohne tiefe technische Programmierkenntnisse muss in den jeweiligen Sprachen gegeben sein.

- Flexibles Design und Layout

  Design und Layout müssen so gewählt werden, dass eine Erweiterung der Inhalte ohne umfangreiche Designänderungen vorgenommen werden kann.

- Intuitive Benutzerführung

  Bei der Zielgruppe handelt es sich um Anwender, die im Großen und Ganzen inhaltlich mit der Thematik aber nicht zwangsläufig mit dem Umgang solcher Systeme und Anwendungen vertraut sind. Aufgrund der Komplexität der Inhalte ist daher auf eine benutzerfreundliche und intuitive Benutzerführung zu achten.

- Technische Erweiterbarkeit

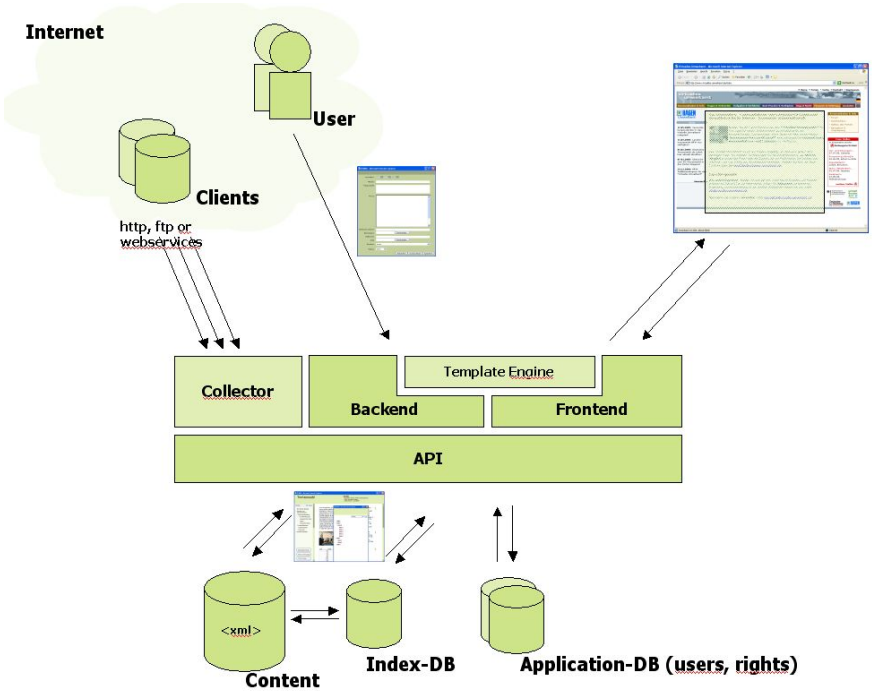
  Ausbaufähigkeit um weitere modulare Teilbereiche

2.3. Technische Basisdaten

Für die Umsetzung des Projektes wird die Software i-logic® verwendet. i-logic® ist eine Content Management Software und ein Applikationsframework auf php-Basis.

Die Basissoftware i-logic® ermöglicht grundsätzlich die Installation sowohl auf einem Windows-Server wie auch auf einem Linux-Server.

Aufgrund möglicher Erweiterungen und voraussichtlich steigendender Anforderungen an das System wird die Anwendung für den Betrieb auf einem Linux-Server entwickelt.



Funktionsweise i-logic

Zum Einsatz kommt der Apache Webserver ab Version 1.3.x. Dieser ist für alle gängigen Plattformen erhältlich. Für den reibungslosen Betrieb von i-logic® ist die Scriptsprache PHP in der Version ab 4.x notwendig.

Als Datenbank wird die MySQL-DB ab Version 4.x verwendet. Diese ist für alle Plattformen frei erhältlich. i-logic® unterstützt neben MySQL auch Oracle ab Version 8.1.5 und PostgreSQL ab Version 7.1.2.

Der Client muss über einen Internet-Browser verfügen, es ist kein zusätzliches Plug-in notwendig.

2.3.1. Datenformat

Für eine erfolgreiche Verwendung und Implementierung von Inhalten aus bereits existierenden Datenbanken sind mehrere Faktoren ausschlaggebend. Die rechtliche Verfügbarkeit und Zulässigkeit sowie die technischen Gegebenheiten der Daten werden in den Partnerländern individuell evaluiert. Die Formate und Datenhaltung einzelner Datenbanken werden abgefragt und mit den Datenbankbetreibern abgestimmt. Ziel ist eine Kooperation zwischen den Datenbankbetreibern und dem EUWAS Portal im Hinblick auf leicht implementierbare Datenformate.

3. Fragen zur technischen Implementierung

1. Über welche Wege werden die Daten dargestellt?

öffentlich	nicht öffentlich
☐ Web-Application	☐ Web-Application
☐ Client /Server (z.B. kommunale Systeme)	☐ Client /Server (z.B. kommunale Systeme)
☐ Client-Anwendung (Access, Excel etc.)	☐ Client-Anwendung (Access, Excel etc.)

2. Was für Daten werden dargestellt?

☐ Grafik ☐ Video ☐ Ton ☐ Texte

☐ Tabellen ☐ Anhänge ☐ Tabellen ☐ Anhänge

☐ Andere .....

3. Wie aktuell sind die Daten? Wie häufig werden sie editiert?

☐ nie ☐ wöchentlich ☐ jährlich

☐ täglich ☐ monatlich

4. Auf welchen Data sources basiert Ihr System? Wie werden die für uns relevanten Daten gehalten?

☐ Datenbanken Welche? .....

☐ Textfiles (z.B. CSV, TXT, XML) Welche? .....

☐ Andere Dokumente Welche ? .....

☐ Statisches HTML

5. Welche Technologie wird verwandt (z.B. Java, C, C++, PHP, Perl)

6. Wie kommt man an die Daten? Was für Zugriffsrechte haben wir?

☐ http ☐ ftp ☐ scp ☐ webservice XML (RPC)

6.A Bevorzugen Sie push or pull services?

Implementierungskriterien

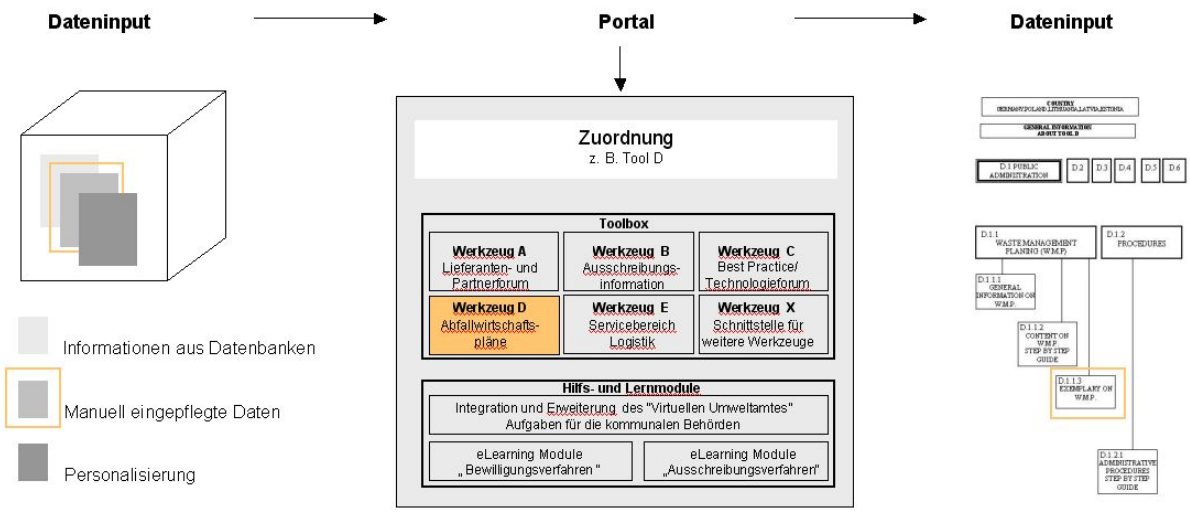
2.4. Dateninput

Das Ziel des EUWAS Projektes ist, Daten und Informationen aus dem Bereich der Abfallwirtschaft aus unterschiedlichen Quellen in einem Portal zusammenzufassen, um eine Informations- und Dienstleistungsplattform für verschiedene - mit der Abfallwirtschaft - betraute Nutzer anzubieten.

Da das Portal zusätzlich auch mehrsprachig aufgebaut ist und Daten teilweise nur für spezielle Länder von Bedeutung sind, setzt das Konzept eine Struktur voraus, die auf übergeordneter Ebene zuweist, welche Daten an welcher Stelle und durch wen in das Portal einfließen.

Zunächst werden drei Wege der Dateneingabe/ -integration festgelegt:

- Informationen aus Datenbanken: Das Projekt EUWAS bezieht sich inhaltlich auf fachbezogene Informationen zur Abfallwirtschaft. Dazu werden Datenbanken durch Experten in den einzelnen Länder manuell durchsucht und Inhalte demnach bewertet, ob sie einen Mehrwert für das Portal darstellen. Nach einem festgelegten qualitativen, sowie technisch-orientiertem Prozess werden diese dann mit dem Portal konnektiert.
- Manuell eingepflegte Daten: Die Toolverantwortlichen haben die Aufgabe, die Verteilung der Inhalte innerhalb der Portalstruktur, sowie die Relevanz der Inhalte festzulegen. Diese Abstimmungen werden länderspezifisch bzw. länderübergreifend besprochen und determiniert.
- Dateneingabe durch Nutzer: Innerhalb der Toolstruktur werden an bestimmten Stellen Funktionen angeboten, die die Zusammenarbeit der Partner in der europäischen Entsorgungswirtschaft aus dem öffentlichen und privaten Sektor fördern sollen. Dem Benutzer wird angeboten, sich und seine Firma auf der Plattform zu präsentieren. Beispielsweise sind dies das Erstellen eines Produktportfolios, das Veröffentlichen von Ausschreibungen oder die Suche nach einem Partner für ein bestimmtes Projekt.



Dateninput Datenbanken I Manuell I Personalisierung I

2.4.1. Rechte zur Datenpflege/Rechtekonzept

Der Dateninput unterliegt einer Rechte- und Rollenstruktur, die bestimmt, welcher Nutzer an welcher Stelle Zugriff auf welche Daten hat (Administrator, personalisierter Nutzer etc.).

Für die einzelnen Rollen werden die i-logic® spezifischen Rollen benutzt. Das Verwaltungs-Backend wird nur für eine geringe Anzahl von Administratoren zugänglich sein (eventuell eine Person pro Land).

Das Rechtekonzept wird voraussichtlich durch verschiedene Charakteristika bestimmt:

- EUWAS_ROLLEN

Innerhalb einer Rolle haben alle Mitglieder dieser Rolle die gleichen Rechte. Die Speicherung erfolgt an Hand eines eindeutigen Namens (Stage I – Stage IV). Ein User kann mehreren Rollen zugeteilt werden.

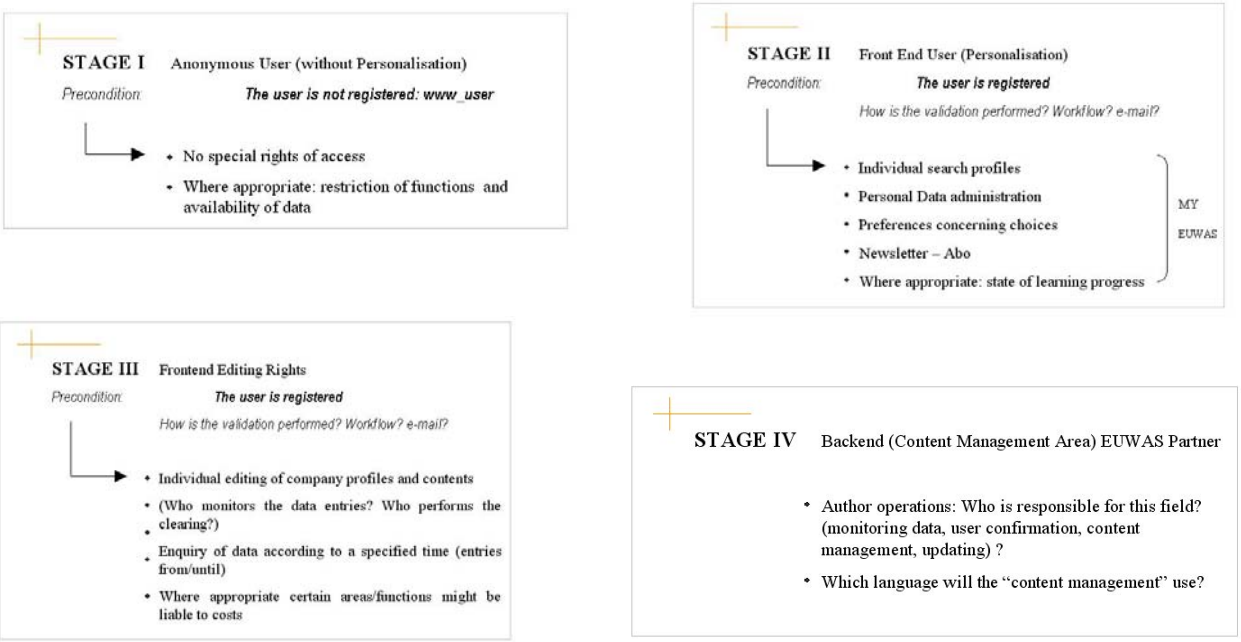
- EUWAS_Rechte

Einzelne Rechte definieren die Gruppenrechte einer Rolle. Ein Recht kann mehreren Rollen zugewiesen werden.

Die Rechte sind bis zu einem gewissen Grad flexibel angelegt und können modifiziert werden. Diese Personalisierung ist u.a. eine Voraussetzung, Teilbereiche des Portals später kostenpflichtig und einer restriktiveren Gruppe von Nutzern zugänglich machen zu können.

- EUWAS_USER

Alle EUWAS Nutzer erhalten ein eindeutiges Login (E-Mail-Adresse) und ein Passwort. Dies ist selbst änderbar und wird verschlüsselt in der Datenbank gehalten.



Rechtestruktur

2.4.1.1. Manueller Dateninput (Content management) - Backend

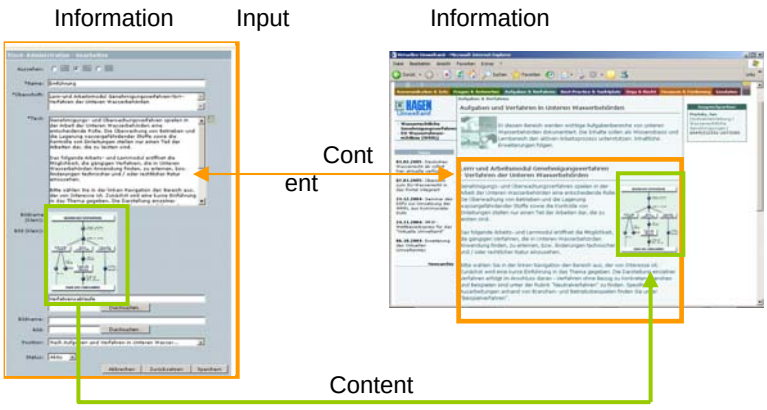
Nach Abschluss der Projektphase besteht ein Hauptziel des EUWAS Projektes darin, die unabhängige Weiterführung der Pflege der Inhalte in den verschiedenen Ländern zu gewährleisten. Dies setzt ein System voraus, das leicht verständlich und logisch strukturiert ist, damit sich Benutzer auch ohne HTML oder XML Kenntnisse zurecht finden. Dies wird mit Hilfe der i-logic® Content Management Software umgesetzt.

Folgende Ziele liegen diesem System zugrunde:

- Trennung von Inhalt und Gestalt/ Design, d.h. Inhalte werden in der Datenbank gehalten und können so an verschiedenen Stellen im Portal unter anderen Designs verwendet werden.
- Pflegbarkeit der Inhalte ohne große Programmierkenntnisse, d.h. der Content manager legt fest: was (welcher Inhalt) - wo (auf welchen Seiten) - wann (zu welcher Zeit) - wem (welchen Nutzern) angezeigt wird.

Das System besteht aus zwei Elementen :

- CMA (Content Management Application)
  Das CMA ermöglicht dem Autor eigenständig und ohne HTML Kenntnisse, Inhalte zu modifizieren, neu anzulegen oder zu löschen.



Inhaltspflege im Backendbereich

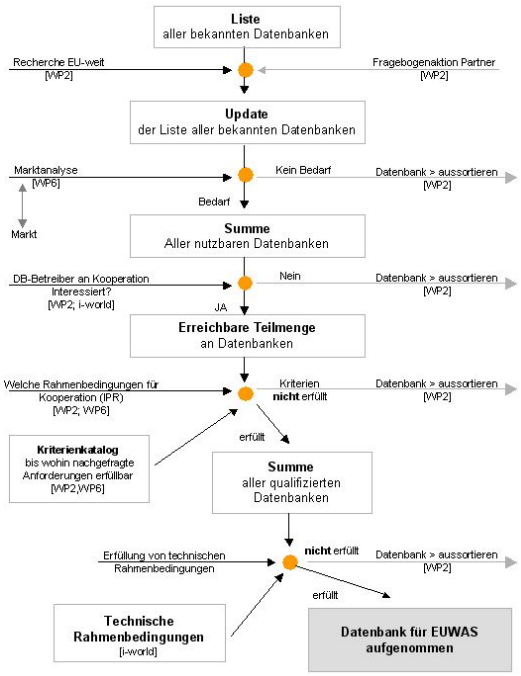
- CDA (Content Delivery Application)
  Das CDA Element nutzt und übersetzt diese Modifikationen und ermöglicht ein Update der Webseite, d. h. einzelne Inhaltselemente werden mittels der Scriptsprache unter Verwendung des Template parsers zu HTML-Seiten generiert und an den Webserver geliefert.

2.4.1.2. Dateninput im Rahmen der Personalisierung

Das EUWAS Portal stellt eine Plattform zur Verfügung, die u.a. die Kommunikation zwischen Abfallwirtschaftsakteuren dahingehend fördert, dass z. B. einzelne Unternehmen sich mit ihrem Produktportfolio und Dienstleistungen präsentieren und gegebenenfalls Partner für nationale oder internationale Projekte finden können. Die Rechtsstruktur sieht vor, dass sich Unternehmen zunächst registrieren müssen. Mit dem Login erhält der Benutzer das Recht seine Firmendaten einzugeben, zu modifizieren oder zu löschen. Um den Missbrauch weitestgehend zu unterbinden, werden diese eingepflegten Daten bewertet und erst nach Durchsicht freigegeben. („Redaktioneller“ Workflow)

2.4.1.3. Anbindung von Fremd-Datenbanken

Auf Grund der Mehrsprachigkeit des EUWAS Portals und der damit verbundenen Datenmenge, werden die Inhalte im Vorfeld auf der Grundlage von Ergebnissen einer durchgeführten Marktanalyse und eigenen Erfahrungswerten der Abfallwirtschaftsexperten in den Partnerländern, gefiltert und bereits im Vorfeld den Tools entsprechend strukturiert sein. Die Notwendigkeit der Anbindung kompletter Datenbanken ist noch unklar. Dennoch bedarf es zu diesem Zeitpunkt einer Vorgehensweise, die Verfügbarkeit der Datenbanken und deren Inhalte zu prüfen, um eventuell wertvolle Inhalte, die z.B. nur in einer Landessprache vorliegen, gegebenenfalls zu übersetzen oder eventuelle Kosten für spezielle Inhalte zu kalkulieren.



Strukturierte Kriterienabfrage

Die Datenabfrage erfolgt nach einer strukturierten Vorgehensweise, in der Schwerpunkte auf inhaltliche, rechtliche und technische Anforderungen gelegt

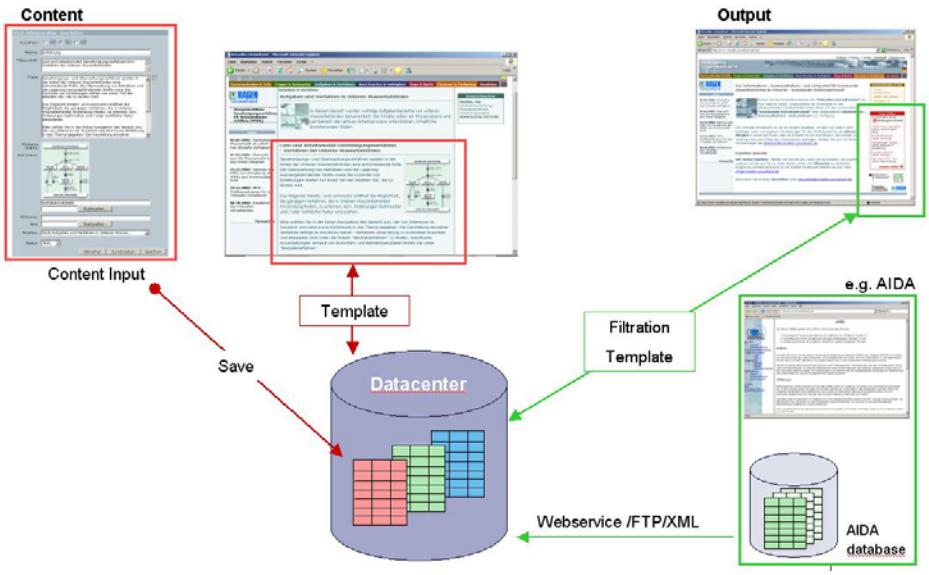
werden. Dabei ist jedes Partnerland für die Informationsrecherche und Bewertung der Inhalte im eigenen Land zuständig. Die Ergebnisse werden an entsprechender Stelle den einzelnen Tools in der jeweiligen Landessprache zugeteilt.

