Texte
54
03
ISSN
0722-186X

Umweltdatenbanken

Nutzung von Metadaten und Standards

Workshop
des Arbeitskreis „Umweltdatenbanken“ der
Fachgruppe „Informatik im Umweltschutz“,
veranstaltet vom Umweltbundesamt am
19. und 20. Mai 2003

Diese TEXTE-Veröffentlichung kann bezogen werden bei
Vorauszahlung von 10,00 €
durch Post- bzw. Banküberweisung,
Verrechnungsscheck oder Zahlkarte auf das

Konto Nummer 4327 65 - 104 bei der
Postbank Berlin (BLZ 10010010)
Fa. Werbung und Vertrieb,
Ahornstraße 1-2,
10787 Berlin

Parallel zur Überweisung richten Sie bitte
eine schriftliche Bestellung mit Nennung
der **Texte-Nummer** sowie des **Namens**
und der **Anschrift des Bestellers** an die
Firma Werbung und Vertrieb.

Der Herausgeber übernimmt keine Gewähr
für die Richtigkeit, die Genauigkeit und
Vollständigkeit der Angaben sowie für
die Beachtung privater Rechte Dritter.
Die in den Beiträgen geäußerten Ansichten
und Meinungen müssen nicht mit denen des
Herausgebers übereinstimmen.

Herausgeber: Umweltbundesamt
Postfach 33 00 22
14191 Berlin
Tel.: 030/8903-0
Telex: 183 756
Telefax: 030/8903 2285
Internet: http://www.umweltbundesamt.de

Redaktion: Fachgebiet II 1.3
Gerlinde Knetsch

Berlin, Oktober 2003

Vorwort

Mit dem Thema Umwelt-Metadaten und der Rolle internationaler Normierungsprozesse für Umweltinformationen beschäftigen sich seit Jahren Gremien, Fachexperten und Anwendungsentwickler von Umweltinformationssystemen. Im Zuge der Entwicklung des ISO-Standards 19115 für Metadaten über Geodaten steht diese Thematik derzeit in der Diskussion.

Das Umweltbundesamt organisierte am 19. und 20. Mai 2003 in Zusammenarbeit mit dem Arbeitskreis „*Umweltdatenbanken*“ des Fachausschusses “*Informatik im Umweltschutz*“ einen Workshop zu dieser Thematik. Etwa 30 interessierte Fachexperten aus Behörden, Institutionen und Forschungseinrichtungen Deutschlands und Österreichs nahmen daran teil.

Ein Schwerpunktthema bildete der Bereich der Standardisierung von Metadaten und speziell die Rolle von ISO 19115 für Umweltinformationen. Einerseits wurden neue Systeme vorgestellt, die den Zugriff auf und die Verwaltung von Geo-Informationen in den Vordergrund stellen, andererseits wurden die Auswirkungen dieses neuen Standards auf existierende Systeme und die Möglichkeiten der Integration in neue Fachanwendungen diskutiert. Eine deutliche Tendenz die auf dem Workshop feststellbar war, zeigte sich in der zunehmenden Anwendung des ISO-Standards 19115 für Fachanwendungen von Geodaten im Umweltbereich so z.B. bei der Entwicklung des Geodaten-Metainformationssystems des Bundes *GeoMIS.Bund*, des Brandenburgischen Metadatenprofils und des Küsteninformationssystems *NOKIS*. Der ISO-Standard sollte jedoch nicht als einzig relevanter Standard betrachtet werden. Umweltfachanwendungen die Metadaten einsetzen, werden auch in Zukunft immer weitergehende Elemente benötigen.

Daneben gab der Workshop interessante Einblicke in aktuelle Anwendungsgebiete des Gesamtkomplexes *Meta-Informationssysteme* und *Umweltdatenbanken*.

Dies betraf die Weiterentwicklung des *Umweltdatenkatalogs* (UDK), des *Umweltobjektkatalogs* Bayerns (UOK), eines Systems zur *Integrierten Mess- und Bilddatenhaltung für Umweltdatenbanken* (rasdaman) sowie der Fachanwendung *Dioxin-Datenbank* des Bundes und der Länder zu einem Web-Service.

Im Ergebnis des Workshops kristallisierten sich verschiedene Ansprüche an Metadaten-Standards heraus, die prinzipiell zwei Anwendungszielen zuzuordnen sind: Einerseits dienen Metadaten in Such- und Recherchesystemen wie dem GeoMIS Bund, UDK, UOK, und GEIN dazu, ihre Einträge zu verwalten und bei Anfragen zugänglich zu machen. Andererseits werden weitergehende Anforderungen an Metadaten gestellt, um beispielsweise für automatisierte Auswertungen oder Modellierungen Umweltdaten verschiedener Quellen miteinander verknüpfen zu können. Damit soll der Integrationsprozess von heterogenen Umweltdaten unterstützt und zielorientiert vorangetrieben werden.

September 2003

Inhalt

Vorwort

Die Entwicklung eines Metainformationsdienstes zu dezentral gehaltenen Geodaten des Bund und der Länder: GeoMIS.Bund®
(Martin Lenk, Jürgen Walther) **1 - 17**

Entwicklung eines Brandenburgischen Metadatenprofils der ISO 19115 und dessen Umsetzung am Beispiel des Deutschen Forschungsnetzes Naturkatastrophen und des behördlichen Datenvertriebs
(P. Köhler, Dr. F. A. Lochter, R. Häner) **19 - 44**

Die Framework Architektur des Umweltobjektkatalogs (UOK) Bayern
(Erich Weihs) **45 - 61**

Die IMPETUS-Metadatenbank Ein Baustein interdisziplinärer Forschung
(Michael Süßer) **63 - 81**

Geografisches Fachinformationssystem zur verteilten Verwaltung von militärischen Altlasten
(Uwe Rüppel, Gerrit Seewald, Michael Petersen) **83 - 108**

Entwicklung eines Web-Services für die DIOXIN-Datenbank des Bundes und der Länder
(Gerlinde Knetsch, Marianne Rappolder, Erich Weihs) **109 - 122**

Integrierte Meß- und Bilddatenhaltung für Umweltdatenbanken
(Peter Baumann) **123 - 136**

Wie bewältigt man Stationaritätsannahmen in der Geostatistik?
(Alexander Brenning, K. Gerald van den Boogaart) **137 - 152**

Anwendung von Metadaten im Küstenzonenmanagement
(Carsten Heidmann, Rainer Lehfeldt, Wassilios Kazakos, Frank Simmering) **153 - 183**

Triggerbasierte Replikationsmechanismen zum Austausch von Grundwasserinformationen in heterogenen Datenbankumgebungen
(Prof. Dr.-Ing. Uwe Rüppel, Dipl.-Ing. Thomas Gutzke Dr.-Ing. Michael Peters, Dipl.-Ing. Gerrit Seewald) **185 - 209**

Erstellung eines semantischen Netzwerkservice (SNS) für das Umweltinformationsnetz Deutschland - German Environmental Information Network (*gein*®)
(Thomas Bandholtz) **211 - 235**

Die Entwicklung eines Metainformationsdienstes zu dezentral gehaltenen Geodaten des Bund und der Länder: GeoMIS.Bund®

Martin Lenk, Jürgen Walther

Geschäfts- und Koordinierungsstelle des Interministeriellen Ausschusses für Geoinformationswesen

Allgemeines

Das Bundeskabinett hat am 17. Juni 1998 dem vom Bundesministeriums des Innern vorgelegten Bericht zur Verbesserung der Koordinierung auf dem Gebiet des Geoinformationswesens und der darin vorgeschlagenen Einrichtung eines "Interministeriellen Ausschusses für Geoinformationswesen" zugestimmt.

Mit seiner Entschließung vom 15. Februar 2001 hat der deutsche Bundestag die Bedeutung von Geodaten und die einer Geodateninfrastruktur eindrucksvoll unterstrichen. Im Antrag wird unter anderem festgestellt, dass Geobasisdaten wesentlich mit dem Einsatz öffentlicher Mittel gewonnen werden und diese eine öffentliche Infrastruktur darstellen.

GeoMIS.Bund als Teil einer Geodateninfrastruktur für Deutschland

Kernbestandteil einer Geodateninfrastruktur Deutschland (GDI-DE) ist die Nationale Geodatenbasis (NGDB), die aus Geobasisdaten (GBD), Geofachdaten (GFD) und deren Metadaten (MD) besteht.

NGDB = {Geobasisdaten, Geofachdaten, Metadaten}

Mit Hilfe der Nationalen Geodatenbasis, eines Geoinformationsnetzwerkes von Diensten und Standards schafft die GDI-DE die Vorraussetzungen für die Gewinnung, Auswertung und Anwendung von Geoinformationen für Nutzer und Anbieter in den öffentlichen Verwaltungen, im kommerziellen und nichtkommerziellen Sektor, in der Wissenschaft und für die Bürger.

GDI-DE = {NGDB, Netzwerk, Dienste, Standards}

Der IMAGI hat auf seiner 7. Sitzung im Oktober 2001 hierzu ein Drei-Stufen Konzept entwickelt. Hierin soll über

1. den Aufbau eines Zugangs zu Geodaten (Metainformationsservice),
2. der Harmonisierung der Objektartenkataloge (einheitliche Fachbegriffe, -kataloge)
3. und der Implementierung (internetbasiertes GeoPortal)

eine Geodateninfrastruktur für Deutschland realisiert werden.

Die erste Stufe ist mit der Fertigstellung von GeoMIS.Bund im Sommer 2003 bereits annähernd realisiert. Die Strategie für die Umsetzung der Stufen zwei und drei werden momentan modellhaft in Pilotprojekten erarbeitet.

Chronologie

Juni 1998:	Einrichtung des IMAGI
Mai 2000:	IMAGI Auftrag: GeoMis.Bund
Nov. 2000:	Expertenanhörung
Mai 2001:	Ausschreibung des Prototypen
Okt. 2001:	Auftragsvergabe Prototyp
März 2002:	Fertigstellung Prototyp
Dezember 2002:	Auftragsvergabe Endprodukt

GeoMIS.Bund – Technik (Übersicht)

Die gesamte Anwendung GeoMIS.Bund basiert im wesentlichen auf jsp, XML, HTML, SOAP und SQL.

Eine im GeoMIS.Bund-Broker eingehende Anfrage wird in eine vorgegebene XML-Struktur gebracht, in einen SOAP-envelope gepackt und nach Einlesen der FMIS-Informationen (Name, IP-Adresse etc.) an die verschiedenen Schnittstellen verteilt (zunächst ca. 25). In der Schnittstelle wird das XML file in eine SQL-Anfrage umgesetzt und mit Hilfe eines JDBC Treibers auf die Datenbank abgesetzt.

Das Ergebnis wird in einer ISO 19115 (Geographic information – Metadata) und ISO 19139 (Geographic information – Metadata – Implementation specification) konformen Struktur transformiert und - in SOAP eingepackt – wieder an den Broker gesendet. Dort werden die Informationen entpackt, mit Hilfe einer XSLT engine und ggf eines Mapservers in HTML umgewandelt und zum user Browser gesendet.

Für die Fachmetainformationssysteme (FMIS) kommen unterschiedlichste Datenbanken zum Einsatz (oracle, SQL-Server, Informix, Access etc.). Die Schemata (ERM) dieser Datenbanken sind sehr heterogen, so dass individuell angepasste SQL statements erstellt werden müssen.