Folgende Kriterien müssen bei der Integration von Inhalten bedacht werden:

- Aktualität der Daten und der Datenhaltung in EUWAS
- Plausibilität/ Konsistenz der Daten
- Intellectual property rights (IPR) der Daten
- Verfügbarkeit der Daten in digitalem Format

Diese Kriterien werden durch eine strukturierte Vorgehensweise und eine länderorientierte Betrachtung vor der Implementierung angewandt.

Eine weitere Hürde stellt gegebenenfalls die technische Ausstattung der Zielgruppe in Osteuropa dar, speziell hier in den KMUs und den kommunalen Einrichtungen.



Datenhaltung- und Architektur (schematisch)

Durch eine Filterung der Inhalte aus den Datenbanken und eine Zuweisung der Daten auf die einzelnen Tools bzw. Bereiche des Portals, ist ein Teilimport der Daten notwendig. Dieser Schritt erfolgt über einen Datenbankimport Assistenten! Die Datenhaltung und das damit verbundene Datenmodell wird nach Abschluss der

ländergesteuerten Datenbank-Evaluation und den Kooperationsgesprächen mit den jeweiligen Datenbankbetreibern modelliert.

2.4.1.4. Konnektivität

Für die Anbindung von heterogenen Datenbanken ist es notwendig den Interaktionsablauf standardisiert abzuwickeln. Mittels standardisierter Kommunikationsmechanismen abgestimmte Inhalte werden in einem ebenfalls abgestimmten Format transportiert.

Eine Möglichkeit sind hier Webservices als eine Komponente, die eine Import Funktionalität über eine veröffentlichte Schnittstelle anbietet und über ein offenes, im Internet verwendetes Protokoll zugreifbar ist. Webservices basieren auf offenen Standards und benutzen bewährte Internet-Protokolle.

Sie ermöglichen eine lose Koppelung zwischen Systemen. Dadurch kann eine große Anzahl von Systemen untereinander kommunizieren, ohne an die Nachteile marginaler Schnittstellenänderungen gebunden zu sein.

Es ist geplant, ein „EUWAS Format“ auf Basis der Metasprache XML selbst zu definieren und dies zur Verfügung zu stellen.

Wo diese Möglichkeit der Konnektierung nicht besteht, ist es desweiteren ebenfalls möglich über „triviale“ Protokolle bzw. Mechanismen Daten zu importieren, wie z. B. ftp, ssh, http, VPN, periodisch manuell etc.

3. Zusammenfassung

Die grenzübergreifende Nutzung und Bereitstellung von Daten und Informationen zum Thema Abfallwirtschaft, die mit Hilfe des EUWAS Portals angestrebt wird, soll die Kooperation und den Wissensaustausch zwischen den Europäischen Nationen fördern und den Weg zu einer grenzübergreifenden Transparenz im Bereich der Abfallwirtschaft unterstützen.

Der Know-how Transfer zwischen den beteiligten Nationen führt dazu, dass Angleichungen und Anpassungen von Standards, wie z.B. auf dem Gebiet der Best Available Technology (BAT) auf einem qualitativ hochwertigen Niveau getroffen werden.

Aufgrund des vernetzenden Charakters zu anderen Projekten wird im Rahmen der technischen Konzeption und Umsetzung auf offene Schnittstellen und Integrationsmöglichkeiten geachtet.

Die strukturierte Haltung und Bereitstellung von Daten für Abfallwirtschaftsakteure in den verschiedenen Ländern spielt bei der erstrebenswerten Vereinigung technischer Ressourcen, sowie der Förderung des Austausches von Erfahrungen und der daraus resultierenden Annäherung von Prozessen eine wesentliche Rolle.

Anbindung der Umweltprobenbank des Bundes (UPB) an ein Web GIS

Martin Stöcker, Institut für Geoinformatik Universität Münster
mstoeck@uni-muenster.de

Stephan Merten, Institut für Geoinformatik Universität Münster
stmerten@uni-muenster.de

Liane Reiche, Institut für Geoinformatik Universität Münster
liane.reiche@uni-muenster.de

Andrea Körner, Umweltbundesamt Dessau
andrea.koerner@uba.de

Nina Brüders, Umweltbundesamt Dessau
nina.brueders@uba.de

Abstract / Einleitung

Im Rahmen der Begleitforschung der **Umweltprobenbank des Bundes (UPB)** wurde vom 01. September 2003 bis 29. Februar 2004 am Institut für Geoinformatik der Universität Münster untersucht, wie eine integrative Nutzung der in der UPB enthaltenen Daten mit weiteren – auch externen – relevanten Geodatenbeständen zu realisieren ist.

Der Schwerpunkt des hier vorgestellten Projektes lag in der Anbindung der Sachinformationen aus der Umweltprobenbank (UPB) an Geobasis- und Geofachdaten auf Basis von modernen GIS- und WEB - Technologien.

Der Beitrag beleuchtet die Ausgangssituation am UBA hinsichtlich eingesetzter Technologien und beschreibt die Konzeption sowie Umsetzung des Entwicklungsvorhabens.

1. Das Umweltbundesamt und die Umweltprobenbank

Eine Kernaufgabe des Umweltbundesamtes (UBA) als wissenschaftliche Umweltbehörde im Geschäftsbereich des Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit (BMU) ist die Ermittlung, Beschreibung und Bewertung des Zustands der Umwelt, um Beeinträchtigungen für Mensch und Umwelt frühzeitig zu erkennen. Aus den erhobenen Informationen generieren die zuständigen Facheinheiten im UBA entsprechende Entscheidungsgrundlagen für

öffentliche Stellen und informieren die Bevölkerung über aktuelle Umweltsituationen, wie beispielsweise die Feinstaubbelastung. Die hierfür notwendigen Erkenntnisse werden durch eigene Forschungen oder die Vergabe von Forschungsaufträgen gewonnen.

Das UBA, welches momentan aus der Zentralabteilung und fünf Fachbereichen besteht, hat seinen zentralen -Sitz in Dessau. Weitere Standorte befinden sich in Berlin, hinzukommen zwei Außenstellen und mehrere Messstationen, die über Deutschland verteilt sind.

Weiterführende Informationen zum Umweltbundesamt sind im Internet unter **www.umweltbundesamt.de** zu finden**.**

Für die UPB werden ökologisch repräsentative Umwelt- und Humanproben gesammelt, auf umweltbelastende oder gefährdende Stoffe untersucht und abschließend unter Tiefkühlbedingungen eingelagert. Die zu untersuchenden Stoffe sind Repräsentanten solcher Stoffe oder Stoffgruppen, von denen eine nachteilige Wirkung auf die Gesundheit von Mensch und Tier vermutet werden kann. Derzeit werden im Routinebetrieb der UPB im Bereich der anorganischen Stoffe 20 Elemente sowie Methylquecksilber, im Bereich der organischen Substanzen 20 Chlorkohlenwasserstoffe (CKW) und 19 polyzyklische aromatische Kohlenwasserstoffe (PAH) untersucht [BMU, 2000].

Sollten mit dem wissenschaftlichen Fortschritt neue Stoffe erkannt werden, die eine eventuell schädliche Wirkung auf Organismen haben, können sie auch retrospektiv in die Untersuchungen aufgenommen werden.

So dienen die in der UPB gesammelten Proben der Erfassung und Dokumentation von großräumigen Umweltbelastungen durch chemische Stoffe, der jährlichen Darstellung des Zustands und der Entwicklung der repräsentativen Ökosysteme sowie der Früherkennung von Gesundheits- und Umweltgefahren.

Zu diesem Zweck werden die Proben unter anderem auf folgende Zielsetzungen hin untersucht:

- Überwachung der Konzentration gegenwärtig bekannter Schadstoffe durch ein systematisches Monitoring der gewonnen Proben vor der Archivierung
- Trendaussagen über lokale, regionale und globale Entwicklungen der Schadstoffbelastungen auf der Grundlage authentischen Materials aus der Vergangenheit
- Bestimmung der Konzentration von Stoffen, die zum Zeitpunkt der Einlagerung noch nicht als Schadstoffe erkannt waren (retrospektives Monitoring)
- Erfolgskontrolle von Verbots- und Beschränkungsmaßnahmen im Umweltbereich.

Für die UPB werden in 13 Probenahmegebieten repräsentative Umweltproben (bspw. Fichten- und Kiefernnadeln) und an vier Standorten Humanproben (bspw. Kopfhaar und Blutplasma) im Auftrag des UBA durch entsprechende Institutionen gewonnen, auf umweltrelevante Stoffe analysiert und anschließend tiefgekühlt eingelagert.

Innerhalb der Probenahmegebiete (PNG), die sich i. d. R. an den Grenzen von Naturschutzgebieten orientieren, wurden zur aussagekräftigen Beprobung repräsentative Flächen, die sog. Gebietsausschnitte (GA), abgegrenzt, in denen sich die eigentlichen Probenahmeflächen (PNF) befinden. Die Abgrenzung der Gebietsausschnitte erfolgte auf Basis von Wassereinzugsgebieten.

Alle im Routinebetrieb der Umweltprobenbank anfallenden Erhebungen werden im Informationssystem der UPB (IS UPB) lückenlos dokumentiert. Basis dieses IS UPB ist eine Oracle 9i Datenbank mit verschiedenen MS-Access Frontends zur Datenpflege.

Weiterführende Informationen zu Inhalten und Struktur der Umweltprobenbank sind im Internet unter **www.umweltprobenbank.de** zu finden**.**

2. Rahmenbedingungen

Aufgrund der Bedeutung von Geoinformationen für die Umweltbeobachtung wird bereits seit mehreren Jahren am UBA auf GI Technologie gesetzt. Hierbei liegt der Schwerpunkt auf ESRI Technologie. Neben den Desktop Produkten ArcView und ArcGIS verfügt das UBA über Lizenzen für die Serveranwendungen ArcSDE und ArcIMS, die für den Betrieb der Applikationen am UBA verwendet werden sollen. Als Datenbankmanagementsystem steht am UBA Oracle zur Verfügung. Weiterhin gab es bereits durch die eingesetzte Hard- und Software eine klare Vorgabe hinsichtlich der eigentlichen Architektur. Auf die Architektur wird in Kapitel 4 näher eingegangen.

Die Datengrundlage für das Web GIS bilden die deutschlandweit über das Bundesamt für Kartographie und Geodäsie (BKG) verfügbaren Geobasisdaten, die dem UBA aufgrund bestehender Verwaltungsvereinbarungen über den Datenaustausch zwischen Bund und Ländern zur Verfügung gestellt werden. Hinzu kommen die am UBA verfügbaren Geofachdaten, welche die Abgrenzung der Probenahmegebiete beschreiben.

Hinsichtlich der weiteren Layoutgestaltung für die Web Applikation waren die Vorgaben des UBA zu berücksichtigen, die sich im Wesentlichen am „BundOnline 2005“ orientieren. Diese im Jahr 2000 durch Bundeskanzler Gerhard Schröder im Rahmen der Expo Eröffnung in Hannover der Öffentlichkeit vorgestellte Initiative

verfolgt das Ziel, auf Bundesebene eGovernment-Applikationen durch die Bereitstellung von Basiskomponenten (z.B. Content Management System) und Standards zu fördern [KBST, 2003]. Auflagen seitens des UBA bestanden in der Nutzung der UBA-Stylesheet-Vorgaben, die Mehrsprachigkeit, da die Applikation sowohl in deutscher, als auch in englischer Sprache verfügbar sein soll, sowie die Nutzung großer Symbole und Button, die auch Sehbehinderten die Nutzung der implementierten Funktionen ermöglichen.

Aufgrund der begrenzten Projektlaufzeit wurden zunächst nur die Daten der 13 Probenahmegebiete des Umweltbereichs berücksichtigt. Die Daten der UPB-Humanproben konnten noch nicht eingebunden werden.

3. Konzeption des Web GIS

Infolge des kurzen Zeitraumes von nur sechs Monaten wurde als Vorgehensweise für die Entwicklung das eXtrem Programming, kurz XP gewählt.

XP als junger Softwareentwicklungsprozess geht auf eine Initiative von Kent Beck, Ron Jeffries und Ward Cunningham zurück [LIPPERT et al., 2002].

XP als Entwicklungsprozess ruht im Wesentlichen auf den folgenden vier Säulen [BECK, ANDRES, 2004]:

- Simplicity

  Durch die Fokussierung auf das Wesentliche mit Blick auf die Funktionalität bei der Entwicklung der Anwendung wird die Möglichkeit geschaffen, innerhalb kurzer Zeiträume funktionsfähige Lösungen zu generieren.

- Communication

  Intensive Kommunikation unter Verwendung aktueller Kommunikationsmittel ermöglicht bei zielgerichteter Verwendung eine erhebliche Reduktion des Dokumentationsaufwandes.

- Feedback

  Um den Anforderungen der Nutzer gerecht zu werden, ist intensives Feedback zwischen Entwicklern und Testern im Rahmen eines iterativen Entwicklungsprozesses mit entsprechenden Prototypen notwendig.

- Courage

  Schlussendlich erfordert die Umsetzung von eXtrem Programming innerhalb von Entwicklungsprojekten eine entsprechende Weitsicht und Courrage auf Seiten der Entscheider.

Zentriert auf die Projektbeteiligten und unter Verwendung der fünf Basisprinzipien (Rapid Feedback, Assume Simplicity, Incremental Change, Embracing Change und Quality Work), die auf den vier oben beschriebenen Säulen ruhen, führt die Anwendung von eXtrem Programming gerade bei kleineren und mittleren Projekten zu hervorragenden Ergebnissen. Voraussetzung hierfür ist allerdings ein intensiv koordinierter, iterativer Entwicklungsprozess mit entsprechend nachhaltiger Einbindung der Anwender, was zu sehr hohen Ansprüchen an die Kommunikations- und Teamfähigkeit der beteiligten Akteure führt [BECK, ANDRES, 2004].

Nachdem im Kick off Meeting der Startschuss für das Projekt gegeben und die jeweiligen Ansprechpartner benannt waren, begann die konzeptionelle Arbeit an der Web Applikation mit der Identifikation der Nutzergruppen. Hierbei kristallisierten sich entsprechend der Recherchemöglichkeit im IS UPB zwei Nutzergruppen heraus. Dies ist zum einen die Öffentlichkeit, die sich generell informieren möchte, zum anderen der Fachanwender, der darüber hinausführende Abfragen durchführen möchte.

Basierend auf diesen beiden identifizierten Nutzergruppen erfolgte die Festlegung der Funktionalitäten und der zu visualisierenden Geodaten, die diesen Nutzergruppen via Web Client angeboten werden sollten. Zwecks Sicherung der Probenahmeflächen der Umweltprobenbank waren sämtliche Funktionalitäten und Geodaten für die Öffentlichkeit zu vermeiden, die eine eindeutige Identifikation der Flächen und somit ein Auffinden der Flächen im Gelände ermöglichen. Für die internen Nutzer sollten allerdings erweiterte Funktionalitäten, wie die Möglichkeit der Streckenmessung und die Möglichkeit der Pufferbildung implementiert werden. Die nachfolgende aufgeführte **Tabelle 1** gibt eine Übersicht über die Funktionalitäten, die der jeweiligen Nutzergruppe zur Verfügung gestellt werden.

Tabelle 1: Übersicht über die Funktionalitäten die den beiden Nutzergruppen via Web Client verfügbar gemacht werden. (x zeigt das Funktion verfügbar ist)

Funktionalität	Nutzergruppe Öffentlichkeit	Nutzergruppe Fachanwender UBA intern
Übersichtskarte ein-/ausblenden	X	X
Zoom in	X	X
Zoom out	X	X
Zurück zur Ausgangsauflösung	X	X
Verschieben	X	X
Informationen über die Probenahmefläche abfragen	X	X
Navigation im Kartenrahmen	X	X
Zurück zur letzten Auflösung		X
Identifizieren der Sachattribute der Probenahmearten		X
Erstellung komplexerer Sachdatenabfragen		X
Suchen nach Attribut		X
Streckenmessung		X
Auswahl via Box		X
Puffer bilden		X
Auswahl aufheben		X
Aktuellen Ausschnitt drucken		X

Innerhalb der Konzeptionsphase erfolgte weiterhin die Vorauswahl der via Web Client zur Verfügung gestellten Geodaten. Unter Berücksichtigung der Prämisse, dass Probenahmeflächen aufgrund der Informationen im Web Client nicht im Gelände aufgefunden werden sollen, wurde auf eine Visualisierung der Geobasisdaten TK 25 (topographische Karte 1:25.000) und TK 50 (topographische Karte 1:50.000) verzichtet.

Die nachfolgend aufgeführte **Tabelle 2** gibt einen Überblick über die ausgewählten Geodaten, die innerhalb der Web Lösung angeboten werden.

Tabelle 2: innerhalb des Web GIS verwendete Geodaten

Verwendete Geodaten	Erläuterung
Räumliche Beprobungseinheiten	Probenahmegebiete, Gebietsausschnitte, Probenahmeflächen bzw. Probenarten
Administrative Grenzen	Bundesländer, Kreise, Gemeinden
CORINE (1992)	Landnutzungsdaten
TÜK500	Topographische Übersichtskarte 1:500.000
TÜK 200	Topographische Übersichtskarte 1:200.000
TK 100	Topographische Karte 1:100.000
Satellitendaten	Landsat-TM

4. Realisierung des Web GIS

Die eigentliche Realisierung erfolgte in den fünf nachfolgend aufgeführten Stufen:

- Einrichtung einer Test- und Entwicklungsumgebung vergleichbar zur Architektur im UBA und Integration einer UPB Instanz.
- Aufbereitung und Integration der notwendigen Geodaten.
- Kartographische Konzeption, Anbindung der UPB und Erstellung der Web GIS-Dienste.
- Entwicklung und Design eines WebGIS Clients, nutzbar für die Öffentlichkeit und für die UBA-Fachnutzer.
- Erstellung von Webseiten zur Integration des Web GIS Clients in den Internetauftritt der Umweltprobenbank und Funktionstests.

Basierend auf den Ergebnissen der Konzeptionsphase und unter Berücksichtigung der Rahmenbedingungen am UBA, wurde am Institut für Geoinformatik eine Test- und Entwicklungsumgebung aufgebaut. Diese ermöglichte es, unter „UBA-Bedingungen“ zu testen und zu entwickeln. Serverseitig als Hardware verwendet wurde eine SUN Blade 1000 mit Solaris 8 als Betriebssystem. Darauf wurden neben dem RDBMS Oracle in der Version 9.0.2 auch die ESRI ArcSDE Version 8.3 und der

ESRI ArcIMS in der Version 4.0.1 installiert. Hinzu kam die Installation des Apache Web Servers in der Version 2.0.43. Abb. zeigt schematisiert die Architektur der Test- und Entwicklungsumgebung.