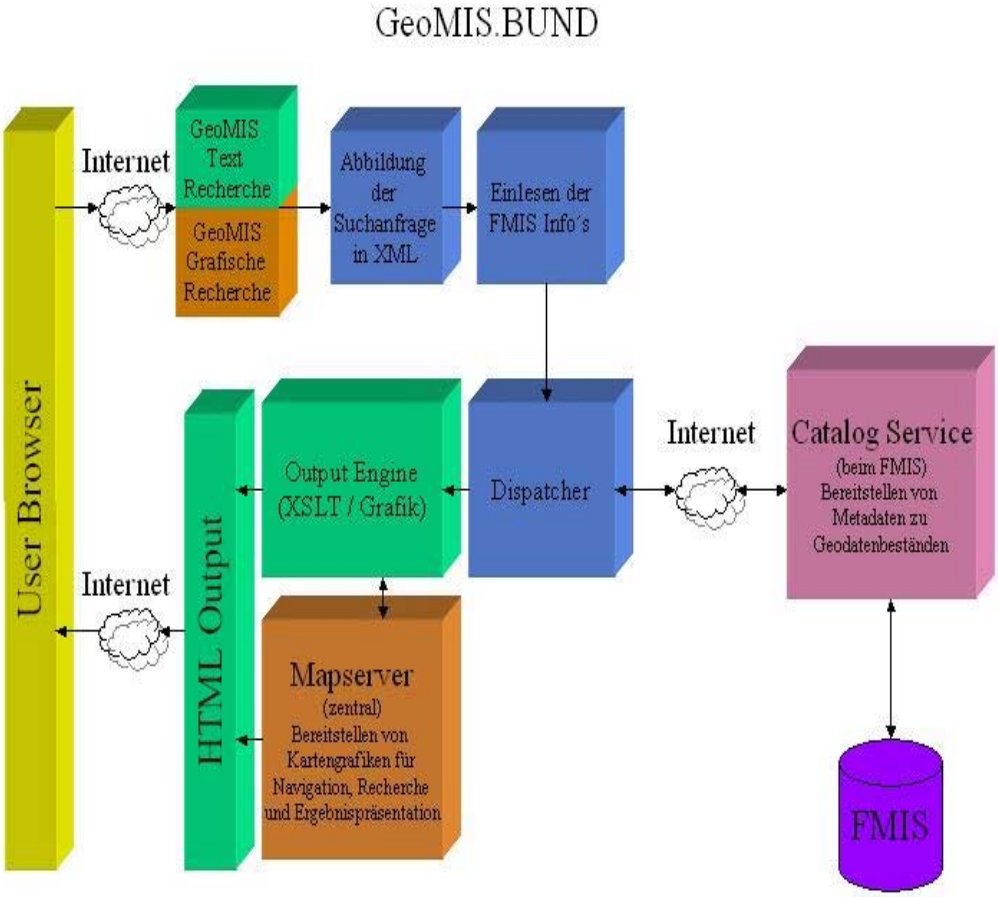


Abb.: 1 Funktionsschema GeoMIS.Bund

Der Zustand der Datenbanken ist recht unterschiedlich. Sowohl in der Aktualität als auch insgesamt bei den Inhalten, den Feldbezeichnungen etc. Problematisch sind u.a. fehlende Koordinatenangaben (incl. Koordinatensystem), Datumsfelder als Textfelder, fehlende Angaben z.B. zur Datenaktualität, Ansprechpartner (Daten und Metadaten) etc.

GeoMIS.Bund als Geographisches Metainformationssystem setzt für die zu suchenden Daten räumliche Angeben für dieselben voraus, so dass z.B. Verbreitungen von Tieren und Pflanzen, Biotopstrukturen aber auch Verbreitungen von Schadstoffen über Koordinaten und/oder geographische Namen erfasst bzw. erfragt werden können.

Fertigstellung GeoMIS.Bund - Ausblick

GeoMIS.Bund wird momentan in der Zusammenarbeit vom Bundesamt für Kartographie und Geodäsie (BKG), der Firma GIStec (Darmstadt) und der Firma wemove (Frankfurt) fertiggestellt. Geplanter Projektabschluss ist für den Sommer 2003 vorgesehen. GeoMIS.Bund wird dann in einer Internet Portal-Umgebung realisiert sein. Neben allgemeinen Diensten der Authentifizierung und Personalisierung werden u.a. Informationen rund um das Thema „Geoinformation“ geboten.

Perspektivisch wird GeoMIS.Bund sämtliche Geo(meta)datenbestände der Bundesverwaltung erreichen (Geobasis- und Geofachdaten). Weiterhin ist an den Anschluss von Metainformationsdiensten der Länder gedacht. In einem Pilotprojekt in Zusammenarbeit mit der Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) sollen bereits bei der Fertigstellung des oben genannten Endproduktes entsprechende Schnittstellen bei ausgesuchten Landesdiensten implementiert werden.

Mit der Entwicklung GeoMIS.Bund kann ein zentraler ISO basierter Metasuchdienst für Geodaten entwickelt und etabliert werden, der insbesondere für fachübergreifende und überregionale Abfragedienste geeignet ist. GeoMIS.Bund ist damit auch grundlegender Bestandteil einer Geodateninfrastruktur für Deutschland und ein wichtiger Baustein für die Entwicklung eines Geodatenportals, dass zukünftig über die Suchdienste hinaus Datenzugriff und Visualisierung ermöglichen sollte.

IMAGI

Die Entwicklung eines Metainformationsdienstes zu dezentral gehaltenen Geodaten des Bundes und der Länder: GeoMIS.Bund ®

Martin Lenk, Jürgen Walther

Geschäfts- und Koordinierungsstelle des Interministeriellen Ausschusses für Geoinformationswesen

Berlin, 19.5.2003

www.imagi.de

IMAGI

Überblick

(1) GeoMIS.Bund als Teil der GDI- DE

(2) Entwicklungsstufen von GeoMIS.Bund

a) Prototyp des Internetbrokers (2002)

b) Endausbau GeoMIS.Bund als Portal (2003)

(3) Zusammenfassung

IMAGI

Auftrag & Organisation

Organisation

- 10 Ressorts der Bundesregierung
- Sitzungen halbjährlich ➔ Beschlüsse
- Geschäfts- und Koordinierungsstelle im BKG in Frankfurt
- Beratung in Expertengruppen
- Infos: www.imagi.de

IMAGI

Strategie & Konzept: GDI – DE

1. Harmonisierung des Zugangs zu den Metainformtionen des Bundes (GeoMis.Bund)
2. Harmonisierung fachlicher Objektarten-kataloge / Standards (technische Integration, Schnittstellen, Standards, ISO19xxx, OGC, ...)
3. Implementierung der NGDB

GeoMIS.Bund-Teil der GDI–DE

- GeoMis.Bund: 1. Schritt zur Geodateninfrastruktur für Deutschland
- „Sicht“ auf NGDB (Nationale Geodatenbasis)
- Geo-Metadaten beschreiben Basis- und Fachdaten
- Integraler Bestandteil von GeoPortal.Bund

Chronologie GeoMIS.Bund

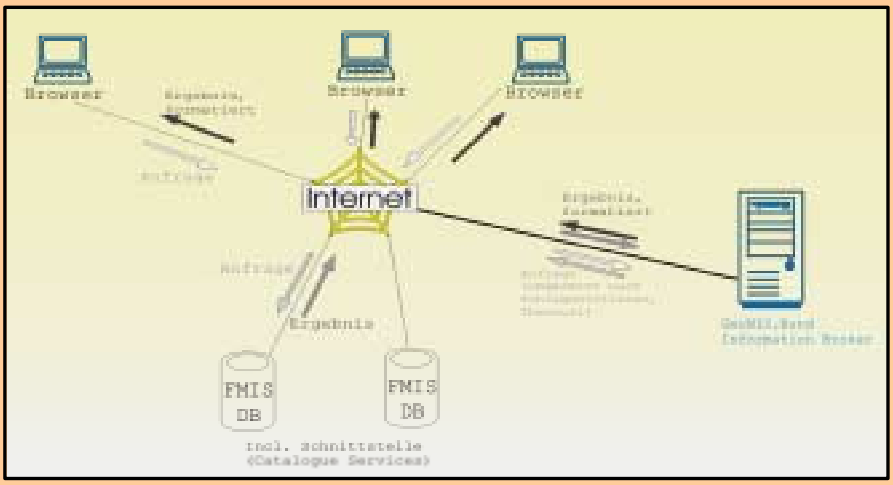
- Mai 2000: IMAGI Auftrag: GeoMis.Bund
- Nov. 2000: Expertenanhörung
- Mai 2001: Ausschreibung des Prototypen
- Okt. 2001: Auftragsvergabe Prototyp
- März 2002: Fertigstellung Prototyp
- Dezember 2002: Auftragsvergabe Endprodukt (Fa. GIStec)
- August 2003: Fertigstellung Endprodukt

http://www.geomis.bund.de

IMAGI

GeoMIS.Bund - brokerbasiert

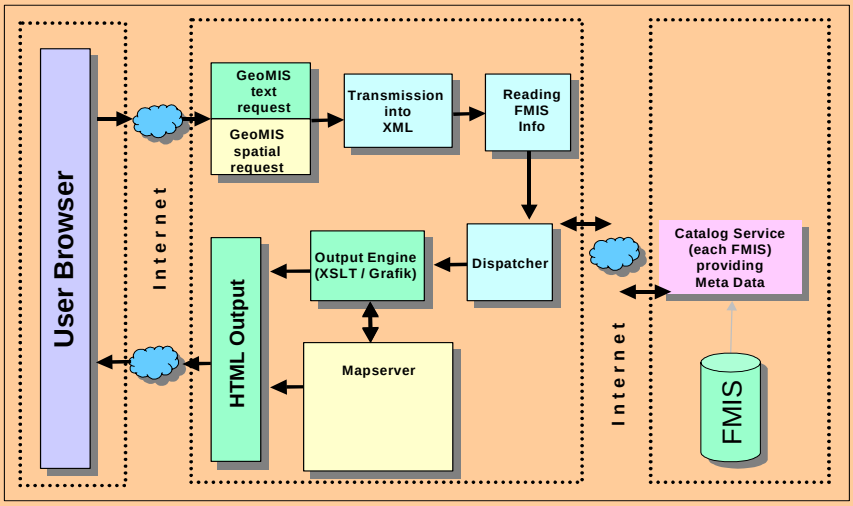
Suche in dezentralen Metadatenbeständen: Ablauf

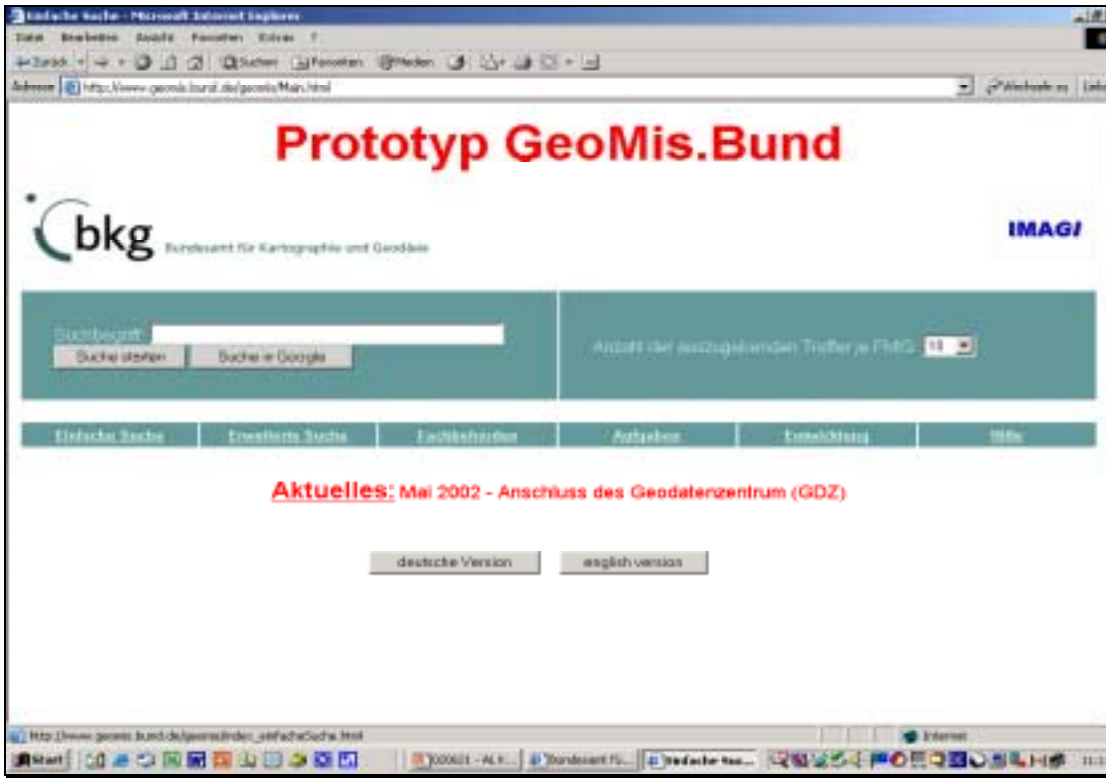


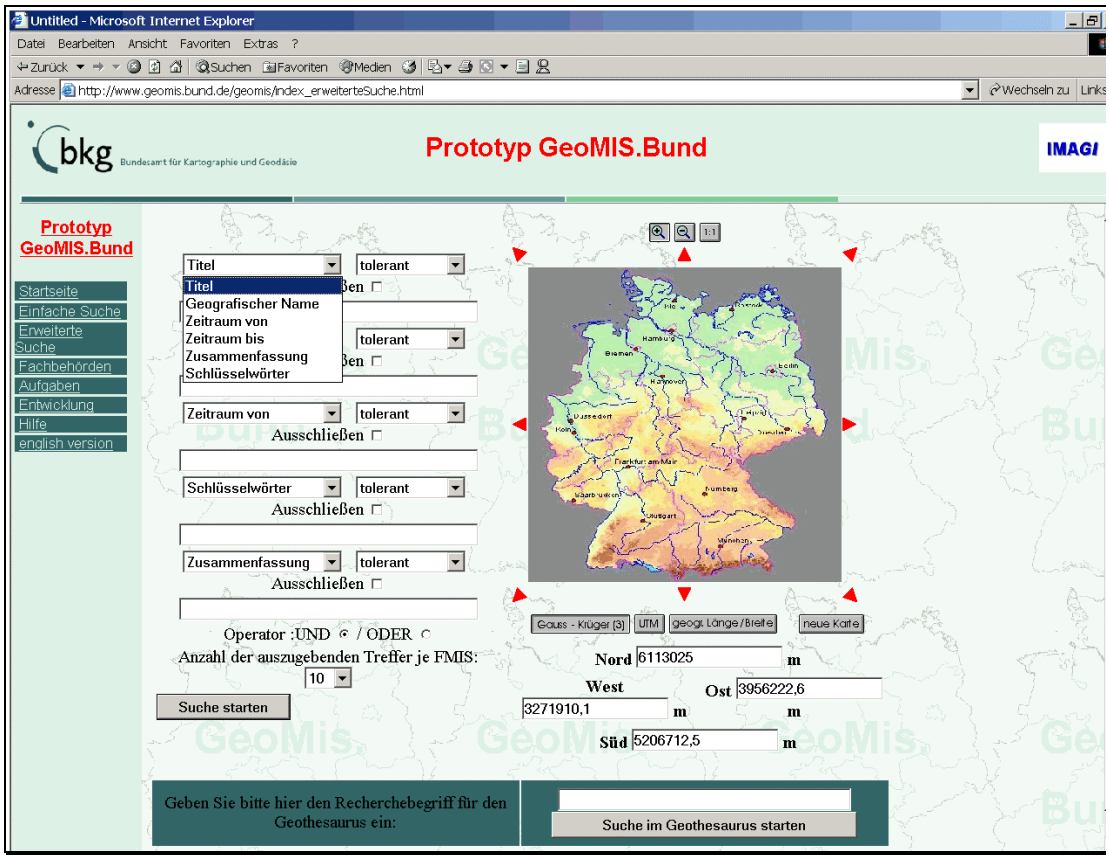
IMAGI

Technischer Aufbau

Architektur




Prototyp GeoMis.Bund
bkg
IMAGI
Aktuelles: Mai 2002 - Anschluss des Geodatenzentrum (GDZ)
deutsche Version
english version


Untitled - Microsoft Internet Explorer
Datei Bearbeiten Ansicht Favoriten Extras ?
Adresse http://www.geomis.bund.de/geomis/index_erweiterteSuche.html
bkg
Bundesamt für Kartographie und Geodäsie
Prototyp GeoMIS.Bund
IMAGI
Prototyp GeoMIS.Bund
Startseite
Einfache Suche
Erweiterte Suche
Fachbehörden
Aufgaben
Entwicklung
Hilfe
english version
Titel
tolerant
Geographischer Name
Zeitraum von
Zeitraum bis
Zusammenfassung
Schlüsselwörter
Ausschließen
Operator :UND / ODER
Anzahl der auszugebenden Treffer je FMIS:
10
Suche starten
Gauss - Krüger (3)
UTM
geogr. Länge / Breite
neue Karte
Nord 6113025 m
West 3271910,1 m
Ost 3956222,6 m
Süd 5206712,5 m
Geben Sie bitte hier den Recherchebegriff für den Geothesaurus ein:
Suche im Geothesaurus starten

Untitled - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück Suchen Favoriten Medien

Adresse http://www.geomis.bund.de/geomis/index_erweiterteSuche.html Wechseln zu Links

bkg Bundesamt für Kartographie und Geodäsie

Prototyp GeoMIS.Bund

IMAGI

Prototyp GeoMIS.Bund

- Startseite
- Einfache Suche
- Erweiterte Suche
- Fachbehörden
- Aufgaben
- Entwicklung
- Hilfe
- english version

Anzahl der Antworten: 0

Deutsches Fernerkundungsdatenzentrum

Anzahl der Antworten: 0

Wasser- und Schifffahrtsverwaltung (F-IT)

Anzahl der Antworten: 78

Titel	Digit. Bundeswasserstraßenkarte 1:2.000 Rhein-Herne-Kanal von km 0 bis km 10 (rhk_000)		
Zeitl. Ausdehnung von	null	Zeitl. Ausdehnung bis	null
Zusammenfassung	fehlt	Schlüsselwörter	Hektometerpunkte;WSV-Aufnahmepunkte;NIV-Festpunkte;Pegelfestpunkte;Ordnungsprofilpunkte;Sonderprofile;Lichtraumprofile;Trigonometrisch Punkte;Katasterfestpunkte;GPS-Referenzstation;Einteilungslinien;Achsen;Fahrrinnen;Profilmeßbezugspunkte;MTHw-Linien

vollst. Anzeige

Titel	Digit. Bundeswasserstraßenkarte 1:2.000 Rhein-Herne-Kanal von km 10 bis km 20 (rhk_001)		
Zeitl. Ausdehnung von	null	Zeitl. Ausdehnung bis	null
Zusammenfassung	fehlt	Schlüsselwörter	Hektometerpunkte;WSV-Aufnahmepunkte;NIV-Festpunkte;Pegelfestpunkte;Ordnungsprofilpunkte;Sonderprofile;Lichtraumprofile;Trigonometrisch Punkte;Katasterfestpunkte;GPS-Referenzstation;Einteilungslinien;Achsen;Fahrrinnen;Profilmeßbezugspunkte;MTHw-Linien

Untitled - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück Suchen Favoriten Medien

Adresse http://www.geomis.bund.de/geomis/index_erweiterteSuche.html Wechseln zu

bkg Bundesamt für Kartographie und Geodäsie

Prototyp GeoMIS.Bund

IMA

Prototyp GeoMIS.Bund

- Startseite
- Einfache Suche
- Erweiterte Suche
- Fachbehörden
- Aufgaben
- Entwicklung
- Hilfe
- english version

Digit. Bundeswasserstraßenkarte 1:2.000 Rhein-Herne-Kanal von km 0 bis km 10 (rhk_000)

Metadaten-Kennung	20021
Metadatensprache	Deutsch
Metadaten erstellt am	null
Kurzbezeichnung	DBWK2 rhk_000
Zusammenfassung	fehlt
Maßstabsbereich	2000
Datensatzsprache	Deutsch
Klassifikation	Binnengewässer
Schlüsselwörter	Hektometerpunkte;WSV-Aufnahmepunkte;NIV-Festpunkte;Pegelfestpunkte;Ordnungsprofilpunkte;Sonderprofile;Lichtraumprofile;Trigonometrische Punkte;Katasterfestpunkte;GPS-Referenzstation;Einteilungslinien;Achsen;Fahrrinnen;Profilmeßbezugspunkte;MTHw-Linien
Thesaurus	
Status	null
Erfassungsgrad	null
Datenqualität, Quelle	fehlt
Datenqualität, Anmerkung	fehlt
Erfassungszeitraum von	null
Erfassungszeitraum bis	null
Datenformat	MicroStation Design-File (DGN)
Datenträger	null
Online-Quelle	fehlt
Nutzungsbedingungen	fehlt
Datenanbieter	Kartenstelle der WSD West
Funktion	fehlt
Ansprechpartner	Paul Proske
Adresse	null

IMAGI Erweiterte Metadatenkernmenge

Kompromiss: ISO 19115 ⇔ Konzeption Bund

- **Kernmenge aus IMAGI-Konzeption: 36 Felder, davon 14 Pflicht, 16 optional, 5 konditional, 1 o.A.**
- **Es fehlen 16 Felder aus der ISO-Kernmenge (2 Pflicht, 2 konditional, 4 optional)**
- **8 Felder aus der Konzeption sind nicht in der ISO-Kernmenge enthalten (6 otional, 1 konditional, 1 o.A.)**

ID	Name	Definition	Typ
4	MD_characterSet	full name of the character coding standard used for the metadata set	C
8	contact	party responsible for the metadata information	M
10	metadataStandardName	name of the metadata standard (including profile name) used	O
11	metadataStandardVersion	version (profile) of the metadata standard used	O
37	spatialRepresentationType	method used to spatially represent geographic information	O
187	referenceSystemIdentifier	name of reference system	C
394	data	reference data for the cited resource	M
40	Data_characterSet	full name of the character coding standard used for the dataset	C

IMAGI Endstufe mit Portalfunktion

Identitätsbereich:

Über uns, Themen, Aktuelles, Publikationen

Servicebereich:

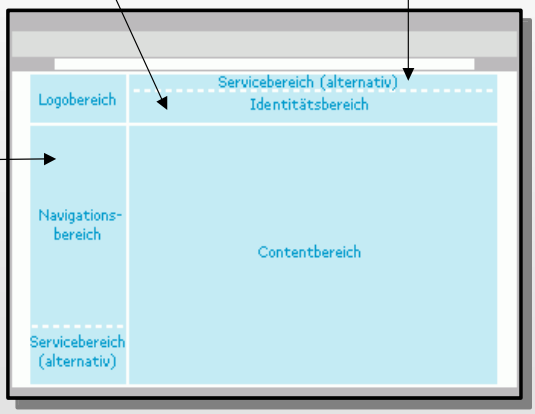
Home, Glossar, Impressum, Hilfe,

Kontakt, Sitemap, Spachumschaltung

Navigationsbereich:

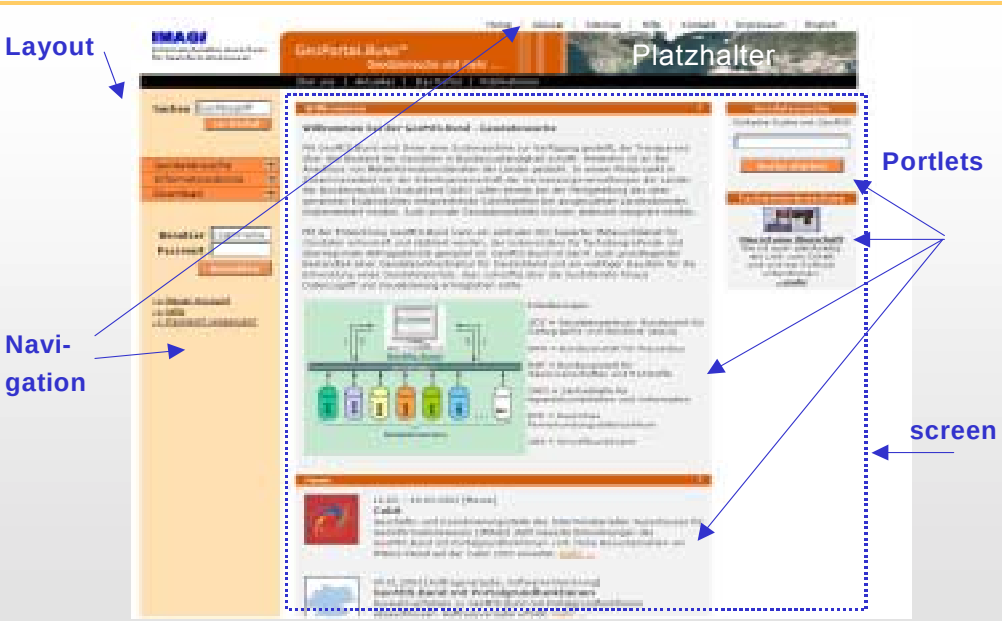
- GeoMis.Bund
  - Über GeoMIS.Bund
  - Einfache Suche
  - Erweiterte Suche
  - Assistent
  - FMIS Info
- Community
  - Forum
  - Jobs
- Infocenter
  - Informationen
  - FAQ
  - Links
  - Fachwörterbuch
- Daten & Downloadcenter