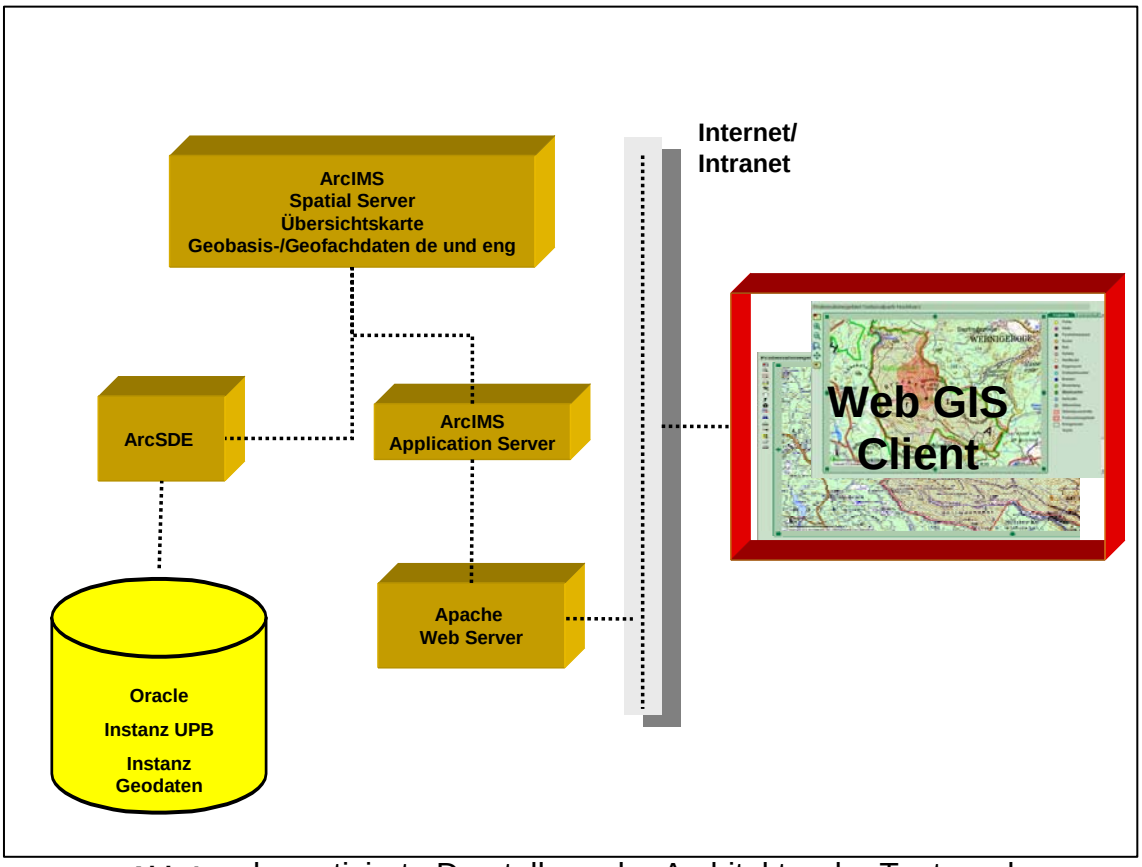


Abb.1: schematisierte Darstellung der Architektur der Test- und Entwicklungsumgebung

Hierbei erfolgt der Datenzugriff auf die eigentliche Geodateninstanz über die ArcSDE und über den ArcIMS Spatial Server, der die Geodatendienste zur Verfügung stellt und das Rückrad des ArcIMS bildet. Während einer Anfrage führt der ArcIMS Spatial Server eine oder mehrer der nachfolgend aufgeführten Funktionalitäten aus und sendet die aufbereiteten Daten als Bild-Dateien via Web Server zum Client:

- Image: übermittelt georeferenzierte Bild-Dateien
- Feature: übermittelt Karten-Objekte
- Query: erlaubt Datenbankabfragen
- Geocode: erlaubt die Georeferenzierung von Adressen
- Extract: generiert Shapefiles aus eine Geodatenbank

Einzelne Anfragen werden über den ArcIMS Application Server verarbeitet. Der Web Server wird hierbei über einen so genannten Connector mit dem Application Server

verbunden. Im Projekt wurde der ArcIMS Servlet Connector verwendet, der zur Kommunikation zwischen Web Server und Application Server das XML Derivat ArcXML gebraucht. Der Application Server leitet die entsprechenden Anfragen an die jeweiligen Geodatendienste des Spatial Servers weiter, überwacht diese und entfernt nicht mehr benötigte Ausgabe-Dateien.

Nach abgeschlossener Installation der Software wurde eine Instanz der Umweltprobenbank, die als Dump verfügbar war, eins zu eins in die Testumgebung übernommen.

Anschließend begann die zweite Stufe mit der Aufbereitung und Integration der in **Tabelle 2** beschriebenen Geodaten. Da die Geodaten deutschlandweit angeboten werden sollten, musste zunächst ein einheitliches deutschlandweites Koordinatensystem gewählt werden. Man entschied sich hier für das Koordinatensystem UTM (Universal Transverse Mercator), Zone 32 N, basierend auf dem European Terrestrial Reference System 1989, welches von den zuständigen Vermessungsinstitutionen der Länder der Europäischen Union als Basis Referenzsystem eingeführt werden soll bzw. bereits eingeführt ist [FLACKE, KRAUS, 2003]. Nachdem die verfügbaren Geodaten, soweit notwendig, projiziert worden sind, wurden sie via Shell Skripte über die angepasste ArcSDE in eine speziell für die vorliegenden Geodaten erstellte und optimierte Oracle Instanz geladen. Hierbei wurden die Vektordaten wie Probenahmegebiete oder Gebietsausschnitte als separate Vektordatensätze, die Rasterdaten wie TÜK 200 als Rasterkataloge in die ArcSDE geladen. Zur Reduktion des Speicherbedarfs wurden bei den Rasterdaten nur die Bereiche in und um die Probenahmegebiete in die Datenbank geladen. Aufgrund von Performanzgewinnen wurde als Schema innerhalb der Oracle Geodateninstanz das von ESRI via SDE zur Verfügung gestellte SDE compressed binary Schema gewählt. Trotz der Performanzverluste der Rasterkataloge gegenüber Rastermosaiken innerhalb der ArcSDE entschied man sich für die Rasterkataloge, da diese einfacher zu pflegen und zu aktualisieren sind. So entfallen hier die bei Mosaiken unter ArcSDE 8 noch notwendigen Arbeiten an der colour map. Basis für die durchgeführten Optimierungen an der ArcSDE und an der Oracle Instanz war der von ESRI zur Verfügung gestellte Konfiguration und Tuning Guide.

Nach erfolgreicher Integration der notwendigen Geo- und Fachdaten in die Test- und Entwicklungsumgebung begann die dritte Stufe mit der kartographischen Konzeption.

Die topographischen Karten und administrativen Grenzen werden nicht gleichzeitig, sondern maßstabsabhängig angezeigt. Werden große Gebiete auf der Karte dargestellt, dann kommen kleinmaßstäbige Karten wie die TÜK500 zum Einsatz, wird

in die Karte „hineingezoomt“, dann werden der Reihe nach die Karten mit größerem Maßstab bis zur TK100 angezeigt.

Für Vektordaten, wie die Landnutzungs- und UPB-Daten, sowie die administrativen Grenzen wurden Legenden erstellt. Dabei konnte auf Vorgaben vom UBA zurückgegriffen werden.

Die UBA-spezifischen Polygone der Kartenlayer „Probenahmeflächen“ und „Gebietsausschnitte“ werden in einem kräftigen Rot mit unterschiedlichen Schraffuren visualisiert. Dadurch können diese Gebiete auch von den Rottönen der TÜK500 unterschieden werden.

Die Probenahmeflächen werden nicht direkt, sondern durch verschiedene Layer, bei denen jeweils die an dem Standort genommenen Probenarten durch unterschiedliche Farben dargestellt werden, abgebildet. Hierzu war es notwendig, den Layer „Probenahmeflächen“ durch eine Verbindung („Join“) über ein eindeutiges Feld mit den Tabellen der einzelnen Probenarten der UPB zu verknüpfen. Dieser „Join“ realisiert somit die Anbindung der UPB- an die Geobasisdaten. Bei Abfragen in der Karte auf die Standorte der Probenarten werden die Attribute direkt aus der UPB benutzt, wodurch ihr Raumbezug analysiert werden kann.

Die darauf aufbauenden ArcIMS Dienste wurden so konzipiert, dass es einen einzigen Dienst gibt, der die Geodaten für alle Probenahmegebiete deutschlandweit zur Verfügung stellt. Für die Anforderung der Zweisprachigkeit der Legenden wurde ein zweiter Dienst mit den gleichen Daten in englischer Sprache erzeugt. Ein dritter Dienst übernimmt die Aufgabe der Übersichtskarte und zeigt daher nur eine Deutschlandkarte mit den Grenzen der Bundesländer an.

Nach der kartographischen Konzeption und der Erstellung der ArcIMS Dienste wurde in Stufe vier der eigentliche Web Client als Schnittstelle zwischen dem Benutzer und den Web GIS-Diensten entwickelt. Er basiert auf dem Standard HTML-Viewer, der vom ArcIMS zur Verfügung gestellt wird. Dieser wurde an das Design bzw. an die Stylesheet-Vorgaben des Umweltbundesamtes angepasst. Neu hinzugefügt wurde die Navigation der Karte in alle Himmelsrichtungen, die das Kartenfenster einrahmt. Auch das Layout und die Funktionen zum Umschalten zwischen Legende und der Auswahl der Karteninhalte wurden verändert.

Basierend auf den Vorgaben aus der Konzeptionsphase wurde der Client so entwickelt, dass es eine gemeinsame Version für die öffentliche Darstellung und für den Fachnutzer gibt. Änderungen, die in beiden Versionen gelten, müssen nur einmal durchgeführt werden. Die jeweilige Version wird durch Übergabe bestimmter Parameter in der URL ausgewählt. Da die oben beschriebenen UPB-WebGIS Dienste das gesamte Gebiet Deutschlands abdecken, wird über den Client die

Auswahl des Kartenausschnitts für das jeweilige PNG vorgenommen. Diese Auswahl geschieht, wie die Auswahl der Sprache, ebenfalls dynamisch über einen bestimmten Parameter in der URL.

Die öffentliche Version des Clients unterscheidet sich von der für die Fachnutzer insbesondere durch einen geringeren Funktionsumfang (siehe hierzu auch **Tabelle 1**). Die Icons, über welche die Funktionen ausgewählt werden können, wurden gegenüber den Symbolen des Standard HTML Viewer vergrößert und an die Farbgebung angepasst. Der Client für die Fachnutzer enthält weitere Funktionen zur Abfrage und Selektion der an die Layer angebundenen Fachdaten.

In der abschließenden fünften Stufe wurden die weiteren Web Seiten zur Einbindung des Web Clients in den Internetauftritt der Umweltprobenbank erstellt und via Test- und Entwicklungsumgebung einschließlich Web Client für Funktionstests zur Verfügung gestellt.

Die Integration des UPB-WebGIS Clients für die öffentliche Darstellung der dynamischen Karten im Internet wurde über verschiedene Zugangsseiten realisiert, über die der Client mit unterschiedlichen Parametern angesprochen werden kann.

Den Einstieg bildet eine Übersichtskarte, auf der alle PNG mit ihrem Namen und dem zugehörigen Ökosystem abgebildet sind. Wenn der Benutzer die Maus über den Namen eines PNG bewegt, wird ein charakteristisches Landschaftsfoto sichtbar (Abb.).

Auf der Karte wurden die PNG so verlinkt, dass der Benutzer über einen Link auf eine Auswahlseite gelangt, die das PNG mit seinen Probenarten näher erläutert und den Link zum UPB-WebGIS Client sowie zu der Recherche der UPB enthält.

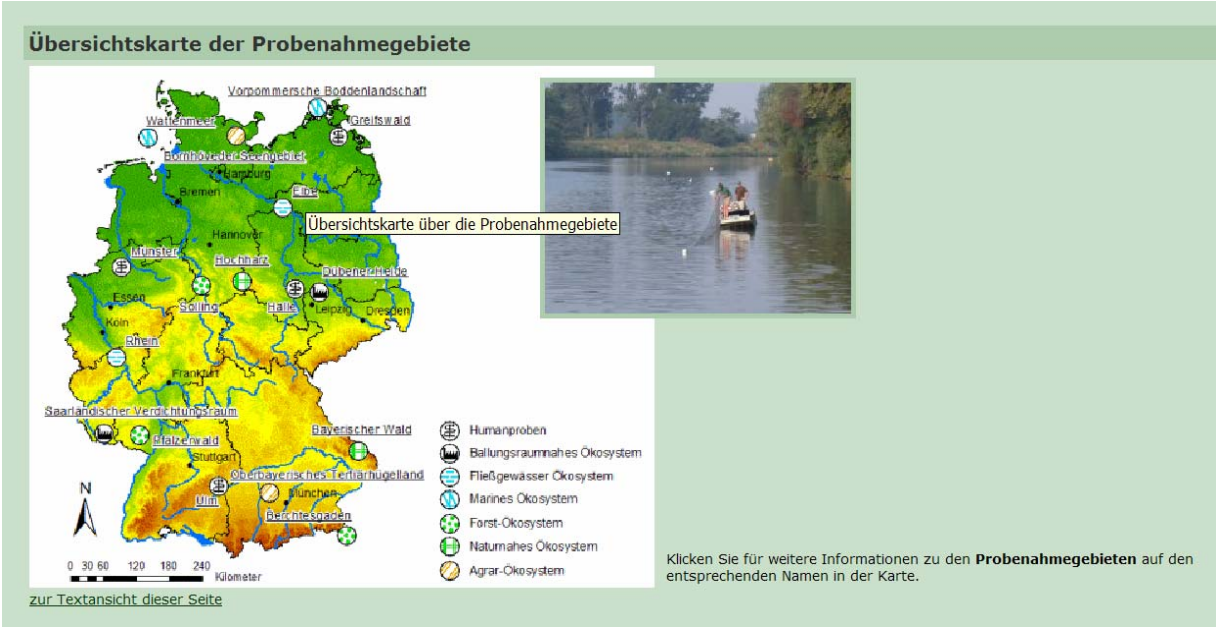


Abb. 2: Übersichtskarte über die Probenahmegebiete

Bei den PNG Wattenmeer, Rhein und Elbe sind die GA weiträumig verteilt. Für diese Gebiete wurde eine eigene Übersichtskarte gestaltet und auf der Auswahlseite abgelegt. Hier wird der Web GIS Client mit der Ausdehnung des GA gestartet.

Die Auswahlseite der PNG Hochharz und Berchtesgaden enthält zusätzlich auch einen Link zu einer 3D-Ansicht des PNG.

Um die Vorgaben seitens des UBA zu realisieren, wurde jede der beschriebenen Webseiten sowohl in einer deutschen als auch in einer englischen Version gestaltet. Abschließend wurden die Web Seiten einschließlich Web Client zum Testen freigegeben. Nach circa drei Wochen intensiver Tests seitens des UBA und Korrektur kleinerer Fehler wurden die Entwicklungen zur Migration und zum Echtbetrieb in der IT Struktur des UBA freigegeben und mit Migrationskonzept und Dokumentation dem UBA übergeben.

5. Migration der Entwicklungen ins UBA

Zum Ende des Projektes wurde gemeinsam mit den IT Verantwortlichen des UBA die Migration der entwickelten Applikationen in die IT Struktur des UBA vorgenommen. Dies geschah auf der Basis eines detaillierten Migrationskonzeptes, welches die verschiedenen Schritte beschreibt und die notwendigen Skripte enthält.

Nachdem am UBA eine entsprechende Oracle Instanz für die Geodaten mit ArcSDE Anbindung installiert war, erfolgte zunächst die Anpassung und Optimierung sowohl der Datenbank, als auch der ArcSDE Installation. Neben der Übergabe von SQL-Skripten, die innerhalb der Oracle Instanz die notwendigen Tablespaces und Nutzer anlegten erfolgt die Übergabe eines Parameterfiles (dbtune-File) für die ArcSDE, der Informationen zur Speicherung der verschiedenen Geodaten (Zuweisung von Tablespaces für bspw. Daten und Indizes, Größe der übertragenen Datenpakete, Art der zu verwendenden Schemata, usw.) enthält.

Anschließend begann das Einlesen der notwendigen Geodaten mittels der bereits im Rahmen des Projektes erstellten Shell Skripte. Nachdem so die Datengrundlage geschaffen war, wurden die notwendigen ArcIMS Dienste aufgesetzt, der Client installiert und die Web Seiten in die Web Seiten der Umweltprobenbank integriert. Das entwickelte Web GIS ist nunmehr via Internet unter der Web Adresse der Umweltprobenbank (http://www.umweltprobenbank.de) z. Zt. noch unter dem Link „Aktuelles“ erreichbar.

6. Zusammenfassung und Ausblick

In diesem Beitrag wurde die Anbindung der Umweltprobenbank des Bundes an ein Web GIS vorgestellt und die einzelnen Phasen des Projektes, dass am Institut für Geoinformatik der Universität Münster realisiert wurde, beleuchtet. Speziell die gewählte Vorgehensweise des eXtreme Programmings führte zu einer sehr gut auf die Nutzerbedürfnisse angepassten Lösung. Durch die Verwendung der am Markt etablierten GIS Produkte der Firma ESRI ist eine effektiv erweiterbare, skalierbare Web GIS Lösung entstanden, die als Basis für weitere Entwicklungen sehr gut geeignet ist.

Im Laufe des beschriebenen Vorhabens wurde die immanent wichtige Bedeutung von Geobasis- und Geofachdaten zur Bearbeitung von Fach-Fragestellungen des UBA erneut deutlich. Wie bereits in einem frühren Vorhaben angeregt, stellt die Verknüpfung der in der UPB vorgehaltenen Sachinformationen mit den verfügbaren Geoinformationen, einen Mehrwert sowohl für UBA-interne, als auch für externe Nutzer dar [STREIT ET AL., 2002].

Literaturverzeichnis

[BECK, ANDRES, 2004]

BECK, K.; ANDRES D.: eXtreme Programming explained. Second Edition, Addison Wesley, 2004.

[BMU, 2000].

BMU, Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (2000) Umweltprobenbank des Bundes – Konzeption. Umweltbundesamt, Berlin.

[FLACKE, KRAUS, 2003]

FLACKE, W., KRAUS, B.: Koordinatensysteme in ArcGIS - Praxis der Transformationen, Halmstadt, 2003.

[KBST, 2003]

KBST (KOORDINIERUNGS- UND BERATUNGSSTELLE DER BUNDESREGIERUNG FÜR INFORMATIONSTECHNIK IN DER BUNDESVERWALTUNG IM BUNDESMINISTERIUM DES INNEREN): Saga – Standards und Architekturen für eGovernment-Anwendungen, Version 1.1.Schriftenreihe der KBSt. Berlin, 2003.

[LIPPERT et al., 2002]

LIPPERT, M.; ROOK, S.; WOLF, H.: Software entwickeln mit XP eXtreme Programming. Erfahrungen aus der Praxis. Heidelberg, 2002.

[STREIT ET AL., 2002]

STREIT, U.; VON DER WEIDEN, A.; SCHMIDT, B.; MÜLLER, A.; ALTGOTT, K.; GERLACH, N.; AHMANN, P.; ARP, K.; STARKE, A.; S. VIENKEN: Entwicklung von Strategien zur Auswertung, Darstellung und Bewertung von Umweltdaten unter Anwendung von geographischen Informationssystemen (unveröffentlichter Abschlußbericht UPB GIS). Münster, 2002.

www.POP-DioxinDB

Ein Web-Service mit XML-Technologie für die Dioxin-Datenbank des Bundes und der Länder

Nina Brüders, Umweltbundesamt, nina.brueders@uba.de

Gerlinde Knetsch, Umweltbundesamt, gerlinde.knetsch@uba.de

Erich Weihs, Bayerisches Staatsministerium f. Umwelt, Gesundheit und Verbraucherschutz, erich.weihs@stmugv.bayern.de

Rene Pöschel, deborate GmbH, rene.poeschel@deborate.de

Abstract

Die Dioxin-Datenbank enthält derzeit Daten von etwa 220 kompartiment-spezifischen Messprogrammen zu Zustandsdaten der Belastung der Umwelt mit Dioxinen und anderen chlororganischen Verbindungen. Um diese Daten verschiedenen Nutzergruppen zugänglich zu machen, hat das Umweltbundesamt in Kooperation mit dem Bayerischen Staatsministerium für Umwelt, Gesundheit und Verbraucherschutz das Projekt „www.POP-DioxinDB“ initiiert. Der Lösungsansatz basiert auf reiner XML-Technologie. Der Web-Service wird im Herbst 2005 fertig gestellt und unter www.POP-DioxinDB.de im Internet erreichbar sein.

1. Einleitung

1.1. Fachlicher Hintergrund

Für die Bewertung der Belastung der Umwelt - einschließlich des Menschen - mit Organochlorverbindungen werden von Bund und Ländern in Deutschland eine Reihe von verschiedenen Messprogrammen durchgeführt, die mit unterschiedlichen Zielsetzungen und Rahmenbedingungen durchgeführt werden. Mit Beschluss der 37. Umweltministerkonferenz 1991 wurde die Bund/Länder-Arbeitsgruppe DIOXINE u.a. damit beauftragt, die „zentrale Dokumentation und Auswertung von Ergebnissen von Untersuchungsprogrammen zur Dioxinbelastung der Umwelt, die durch Bund und Länder initiiert werden,“ zu verbessern. Dies wurde mit dem Aufbau eines zentralen Datenbanksystems im Umweltbundesamt realisiert, an dem das ehemalige

Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin beteiligt ist.

Die Aufgabe der Dioxin-Datenbank ist sowohl die Aufnahme und Sammlung der in der Bundesrepublik erhobenen Messdaten zu den angesprochenen Verbindungen, als auch die Auswertung dieses Datenpools hinsichtlich Belastungsgrad von Kompartimenten, zeitlichen und räumlichen Trendaussagen, Aussagen zu Kompartimentübergängen, usw.. Diese Bewertungen sollen letztlich für Vorschläge zur Ableitung von Grenz- und Richtwerten, für die Ermittlung eines weiteren Datenbedarfs sowie zur Erfüllung von nationalen und internationalen Verpflichtungen zur Dokumentation des Zustandes der Umwelt herangezogen werden können.

Die Dioxin-Datenbank des Bundes und der Länder ist ein positives Beispiel für die Zusammenarbeit zwischen Bund und Ländern. Sie ist ein medienübergreifendes Instrument, in dem in Deutschland erhobene Daten aus der Umwelt sowie Lebensmittel- und Humandaten einschließlich der zu ihrer Bewertung notwendigen Metadaten dokumentiert sind.

Um die Daten aus diesen heterogenen Messprogrammen interpretieren und miteinander in Beziehung setzten zu können, ist neben der Speicherung der eigentlichen Messwerten die Aufnahme umfangreicher Zusatzinformationen wie Ort der Probenahme, Probenahmemethode, Analysemethode, Labordaten etc. notwendig. Durch die Erfassung von Rohdaten mit zusätzlichen Informationen können die Daten flexibel berechnet und mit durch andere Methoden erhobenen/berechneten Daten verglichen werden.

Die Datenbank enthält Expositionsdaten von etwa 220 kompartiment-spezifischen Messprogrammen zu über 12.000 Proben für die Umweltkompartimente Boden, Wasser, Luft, für den biotischen Bereich (Pflanze, Tier), für Abfall, Wertstoffe, Reststoffe, Zubereitungen und Erzeugnisse, für Haus- und Dachbodenstäube, für Futtermittel und Lebensmittel sowie für den Humanbereich. Darüber hinaus sind umfangreiche Informationen zu Probenahme, Analytik und Standortbeschreibung enthalten.

Grundlage für den Datenaustausch zwischen Bund und Ländern stellt die Verwaltungsvereinbarung zwischen Bund und Ländern über den Datenaustausch im Umweltbereich in der Fassung vom März 1996, mit dem Anhang II.3. „Austausch von

Daten zu polyhalogenierten Dibenzodioxinen und Dibenzofuranen sowie weitere chlororganische Stoffe“, dar.

Eine Datenlieferung ist zeit- und arbeitsaufwendig, da die Daten in der Regel in unterschiedlichen Formaten vorliegen und angepasst werden müssen. Um vorhandene Daten aufbereitet zur Verfügung zu stellen, fehlen oft die notwendigen Ressourcen.

Um den Datenaustausch zu vereinfachen und datenliefernden Stellen einfachen und geschützten Zugriff auf die von ihnen gelieferten Daten zu gewähren, hat das Umweltbundesamt in Kooperation mit dem Freistaat Bayern ein Projekt "www-POP-DioxinDB" initiiert. Der Lösungsansatz basiert auf reiner XML-Technologie. Durch dieses zukunftsweisende Konzept auf der Basis offener Standards wird, für verschiedene Nutzergruppen (Öffentlichkeit, datenliefernde Stellen und Fachnutzer) ein Zugriff auf die Dioxin-Datenbank über Web-Technologien ermöglicht. Diese webbasierte Anwendung wird im Herbst 2005 unter www.POP-DioxinDB.de im Internet erreichbar sein.

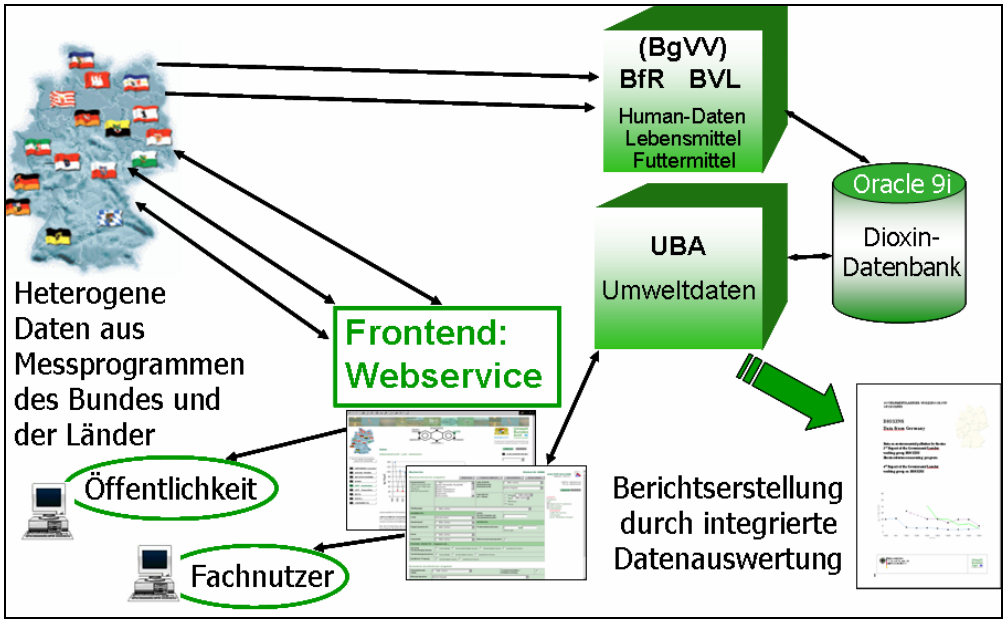


Abb.1: Schematische Darstellung des Datenflusses zwischen Bundes- und Länderbehörden durch die Einführung des Webservice

1.2. Technischer Hintergrund

Die Dioxin-Datenbank läuft seit 1998 als Client-Server-Anwendung im Wirkbetrieb. Der Datenserver läuft unter dem Betriebssystem UNIX mit dem Datenbanksystem Oracle Version 9. Die Client-Anwendung unter WINDOWS NT wurde auf Basis von

MS ACCESS 97 programmiert. Die Verbindung zwischen Client und Server erfolgt mit dem standardmäßig verfügbaren ODBC-Treiber von Oracle.

1.3. Ziel des Projektes

In dem Projekt www.POP-DioxinDB wurde eine webbasierte Anwendung mit XML-Technologie entwickelt. Damit wird:

- der im Anhang II 3 der Verwaltungsvereinbarung zum Datenaustausch Absatz (3) geforderte Zugriff der Länder „..auf alle von ihnen selbst übermittelten Daten und Informationen“ in Zusammenarbeit mit den Ländern und BMU realisiert
- für die Öffentlichkeit ein statisches Internetangebot zu ausgewählten Ergebnissen aus der Dioxin-Datenbank und allgemeinen Informationen zum Thema Dioxine geschaffen
- die Dioxin-Datenbank in das German Environment Information Network (GEIN) eingebunden, das die Vernetzung von Informationen zum Thema Dioxine mit anderen Anbietern unterstützt
- der direkte Zugriff auf die in der zentralen Dioxin-Datenbank im Umweltbundesamt vorliegenden Zustandsdaten für bestimmte Nutzergruppen erstellt

Schwerpunkt des Projektes ist es, den verschiedenen Nutzern abhängig von ihrer Rolle unterschiedliche Möglichkeiten des Datenzugriffs über das Internet anzubieten. Das System ist so vorbereitet, dass auch die Anforderungen aus dem novellierten Umweltinformationsgesetz abgedeckt werden können. Wir unterscheiden im Wesentlichen drei Zugriffspfade:

1. Internetangebot für die Öffentlichkeit ohne direkten Zugriff auf die Dioxin-Datenbank

Für die Öffentlichkeit wird ein Informationsangebot bereitgestellt, das keinen direkten Zugriff auf die Dioxin-Datenbank zulässt aber laufend aktualisiert wird.

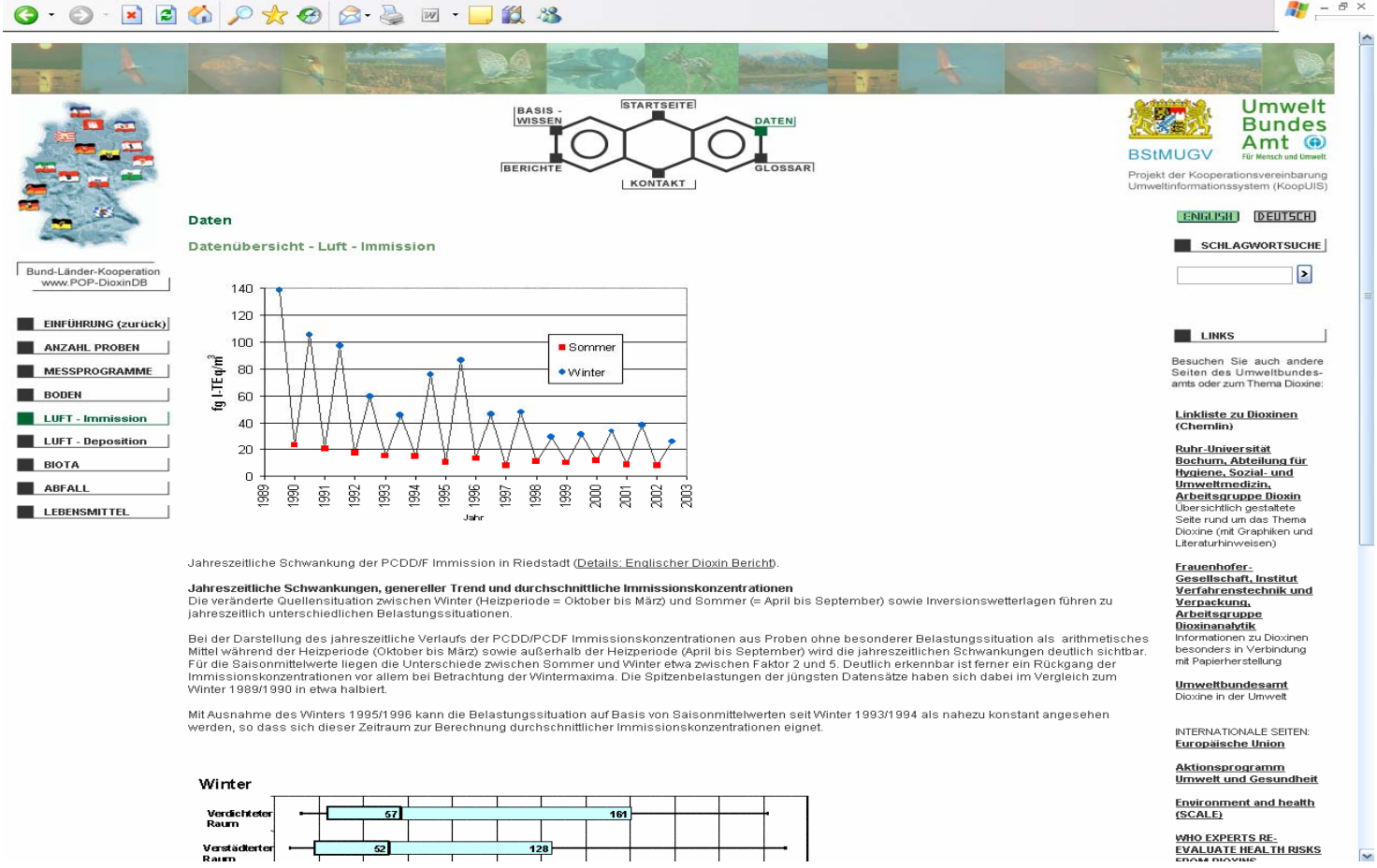


Abb.2: Öffentlicher Informationsbereich

Dieses Informationsangebot enthält folgende Inhalte:

- Allgemeine Anmerkungen zur DIOXIN-DB des Bundes und der Länder (z.B. Anlass, Auftrag, Zielstellung, Informationsfluss)
- Basiswissen zur Dioxin-Problematik (Einführung, Vorkommen, Entstehung, Toxizität, Bewertung),
- Einstellen von Berichten (Forschungsberichte und Veröffentlichungen des UBA und des ehemaligen BgVV, Berichte der B/L-AG Dioxine),
- Glossar – fachliche Erläuterungen und Erklärung von Abkürzungen
- Links zu Seiten im World Wide Web insbesondere zu Länderinformationen und zu weiterführenden Informationsangeboten über Dioxine
- Online Hilfe zum Navigieren und für kurze fachliche Erklärungen
- Informationen über den Zustand der Umwelt auf der Basis des 3. und 4. Berichtes der Bund/Länder-Arbeitsgruppe DIOXINE in Form von Tabellen, und

Diagrammen, mit erläuternden Texten. (z.B. „Immissionsdaten“ – „zeitliche Trends“).

2. Web- Service für die Datenlieferanten mit direkten selektiven Zugriff auf die selbst gelieferten Daten in der Dioxin-Datenbank

Die datenliefernden Stellen werden einen gesicherten Zugang zur Einsichtnahme in die von ihnen gelieferten Daten erhalten, z.B. für die Qualitätssicherung. Das Web-Angebot ist aus folgenden Komponenten aufgebaut:

- Zugriff auf die eigenen gelieferten Daten
- Zugriff auf die von der Datenbank daraus berechneten Daten

Dabei erfolgt der Datentransfer gesichert und das Herunterladen ist ohne zusätzliche Installationen einer Anwendung in den allgemein verfügbaren Formaten (z.B. HTML, Excel) möglich.

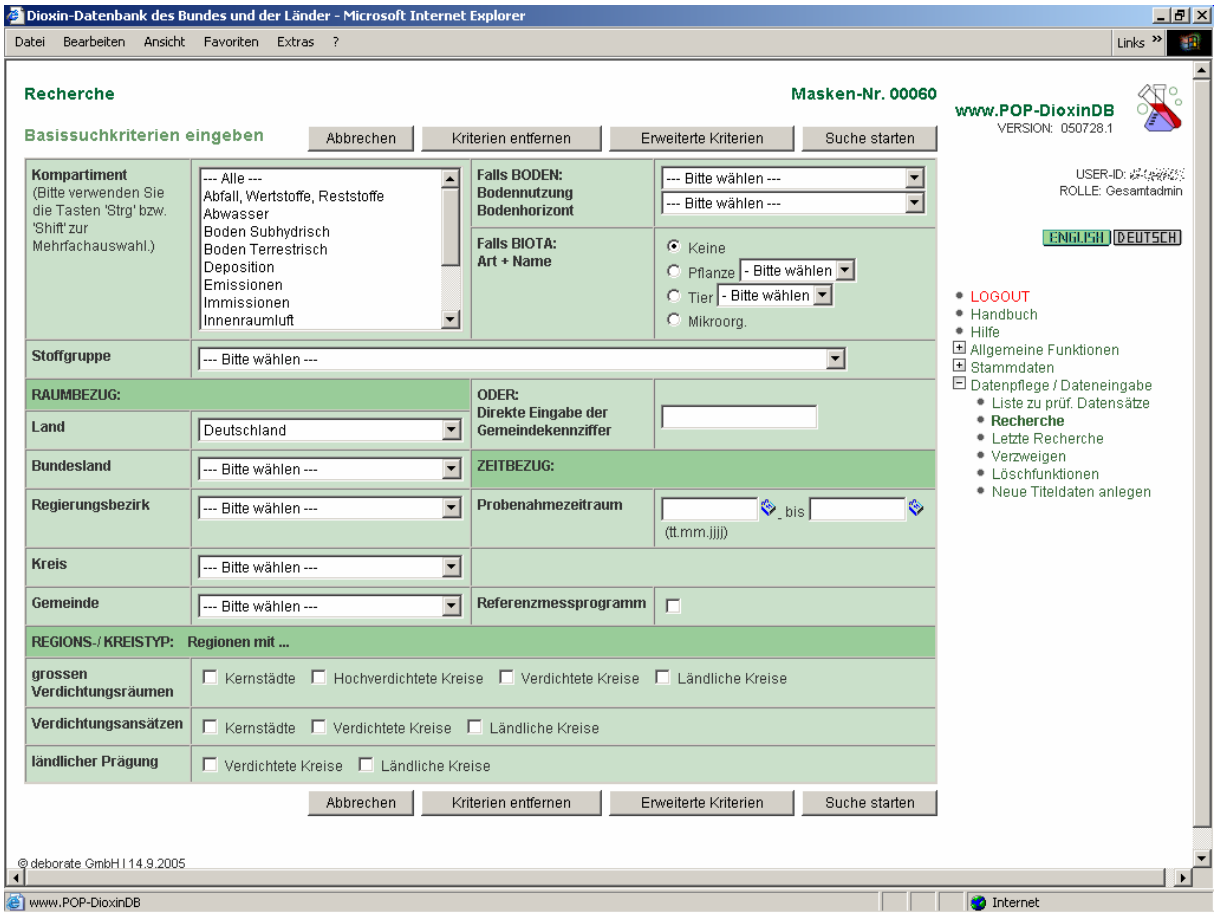


Abb.3: Recherchemaske für Fachnutzer und Datenlieferanten

3. Web-Service für Fachnutzer mit direkten partiellen Zugriff auf ausgesuchte und anonymisierte Daten der Dioxin Datenbank

Den Fachnutzern wird der gesicherte Zugriff auf Daten durch das UBA nach der Zustimmung der zuständigen Landesbehörden mittels Passwort erteilt. Das zu Grunde gelegte Rechtskonzept wird die erwarteten Anforderungen aus dem novellierten UIG erfüllen können. Das Web-Angebot ist aus folgenden Komponenten auf gebaut:

- Downloaden der Rechercheergebnisse
- Anonymisierte Darstellung der datengeschützten Informationen (z.B. Raumbezug, Institutionen, Produktnamen)
- Suchmasken, orientiert an den Recherchemasken der Access-Anwendung
- Ergebnispräsentation der Daten in übersichtlicher Form
- Link-Seite mit den e-Mail Adressen der Dioxin-Koordinatoren der Länder bzw. der Datenlieferanten für eine direkte Kontaktaufnahme und Antragsformular für die Zugriffsbewilligung

Die Schaffung der Möglichkeit dezentraler Dateneingabe und –import soll perspektivisch dazu führen, die Qualitätsmaßstäbe für einen Web-Service arbeitsteilig zu definieren. Die Fachverantwortung der Daten liegt bei den datenliefernden Stellen und die Administrationsverantwortung der Datenbank bei der Bundesbehörde.

1.4. Technische Lösung

Der Webservice ist konsequent auf Basis von XML konzipiert. Sowohl die Darstellung und Bearbeitung der Daten im Frontendbereich, die Datenübertragung zwischen den Servern, als auch die Verarbeitung im Backendbereich (Web-Services) erfolgt ausschließlich über die XML-Technologie. Alle Server-Schnittstellen basieren auf SOAP. Als Datenbanksystem wurde die bestehende ORACLE-Datenbank integriert. Die Anbindung des relationalen Datenbanksystems an die XML-Implementierung der Web-Services erfolgt über eine Mapping-Komponente. Dies kapselt die Datenbank gegenüber den Web-Services und sichert somit die Austauschbarkeit gegen eine andere relationale oder XML-Datenbank. Weiterhin

sind die Daten Datenbankunabhängig durch XML definiert, was ebenfalls einen System- und Versionswechsel vom Datenbankensystem vereinfacht.

In Kombination mit der Java Web Services Technologie konnte dieses zukunftsweisende Konzept auf der Basis offener Standards integriert werden. Der Lösungsansatz bietet gegenüber anderen webbasierenden Lösungen im Wesentlichen folgende Vorteile:·

- Systemunabhängigkeit
- Plattformunabhängigkeit
- Einfacher Datenaustausch
- Hohe Skalierbarkeit ·
- Trennung zwischen Anwendungslogik und der Benutzerschnittstelle (Masken)
- Vollständige Einbindung der Programmiersprache Java mit allen damit verfügbaren Funktionen und APIs

Im Einzelnen werden folgende OpenSource Komponenten eingesetzt:

- Webserver (Apache, Tomcat (Durchleitungsserver1, ServletEngine-Server2 + 3))
- Parser- und Stylesheet- Prozessoren (Xalan, Xerces)
- Datenkonvertierung zwischen relationaler Datenbank und XML-Dokumenten (Exolab CASTOR (Marshalling u. Unmarshalling))
- Servlet Engine (TOMCAT)
- Entwicklungsframework für XML-Anwendungen (StrutsCX – XML - Ergänzung des Frameworks Apache Struts zum Erstellen von Web- Applikationen mit Java nach dem MVC- Konzept)
- Excel-Export, Auswertungen und Statistiken (Jakarta POI für die Erzeugung von Exceldateien)
- Testverfahren, Betriebsüberwachung und Fehlermanagement (JMeter und Junit, CVS (Versionsmanagement, intern), Log4J (Applikationslogging)

Die Architektur der Anwendung berücksichtigt die Sicherheitsrichtlinien des UBA und ist im folgenden Schaubild dargestellt.