Style – Guide BundOnline 2005



IMAGI

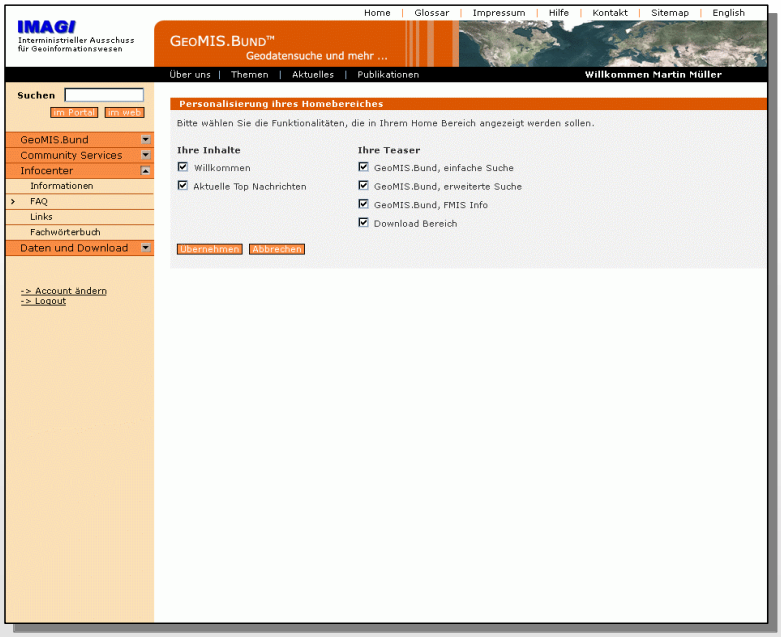
Aufbau GeoMIS/Portal



IMAGI

Umsetzung GeoMIS/Portal

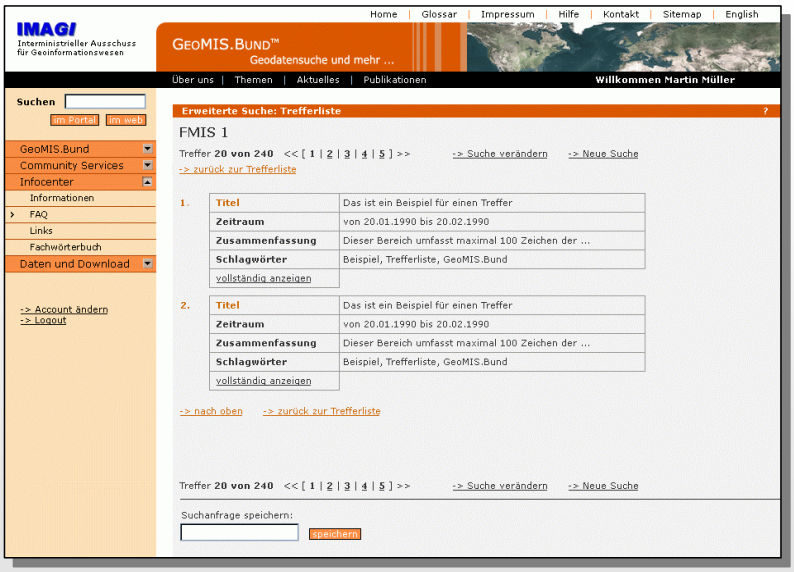
Beispiel: Personalisierung



IMAGI

Umsetzung GeoMIS/Portal

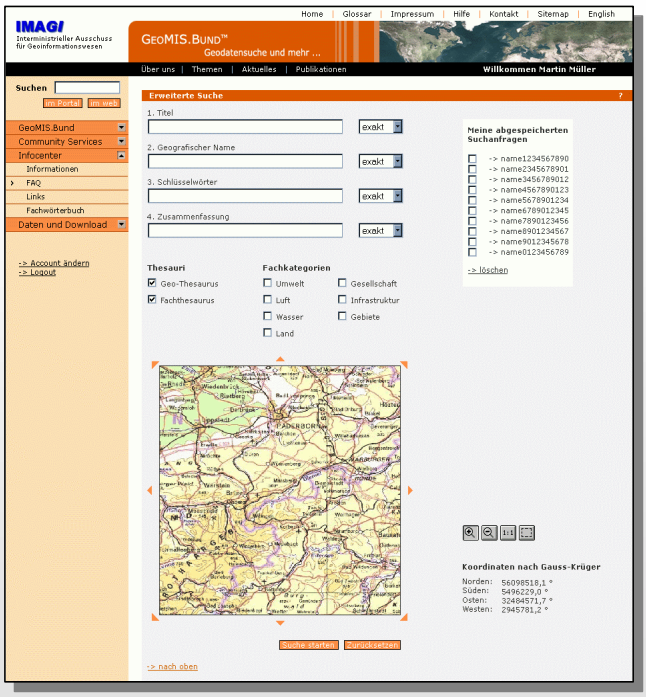
Beispiel: Ergebnisformular



IMAGI

Umsetzung GeoMIS/Portal

Beispiel: Erweiterte Suche



IMAGI Zusammenfassung: GeoMis.Bund

Attribute: GeoMIS.Bund (1)

- internetbasiertes, selbsterklärendes Recherchewerkzeug (Broker-Lösung)
- plattformunabhängiger Zugang für Nutzer und modulare Architektur (Java-Servlets) des GeoMIS.Bund
- räumliche, thematische und zeitbezogene Recherche
- Berücksichtigung semantischer Fragestellungen
- einfache Benutzerführung
- Übersichtliche, dynamische Darstellung der Suchergebnisse

IMAGI Zusammenfassung: GeoMis.Bund

Attribute: GeoMIS.Bund (2)

- Weiterleitung zu den Details bzw. vollständigen Trefferlisten
- Tabellarische Detailansicht gemäß ISO 19115
- Schnittstellen werden zur Verfügung gestellt
- ISO/OGC – konform
- deutsch- und englischsprachige Realisierung
- Integration des GeoMIS.Bund in ein zukünftiges GeoPortal.Bund
- Darstellung aller Daten auf einer einheitlichen Oberfläche

IMAGI Zusammenfassung: GeoMis.Bund

Attribute: GeoMIS/Portal (1)

- Basisversion frei von aktiven Inhalten (ActiveX, Javascript etc.)
- Plattformunabhängiges, open source framework Jakarta
- Java Servlet / Java Server Pages Technologie
- Verwendung von Portlets
- Möglichkeit der Personalisierung
- Portierbar auf alle Plattformen ohne zusätzliche Software
- BundOnline 2005 konform: SAGA V1.1, Integration Basismodule, Barriere frei, Vorbereitung CMS
- Umfangreicher Service - Dienste

IMAGI Zusammenfassung: GeoMis.Bund

Attribute: GeoMIS/Portal (2)

- Fertigstellung bis Mitte 2003 anvisiert
- Bis zu 50 dezentrale Geo-Metainformationssysteme in der ersten Ausbaustufe
- Metadaten von Bund, Ländern und Privatwirtschaft
- Zusammenarbeit mit AdV
- Einbindung von GeoMis.Bund in ein zu entwickelndes GeoPortal.Bund
- Integration in ESDI

IMAGI

Vielen Dank

Weitere Informationen

- www.imagi.de
- www.geomis.bund.de
- www.bkg.bund.de
- www.geodatenzentrum.de

Entwicklung eines Brandenburgischen Metadatenprofils der ISO 19115 und dessen Umsetzung am Beispiel des Deutschen Forschungsnetzes Naturkatastrophen und des behördlichen Datenvertriebs

P. Köhler, GeoForschungsZentrum Potsdam (GFZ, p.koehler@gfz-potsdam.de)
Dr. F. A. Lochter, Landesamt für Geowissenschaften und Rohstoffe Brandenburg (LGRB, lochter@lgrb.de)
R. Häner, GeoForschungsZentrum Potsdam (GFZ, rhaener@gfz-potsdam.de)

Einführung

Mit einer „Gemeinsamen Erklärung“ (http://katalog.lgrb.de/lgrb/ui/dhtml/gib.html) des Landesvermessungsamtes Brandenburg (heute: Landesvermessung und Geobasisinformation Brandenburg, LGB), des Landesamtes für Geowissenschaften und Rohstoffe Brandenburg (LGRB), des Landesumweltamtes Brandenburg (LUA) sowie des GeoForschungsZentrums Potsdam (GFZ) ist 2001 eine Initiative zum Aufbau einer Geodateninfrastruktur Brandenburg (GIB) ins Leben gerufen worden.

Ihr Ziel ist die Schaffung von Synergien in und zwischen den Einrichtungen und Behörden sowie bei der Erschließung und Nutzung von Geodaten und von Know-how im Bereich Geoinformation und Geo-Informationstechnologie. Zu erwartende Mehrwerte sind

- die einrichtungsübergreifende und interdisziplinäre Nutzung und der ungehinderte Austausch von Geodaten,
- die Schaffung wesentlicher Voraussetzungen für die erfolgreiche Einführung von eGovernment-Prozessen,

- der Aufbau eines Geodatenmarktes in Brandenburg,
- die Einrichtung einer Kommunikations- und Kompetenzplattform für Verwaltung, Wirtschaft, Wissenschaft und Gesellschaft.

Ein Anwendungsfeld, dessen Abhängigkeit von einer funktionierenden Geodateninfrastruktur (GDI) bereits mehrfach demonstriert wurde, ist die Vorsorge und Bewältigung von Katastrophen. Die Existenz einer aktuellen und qualitativen Daten- und Informationsbasis ist hier wie kaum in einem anderen Bereich die Grundlage der Entscheidungsfindung. Neben der mangelnden Integration heterogener Daten (Geobasisdaten, Umweltdaten, sozioökonomische Daten etc.) fehlt es bereits häufig an einem umfassenden Überblick über vorhandene Daten und Informationen. Die Dokumentation von Daten anhand von Metadaten sowie deren Bereitstellung in Metainformationssystemen bieten die Möglichkeit des zentralen Zugriffs auf dezentral vorhandene Daten, wie das Beispiel des Deutschen Forschungsnetzes Naturkatastrophen (DFNK) zeigt. Davon profitieren jedoch ebenfalls kommerzielle Datenanbieter sowie Einrichtungen der öffentlichen Hand, dargestellt am Beispiel des Landesamtes für Geowissenschaften und Rohstoffe Brandenburg.

Erstellung eines Brandenburgischen Metadatenprofils auf der Basis von ISO 19115

Diese Thematik greift die „Special Interest Group“ (SIG) Metadaten der Initiative zum Aufbau einer Geodateninfrastruktur in Brandenburg auf. Hier befinden sich Experten der o.g. Einrichtungen und des Ministeriums für Landwirtschaft, Umweltschutz und Raumordnung Brandenburg (MLUR) sowie der Firma Delphi IMM in intensivem Dialog miteinander. Mitglieder der SIG waren beispielsweise eingebunden in das Hearing des Bundesamtes für Kartographie und Geodäsie (BKG) zum Aufbau des GeoMIS.Bund, über Jahre aktiv beteiligt an den Normungsaktivitäten des Open GIS Consortium (OGC), sind in Expertengremien des Umweltdatenkatalogs (UDK) und Arbeitsgruppen des Deutschen Dachverbands für Geoinformation (DDGI) vertreten oder Mitautoren des „GSDI Cookbook“ zum Aufbau einer Global Spatial Data Infrastructure (NEBERT 2001).

Ihnen gemeinsam ist der Bedarf eines allgemein anerkannten Metadatenformats. Ziel aller beteiligten Einrichtungen ist die Erstellung eines brandenburgischen Metadatenprofils basierend auf und konform zu den Standards der International Organisation for Standardisation (ISO).

Vorgehensweise

Die Erarbeitung des Metadatenstandards der ISO, ISO 19115, wird teilweise in Personalunion in der ISO-Gruppe und der Arbeitsgruppe Metadaten des OGC vorangetrieben, die bevorstehende Verabschiedung ist seit nun mehr zwei Jahren angekündigt. Die Langwierigkeit des Prozesses ist einerseits in der Komplexität und dem Anspruch des Standards, andererseits in den unterschiedlichen Interessen der Mitglieder bzw. der Erzeuger und Nutzer von Metadaten aus verschiedensten Anwendungsbereichen begründet. Da bisher kein verbindlicher Standard verabschiedet werden konnte, basiert die Arbeit der SIG Metadaten auf den Versionen des Draft International Standard (DIS) sowie des aktuellen Final Draft International Standard (FDIS, Stand 23. März 2003) der ISO 19115 (ISO/TC 211 2003).

Bereits im Herbst 2001 wurde durch LGRB und LGB ein erster Entwurf eines gemeinsamen Profils der ISO 19115 (DIS) vorgelegt. Im Rahmen der SIG Metadaten wurde nun in Abstimmung aller beteiligten Einrichtungen ein Brandenburgisches Metadatenprofil basierend auf den aktuellen Dokumenten der ISO erarbeitet. Die Diskussion beschränkte sich zunächst auf eine für Brandenburg verbindliche Auswahl von Attributen und Elementen, d.h. es wurden diejenigen Attribute und Elemente sowie deren Ausprägungen festgelegt, die zur Beschreibung von Geodaten im Rahmen der brandenburgischen Geodateninfrastruktur gewährleistet sein müssen.

Die inhaltliche Diskussion um das Metadatenprofil ist im April 2003 abgeschlossen worden. Anschließend erfolgte die Entwicklung eines „Brandenburger Application Schemas“, das eine Untermenge der zukünftigen ISO 19115-Norm abbildet. Eine weitere Zielsetzung der SIG ist schließlich die Abstimmung von Optionslisten, unter anderem für Schlagwortkataloge.

Ergebnis: ISO 19115-konformes Metadatenprofil für Brandenburg

Das ISO/DIS 19115-Modell definiert mehr als 300 unterschiedliche Metadatenelemente. Diese sind in M (Mandatory) = „verpflichtende“, O (Optional) = „optionale“ und C (Conditional) = „bedingte“ Elemente unterteilt. ISO 19115 baut auf dem sog. Kern („essential profile“, „core“), den weiterreichenden („comprehensive“) und den erweiterten („extended“) Metadatenelementen auf.

Das Metadatenmodell der SIG Metadaten benutzt lediglich Kernelemente und einige weiterreichenden Metadatenelemente der ISO 19115 (DIS) und ist vollständig mit ihr kompatibel, d.h. es

- entspricht der durch die ISO/DIS 19115 vorgegebenen Semantik,
- nutzt die von der ISO/DIS 19115 definierten Datenelemente „eindeutig“ (in Nomenklatur und Inhalt) und
- erhält die von der ISO/DIS 19115 definierten Verbindungen innerhalb der „ISO 191xx-Familie“.

Das ISO/DIS 19115-konforme Metadatenmodell der Initiative Geodateninfrastruktur Brandenburg wird in XML ausgeführt. Vorlage ist das „Application Schema“ der National Imagery and Mapping Agency (NIMA), das folgende Schemata beinhaltet:

- GML Geometry (Erweiterungen für den Raumbezug)
- ISO 19103 (grundlegende Datentypen)
- ISO 19109 (Regeln für Anwendung eines „Application Schema“)
- ISO 19115 (Geographische Information, Metadaten)
- ISO 19118 (Codierungsregeln)
- ISO 639-2 (Ländercodes)
- ISO 4217 (Währungscodes)

-

Das Application Schema der NIMA als Grundlage für das Brandenburgische ISO 19115-Profil ist dabei in folgender Hinsicht reduziert worden:

- ISO 19109 und GML Geometry entfallen.
- Die Codelisten der ISO 4217 und ISO 639-2 wurden reduziert auf im Brandenburger Rahmen relevante Begriffe.
- Alle nicht genutzten Elemente der ISO 19115 wurden, soweit dies nicht die Vorschriften der Norm verletzt, entfernt.

Darüber hinaus sollten jedoch Struktur und Eigenschaften der ISO 19115-Elemente keinesfalls verändert werden.

Abstimmung von Optionslisten und Schlagwortkatalogen

Die Festlegung auf Elemente und Attribute allein garantiert jedoch nicht, dass allen Anforderungen, sowohl den der Datenprovider als auch den der Nutzer, genüge getan wird. Die Autoren mussten in ihrer Arbeit erkennen, dass zum Angebot eines geeigneten nutzerfreundlichen Metainformationssystems auch abgestimmte Schlagworte gehören. Wenn beispielsweise der Nutzer nach dem Schlagwort „Geologie“ sucht, in den Metadaten jedoch „geologische Karte“ abgelegt ist, wird das passende Datenangebot nicht gefunden werden.

Daher wurden bereits zu einem früheren Zeitpunkt zwischen LGRB und LGB aber auch zwischen den Staatlichen Geologischen Diensten (SGD) thematische Schlagworte zu ausgewiesenen Themenbereichen abgestimmt. Schwerpunkt dabei war nicht die begriffliche Abbildung aller relevanten geowissenschaftlichen Aspekte, sondern das Ziel, jedes Datenangebot durch mindestens ein Schlagwort charakterisieren zu können. Dementsprechend ist die Abstimmung gemeinsam nutzbarer Schlagwortlisten ein weiteres Thema im Metadatenkreis der Brandenburgischen Geodateninfrastruktur.

Erfassung ISO 19115-konformer Metadaten

Die Erfassung der Metadaten ist zur Zeit noch nicht automatisiert in dem Sinne, dass bestehende Daten älterer Formate oder aus anderen Systemen ohne weiteres konvertiert werden können. Auch die Erstellung von Seriendateien mit jeweils nur punktuell abweichenden Einträgen wird noch nicht unterstützt. Schnittstellen zu den häufig genutzten Spreadsheet-Applikationen wie z.B. MS Excel sind nicht vorhanden.

Sowohl LGRB als auch LGB betreiben jedoch bereits OGC- und ISO-konforme WWW-Produktkataloge. Am Geologischen Landesamt wurde zunächst ein einfacher Texteditor zur Metadatenerfassung genutzt. Da hierbei jedoch auch Formatelemente eingegeben und deren Hierarchie und Reihenfolge eingehalten werden müssen, ist aufgrund der Komplexität der XML-Syntax und des Metadatenformats die Fehleranfälligkeit sehr groß. Von der Firma Delphi IMM wurde daher ein spezieller Editor entwickelt, der mit einer nutzerfreundlichen Oberfläche stets sowohl syntaktisch richtige, sog. „wohlgeformte“, als auch regelgerechte, sog. „gültige“ XML-Dokumente erzeugt. Mit Hilfe des Editors lassen sich ebenfalls Templates für Serien von Metadatensätzen verwenden. Die Komplexität der XML-Struktur bleibt dem Nutzer verborgen, so dass auch Mitarbeiter ohne XML-Kenntnisse Metadaten erfassen und sich ganz auf den Inhalt der Dokumente konzentrieren können. Für die Beschlagwortung von Lokationen wurde das offizielle Ortsnamensverzeichnis von Brandenburg eingearbeitet, des weiteren werden Schlagworte für den Paläozeitbezug und für die thematische Datenbeschreibung zur Verfügung gestellt. Die Software lädt die Schlagwortlisten aus Konfigurationsdateien, die der Nutzer selbst verändern kann.

Folgende Möglichkeiten einer Eingabe sind somit derzeit realisiert:

- Editor der Firma Delphi IMM, basierend auf der Document Type Definition (DTD) der ISO 19115, Stand August 2002 (entspricht dem aktuellen ISO/DIS 19115),
- frei verfügbare XML-Editoren
- Altova XML-Spy und andere kommerzielle Editoren.

-

Ein Werkzeug zur automatisierten Generierung von ISO 19115-konformen Metadaten über einen Editor und diverse Schnittstellen zu anderen Systemen befindet sich zurzeit am GFZ Potsdam in der Entwicklung. Ebenso sollen Visualisierung und Validierung der Metadaten unterstützt werden.

Systemtechnische Anwendung des Brandenburgischen ISO 19115-Profils

Applikationen zur Nutzung der Metadaten ermitteln und verarbeiten i.d.R. lediglich die im spezifischen Kontext interessierenden und bekannten Fragmente aus den Metadaten. Sie benötigen nicht notwendigerweise eine vollständige Validierung der Dokumente, nur die Gültigkeit hinsichtlich der Struktur und der Datentypen gemäß ISO ist unerlässlich. In den meisten Fällen werden Informationen an den Nutzer durch Katalogservices (s.u.) lediglich weitergegeben, wenn beispielsweise eine Auswahl aus Optionslisten eingetragen wurde. Eine Validierung ist hier nicht nötig. Für regionale oder abweichende Ansprüche kann das Metadatenmodell der ISO 19115 erweitert werden, wenn gewährleistet ist, dass zusätzliche, abweichende Informationen verarbeitet werden können. Insofern können bspw. bestehende Optionslisten erweitert werden, auch wenn die Gültigkeit gegenüber der ISO 19115-Norm streng genommen nicht mehr gegeben ist. Ein Beispiel zeigt folgender Ausschnitt aus dem Application Schema der NIMA bzgl. des Datentyps Längeneinheit nach ISO 19115:

```
<xsd:simpleType name="LengthUnitNameType">
  <!-- Beschränkung der Wertemenge “Zeichenkette” auf eine Liste von
vorgegebenen Begriffen --> <xsd:restriction base="iso19103:CharacterString">
    <xsd:enumeration value="kilometer"/>
    <xsd:enumeration value="meter"/>
    <xsd:enumeration value="decimeter"/>
    <xsd:enumeration value="mile"/>
    <xsd:enumeration value="yard"/>
    <xsd:enumeration value="foot"/>
    <xsd:enumeration value="nanometer"/>
    <xsd:enumeration value="micrometer"/>
  </xsd:restriction>
</xsd:simpleType>
```

Im Falle einer Erweiterung der Optionsliste um einen Wert, z.B. um

```
<xsd:enumeration value="centimeter"/>
```

würde eine Überprüfung auf Gültigkeit keine Übereinstimmung mit der ISO 19115-Norm erzielen. Eine Anwendung zur Auswertung der Metadaten muss davon ausgehen, dass die Bezeichnung dieser Optionslisten und deren Datentyp mit dem der ISO19115-Norm übereinstimmt, nicht aber die Wertemenge, wenn ein Datensatz bspw. in Zentimetern erfasst wurde. In dieser Hinsicht sind Abweichungen von der Norm also durchaus sinnvoll und zulässig.

Schlagwortlisten erweitern das ISO 19115-Modell dahingehend, dass sog. „Parsable Character Data“ (PCDATA), frei wählbare Zeichenketten, auf bestimmte Begriffe zur Auswahl eingeschränkt werden. An der Gültigkeit der Norm ändert dies nichts, da lediglich eine Untermenge aus den Vorschriften genutzt wird.

Beispiel „Keyword“: Nach ISO ist die Verwendung beliebiger Zeichenketten zur Beschreibung von Daten mit spezifischen Schlüsselwörtern erlaubt. Im Brandenburgischen Profil wird eine Menge von für Brandenburg relevanten Begriffen vorgegeben:

Datentyp: Keyword (ISO 19115)

```
<!-- ISO19115 Norm  -->
<!-- hier wird der Datentyp auf eine frei wählbare Zeichenkette festgelegt der Mindestlänge 1 -->
<xsd:simpleType name="MD_Keyword_CodeList">
    <xsd:restriction base="xsd:string">
        <xsd:minLength value="1"/>
    </xsd:restriction>
</xsd:simpleType>
```

Datentyp: Keyword (Brandenburger Profil)

```
<!-- Erweiterung für das Brandenburger Profil -->
<!-- Festlegung des Datentyps auf einen aus einer vorgegebene Menge von Zeichenketten frei wählbaren Begriff -->
<xsd:simpleType name="MD_Keyword_CodeList">
    <!-- Beschränkung der Wertemenge “frei wählbare Zeichenkette” auf eine Liste vorgegebener Begriffe -->
    <!-- Ableitung aus einer frei wählbare Zeichenkette -->
    <xsd:restriction base="nonNullStringType">
        <xsd:enumeration value="Geologie"/>
        <xsd:enumeration value="Geographie"/>
        <xsd:enumeration value="Geophysik"/>
        <xsd:enumeration value="Geochemie"/>
    </xsd:restriction>
</xsd:simpleType>
```

Diese Erweiterung um eine Schlagwortliste ermöglicht zum einen die einheitliche Charakterisierung von Daten mittels Schlüsselwörtern und zum anderen die Führung des Nutzers bei seiner Datenrecherche (s.o.).

Die Nutzbarkeit der Metadaten gemäß des Brandenburger Profils soll noch 2003 in einem Testbed erprobt werden. Dazu werden in den beteiligten Einrichtungen Metadaten für ein konkretes Anwendungsszenario über Katalogservices bereitgestellt. Ein Katalogservice im Sinne der Spezifikationen des Open GIS Consortiums (OGC) ist ein WWW-Service, der auf eine Anfrage in einem definierten Format (z.B. CQL – Common Query Language) in einer definierten Art und Weise antwortet. Demnach spezifiziert zunächst ein Nutzer, der bestimmte raumbezogene Daten sucht, mit Hilfe einer grafischen Nutzeroberfläche diverse Suchkriterien. Die Anfrage gelangt über einen Katalog-Gateway zu den Katalog Servern, die existierende Metadatensätze zu Geodaten bereitstellen und aus diesen die der

Suchanfrage des Nutzers entsprechenden Ergebnisse selektieren. Eine vereinfachte Darstellung der Arbeitsweise soll folgende Abbildung verdeutlichen:

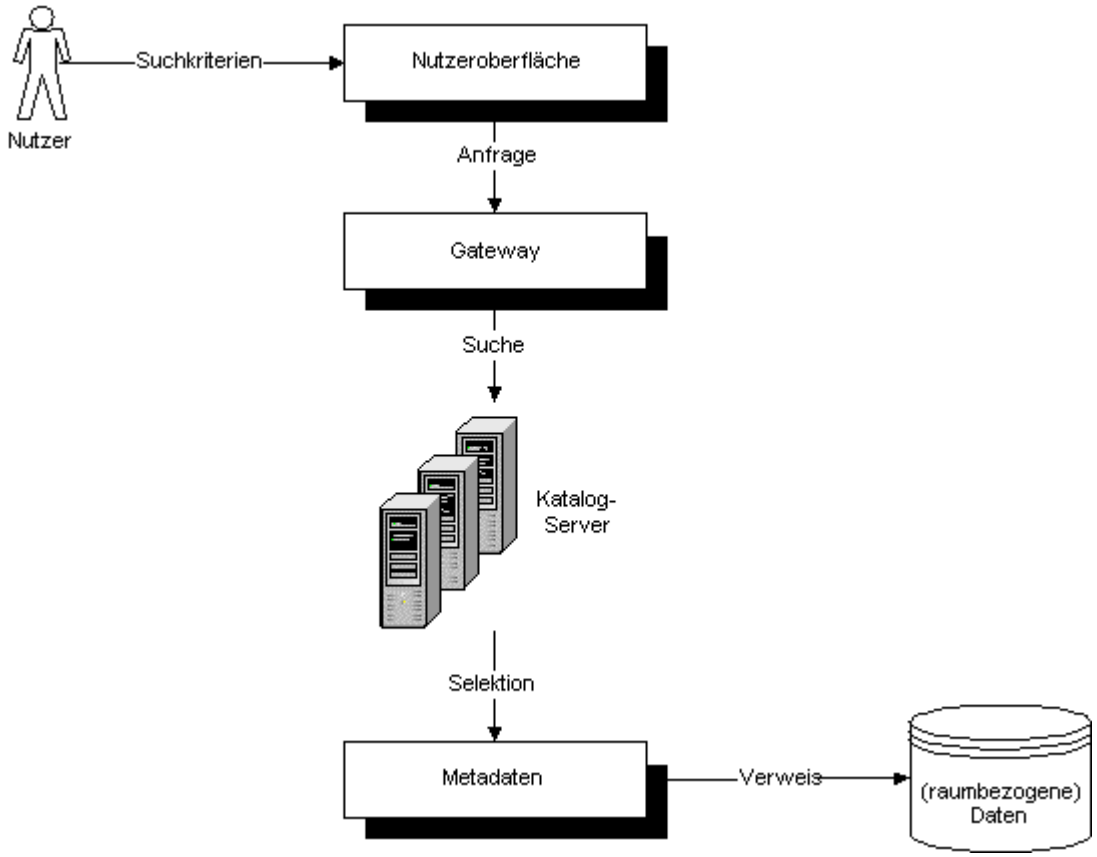


Abbildung 1: Funktionsweise von Anfragen an Katalogsystemen (KALMES 2000, S. 28, verändert nach NEBERT 2001, S. 45)

Austausch mit anderen Metadateninitiativen

Über die in der SIG Metadaten der Geodateninfrastruktur Brandenburg vertretenen Umwelteinrichtungen ist eine enge Kooperation mit der Expertengruppe des Umweltdatenkatalogs (UDK) gewährleistet. Der ebenfalls enge Kontakt zu den Staatlichen Geologischen Diensten, welche im Begriff sind, einen gemeinsamen WWW-Katalog zu installieren, besteht über das Landesamt für Geowissenschaften und Rohstoffe. Das Geoforschungszentrum Potsdam stellt den Austausch mit der Initiative GDI NRW sowie der Fachgruppe Geodaten des Deutschen Dachverbandes für Geoinformation (DDGI) her, während Landesvermessung und Geobasisinformation Brandenburg die Verbindung zur Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder (AdV) gewährleistet. Im Folgenden wird auch der Kontakt zum Bundesamt für Kartographie und Geodäsie und dem Vorhaben GeoMIS.Bund gesucht werden.