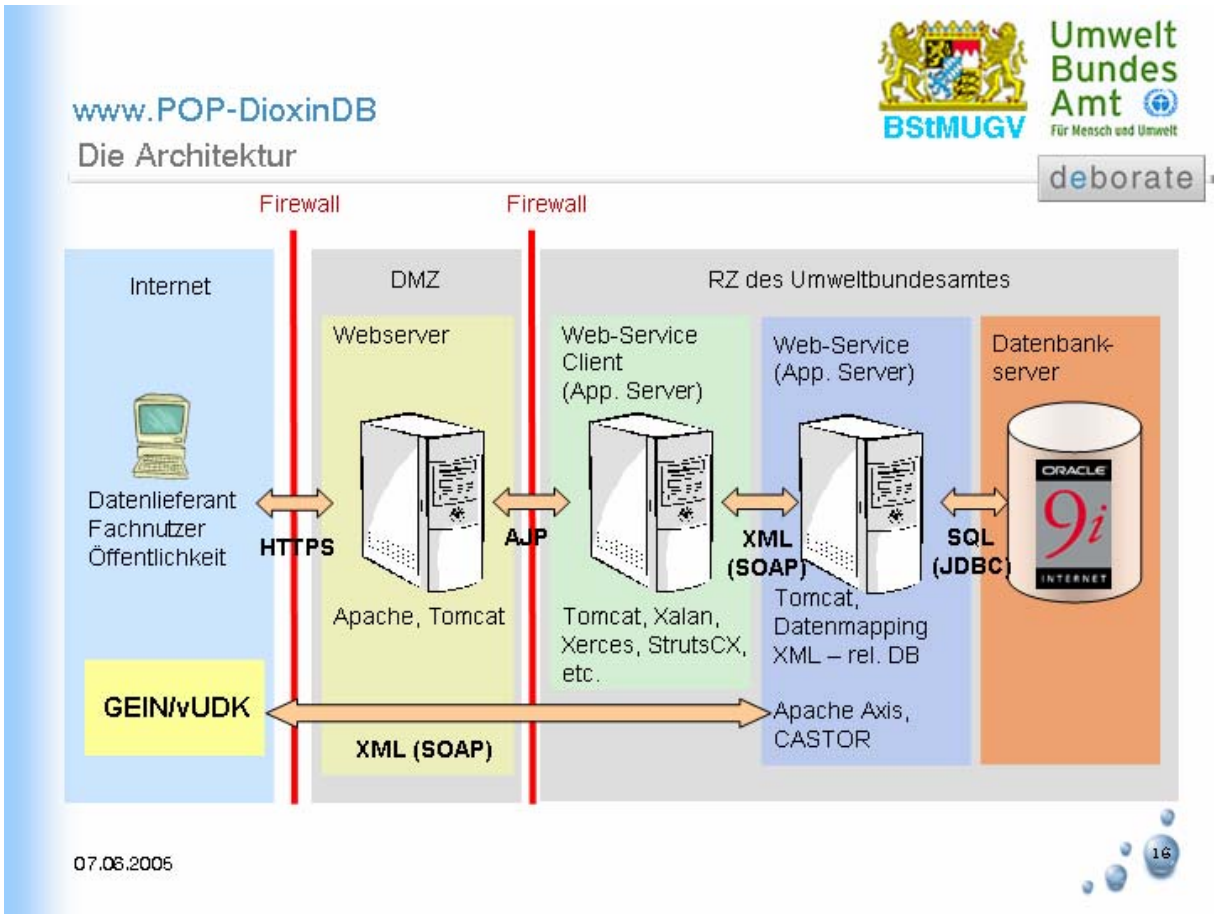


Abb. 4: Architektur zu www.POP-DioxinDB

Aus Sicherheitsgründen befindet sich keinerlei Anwendungslogik auf dem Webserver in der DMZ. Dies hat den Vorteil, dass es nur einen einzigen Installationsort der Anwendung gibt. Für spätere Erweiterungen ist die Anwendung nur auf einem Server anzupassen und für alle Anwender sofort verfügbar. Externe Systeme, wie gein® oder vUDK, erhalten über eine XML-Schnittstelle einen direkten Zugriff auf definierte Schnittstellen des Webservices.

Literaturverzeichnis

Deborate GmbH: DV-technisches Feinkonzept für die Umsetzung der Dioxin-Datenbank des Bundes und der Länder. (unveröffentlicht), 2004.

Knetsch, G., Weihs., Rappolder, M.,: Entwicklung eines Web-Services für die Dioxin-Datenbank des Bundes und der Länder. Umweltdatenbanken – Nutzung von Metadaten und Standards UBA-Texte 54/03, pp. 342-392, 2003.

Vom Luftbild zur FFH-Kartierung: Kartierung der terrestrischen Bereiche des niedersächsischen Nationalpark Wattenmeer mit dem Freien Geoinformationssystem GRASS GIS

Manfred Redslob, GDF Hannover, redslob@gdf-hannover.de
Jörg Petersen, nature-consult, petersen@nature-consult.de
Otto Dassau, GDF Hannover, dassau@gdf-hannover.de
Hans-Peter Dauck, nature-consult, dauck@nature-consult.de

1 Abstract / Einleitung

Auf Grundlage der FFH-Richtlinie sind die Mitgliedsstaaten verpflichtet, Arten und Lebensräume gemeinschaftlichen Interesses im Rahmen des NATURA 2000-Netzes auszuweisen und für diese einen günstigen Erhaltungszustand zu sichern und nachzuweisen. Für diesen Nachweis wurde 2004 eine flächendeckende Kartierung der terrestrischen Gebiete des Nationalparks Niedersächsisches Wattenmeer durchgeführt. Ein wichtiger Bestandteil der Vegetationserfassung waren die Analysemethoden der Fernerkundung, die im Wesentlichen mit dem freien GIS GRASS durchgeführt worden sind. Für die Klassifizierung wurde ein hierarchischer Ansatz gewählt. Im Vorfeld der Klassifizierung sind die nicht zu klassifizierenden Bereiche ausmaskiert und die verbleibenden Flächen in vegetationskundlich definierte Gruppen (Serien) aufgeteilt worden. Nach der Festlegung von Trainingsflächen erfolgte schließlich die eigentliche Klassifizierung. Hierbei wurde der SMAP-Algorithmus eingesetzt, da dieser auch Nachbarschaftsbeziehungen zwischen den Pixeln in die Klassenzuordnung einfließen lässt. Die Ergebnisse wurden zur Eliminierung der Kleinstflächen durch empirisch ermittelte Filterabläufe geglättet und abschließend vegetationskundlich überarbeitet. Auf der Grundlage der so entstandenen Feldkarten konnte der Nationalpark (ca. 25.500 ha) in nur drei Monaten Geländearbeit flächendeckend kartiert werden.

2 Vorbemerkung

Am 21. Mai 1992 erließ die Europäische Union - damals noch die Europäische Wirtschaftsgemeinschaft – die Fauna-Flora-Habitat-Richtlinie (RL 92/43/EWG) (BfN 1998). Die FFH-Richtlinie verfolgt das Ziel, die biologische Vielfalt auf dem Gebiet der Europäischen Union durch ein nach einheitlichen Kriterien ausgewiesenes Schutzgebietssystem (NATURA 2000) nachhaltig zu schützen und zu erhalten. Alle Mitgliedsstaaten sind durch die Richtlinie verbindlich verpflichtet, die im Anhang der Richtlinie aufgelisteten Lebensräume gemeinschaftlichen Interesses in den alpinen, kontinentalen und atlantischen biogeographischen Regionen ihres Hoheitsgebietes als Basis eines europaweiten Schutzgebietsnetzes zu melden und für diese nachhaltig einen „günstigen Erhaltungszustand“ zu gewährleisten.

Im Juni 1995 verstreicht die erste Frist zur Übermittlung der nationalen Gebietslisten an die EU Kommission, ohne dass Deutschland ein einziges Gebiet gemeldet hat. Wegen nicht ausreichender Meldungen wird Deutschland schließlich 1998/99 von der EU Kommission vor dem Europäischen Gerichtshof verklagt und am 11. September 2001 von diesem verurteilt. Das in der Folge im April 2003 eröffnete Zwangsgeldverfahren wird jedoch unter der Bedingung, dass Deutschland das Meldeverfahren mit dem Ziel forciert, es im Januar 2005 abzuschließen, nicht aktiv weiter getrieben. Daraufhin werden von Deutschland 4588 Gebiete, das entspricht im terrestrischen Bereich ca. 9,3% der Landesfläche, in Brüssel vorgelegt (Stand 28. Januar 2005) [1].

Zur Absicherung eines nachhaltig „günstigen Erhaltungszustandes“ verpflichtet die FFH-Richtlinie in Artikel 11 die Mitgliedsstaaten zu einer allgemeinen Überwachung der Arten (vgl. Anhang II, IV und V FFH-RL) und Lebensräume (vgl. Anhang I FFH-RL) gemeinschaftlichen Interesses. Dieses Monitoring ist laufend durchzuführen und unterliegt der „Berichtspflicht“: Im Rahmen der FFH-Richtlinie sind die Mitgliedsstaaten verpflichtet, alle sechs Jahre über den Zustand der in ihrem Zuständigkeitsbereich liegenden Teile des NATURA 2000-Netzes liegenden Bestandteile Bericht (vgl. Artikel 17 FFH-RL) zu erstatten. Hierbei ist sicher zu

stellen, dass die erhobenen Daten vergleichbar sind und eine ähnliche Struktur aufweisen. Die EU Kommission erstellt auf der Grundlage der nationalen Berichte einen zusammenfassenden Bericht, der zur Erfolgskontrolle unter anderem eine Bewertung des Erhaltungszustandes der entsprechenden Gebiete beinhaltet.

3 Anforderungen an das Projekt

Vor dem Hintergrund der bestehenden „Berichtspflicht“ hat das Büro „nature-consult“, Hildesheim [2] in Zusammenarbeit mit der „GDF Hannover“ [3] 2004 im Auftrag der Nationalparkverwaltung Niedersächsisches Wattenmeer [4] eine flächendeckende Kartierung der terrestrischen Bereiche des Nationalparks Niedersächsisches Wattenmeer durchgeführt. Ziel des Projektes war vor allem die Erfassung der FFH-Lebensraumtypen in ihrem aktuellen Erhaltungszustand, um der FFH-Berichtspflicht genüge zu tun. Hierbei ist aus Gründen der verbindlich vorgegebenen europaweiten Vergleichbarkeit der Daten die Erfassung der in der FFH-Richtlinie definierten FFH-Lebensraumtypen vorgegeben. Weiterhin wurden auf der Grundlage der Ergebnisse der TMAP-Arbeitsgruppen Salzwiesen und Dünen ergänzend TMAP-Vegetationseinheiten (**T**rilateral **M**onitoring and **A**ssessment **P**rogram, vgl. Bakker et al., 2005, Petersen & Lammerts 2005) und Biotoptypen (vgl. Drachenfels 2004) erfasst. Insgesamt umfasste die vegetationskundlich zu kartierende Fläche ca. 25.500 ha.

Nicht zuletzt das Ausmaß der zu kartierenden Gesamtfläche legte die Überlegung nahe, die Vegetationskartierung durch Methoden der Fernerkundung zu unterstützen. Um dies zu erreichen, ist eine interdiziplinäre Entwicklung der einzusetzenden Verfahren in einer frühen Phase des Projektes zwingend erforderlich. Das Projekt basiert folglich auf dem Ansatz einer integralen Entwicklung der einzusetzenden Bildverarbeitungsstrategie (s. Redslob, 1999), bei der das Fachwissen der Vegetationsspezialisten mit dem Fachwissen der Fernerkundungsspezialisten frühzeitig zusammengeführt wird. Auf diese Weise können die fachinhaltlichen Ansprüche an die Kartierung (Erfassung der FFH-Lebensraumtypen, TMAP-Typen und Biotoptypen) mit dem Informationsgehalt der verfügbaren Fernerkundungsdaten (geometrische und spektrale Auflösung,

Klassifizierungsalgorithmus) abgeglichen werden, um die optimale Klassifizierungsstrategie zu entwickeln. Auf der Grundlage von Testklassifizierungen und den zu erwartenden Ergebnissen erfolgt schließlich die Gesamtplanung des Projektverlaufes, in der festgelegt wird, welche Arbeitsschritte für die Bildvorverarbeitung, die Klassifizierung, die vegetationskundliche Nachbearbeitung, die Verifizierung bzw. Kartierung im Gelände, sowie für die Aufbereitung der Erfassung als Vektordatensatz notwendig werden. Für die Verarbeitung der Bild- und Vektordaten wurde je nach Teilaufgabe und Programmeignung sowohl freie, als auch proprietäre GIS-Software eingesetzt. Insofern ist eine hohe Interoperabilität zwischen den eingesetzten Geoinformationssystem eine grundlegende Voraussetzung für die Bearbeitung.

Insgesamt konnte in dem im Folgenden noch näher beschriebenen Pilotprojekt durch die im gesamten Projektverlauf enge interdisziplinäre Zusammenarbeit zwischen den beteiligten Vegetations-, GIS- und Fernerkundungsspezialisten eine zukunftsorientierte Methodik getestet und optimiert werden, bei der Fernerkundungsmethoden sehr effizient als unterstützendes Element zur großflächigen Vegetationserfassung eingesetzt wurden. Vor dem Hintergrund der bestehenden FFH-Berichtspflicht über den Erhaltungszustand der NATURA 2000-Flächen kommt dem Einsatz, aber auch der gezielten Weiterentwicklung von modernen Methodik eine hohe Bedeutung zu, um kostengünstige, aber qualitativ hochwertige Vegetationserfassungen durchführen zu können.

4 GRASS GIS

Die GDF Hannover setzt bei ihrer Arbeit konsequent auf den Einsatz Freier GIS-Lösungen. Eine Definition Freier Software sowie eine Beschreibung der Lizenzmodelle erfolgt u.a. bei Grassmuck 2002 [5]. Grundsätzlich bietet Freie Software dem Nutzer gegenüber der Verwendung proprietärer Produkte zahlreiche Vorteile:

- Es besteht kein Kaufzwang für Updates.
- Es besteht keine Hersteller- oder Händlerbindung.
- Wartungsverträge und Support können frei gestaltet werden.
- Lizenzgebühren fallen nicht an.
- Es besteht eine hohe Interoperabilität zu leistungsstarker freier und proprietärer Software.
- Es besteht der volle Einblick, insbesondere bei sicherheitsrelevanten Systemen.

Bei der Bearbeitung des Projektes wird GRASS GIS (**G**eographic **R**esources **A**nalysis **S**upport **S**ystem) (Neteler & Mitasova, 2004) [6] und weitere freie GIS Applikationen eingesetzt, da GRASS als kraftvolles GIS-Werkzeug alle Anforderungen an die Methoden der modernen Fernerkundung erfüllt. Hinsichtlich der Leistungsfähigkeit braucht GRASS GIS einen unmittelbaren Vergleich mit proprietären Produkten zur Bildverarbeitung, wie beispielsweise ERDAS, nicht zu scheuen. GRASS GIS stellt also – im Gegensatz zu anderen GIS-Produkten, die diesen Funktionsbereich bestenfalls über zusätzlich zu erwerbende Applikationen realisieren – von vorne herein sehr umfassende Tools zur Bildverarbeitung zur Verfügung. Die hohe Interoperabilität und der modulare Aufbau von GRASS GIS erlaubt es zudem, weitere Softwareprodukte in den notwendigen Arbeitsablauf zu integrieren. Darüber hinaus ermöglichen die zahlreichen vom Programm unterstützten GIS-Formate eine hohe Flexibilität beim Datenaustausch mit proprietärer Software und somit auch bei der Kooperation zwischen den beteiligten Projektpartnern sowie dem Auftraggeber.

Die Entwicklung von GRASS GIS begann 1982 durch das **C**onstruction **E**ngeneering **R**esearch **L**ab (CERL) der US Army in Champaign, Illinois. Seither unterliegt die Software, anfänglich unter Beteiligung amerikanischer Behörden und dann durch Universitäten, Forschungsinstitute und kommerzielle Einrichtungen einer ständigen Weiterentwicklung. Mitte der 1990er Jahre wird diese bei der US Army eingestellt.

Mit der Einstellung wird der Source Code von GRASS GIS nach amerikanischem Recht mit weltweiter Wirkung offen gelegt. Seit 1998 wird die nunmehr weltweit mögliche Weiterentwicklung durch den Mitgesellschafter der GDF Hannover, Herrn Neteler, zunächst an der Universität Hannover und heute am ITC-irst [6] in Trento, Italien, koordiniert. Im Jahr 1999 wird GRASS GIS (Version 5.0) unter die am weitesten reichende Lizensierung für Freie Software, die GNU GPL (General Public Lizense) gestellt. Es wird geschätzt, dass in die Entwicklung von GRASS GIS seit der Version 4.1 (1993) mehrere Millionen Dollar investiert worden sind. Als stabile Version steht GRASS GIS in Kürze als Release 6.0.1 zur Verfügung.

GRASS GIS ist unter zahlreichen Betriebssystemen (GNU/Linux, MS-Windows-Versionen, Mac OSX ...) und auch auf Handhelds (HP/iPAQ, Sharp/Zaurus) einsetzbar. Dem Anwender stehen mit über 300 Modulen alle notwendigen Funktionalitäten zur Eingabe, Verwaltung, Analyse und Präsentation räumlicher Daten zur Verfügung. Durch Strukturierung der GRASS-Datenbasis in sogenannte „Locations“ und „Mapsets“ sowie eine entsprechende Vergabe von Lese- und Schreibrechten kann GRASS uneingeschränkt im Multiuser-Betrieb eingesetzt werden. Somit ist es möglich, von mehreren Arbeitsplätzen aus gleichzeitig eine Datenbasis abzufragen und zu bearbeiten. Die modulare Struktur des Systems erlaubt hierbei einen überaus flexiblen Einsatz der jeweils benötigten Funktionalitäten.

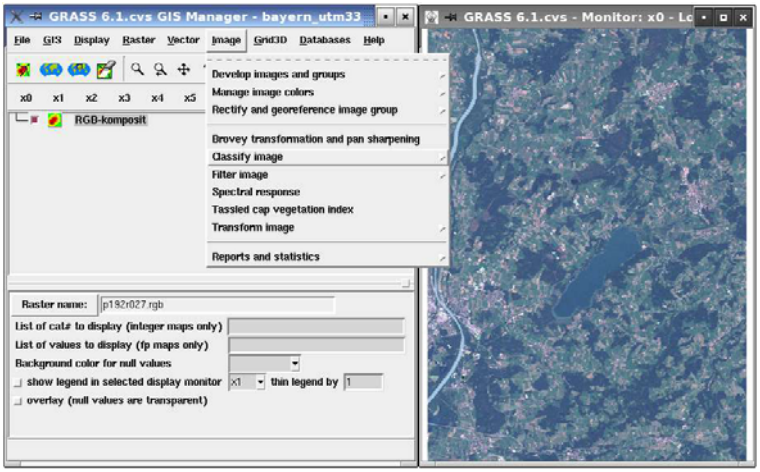


Abb.1: Tcl/TK basierter GIS-Managger

Die Ansprache der Module kann wahlweise über Kommandozeilen oder graphisch mit dem von GRASS bereitgestellten GIS-Manager erfolgen. Als besonderes Feature von GRASS GIS ist das 3D-Visualisierungstool NVIZ zu nennen, mit dem Geländemodelle interaktiv betrachtet und Animationen erstellt werden können. Neben diesem GRASS eigenen Graphical User Interface (Abb.1) existieren weitere Freie Softwareprojekte, die wie beispielsweise Quantum GIS [7] untereinander unmittelbar kommunizieren und die GIS-Funktionalitäten von GRASS in eine neue graphische Oberfläche einbinden (Abb. 2). Quantum GIS unterliegt derzeit einer rasante Weiterentwicklung und wird sich mit wachsendem Funktionsumfang künftig zu der Benutzeroberfläche für GRASS GIS entwickeln. GRASS GIS dürfte so einen starken Verbreitungsschub erhalten, da sich die Quantum-Oberfläche eng am „look and feel“ der propietären Produkte des GIS-Marktführers ESRI orientiert. Umsteiger können sich so problemlos den Leistungsumfang von GRASS erschließen, ohne sich mit einer völlig neuen Benutzeroberfläche auseinander setzen zu müssen. Insbesondere der Kartengenerator von Quantum GIS schließt eine bislang bestehende Lücke im Leistungsspektrum Freier GIS-Software.

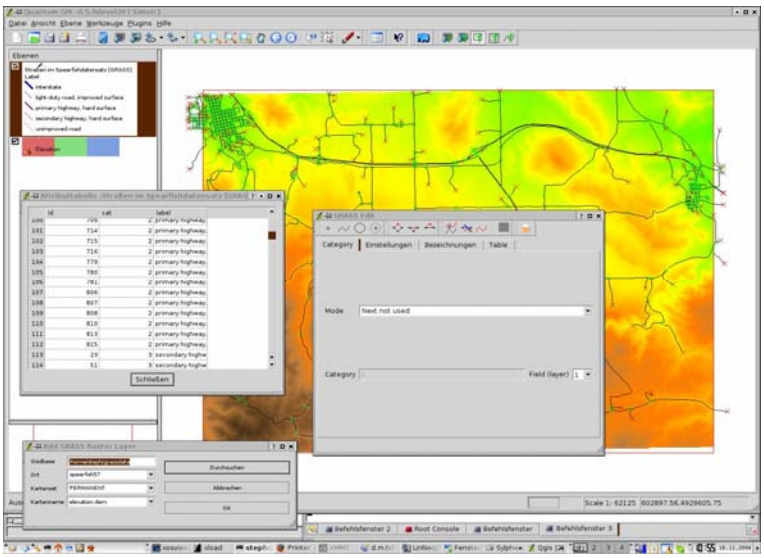


Abb. 2: Quantum GIS mit GRASS Support

Hinsichtlich des Raumbezuges werden von GRASS GIS durch die Nutzung der Projektionsbibliothek PROJ4 [8] über 120 Projektionen unterstützt. Als Geometriedaten können 2D-Rasterdaten, 3D-Voxeldaten, 2- und 3D-Vektordaten sowie multidimensionale Punktdaten eingesetzt werden. Die aktuelle GRASS Version beinhaltet eine völlig neue, topologische, 2D/3D-Vektorarchitektur (vgl. Abb. 3) und unterstützt Netzwerkanalysen. Die Vektorgeometrie kann wahlweise im nativen Format vorgehalten werden, oder über die eingebundene OGR-Bibliothek [9] in alle gängigen GIS-Formate geschrieben werden. Durch OGR ist gleichfalls die Möglichkeit zum Einbinden der räumlichen Datenbank PostGIS gegeben. Dort werden sowohl Geometrien, als auch Attribute in der relationalen Datenbank PostgreSQL [10] abgespeichert. Als Speicherformat wird das durch das OGC (**O**pen **G**IS **C**onsortium) spezifizierte WKT-Format (**W**ell-**K**nown-**T**ext) verwendet. Über ODBC (**O**pen **D**ata**b**ase **C**onnectivity) können durch die DBMS-Bibliothek (**D**ata**b**ase **M**anagment **S**ystem) auch andere proprietäre Datenbanken angesprochen werden.