Bis zum Sommer diesen Jahres wird die Gruppe eine umfassende Publikation der Ergebnisse ihrer Arbeit liefern und damit alle Gruppen mit ähnlicher Aufgabenstellung zur Diskussion und Kooperation einladen. Nach der bevorstehenden Verabschiedung der ISO 19115 als Standard und der damit verbundenen Aktualisierung der Ergebnisse der Arbeitsgruppe werden diese ohne Einschränkungen zur Verfügung stehen. XML-Schemata, Musterdatensätze und schließlich die Optionslisten werden über das WWW zur Verfügung gestellt werden, Ansprechpartner hierzu ist der Sprecher der SIG Metadaten (lochter@lgrb.de).

Umsetzung des Brandenburgischen ISO19115-Profils im Rahmen des Deutsches Forschungsnetz Naturkatastrophen und im behördlichen Produktvertrieb

Katastrophenbewältigung als Anwendungsszenario für den Einsatz des Brandenburgischen Metadatenprofils

Das „Deutsche Forschungsnetz Naturkatastrophen“ (http://dfnk.gfz-potsdam.de/) ist ein Forschungsverbund von 15 Einrichtungen (Universitäten, Behörden, DWD, Münchener Rückversicherung etc.) mit dem Ziel, wissenschaftliche Grundlagen und neue Methoden für ein modernes Management von Naturkatastrophen zu erarbeiten und potentiellen Nutzern zur Verfügung zu stellen. Finanziert durch das Bundesministerium für Bildung und Forschung und unter Leitung des GeoForschungsZentrums Potsdam konzentriert sich die wissenschaftliche Methodenentwicklung auf die Naturkatastrophen Hochwasser, Waldbrand, Stürme und Erdbeben.

Im Rahmen des DFNK-Informationsmanagements sollen alle relevanten Daten und Informationen über und aus DFNK internen und externen Partnern zugänglich gemacht werden. Zur Umsetzung dieses Ziels wurde ein Internetportal aufgebaut und ein Katalogservice eingerichtet. Während das Portal den WWW-basierten Zugang zu allgemeinen und internen Informationen und Dokumenten für die interessierte Öffentlichkeit sowie die Teilnehmer des Projektes eröffnet, unterstützt der Katalogservice den Überblick über die im DFNK vorhandenen und dezentral gehaltenen Daten bzw. eine gezielte Datenrecherche. Diese kann nach den Suchkriterien Themenbezug (Schlagworte), Zeitbezug (Zeitrahmen, für den die Daten gültig sein sollen) und Raumbezug (Raumausschnitt, den die Daten abdecken sollen) erfolgen.

Dieses sog. „DFNK-Clearinghouse“ ist ein auf Metadaten basierender Informationsdienst. Die Umsetzung erfolgte nach dem „Open Source“-Ansatz und darüber hinaus komponentenbasiert, was den Vorteil hat, dass bei Bedarf zusätzliche Komponenten entwickelt und integriert sowie vorhandene Module angepasst, umgekehrt jedoch auch Komponenten entfernt werden können. Folgende Komponenten sind realisiert worden und in Abbildung 2 dargestellt:

- Kernsystem: Metadaten-Editor, Metadaten-Speicher, Recherchesystem und WWW-Server
- Visualisierungskomponente (Map Server)
- Nutzerschnittstelle
- Thesaurusmodul

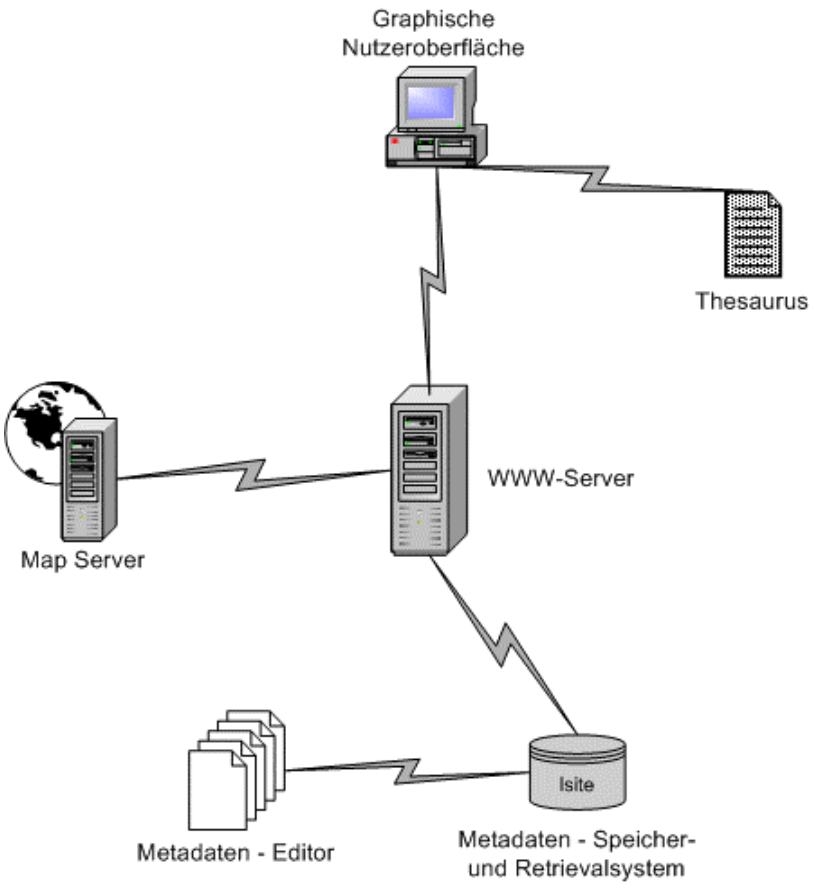


Abbildung 2: Architektur des DFNK-Clearinghouse

Die Metadaten, die bislang im Directory Interchange Format (DIF) vorlagen und zu deren Erfassung ein Editor auf der Basis von MS Access entwickelt worden ist, umfassen Angaben zum Raum-, Zeit- und Themenbezug der Daten sowie zum

Erzeuger, zur Datenqualität etc. und beinhalten schließlich einen Verweis auf den jeweiligen Eigner und Ort der entsprechenden Daten. Die Umstellung auf ISO 19115 ist bereits frühzeitig beschlossen worden. Zunächst wurde ein Mapping des bisherigen DIF-Datenmodells vorgenommen bis hin zu einem entsprechenden ISO-Profil auf der Basis der DIS-Version von August 2001. Nahezu alle bislang genutzten DIF-Attribute ließen sich auch weiterhin abbilden, in wenigen Fällen war eine konzeptionelle Überarbeitung des Datenmodells notwendig. Im Rahmen der aktiven Beteiligung an den Aktivitäten der SIG Metadaten erfolgte anschließend der Abgleich des DFNK-Modells mit dem Brandenburgischen ISO 19115-Profil. Die Übertragung dieses Profils in ein solches Anwendungsprojekt bietet für das Projekt einen wertvollen Input und leistet zum anderen für die SIG einen Beitrag zur Validierung ihrer Ergebnisse in der Praxis.

Stellt man sich exemplarisch die Abfrage nach aktuellen Wasserstandspegeln im Fall eines Hochwasserereignisses vor, bietet sich dem Nutzer des DFNK-Clearinghouse die Möglichkeit der alphanumerischen Recherche z.b. über die Auswahl oder Eingabe geeigneter Suchbegriffe bzw. der graphischen Recherche über einen integrierten Map Server in Form digitaler Karten. Hier kann der Nutzer den gewünschten Raumausschnitt durch Markierung eines Rechtecks (Einbettungsrechteck) geographisch eingrenzen und darüber hinaus sowohl einen Schlagwortkatalog zur thematischen als auch ein Ortsnamensverzeichnis zur raumbezogenen Spezifizierung nutzen.

Die Pegelmesspunkte in dem gewählten Raumausschnitt werden schließlich aufgelistet oder als verweissensitive Symbole auf einer Karte dargestellt. Die entsprechenden Metadaten können bislang im Format DIF und zukünftig auch als ISO 19115-konforme XML-Information abgerufen werden. Die aktuellen Pegelstandsdaten der Bundesanstalt für Gewässerkunde in Koblenz beispielsweise stehen schließlich über einen Hyperlink in tabellarischer und graphischer Form zur Verfügung.

Anwendungsszenario Produktvertrieb des Landesamtes für Geowissenschaften und Rohstoffe Brandenburg

Zum 10. Jahrestag des Landesamtes für Geowissenschaften und Rohstoffe Brandenburg am 09. April 2002 sowie zum „Jahr der Geowissenschaften“ hat das LGRB, als erstes geologisches Landesamt der Bundesrepublik Deutschland einen Internet-Katalogservice verfügbar gemacht (http://katalog.lgrb.de), mittlerweile sind

ca. 10000 Hits pro Monat zu verzeichnen. Das System dient primär dem Suchenden, das Konzept ist auf die Nutzer des LGRB ausgerichtet. Der Kunde soll über einen Standard-Internetbrowser (ohne zusätzliche Plug-ins) räumlich und inhaltlich nach Geodaten-Produkten recherchieren und diese einfach bestellen können.

Der Katalogservice besteht aus Modulen, die über klar definierte Schnittstellen miteinander arbeiten. Derzeitig sind folgende Module installiert:

- OGC-konformer WWW-Map Server (Internet-Landkarten),
- OGC-konformer WWW-Katalogserver (Haltung/Recherche der ISO-basierten Metadaten),
- WWW-Portal (Nutzeroberfläche im Internet).

Als Map Server wird ein „Open Source"-Standardprodukt eingesetzt, welches frei im WWW verfügbar und konform zu den Standards des OGC ist. Der Katalogserver ist ein WWW-Service, der über eine international genormte Abfragesprache Recherchebedingungen zu Raum-, Sach- und Zeitbezug über Produkte entgegennimmt, eine Produktrecherche startet und die Treffer konform zu ISO-Standards liefert. Er ist das Metainformationssystem zu den Produkten. Da ein solcher Katalogserver ein klar definiertes Verhalten hat, kann er nicht nur in der Infrastruktur des LGRB sondern auch aus anderen Projekten abgefragt werden. So wird nur ein Metainformationssystem gepflegt, das aber für mehrere Projekte aus dem WWW heraus abfragbar ist

Die Einstellung der Metadaten in den Katalogservice erfolgt – ebenso im Deutschen Forschungsnetz Naturkatastrophen – je nach Bereitstellung durch den Administrator des Katalogservers. Dafür wurden umfangreiche Routinen erarbeitet, die sowohl das Einchecken als auch die Qualitätskontrolle der Metadaten übernehmen. Korrupte Dokumente werden nicht eingecheckt und der Administrator erhält die Liste der korrupten Dokumente mit Angabe des Fehlers. In Diskussion ist die Möglichkeit, über ein WWW-Interface die Dokumente einzuchecken. Dies hat aber auch erhebliche Nachteile, weshalb diese Diskussion noch nicht abgeschlossen ist.

Zur Wartung der Metadaten sind Werkzeuge in der Planung, die über Stylesheet-Transformationen (XSLT) einfache Änderungen an ganzen Gruppen von Dokumenten erlauben. Ändert sich z.B. der Preis der Produktgruppe „Geologische

Karten“ so ist dieser in einem Schritt an 275 Metadokumenten zu aktualisieren. Hier wäre eine händische Bearbeitung jedes Dokuments aufwendig und fehlerintensiv.

Die gesamte am LGRB im Einsatz befindliche und erprobte Technologie wurde unter „Open Source“-Gesichtspunkten erstellt und kann kostenfrei nachgenutzt werden. So werden die in Brandenburg entwickelten Module des WWW-Katalogservice im Vorhaben „INFOGEO.DE“ wieder verwendet (CZEGKA & LOCHTER 2003). Hier erarbeiten die IT-Verantwortlichen der Staatlichen Geologischen Dienste (SGD) in einer Arbeitsgruppe der Länder Niedersachsen, Mecklenburg-Vorpommern, Baden Württemberg und Brandenburg gemeinsam mit der Bundesanstalt für Geowissenschaften und Rohstoffe (BGR) eine erste Version eines gemeinsamen WWW-Portals (http://www.infogeo.de). Dort sollen

- allgemeine Informationen über die Aufgaben der SGD, aber insbesondere
- Angaben über Leistungen/Produkte der SGD und deren Bezugsbedingungen

bereitgestellt werden. Damit wird dieses Portal zentraler Einstiegspunkt für alle geologischen Dienste in Deutschland. Die Produkte und Leistungen der SGD sollen über einen WWW-Katalogservice zur Verfügung gestellt werden. Die mögliche Architektur dieser Lösung ist in Abbildung 3 dargestellt.

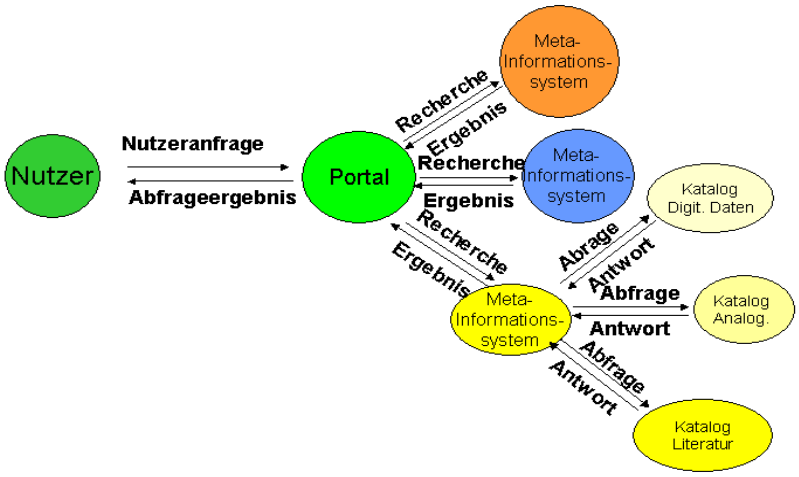


Abbildung 3: Skizze einer möglichen Architektur des Portals der SGD „INFOGEO.DE“

Dabei wird auf die Einhaltung der ISO 19115 (Mindestdatensatz gemäß "Essential Profile") zur Erstellung der Metadaten geachtet. Das LGRB bietet allen Interessierten

die kostenneutrale Nutzung seiner Software an, die einzelnen Module werden gemeinsam und kostenteilig weiterentwickelt.

Ausblick

Nach erfolgreicher Verabschiedung des Standards ISO 19115 wird eine Übertragung der Ergebnisse der SIG Metadaten auch auf zukünftige Projekte und Anwendungen angestrebt. So ist eine Verwendung im Rahmen des „Center for Disaster Management and Risk Reduction Technology“ (CEDIM), einer gemeinschaftlichen Einrichtung des GFZ Potsdam und der Universität Karlsruhe, bereits vorgesehen. Katastrophenforschung und -management ist jedoch als nur eines von vielen Anwendungsfeldern zu sehen, wo die Dokumentation von Daten auf der Basis standardisierter Metadaten messbare Mehrwerte liefern kann. Werkzeuge wie das dargestellte Clearinghouse eignen sich auch in anderen Fachgebieten für die dynamische Verschneidung von Geodaten unterschiedlicher Herkunft und Ausprägung zur Abdeckung eines komplexen Informationsbedarfes. Weitere Anwendungsszenarien können z. B. Tourismus-Informationssysteme oder Bürgerinformationssysteme zur Nutzbarkeit von Geothermie sein.

Es ist festzustellen, dass vernetzte Produktkataloge für Geodaten i.A. von größerer Bedeutung sind als singuläre Angebote eines Providers. So wird ein gebündeltes themenspezifisches Angebot (z.B. Topographische Daten) mehr Beachtung finden als ein Portal eines einzelnen Anbieters. Der Grundsatz der dezentralen Erfassung aller beschreibenden Produktinformation (Metadaten) bei den spezifischen und kompetenten Datenprovidern kann eingehalten werden, obwohl der Nutzer über ein Portal recherchieren kann. In den USA wird derzeit das Geospatial One-Stop (GOS) Portal (http://ip.opengis.org/gos-pi/) vorbereitet. Es wird einen Schlüssel in der eGovernment -Strategie einnehmen und aktiv vom OGC unterstützt werden. Die brandenburgischen Metainformationssysteme sollen in kommerzielle, wissenschaftliche und administrative Netzwerke eingebunden werden. Ihre Weiterentwicklung soll sich darüber hinaus an internationalen Entwicklungen, z.B. INSPIRE und GSDI, orientieren.

Entwickler von Applikationen werden auf diese Art und Weise in der Lage sein, ihre Anwendungen über WWW-Katalogservices dynamisch und aktuell mit Daten zu versorgen. Grundlage sollte letztendlich die Gründung auf anerkannten Standards sein, um den ungehinderten Austausch und die anwendungsunabhängige und übergreifende Nutzung von Daten und Informationen im Sinne einer Infrastruktur für Geoinformation und Geoinformationsdienste zu gewährleisten. Der Aufbau der Geodateninfrastruktur Brandenburg ist damit eine Investition ähnlich dem Aufbau einer Verkehrsinfrastruktur (Straßen, Flusswege, Luftlinien), die eine Vielzahl von (Inter)Aktionen erst ermöglicht. Der Aufbau einer Nationalen Geodateninfrastruktur ist eine originäre Aufgabe des Staates, die durch derartige regionale Aktivitäten und die Einbeziehung der Industrie unterstützt werden kann und einen wichtigen Faktor

Literatur:

CZEGKA, W., LOCHTER, F. A. (2003): Integration durch Standardisierung und Modularisierung: Die Einbindung eines geologischen Dienstes in nationale Geodatenintiativen am Beispiel des LGRB – Beitrag zur AGIT 2003, Salzburg (in Vorbereitung)

ISO/TC 211 (2003): ISO/FDIS19115 Geographic information – Metadata, Genf

KALMES, P. (2000): Verwaltung zweidimensionaler Geodaten mit Hilfe von Metadaten und ihre Visualisierung im World Wide Web – Unveröffentlichte Diplomarbeit, Universität Trier

NEBERT, D., Hrsg. (2001): Developing Spatial Data Infrastructures: The SDI Cookbook, http://www.gsdi.org/pubs/cookbook/cookbook0515.pdf

ISO 19115: Entwicklung und Einsatz eines Brandenburgischen Profils

am Beispiel des Deutschen Forschungsnetz Naturkatastrophen und des behördlichen Produktvertriebs

P. Köhler, F. A. Lochter & R. Häner
19. Mai 2003

Gliederung des Vortrags

Aufbau einer Geodateninfrastruktur Brandenburg

Entwicklung eines Brandenburgischen ISO 19115-Profils

Anwendungsszenarien: Katastrophenbewältigung und Produktvertrieb

Ausblick

GFZ POTSDAM

Geodateninfrastruktur Brandenburg

Ziele

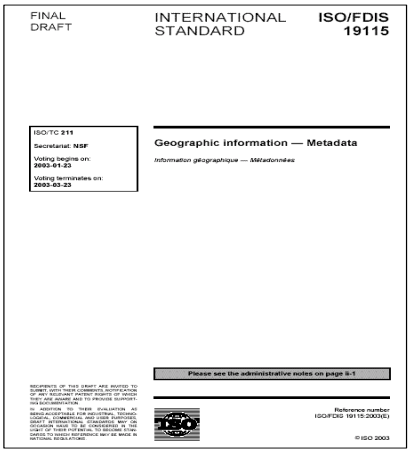
- einrichtungsübergreifender Austausch und ungehinderte interdisziplinäre Nutzung von Geodaten
- Vorbereitung wesentlicher Voraussetzungen für die erfolgreiche Umsetzung von eGovernment
- Aufbau eines transparenten Geodatenmarktes in Brandenburg
- Einrichtung einer Kommunikations- und Kompetenz-plattform für Verwaltung, Wirtschaft, Wissenschaft und Gesellschaft

GFZ POTSDAM

Entwicklung eines Brandenburgischen ISO 19115-Profils

Vorgehensweise

- „Special Interest Group“ Metadaten: Delphi IMM, GFZ, LGB, LGRB, LUA, MLUR
- Arbeitsgrundlage: ISO/DIS 19115, ISO/FDIS 19115
- Diskussion und Festlegung von Elementen und Attributen zur Beschreibung von Geodaten
- Entwicklung eines „Branden-burgischen Application Schemas“ auf Grundlage der NIMA


FINAL DRAFT

INTERNATIONAL STANDARD

ISO/FDIS 19115

ISO/TC 211

Secretariat: NSF

Voting begins on: 2003-01-23

Voting terminates on: 2003-03-23

Geographic information — Metadata

Information géographique — Métadonnées

Please see the administrative notes on page ii-1

Reference number ISO/FDIS 19115:2003(E)

© ISO 2003

GFZ POTSDAM Entwicklung eines Brandenburgischen ISO 19115-Profils

Ergebnisprofil

- Brandenburgisches Metadatenprofil konform zur ISO 19115
  - Nutzung des „essential profile“ und ausgewählter weiterreichender Elemente
  - Entsprechung der vorgegebenen Semantik
  - Erhalt der definierten Verbindungen innerhalb der ISO 191xx-Familie
- Ausführung in XML und gemäß Brandenburger Application Schema
- geplant: Abstimmung von Optionslisten und Schlagwortkatalogen

GFZ POTSDAM Entwicklung eines Brandenburgischen ISO 19115-Profils

Systemtechnische Umsetzung

- Metadatenerfassung über
  - Editor der Firma Delphi IMM
  - frei verfügbare XML-Editoren
  - Altova XML-Spy und andere kommerzielle Editoren
  - In Vorbereitung: Werkzeug zur automatisierten Generierung von ISO-konformen Metadaten (GFZ Potsdam)
- Systemtechnische Umsetzung über Katalogservices auf der Basis des Brandenburger Application Schemas
- geplant: Testbed unter Beteiligung aller Partner

GFZ POTSDAM Entwicklung eines Brandenburgischen ISO 19115-Profils

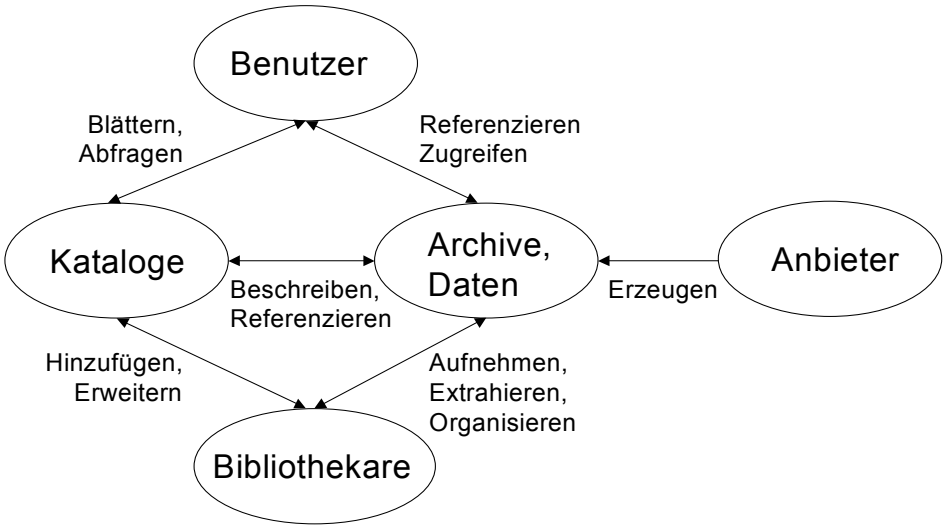
Umsetzung in Katalogservices

Katalogservice: Gesamtheit aller Schnittstellen zur Unterstützung des Managements von und der Informationsgewinnung aus raumbezogenen Daten durch:

- Organisation und Management der Daten (Metadaten)
- Unterstützung beim Auffinden von Daten (Retrieval)
- Unterstützung bei der Bereitstellung von Daten und Metadaten (Metainformationssystem)