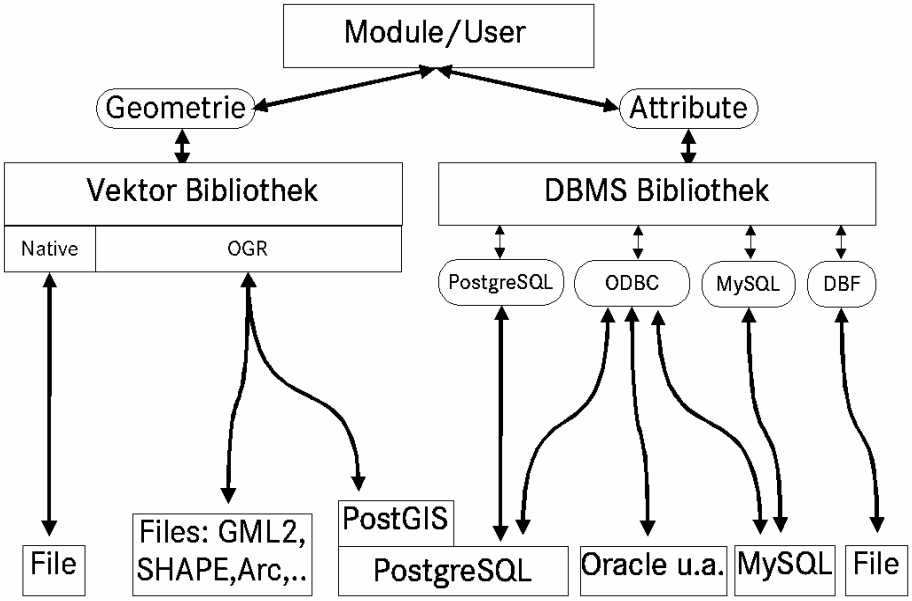


Abb. 3: Vektordatenmanagment in GRASS GIS

In dem hier beschriebenen Projekt ist jedoch vor allem der Leistungsumfang von GRASS GIS hinsichtlich der Verarbeitung und Analyse von Fernerkundungsdaten, also 2D-Rasterdaten, von Bedeutung. Durch die Einbindung der GDAL-Bibliothek [9] ist es bei den Rasterdaten ebenfalls möglich, zahlreiche Formate zu lesen und zu

schreiben. Tabelle 1 zeigt zusammenfassend die wesentlichen Freien Sotwarekomponenten, die bei der Projektbearbeitung zum Einsatz kamen:

Tabelle 1: Im Projekt eingesetzte Freie Softwarekomponenten

Software	***Verwendung***
GRASS GIS	Geoinformationssystem - Verwaltung der räumlichen Daten - Analyse und Modellierung - Bildverarbeitung - Visualisierung und Kartenerstellung
Quantum GIS	Geoinformationssystem / Datenviewer - Graphische Benutzeroberfläche
GDAL	Graphikbibliothek - Im- und Export von Rasterdaten
OGR	Graphikbibliothek - Im- und Export von Vektordaten
PROJ4	Projektionsbibliothek - Lagegenaue und deckungsgleiche Projektion unterschied-licher Datensätze
PostgreSQL	Relationale Datenbank - Attributverwaltung

5 Kartierung der terrestrischen Bereiche des niedersächsischen Nationalpark Wattenmeer

Die Kartierung der gesamten terrestrischen Bereiche des Nationalparks Niedersächsisches Wattenmeer erfolgte auf der Basis digitaler Luftbilder. Im Vorfeld der eigentlichen Geländearbeit wurden diese Daten klassifiziert und fachlich bearbeitet. Damit konnte den Kartierern eine zuvor generierte Arbeitskarte mit den Klassifizierungsergebnissen an die Hand gegeben werden, in der die Geometrien und soweit möglich auch die Attributierung der zu erfassenden Vegetationseinheiten dargestellt wurde. Die Arbeit vor Ort konnte so gegenüber konventionellen Geländeerhebungen deutlich effektiviert werden, da keine Kompletterfassung (Abgrenzung und inhaltliche Zuordnung aller Flächen) notwendig wurde. Der

Schwerpunkt der Geländearbeit lag somit auf der Verifizierung und ggf. Anpassung der klassifizierten Geometrien und Attribute. Die gewählte Methodik besitzt eine hohe Effektivität. Dadurch und durch die Vergleichbarkeit der bereitgestellten Ergebnisse ist sie prädestiniert für Monitoringprogramme. Die Vorteile dieser Methodik entsprechen somit der FFH-Verpflichtung zum Flächenmonitoring und werden im Grundsatz auch in anderen Bundesländern eingesetzt (vgl. DÜVEL & FRICK, 2005).

5.1 Daten

Grundlage der Klassifizierung waren digitale Luftbilder der Kameras HRSC-AX (**H**igh **R**esolution **S**tereo **C**amera – **A**irborne E**x**tended) sowie DMC (**D**igital **M**odular **C**amera) aus den Jahren 2002 und 2003. Die HRSC-AX-Daten decken hierbei die Wurster und Butjadinger Küste sowie die Inseln ab, die DMC-Daten die restlichen Küstengebiete des Nationalparks. Die technischen Parameter der beiden Kamerasysteme können Tab. 2 entnommen werden.

Tabelle 2: Technische Spezifikation der eingesetzten Sensoren

Parameter	***HRSC-AX***	***DMC***
Geometrische Auflösung	32 cm	24 cm
Lagegenauigkeit	+/- 32 cm	+/- 100 cm
Radiometrische Auflösung	unsigned 16 bit	16 bit
	Blau: 440 – 510 nm	Blau: 400 – 580 nm
	Grün: 520 – 590 nm	Grün: 500 – 650 nm
	Rot: 620 – 680 nm	Rot: 590 – 675 nm
	Infrarot: 780 – 850 nm	Infrarot: 675 – 850 nm
	Panchromatisch: 520 – 760 nm	Panchromatisch: 400 – 850 nm
Datenformat	GeoTIFF	ERDAS/Img

5.2 Klassifizierungsmethode

Basierend auf den Erfahrungen aus anderen, bereits durchgeführten Projekten wurde ein hierarchischer Klassifizierungsansatz (vgl. REDSLOB, 1999) gewählt. Ziel dieses Ansatzes ist es, zum einen nicht zu klassifizierende Bereiche aus der Analyse vorab auszuschließen, und zum anderen die zu klassifizierenden Bereiche möglichst

im Vorfeld in eindeutig definierte Teilbereiche zu differenzieren. Damit reduzieren sich die in den einzelnen Serien zu klassifizierenden (Vegetations-)Einheiten. Als Resultat ist so eine deutlich verbesserte Trennung der Klassen und Zuweisung der Attribute zu erwarten. Die Hauptarbeitsschritte des gewählten Ansatzes gliedern sich wie folgt:

- Ausmaskierung nicht zu klassifizierender Bereiche (Siedlungen, Priele etc.),
- Differenzierung der zu klassifizierenden Flächen durch inhaltlich definierte Serienbildung,
- Festlegung der Trainingsflächen,
- Klassifizierung,
- Filterung der Ergebnisse und
- digitale vegetationskundliche Nachbearbeitung.

5.2.1 Ausmaskierung

Vor Beginn der Klassifizierung wurden zunächst alle außerhalb des Nationalparks liegenden Bereiche ausmaskiert. Unter Nutzung von u.a. ALK-Datensätzen sind zudem innerhalb der Nationalparkgrenzen liegende Verkehrswege, Parkplätze, Flugplätze, Ortslagen und Einzelhöfe sowie Flüsse, Priele und Stillgewässer ebenfalls ausmaskiert worden (vgl. Abb. 4).

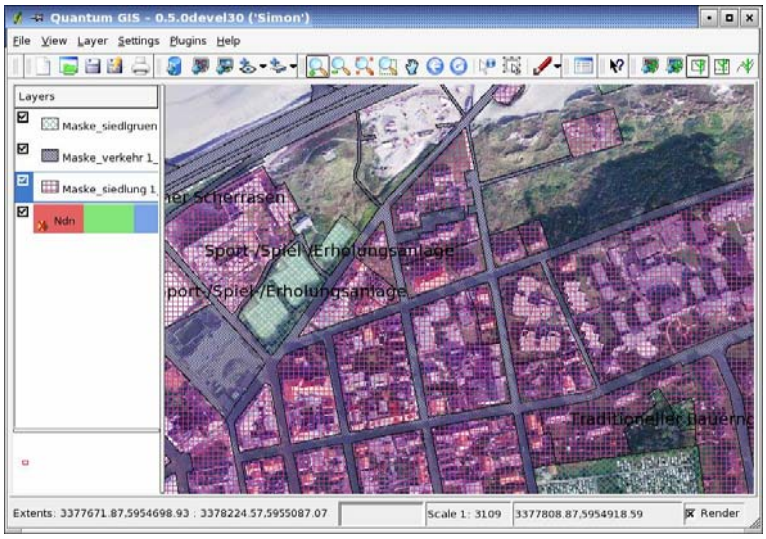


Abb. 4: Ausmaskierung nicht zu klassifizierender Bereiche

Alle weiteren Klassifizierungen beziehen sich somit ausschließlich auf die verbleibenden Flächen. Fehlklassifizierungen können durch diese vorbereitenden Arbeitsschritte minimiert werden.

5.2.2 Differenzierung durch Serienbildung

Zur weiteren Optimierung der Klassifikation erfolgte eine Gruppierung der zu differenzierenden Vegetationseinheiten zu so genannten Serien (vgl. Abb. 5). Dabei wurden die Einheiten der Dünen- und Dünentäler, Salzwiesen sowie des Grünlandes durch die Vegetationsökologen vorab voneinander getrennt, um diese dann jeweils einzeln klassifizieren zu können.

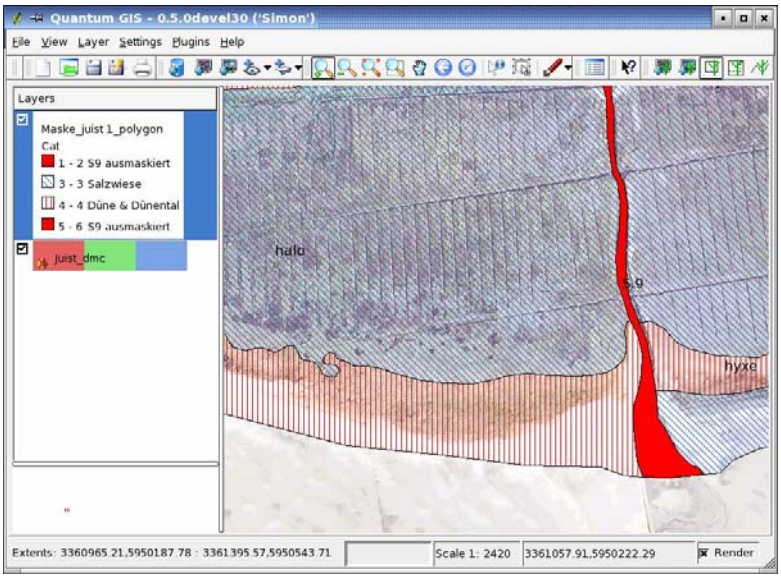


Abb. 5: Trennung der vegetationskundlich definierten Serien

5.2.3 Festlegung von Trainingsflächen und Resampling

Für die Güte der Klassifizierung ist die Auswahl der Trainingsgebiete von entscheidender Bedeutung. Die Trainingsgebiete sollen idealerweise die spektralen Eigenschaften der Vegetationseinheiten als homogene Stichproben möglichst repräsentativ widerspiegeln. Auf Grundlage von hochauflösenden RGB-Komposits wurden diese Flächen von den ortskundigen Vegetationsspezialisten ausgewählt und digitalisiert. Als Probelauf wurden zahlreiche Testklassifizierungen durchgeführt, deren Ergebnisse wieder in den Prozess der Auswahl von Trainingsflächen eingeflossen sind. Auf diese Weise erfolgte ein iterativer Lernprozess bei den Bearbeitern, der zu einer Optimierung der Auswahl von Trainingsgebieten führt.

Darüber hinaus wurden die Testklassifizierungen genutzt, um zu ermitteln, inwieweit die hohe Bodenauflösung der eingesetzten Daten verringert werden kann, ohne dass es zu relevanten Informationsverlusten bei den Ergebnissen kommt. Im Ergebnis zeigte sich, dass bei einem Resampling auf die jeweils doppelte Auflösung der Daten nahezu identische Klassifizierungsergebnisse wie mit den Originalauflösungen der Daten erzielt werden konnte. Hiermit war es möglich, die erforderliche Rechenzeit für die Klassifizierungen auf ein Viertel zu reduzieren.

5.2.4 Gesteuerte Klassifizierung

Für die eigentliche Klassifizierung mit GRASS GIS wurde das GRASS-Modul „i.smap“ eingesetzt. Dieses Modul besteht im Wesentlichen aus dem multispektralen Segmentierungsalgorithmus „Sequential Maximum a Posteriori“ und greift auf das „Multiscale Random Field“ Modell (MSRF) zurück (vgl. BOUMAN & SHAPIRO, 1996, JUNEK, 2004 u.a.). Der gewählte Klassifizierungsalgorithmus arbeitet im Gegensatz zu den meisten anderen Agorithmen nicht pixelweise, sondern bezieht als Pyramidenklassifizierung auch Nachbarschaftsbeziehungen in die Klassifizierung mit ein. In einem ersten Schritt werden auf der Basis der Trainingsgebiete Signaturen erstellt. Durch das Prinzip des Multiscale Random Field Model werden den Pixeln mit dem auf unterschiedlichen Bildauflösungen arbeitenden Algorithmus (vgl. Abb. 6) statistische Parameter zugewiesen.

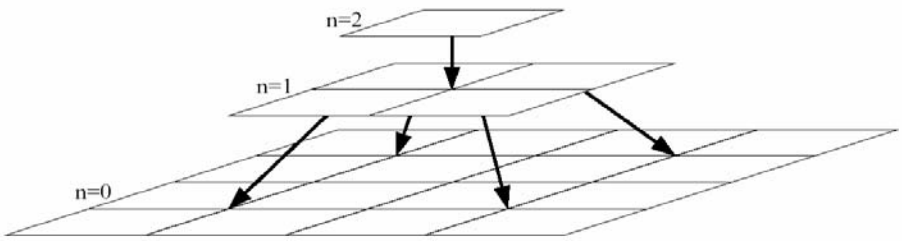


Abb. 6: Pyramidenstruktur des Multi Scale Random Field Model (BOUMAN & SHAPIRO, 1996)

Diese Zuweisung mündet in der Bildung von Unterklassen, denen die Pixel zugewiesen werden. Die Zugehörigkeit zu einer über die Trainingssignaturen vordefinierten Klasse erfolgt schließlich durch das parameterfreie Verfahren SMAP. Hier wird den vorab ermittelten Unterklassen entsprechend ihrer Gewichtung eine a-priori-Wahrscheinlichkeit zur Zugehörigkeit zu einer Hauptklasse zugewiesen. Der beschriebene Ansatz führt zu einer deutlichen Verbesserung der Klassifizierungsergebnisse gegenüber rein pixel-orientierten Algorithmen.

5.2.5 Nachbearbeitung der Klassifizierungsergebnisse

Die Klassifizierungsergebnisse müssen nicht zuletzt aufgrund der hohen Bodenauflösung der eingesetzten Daten und der naturgemäß auftretenden Fehlklassifizierungen durch Mischpixel im Randbereich unterschiedlicher Flächenzustände, vor dem Einsatz im Gelände abschließend gefiltert werden. Kleinstflächen wurden hierbei rasterseitig mit GRASS GIS unter Einsatz des Nearest Neighbour Algorithmus durch „Moving Windows“ eliminiert. Insgesamt wurde ein neun-stufiges Filterverfahren eingesetzt, das in enger Zusammenarbeit der Vegetationskunde mit der Fernerkundung empirisch entwickelt worden ist.

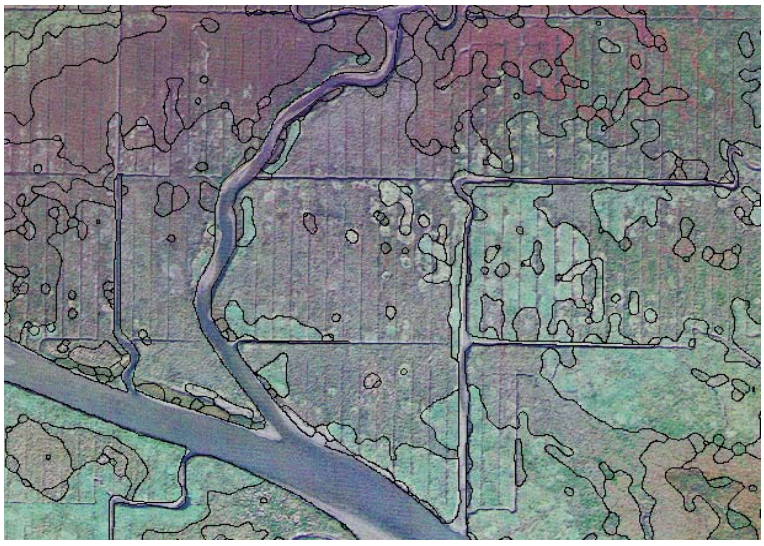


Abb. 7: Ergebnis der überwachten SMAP-Klassifizierung (Salzwiesen)

Nach diesem Filterprozess wurden die einzelnen, bis hier hin getrennt verarbeiteten Serien, wieder zusammengefügt und ins Vektorformat überführt. Abb. 7 zeigt einen Ausschnitt der Ergebnisse. Der Export der Ergebnisse erfolgte im ESRI-Shapeformat. Diese Vektoren wurden abschließend vektorbasiert mit ArcGIS erneut gefiltert. Damit wurden die klassifizierten Geometrien auf eine für die Feldarbeit sinnvolle Mindestflächengröße von 200 m^2 optimiert. Der letzte Arbeitsschritt zur Fertigstellung der Feldkarten umfasste die notwendige digitale vegetationskundliche Überarbeitung.

5.3 Geländeerhebung

Die so erstellten Erhebungskarten bildeten die entscheidende Arbeitsgrundlage der Feldarbeit bzw. Verifizierung/Kartierung. Die Geländearbeit konnte auf der nun vorliegenden Basis ausgesprochen effizient durchgeführt werden und führte zu Vegetationskarten (vgl. Abb. 8) im Maßstab 1:3.000. Die Insel Borkum mit einer Gesamtfläche von ca. 3.830 ha wurde beispielsweise von vier Kartierern in nur vier Tagen vollständig erfasst. Für diese flächendeckende Bearbeitung wurden 126 analoge Geländekarten (DIN A4) angefertigt und bearbeitet.

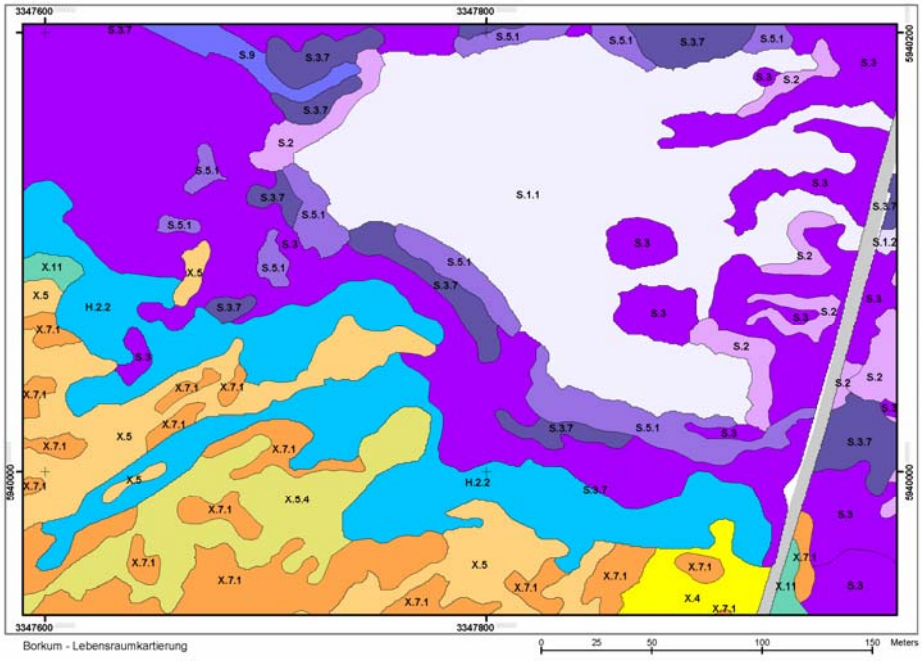


Abb. 8: Fertige Vegetationskarte (Borkum, Ausschnitt)

Abschließend erfolgten die Digitalisierung der Ergebnisse der Feldarbeit, eine Verknüpfung mit der bestehenden Datenbasis sowie eine GIS-technische und fachliche Kontrolle der Ergebniskarten.

6 Ausblick

Nur durch gute Teamarbeit und Projektplanung, fachübergreifende Zusammenarbeit sowie hohe Motivation der Vegetations-, GIS- und Fernerkundungsspezialisten war es überhaupt möglich, eine qualitativ sehr hochwertige Kartierung der gesamten terrestrischen Bereiche des Nationalparks Niedersächsisches Wattenmeer in einem so kurzen Zeitrahmen (Hauptkartierzeit: drei Monate) nicht nur effizient, sondern auch kostengünstig durchzuführen.

Im Ergebnis ergibt die dargestellte Methodik fertige Vegetationskarten mit sehr hoher Genauigkeit in Bezug auf Lage und Vegetation. Die Fernerkundung kann dabei den Kartierern vor Ort mit seiner Erfahrung nicht ersetzen, wohl aber seine Arbeit erleichtern und unterstützen bzw. den gesamten Erfassungsaufwand effizienter und somit kostengünstiger gestalten.

Hinsichtlich den Anforderungen und Berichtspflichten, die aus der FFH-Richtlinie resultieren, ist es nahe liegend, auf eine Weiterentwicklung des vorgestellten Verfahrens hinzuwirken. Dies sollte mit dem Ziel erfolgen, die Methodik als Basis für ein langfristiges Monitoring der NATURA 2000-Flächen zu etablieren. So kann bei künftigen Erfassungen auf den bereits vorliegenden Geometrien aufgebaut werden, indem sie beispielsweise zur Identifizierung geeigneter Trainingsflächen herangezogen werden. Auch eine Bilanzierung der Veränderung der Flächenzustände zwischen den Erfassungszeiträumen ist anhand dieser Datensätze schnell und exakt möglich, um beispielsweise den geforderten „günstigen Erhaltungszustand“ nachzuweisen und dessen Bewertung durch die EU Kommission zu unterstützen.

7 Literaturverzeichnis

7.1 Literatur

BAKKER, J., BUNJE, J., DIJKEMA, K., FRIKKE, J., HECKER, N., KERS, B., KÖRBER, P., KOHLUS, J. & M. STOCK (2005):

Salt Marshes - Chapter 7, Wadden Sea Quality Status Report. Common Wadden Sea Secretariat, Trilateral Monitoring and Assessment Group, Wilhelmshaven, S .165 – 180

BFN (BUNDESAMT FÜR NATURSCHUTZ) (HRSG.) (1998):

Das europäische Schutzgebietssystem NATURA 2000, BfN-Handbuch zur Umsetzung der Fauna-Flora-Habitat-Richtlinie und der Vogelschutzrichtlinie.- Schriftenreihe für Landschaftspflege und Naturschutz, Bonn Bad Godesberg, Heft 53, 560 S.

BOUMAN, C. & S. SHAPIRO (1996):

A Multiscale Random Field Model for Bayesian Image Segmentation.- In: IEEE Trans. On Image Processing, 3(2), S. 162 -177

DÜVEL, M. & A. FRICK (2005):

Ersterfassung der Lebensraumtypen nach Anhang I der FFH-RL / Anwendung von hochauflösenden Satellitendaten bei der Kartierung und Bewertung.- In: Natur und Landschaft, 5 (80), S.196

DRACHENFELS, O. V. (2004):

Kartierschlüssel für Biotoptypen in Niedersachsen. - Naturschutz und Landschaftspflege in Niedersachsen A/4, Hildesheim, 240 S.

GRASSMUCK, V. (2002):

Freie Software zwischen Privat- und Gemeineigentum.- BPB (BUNDESZENTRALE FÜR POLITISCHE BILDUNG (HRSG.): Themen und Materialien, Kevelaer, 438 S.

JUNEK, L, (2004):

Ermittlung der Gebietsretention anhand GIS-gestützter Auswertungen von CIR-Luftbildern – ein Beitrag zur Abschätzung von Hochwassergefährdung.- Unveröffentlichte Diplomarbeit am Geographischen Institut der Universität Hannover, 80 S.

NETELER M. & MITASOVA H. (2004):

Open Source GIS: A GRASS GIS Approach.- 2nd Edition.- Kluwer Academic Publishers / Springer, Boston, SECS, Volume 773, 424 S.

PETERSEN, J. & LAMMERTS, E.-J. (2005):

Dunes. - Chapter 9.2, Wadden Sea Quality Status Report. Common Wadden Sea Secretariat, Trilateral Monitoring and Assessment Group, Wilhelmshaven. S. 249 - 266

REDSLOB, M. (1999):

Radarfernerkundung in niedersächsischen Hochmooren.- Ibidem Verlag, Stuttgart, 278 S.

7.2 Internet-Referenzen

[1] Bundesanstalt für Naturschutz: http://www.bfn.de

[2] nature-consult: http://www.nature-consult.de

[3] GDF Hannover bR: http://www.gdf-hannover.de

[4] Nationalparkverwaltung Niedersächsisches Wattenmeer: http://www.nationalpark-wattenmeer.niedersachsen.de

[5] Bundeszentrale für politische Bildung: http://freie-software.bpb.de

[6] GRASS GIS / ITC-irst: http://grass.itc.it

[7] Quantum GIS: http://www.qgis.org

[8] PROJ4: http://www.proj.org

[9] OGR / GDAL: http://www.gdal.org

[10] PostgreSQL: http://www.postgresql.org

AGXIS –
Ein Konzept für eine generische
Schnittstellenbeschreibung

Ulrich Hussels, RISA Sicherheitsanalysen GmbH, risa@risa.de

Abstract / Einleitung

Ein grundsätzliches Problem von Umweltdatenbanken ist die Ersterfassung bzw. Übernahme von Daten aus bestehenden Datensammlungen. Die Ersterfassung von größeren Datenbeständen ist generell sehr aufwändig und bei der Datenübernahme aus bestehenden Datensammlungen besteht das Problem darin, dass diese Daten bereits unter bestimmten Prämissen gesammelt wurden, die in der Regel nicht explizit dokumentiert sind. Dadurch sind selbst Daten gleicher Struktur oft nicht miteinander vergleichbar. Daher ist es immer wichtig, das gesamte hinter einer Datensammlung stehende Modell zu kennen, um die Daten richtig erfassen bzw. interpretieren zu können.

Um also geeignete Schnittstellen für eine Datenerfassung oder eine Datenübernahme herzustellen, sollte zunächst ein fachliches, d. h. problem- und anwenderorientiertes Datenmodell formuliert werden. Aus diesem sollten dann die Schnittstelle bzw. eine Menge von zusammengehörigen Teil-Schnittstellen abgeleitet werden. Die Schnittstellenformate und die Modellbeschreibung sollten sowohl vom Menschen als auch von der Maschine zu interpretieren sein. Als Grundlage für die technische Realisierung solcher Schnittstellen bietet sich XML an.

Konkret geht es darum, umweltbezogene Unternehmensdaten von diesen zu erfassen, ggf. über mehrere (administrative) Stationen, in denen diese Daten geprüft werden können, weiter zu leiten und schließlich in einer gemeinsamen Datenbank zusammenzufassen (ETL-Vorgang). Manche Unternehmen sind in der Lage zu diesem Zweck vorhandene Datenbanken ’anzuzapfen’, andere müssen die Daten

erst erheben oder es gibt eine Mischung von Beidem. In jedem Fall unterscheidet sich die Interpretation dessen, was erfasst werden soll, von Unternehmen zu Unternehmen. Dies ist insbesondere dann der Fall, wenn es sich um unterschiedliche Branchen handelt. Ist der Zweck der Datensammlung bekannt, hilft dies schon weiter. Besser ist es, wenn schon bei der Erfassung deutlich wird, in welchem Zusammenhang die Daten schließlich abgelegt und ausgewertet werden sollen. Idealer Weise enthält das Datenmodell folglich alle für die Interpretation und den Vergleich der Daten notwendigen Informationen. Für Messwerte sind dies z. B. sämtliche Randbedingungen, die im Zusammenhang mit dem Gesamtmodell bei den abgefragten Quellen variieren können und nicht nur solche Randbedingungen, die für die Daten liefernde Stelle für relevant gehalten werden.

1. Aufgabe

Die Aufgabe, für die hier eine Lösung beschrieben wird, besteht in der Entwicklung eines Werkzeugs zur Definition eines fachlichen (problem- und anwenderorientierten) Datenmodells für Umweltfragestellungen einschließlich der darauf möglichen formalen und inhaltlichen Prüfungen sowie der Festlegung von Formatierungsinformationen zum Zweck der Visualisierung der Daten für den Menschen.

Umweltfragestellungen stellen deswegen eine Besonderheit dar, weil die Zusammenhänge wesentlich komplexer als z. B. bei normalen Verwaltungsaufgaben sind. Ein vollständiges Modell der Umwelt ist prinzipiell nicht herstellbar und zudem sind administrative Aspekte, wie hier die Erfüllung von Umweltberichtspflichten, mit den eigentlichen Fachdaten zu verknüpfen.

Eine weitere, von außen aufgeprägte wesentliche Randbedingung, die es erfordert ein spezielles Werkzeug einzusetzen, ist die, dass zwischen dem Zeitpunkt der endgültigen Festlegung des Datenmodells und dem ersten Einsatz der Schnittstellen oft nur eine sehr kurze Zeit liegen darf. Entscheidungen werden spät gefällt, Termine müssen aber eingehalten werden. Erfahrungsgemäß wird das Datenmodell zuvor auch in kurzen Abständen häufig geändert, was ohne ein entsprechendes Werkzeug nicht konsistent durchführbar wäre.

Die genannte Aufgabe ist, wie bereits erwähnt, nicht unabhängig vom fachlichen Kontext, hier dem administrativen Bereich des Umweltschutzes, zu sehen. Die Modelle müssen neben dem administrativen Bereich umfangreiche technisch-naturwissenschaftliche Daten sowie die oft komplexen Abhängigkeiten von Umweltdaten abbilden. Dabei kann weder vorausgesetzt werden, dass die Datenlieferanten besondere dv-Kenntnisse besitzen, noch dass sie mehr Fachkenntnisse besitzen, als für ihren Tätigkeitsbereich im Umweltschutz notwendig ist. Die voraussichtlich komplexe Gesamtstruktur muss also in leicht verständliche ’Häppchen’ zerlegt werden, unter denen sich ’jeder’ etwas vorstellen kann. Dennoch muss es möglich sein, daraus eine beliebig komplexe Gesamtstruktur, also ein komplexes Datenmodell zu konstruieren.

Von dem unter diesen Randbedingungen formulierten fachlichen Datenmodell sind dann die Schnittstellen zur Datenerfassung und Datenübernahme, die Prüfregeln für ein Prüfprogramm und die Formatierungsvorgaben für die Visualisierung abzuleiten.

Die Daten liefernden Stellen sollen ihre Daten für die Schnittstelle aufbereiten, diese auf formale und inhaltliche Korrektheit (soweit es die Prüfungen vorsehen) prüfen, auf Wunsch bearbeiten und ggf. formatiert ausdrucken können.

Die empfangende Seite soll ebenfalls die Korrektheit der Datei prüfen und deren Inhalt bearbeiten und ggf. formatiert ausdrucken können, ohne dass die Datei in eine Datenbankanwendung eingelesen werden muss.

2. Lösung

Die Lösung besteht in einer weiteren Verallgemeinerung des RISA-GEN Ansatzes [1], welcher vor vier bzw. zwei Jahren an gleicher Stelle vorgestellt wurde und der in der Zwischenzeit mehrfach, zuletzt für die Datenerfassung zum Emissionshandel, eingesetzt wurde.

RISA-GEN verwendet einen objektrelationalen Ansatz zur Parametrisierung eines fachlichen Datenmodells. Die Praxis der letzten Jahre hat aber gezeigt, dass der Anwender anstelle beliebiger Relationen zwischen den Objekten eine Hierarchisierung der Datenstrukturen bevorzugt. Der hier skizzierte Nachfolger AGXIS (Advanced Generic XML Interface System) verwendet daher einen hierarchisch-objektrelationalen Ansatz. Die zusätzliche Möglichkeit, explizit

Hierarchien zu definieren, kommt sowohl dem menschlichen Denken als auch dem vorgesehenen XML-Format entgegen, wobei es sich bei den nun im Konzept ergänzten Hierarchien streng genommen lediglich um gerichtete Graphen handelt, deren unmittelbare Umsetzung in XML nicht möglich ist. Die Beschränkung auf Baumstrukturen würde für die oben genannte Aufgabenstellung jedoch nicht ausreichen.

Zur Strukturierung des Datenmodells wird das Strukturelement ’Formular’ eingeführt. Dieser Begriff wurde gewählt, weil er relativ anschaulich ist und der Vorstellung einer strukturierten Datenerfassung nahe kommt. Formulare bestehen nach allgemeinem Verständnis aus einer Menge von Feldern zur Erfassung von zusammengehörigen Daten, wobei alle zu einem Formular passenden Datensätze der gleichen (nämlich der Formular-) Struktur folgen. Ein Zugriff auf die Daten kann nur über das ausgefüllte Formular, d. h. über den Datensatz erfolgen. Formulare können zudem Hierarchien bilden. Ein Unterformular kann eine 1:n-Beziehung zu seinem übergeordneten Formular realisieren. Diese anschauliche Interpretation stellt die Basis des Konzeptes dar. Es eignet sich besonders gut für die Datenerfassung durch den Menschen und ist (auch) in der (Umwelt-) Verwaltung millionenfach bewährt.

Ein Datenmodell besteht aus einer beliebigen Zahl von unterschiedlichen Formularen (Formulartypen), die jedoch alle miteinander verknüpft sein müssen. (Sonst sind es mehrere unabhängige Modelle). Die Struktur eines Formulars wird über die Parameter des generischen Modells festgelegt.

Jedes ausgefüllte Formular stellt eine Instanz eines Formulartyps (einen Datensatz) dar und wird über einen vom Dateninhalt unabhängigen Schlüssel beschrieben.

Formulare können ’hierarchisch’ zu einem gerichteten zyklischen Graphen sowie relational über Zeiger miteinander verknüpft sein. In beiden Fällen enthalten die Formulare Felder mit den entsprechenden Referenzen. Entscheidend für die Praktikabilität des Konzepts sind die Regeln für die Vergabe der Identifikationen für die Instanzen bzw. bei hierarchisch untergeordneten Formularen die Identifikationen der Subinstanzen. Diese muss im konkreten Fall der Denkweise des Menschen entsprechen und gleichzeitig allgemeingültig (generisch) sein. Die Identifikation einer Instanz bzw. Subinstanz erfolgt in diesem Konzept über Kurzbezeichnungen, die bezüglich eines jeden Hierarchievorgängers der (Sub-) Instanz innerhalb seines

Formulartyps eindeutig sein muss. Auf der obersten Hierarchieebene müssen die Instanz-Kurzbezeichnungen demnach bezüglich des Formulartyps eindeutig sein. Auf den darunter liegenden Ebenen ergibt sich die Eindeutigkeit jeweils über einen (von möglicherweise mehreren) Pfad(en) zur Subinstanz, nicht durch die Kurzbezeichnung der Subinstanz allein.

Die Aufgabe der Software (AGXIS) ist es, aus dem gerichteten und unter Umständen auch zyklischen Graphen eine XML-Gesamt- bzw. mehrere XML-Teil-Schnittstellen-strukturen zu erzeugen, wobei innerhalb der Schnittstellenstrukturen im Allgemeinen mehrere XML-Hierarchien auftreten, die aus der Auflösung der komplexen fachlichen Strukturen entstehen. Für diese Auflösung werden nachvollziehbare Regeln verwendet. Die Strukturen der Schnittstellen werden in XML-Schema-Dateien abgelegt.

Weiterhin werden das eigentliche fachliche Datenmodell, formale und inhaltliche Prüfregeln auf diesem Modell sowie Formatierungsregeln für die Darstellung in definierten und dokumentierten XML-Strukturen in Dateiform abgelegt. Mit Hilfe dieser Dateien lassen sich die zugehörigen Schnittstellen-Datensätze prüfen, in Berichtsform ausdrucken sowie editieren.

Bei Anwendung dieses Konzepts ist es auch Dritten möglich, Softwaremodule für diese Schnittstellen zu erstellen, bevor das konkrete Fachdatenmodell vorliegt. Damit verkürzt sich auch dort die Zeit zur Realisierung des Im- bzw. Exports.

3. Modellierungsregeln

Basiselement für die Modellierung eines AGXIS-Datenmodells ist das Formular (der Formulartyp). Ein Formular enthält beliebig viele Felder. Felder bestehen jeweils aus einer Feldbezeichnung und einem Antwortdatentyp. Bei der Modellierung kann festgelegt werden, ob ein Feld maximal eine zulässige Antwort oder eine geordnete Liste von gleichzeitig gültigen Antworten erlaubt. Felder können innerhalb eines Formulars zu Tabellen kombiniert werden. In diesem Fall stellen die Felder die Tabellenspalten und die Antworten die Tabellenzeilen dar.

Bei den Datentypen werden Standard-Datentypen (Ganzzahl, Kommazahl, Text, Zeitpunkt) und Zeiger auf (andere) ausgefüllte Formulare unterschieden.

Ein ausgefülltes Formular ist eine Instanz des Formulars und besitzt eine Identifikation (eine Kurzbezeichnung bzw. einen Namen).

Formulare können an oberster Stelle einer Hierarchie stehen oder Unterformulare (mehrerer) anderer Formulare sein. Unterformulare besitzen besonders gekennzeichnete Felder mit Zeigern auf die übergeordneten Formulare. Die Instanzen von Unterformularen müssen mindestens eine übergeordnete Instanz besitzen. Die Identifikation (Kurzbezeichnung, Name) einer (Sub-) Instanz muss innerhalb desselben Formulars bezüglich jeder übergeordneten Instanz eindeutig sein.

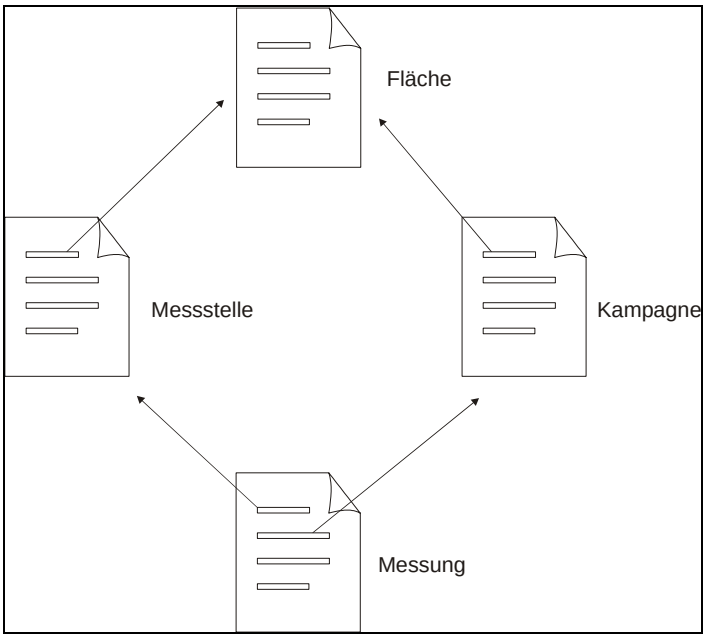
Für die Antworten können Einschränkungen (z. B. Wertebereiche) definiert werden. Diese können an Bedingungen geknüpft werden (z. B.: Wenn Branche = X, dann muss Wert > Y sein).

Ferner können für Formulare Formatierungsregeln angegeben werden.

Um es bei größeren und komplexeren Modellen den Daten liefernden Stellen einfacher zu machen, ihre Daten einzugeben bzw. entsprechend der Schnittstellendefinition aufzubereiten, können Teil-Datenmodelle als Teilmengen des gesamten Datenmodells definiert werden. Zu jedem Teil-Datenmodell gehört eine Teil-Schnittstelle, die sich von der Struktur her erheblich von den anderen (Teil-) Schnittstellen unterscheiden kann, wenn im Teil-Datenmodell z. B. ganze Formulare entfallen.

4. Beispiel

Zum besseren Verständnis der häufig auftretenden Fragestellungen wird im Folgenden eine sehr einfache und trotzdem schon nicht mehr triviale Datenstruktur erläutert.



Bild

Abb 1: Einfache Datenstruktur

Es sollen Messungen verwaltet werden, die entweder einer Messstelle (Ortsbezug) und/oder einer Messkampagne (Zeit- und Flächenbezug) zugeordnet werden können. Sowohl Messstellen als auch Messkampagnen können Flächen (z. B. Beprobungsflächen) zugeordnet werden.

Um nun nicht jeder Messung, jeder Messstelle und jeder Kampagne eine eigene, über das ganze Modell eindeutige Identifikation geben zu müssen, wird die Fläche als oberste Formularebene eingeführt. Messstellen und Kampagnen sind direkt darunter angeordnete Formulare und die Messungen sind wiederum diesen beiden untergeordnet.

Je nach Verfügbarkeit der Information kann bei einer Messung angebenen werden, zu welcher Messstelle oder zu welcher Kampagne diese gehört. Evtl. sind auch beide Informationen verfügbar. Um eine flächenbezogene Auswertung vornehmen zu können, reicht jedoch eine der beiden Angaben aus. Bei der Übernahme der Daten aus bestehenden Datenbanken kann die jeweils andere Information nicht mehr ermittelt werden.

Die Identifikation einer Messung muss nur bezüglich seiner übergeordneten Strukturebene (also bezüglich der Messstelle und/oder der Messkampagne) eindeutig sein. Die Identifikationen der Messstellen und der Messkampagnen

müssen nur bezüglich der Flächen eindeutig sein. Dies entspricht auch der Sichtweise bei der Datenerfassung.

Diese Datenstruktur lässt sich nicht als Baum in XML abbilden. Die Struktur ’Messung’ kann zwar mehrfach auftreten, aber die einzelnen Datensätze müssten ggf. auch mehrfach erscheinen.

AGXIS würde hieraus eine Schnittstellenstruktur bilden, die die Messungen per xpath-Ausdruck an die Messstellen und die Messkampagnen anbindet und auf der obersten Ebene neben die Flächen setzt.

Ein Teil-Datenmodell, welches keine Messkampagnen kennt, würde die Messungen in der Schnittstellenstruktur dagegen unterhalb der Messstellen anordnen können.

Innerer Aufbau des Formulars ’Messung’:

Feld 1: Zeiger auf eine Messstelle (max. eine Ausprägung)

Feld 2: Zeiger auf eine Messkampagne (max. eine Ausprägung)

Feld 3: Zeitpunkt der Messung (max. eine Ausprägung)

Feld 4: Parameter (Kurztext; beliebig viele Ausprägungen)

Feld 5: Wert (Kommazahl; beliebig viele Ausprägungen)

Feld 6: Einheit (Kurztext; beliebig viele Ausprägungen)

Die Felder 4, 5 und 6 bilden eine Tabelle.

Bedingungen für eine Messung:

Feld 1 oder Feld 2 muss ausgefüllt sein.

Feld 3 muss ausgefüllt sein.

Wenn Feld 5 in einer Zeile ausgefüllt ist, müssen die Felder 4 und 6 ebenfalls ausgefüllt sein.

Wenn Feld 6 mit ’ppm’ ausgefüllt ist, dann muss Feld 5 mit einem Wert größer oder gleich 0 ausgefüllt sein.

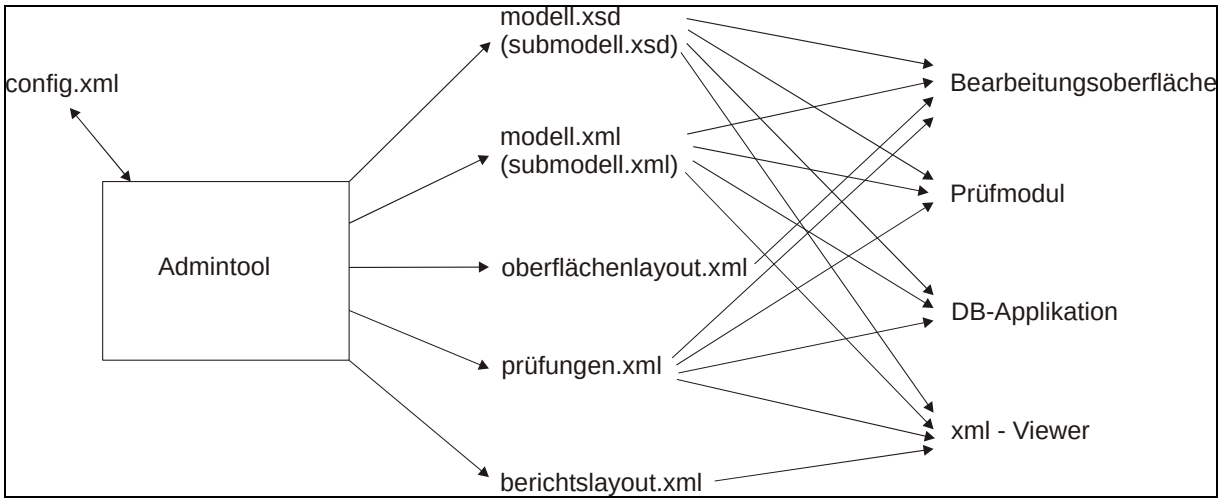
Formatierungsinformation für die Anzeige der Daten:

Feld 4 soll 60 %, Feld 5 20 % und Feld 6 20 % der Breite der Tabelle ausmachen.

5. Technische Realisierung

AGXIS besteht aus vier Grundkomponenten:

Die **zentrale Komponente** ist die Administrationskomponente (Admintool), mit der das Datenmodell, die Teil-Datenmodelle, die Prüfregeln und die Formatierungsregeln festgelegt werden und welches die Schnittstellendefinitionen (Schema-Dateien) erzeugt. Mit dieser Komponente arbeitet nur der Modellentwickler. Vorhandene Modellbeschreibungen können eingelesen, verändert und in Form von Dateien wieder abgespeichert werden.



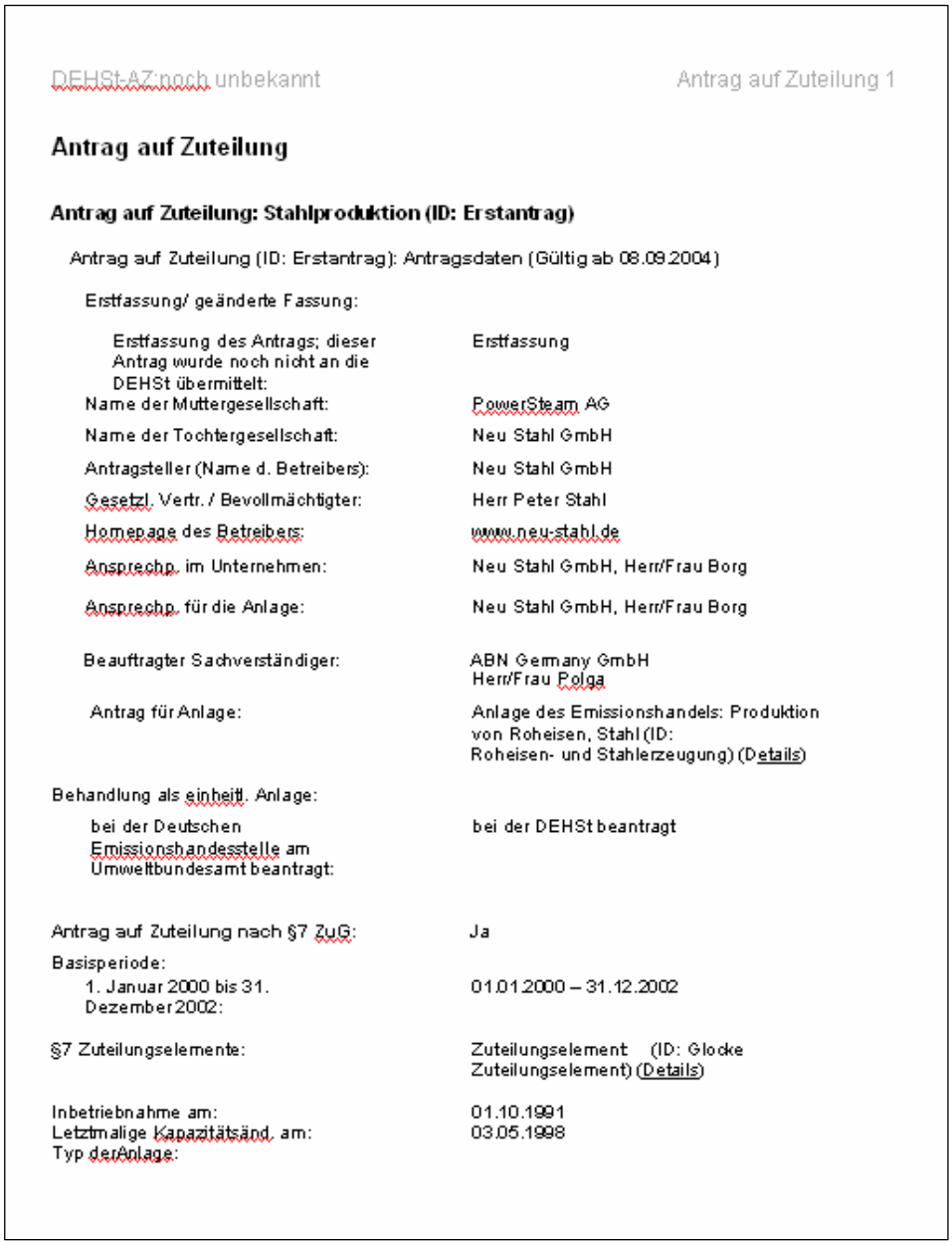
Bild

Abb. 2: Realisierungskomponenten

Weitere Komponenten sind:

Die **Prüfkomponente** (Prüfmodul) liefert eine Liste von Meldungen über nicht eingehaltene Bedingungen, wobei immer zuerst gegen das Schema und anschließend gegen zusätzliche Bedingungen, denen die Schnittstellendatei genügen muss, geprüft wird. Diese Bedingungen werden in Form einer zusätzlichen XML-Datei von der zentralen Komponente zur Verfügung gestellt.

Die **Visualisierungskomponente** (xml - Viewer) bereitet die XML-Schnittstellendateien zu einem für den Menschen lesbaren und verständlichen PDF-Dokument auf. Dazu muss die Schnittstellendatei mindestens dem geforderten Schema entsprechen. Mit Hilfe zusätzlicher Informationen aus der Modellbeschreibung wird daraus ein aufbereitetes Dokument, welches den Inhalt der Schnittstellendatei wiedergibt. Beim Antrag auf kostenlose Zuteilung von CO2-Emissionszertifikaten sah die (hart kodierte) Umsetzung wie in Abbildung 03/04 aus.



DEHSt-AZ:noch unbekannt Antrag auf Zuteilung 1

Antrag auf Zuteilung

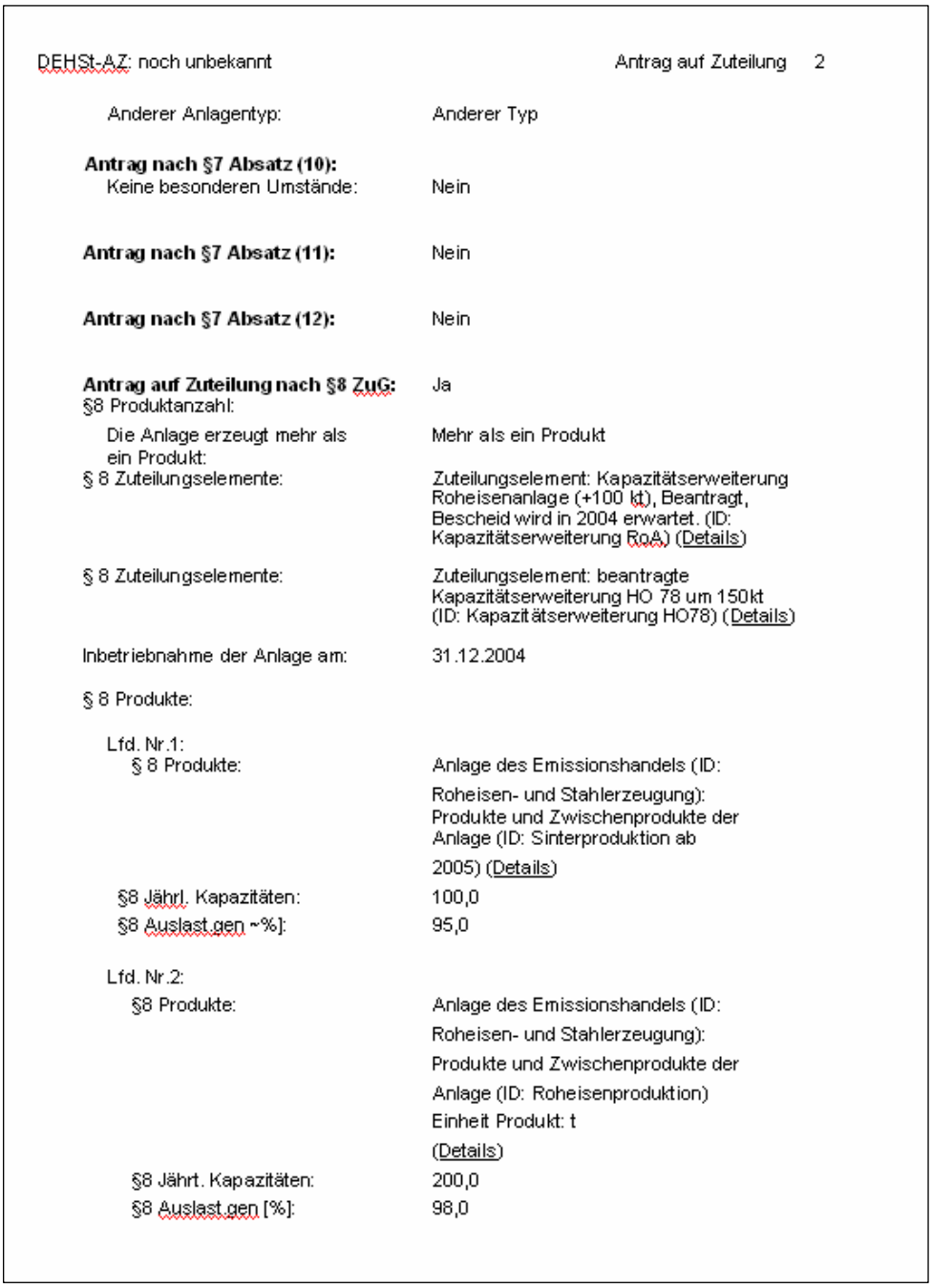
Antrag auf Zuteilung: Stahlproduktion (ID: Erstantrag)

Antrag auf Zuteilung (ID: Erstantrag): Antragsdaten (Gültig ab 08.09.2004)

Erstfassung/ geänderte Fassung:	
Erstfassung des Antrags; dieser Antrag wurde noch nicht an die DEHSt übermittelt:	Erstfassung
Name der Muttergesellschaft:	PowerSteam AG
Name der Tochtergesellschaft:	Neu Stahl GmbH
Antragsteller (Name d. Betreibers):	Neu Stahl GmbH
Gesetzl. Vertr. / Bevollmächtigter:	Herr Peter Stahl
Homepage des Betreibers:	www.neu-stahl.de
Ansprechp. im Unternehmen:	Neu Stahl GmbH, Herr/Frau Borg
Ansprechp. für die Anlage:	Neu Stahl GmbH, Herr/Frau Borg
Beauftragter Sachverständiger:	ABN Germany GmbH Herr/Frau Polga
Antrag für Anlage:	Anlage des Emissionshandels: Produktion von Roheisen, Stahl (ID: Roheisen- und Stahlerzeugung) (Details)
Behandlung als einheitl. Anlage:	
bei der Deutschen Emissionshandesstelle am Umweltbundesamt beantragt:	bei der DEHSt beantragt
Antrag auf Zuteilung nach §7 ZuG:	Ja
Basisperiode:	
1. Januar 2000 bis 31. Dezember 2002:	01.01.2000 – 31.12.2002
§7 Zuteilungselemente:	Zuteilungselement (ID: Glocke Zuteilungselement) (Details)
Inbetriebnahme am:	01.10.1991
Letztmalige Kapazitätsänd. am:	03.05.1998
Typ der Anlage:	

Bild

Abb. 3: Beispielseite 1 eines Berichts



DEHSt-AZ: noch unbekannt — Antrag auf Zuteilung 2

Anderer Anlagentyp:	Anderer Typ
Antrag nach §7 Absatz (10): Keine besonderen Umstände:	Nein
Antrag nach §7 Absatz (11):	Nein
Antrag nach §7 Absatz (12):	Nein
Antrag auf Zuteilung nach §8 ZuG:	Ja
§8 Produktanzahl: Die Anlage erzeugt mehr als ein Produkt:	Mehr als ein Produkt
§ 8 Zuteilungselemente:	Zuteilungselement: Kapazitätserweiterung Roheisenanlage (+100 kt), Beantragt, Bescheid wird in 2004 erwartet. (ID: Kapazitätserweiterung RoA) (Details)
§ 8 Zuteilungselemente:	Zuteilungselement: beantragte Kapazitätserweiterung HO 78 um 150kt (ID: Kapazitätserweiterung HO78) (Details)
Inbetriebnahme der Anlage am:	31.12.2004
§ 8 Produkte:	
Lfd. Nr.1: § 8 Produkte:	Anlage des Emissionshandels (ID: Roheisen- und Stahlerzeugung): Produkte und Zwischenprodukte der Anlage (ID: Sinterproduktion ab 2005) (Details)
§8 Jährl. Kapazitäten:	100,0
§8 Auslast.gen ~%]:	95,0
Lfd. Nr.2: §8 Produkte:	Anlage des Emissionshandels (ID: Roheisen- und Stahlerzeugung): Produkte und Zwischenprodukte der Anlage (ID: Roheisenproduktion) Einheit Produkt: t (Details)
§8 Jährt. Kapazitäten:	200,0
§8 Auslast.gen [%]:	98,0

Bild

Abb. 4: Beispielseite 2 eines Berichts

Die **Bearbeitungskomponente** (Bearbeitungsoberfläche) ist ein generischer Editor für die XML-Schnittstellendateien. Es wird jedoch nicht direkt auf dem XML-Schema der Schnittstellendatei gearbeitet, sondern auf der Struktur des fachlichen (Teil-)

Datenmodells. Dazu benötigt die Bearbeitungskomponente zusätzlich zur Schnittstellendatei auch die AGXIS-Datenmodellbeschreibung.

Alle Komponenten werden in JAVA implementiert, um Betriebssystem unabhängig eingesetzt werden zu können.

Grundsätzlich ist auch eine **Datenbankanwendung** (DB-Applikation) zur gemeinsamen Verwaltung und Auswertung größerer Zahlen von Schnittstellendatensätzen vorgesehen, sofern nicht auf bereits vorhandene Datenbankanwendungen aufgesetzt werden soll. Diese ist aber nicht Bestandteil des eigentlichen Schnittstellenkonzepts.

6. Zusammenfassung

Mit Hilfe weniger Modellierungsregeln auf Basis des aus praktischen Überlegungen gewählten Modellierungselements ’Formular’ erlaubt das vorgestellte Konzept die Definition eines anschaulichen fachlichen Datenmodells, aus dem jederzeit Schnittstellen für eine Datenerfassung abgeleitet werden können.

Veränderungen am Datenmodell können in kürzester Zeit zu den Schnittstellen propagiert werden.

Fehlinterpretationen bei der Bedienung der Schnittstellen können eingeschränkt werden, da ein (umfassendes) fachliches und vom Menschen lesbares Datenmodell zugrunde liegt. Ferner können die Schnittstellendateien mehr als nur formal (also gegen das Schema) geprüft und in einer für den Menschen lesbaren Form aufbereitet werden.

Bei Kenntnis des generischen (Meta-) Modells können Anwendungen, die die AGXIS-Schnittstellen verwenden sollen, bereits entwickelt werden, bevor ein konkretes Schnittstellenformat existiert.

7. Ausblick

Derzeit wird das Konfigurationswerkzeug (zentrale Komponente) mit Hilfe von RISA-GEN erstellt. Aufbauend darauf werden die Prüfkomponente, die Darstellungs-komponente und die Dateibearbeitungssoftware entwickelt.

Eine auf diesem Konzept basierende Datenbankapplikation zu entwickeln, ist erst für einen späteren Zeitpunkt geplant, da das Konzept zunächst Schnittstellen zwischen verschiedenen Organisationen (Unternehmen und Behörden) definieren soll, die ihre Daten unabhängig voneinander und unabhängig von diesem Konzept verwalten.

Das erste Ziel ist die Unterstützung von Umwelt-Berichtspflichten gegenüber der EU, z. B. in Verbindung mit der IVU-Richtlinie. Hierzu wird aus den jeweiligen Richtlinien und weiteren (in der Regel administrativen) Einflussgrößen mit Hilfe von AGXIS ein Datenmodell entwickelt, welches es gestattet, die geforderten Berichte zu erstellen. Die Datenerfassung erfolgt dann mit Hilfe der daraus generierten Schnittstellen. Da die Zahl der Berichtspflichten relativ groß ist und sich die Berichtspflichten auch regelmäßig (zumindest geringfügig) ändern, könnte ein Werkzeug wie AGXIS den Aufwand für die Bedienung der Schnittstellen reduzieren helfen. Für die Daten liefernden Stellen ergäbe sich ebenfalls ein Vorteil, wenn nicht jede Datenabfrage wieder individuell und nach einer anderen Systematik erfolgen würde.

Literaturverzeichnis

[1] Hussels, U.; Nagel, J.: Vorstellung eines generischen Ansatzes zur Erstellung und Pflege von Umweltdatenbanken. In: Wittmann, J.; Bernard, L. (Hrsg.) : Simulation in Umwelt- und Geowissenschaften : Workshop Münster 2001. Aachen : Shaker Verlag, 2001, S. 173-178, ISBN 3-8265-9251-4