GFZ POTSDAM Entwicklung eines Brandenburgischen ISO 19115-Profils

Katalogservice: Funktionsweise



GFZ POTSDAM Entwicklung eines Brandenburgischen ISO 19115-Profils

Austausch mit anderen Initiativen

- Umweltdatenkatalog (UDK)
- GDI NRW
- Deutscher Dachverband für Geoinformation e.V. (DDGI)
- Arbeitsgemeinschaft der Vermessungsverwaltungen (AdV)
- Staatliche Geologische Dienste (SGD)
- Bundesamt für Kartographie und Geodäsie (BKG)

➡ **Generell: Uneingeschränkte Bereitstellung der Ergebnisse!**

GFZ POTSDAM Anwendungsszenarien

Deutsches Forschungsnetz Naturkatastrophen

- Interdisziplinärer Ansatz: Verknüpfung von Natur-, Ingenieur- und Sozialwissenschaften sowie der Informatik
- 15 Partner: Universitäten, Forschungseinrichtungen, Deutscher Wetterdienst (DWD), Münchener Rück etc.
- Ziel: Entwicklung wissenschaftlicher Grundlagen und neuer Methoden für ein modernes Risikomanagement zum Schutz von Leben und Eigentum vor natürlichen Katastrophen

DFNK
Deutsches Forschungsnetz Naturkatastrophen

GFZ POTSDAM Anwendungsszenarien

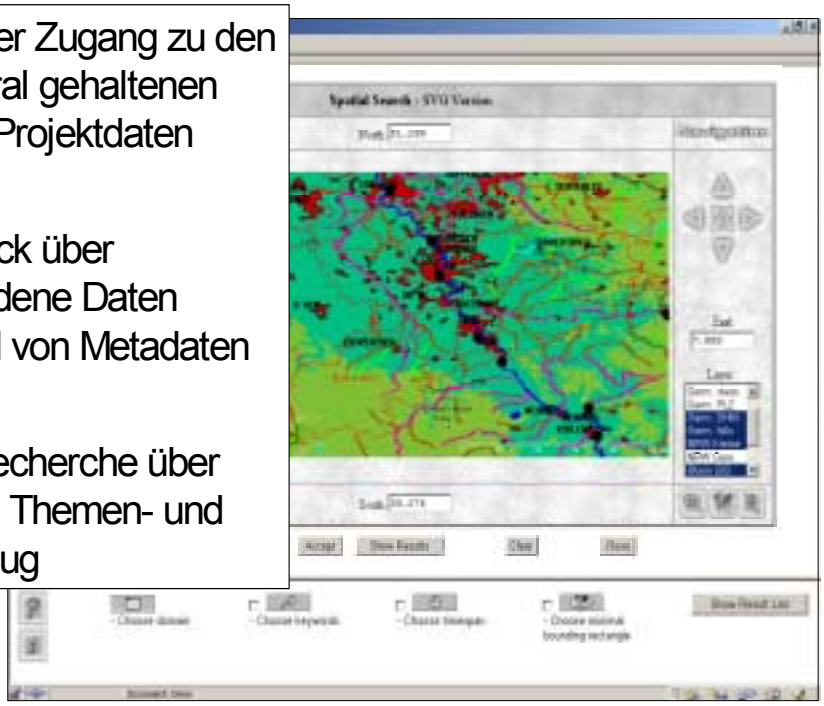
Das DFNK-Clearinghouse

- Kernsystem: Metadaten-Editor, Speicher- und Recherchesystem, WWW-Server
- Visualisierungskomponente (Map Server)
- Nutzerschnittstelle
- Thesaurusmodul

GFZ POTSDAM Anwendungsszenarien

Datenretrieval über Metadaten

- Zentraler Zugang zu den dezentral gehaltenen DFNK-Projektdaten
- Überblick über vorhandene Daten anhand von Metadaten
- Datenrecherche über Raum-, Themen- und Zeitbezug



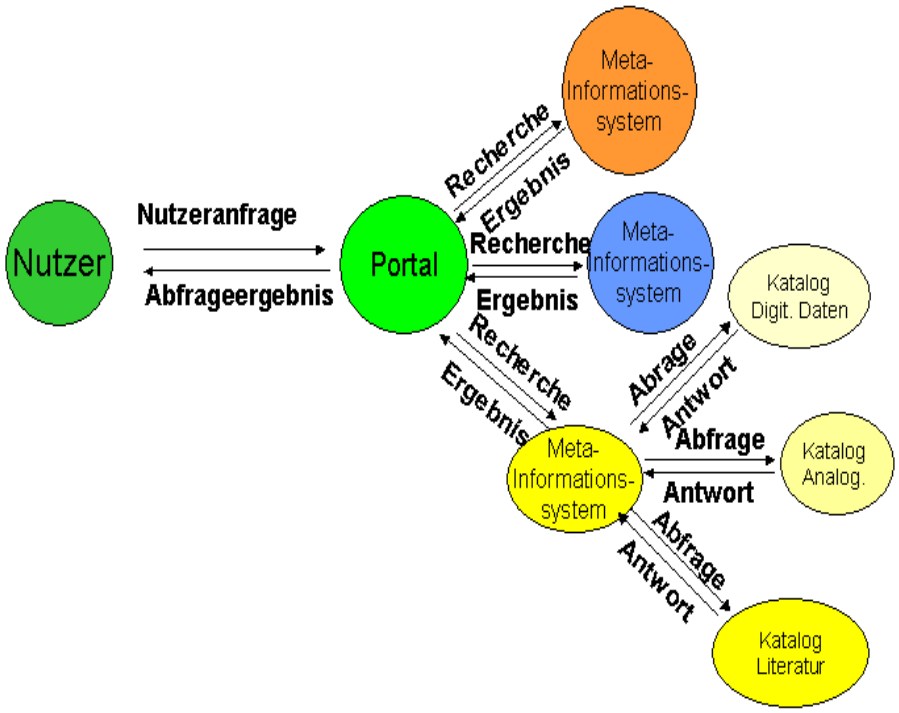
GFZ POTSDAM Anwendungsszenarien

Produktvertrieb des LGRB

- OGC- und ISO-konformer Produktkatalog (http://katalog.lgrb.de)
  - OGC-konformer WWW-Map Server
  - OGC-konformer WWW-Katalogserver
  - WWW-Portal
- Realisierung auf „Open Source“-Basis
- Verwendung und Weiterentwicklung der Technologie-Baussteine im Rahmen des WWW-Portals der Staatlichen Geologischen Dienste „INFOGEO.DE“

GFZ POTSDAM Anwendungsszenarien

Architekturentwurf INFOGEO.DE



Ausblick

- Aktualisierung der Arbeitsergebnisse nach erfolgter Verabschiedung von ISO 19115
- Veröffentlichung und Freigabe des Brandenburgischen Profils und Application Schemas
- Einsatz in Einrichtungen und Projekten, z.B. im Center for Disaster Management and Risk Reduction Technology
- Einbindung brandenburgischer Metainformationssysteme in Netzwerke
- Orientierung an internationalen Entwicklungen wie INSPIRE und GSDI

Die Framework Architektur des Umweltobjektkatalogs (UOK) Bayern

Eric Weihs[1]

Kurzfassung

Der Umweltobjektkatalog (UOK) des Informationssystems für Planung und Umwelt (ISPU) des Bayerischen Staatsministeriums für Landesentwicklung und Umweltfragen, München, ist mittlerweile nicht nur IT- technische Grundlage des Nachweises, wo welche Umweltdaten des Geschäftsbereichs des Bayerischen Staatsministeriums für Landesentwicklung und Umweltfragen zu finden sind. Die durchgängige XML- basierte Frameworktechnologie ermöglicht über das Klassenkonzept der Umweltdatenkataloge hinaus das Monitoring[2] von Verwaltungs-Aufgaben. Damit wird auch eine Abgrenzung zu dem unscharfen Begriff der Metadaten deutlich.

Die Diskussion über die Dateninhalte der Katalog- bzw. Verweissysteme wird in Hinblick auf die Umsetzung des novellierten Umweltinformationsgesetzes (UIG), insbesondere der §4 und 12 erforderlich: Der nach dem UIG zu befriedigende Informationsbedarf der Öffentlichkeit geht über die derzeitigen Dateninhalte der Katalogsysteme deutlich hinaus. Die Katalogsysteme werden ihre Bedeutung nur behalten, wenn sie sich den wachsenden Anforderungen stellen. Die Framework-Technologie des UOK ermöglicht eine kontinuierliche Anpassung an die steigenden Anforderungen.

1. Der Umweltobjektkatalog[3]

Der Nachweis, wo welche Daten über reale Objekte oder Abstraktionen davon unter welchen qualitativen und quantitativen Bedingungen *Medien übergreifend* in Bayern

[1] Bayerisches Staatsministerium für Landesentwicklung und Umweltfragen, München erich.weihs@stmlu.bayern.de

[2] Die beobachtende Beschreibung eines Prozessverlaufs legt die Bezeichnung Monitoring nahe

[3] In Weihs (1998) noch als Objektdatenkatalog bezeichnet

zu finden und zu bearbeiten sind, ist Aufgabe des Umweltobjektkatalogs (UOK). Er ist die zentrale Komponente des Kernsystems des Bayerischen Informationssystems für Planung und Umwelt (ISPU). *Medien übergreifend*[4] deswegen, weil es heute unabdingbar ist, umweltpolitische Fragestellungen aus einer ganzheitlichen Sicht heraus zu beurteilen. Ganzheitlich bedeutet, dass es nicht ausreichend ist, verschiedenen fachlich (letztlich willkürlich) abgegrenzte Medien sequentiell oder parallel zu bearbeiten, sondern im Sinne der geforderten Nachhaltigkeit holistisch zu begreifen und umzusetzen.

Wir hatten 1998 den zentralen Nachweis im Sinne von Coad und Yourdan (1994, 75) als Umwelt*objekt*katalog bezeichnet, da dieser nach den o.a. Kriterien klassifiziert und verweist u.a. auf

- reale Objekte wie z.B. Schutzgebiete, Landschaftselemente oder geologisch schutzwürdige Objekte (Geotopkataster) mittels Beschreibungen, Karten oder Bildern
- Abstraktionen darüber, wie z.B. Datenbestände Dokumentationen und Karten ähnlich dem Umweltdatenkatalog des Bundes und der Länder (UDK)
- Monitoring der Forschungsvorhaben des Geschäftsbereichs
- Monitoring der Twinning und Interreg Vorhaben
- Monitoring der Umweltverträglichkeitsverfahren in Polen (im Rahmen eines Twinning -Projektes)
- Klassifikationssysteme selbst (z.B. Grunddatenkatalog des Bundes und der Länder), GEMET, UBA-Thesaurus, Mikrothesauri[5], strukturierte Schlüssellisten (z.B. Artenlisten))
- Adressen,
- Benutzerprofile, Security etc.

Alle Objekte/Entitäten sind nach einer einheitlichen Methodik behandelt, die weiter unten dargestellt ist.

Die derzeit gebräuchliche Trennung zwischen den Metadaten der Objekte und den Fachdaten ist fließend: Mit zunehmender Detailschärfe der Beschreibung ist ein

[4] nicht nur die klassischen Umweltmedien Luft, Wasser, Landschaft usw. sondern auch Gesundheit, Raumplanung (z.B. UVP), Gesellschaft etc.

[5] Mit Mikrothesauri bezeichnen wir kleinere aber fachlich erforderliche spezialisierte Thesauri. Die fachliche Akzeptanz der Indexierung hängt von deren Bereitstellung ab.

stetiger Übergang von der einfachen Referenzirrung *auf einen* Datenbestand *zu diesem selbst* zu verzeichnen. Aus diesem Grunde weicht die strenge Trennung zwischen Fachdaten und Verweisdaten einer pragmatischen Lösung – insbesondere wenn es sich um weiterführende Einzelverweise auf reale Objekte handelt, die von geschäftsbereichsweitem Interesse sind wie Bild-Dokumentationen von Landschaftselementen oder geologisch schutzwürdigen Objekten, Schutzgebietsverweise usw. (Weihs 1998, 622). Die neuere Entwicklung des UOK zeigt die notwendige Erweiterung der dargestellten Inhalte um Prozesse wie UVP-Verfahren, Verwaltung von Forschungsvorhaben und internationalen Projekten im Rahmen von Twinning und Interreg usw. um diese im Sinne des Monitoring begleiten zu können. Das erforderliche „Bewertungswissen“ nach D. Rosenkranz (2003), zur Lösung ganzheitliche Fragestellungen wird neben Metadaten auch Wissen über Zustandsdaten, Methoden usw. erfordern. Die Katalogsysteme werden sich hier in Verbindung mit dem UIG neu ausrichten müssen, da sich letztlich das Bewertungswissen als Entscheidungsgrundlage bei Entscheidungsträger im Recht der Öffentlichkeit auf Information widerspiegelt.
Die XML kommt im UOK der Realisierung der IT- Applikation zur Umsetzung der oben genannten inhaltlichen Anforderungen durch ihre Flexibilität bei der Modellierung des offenen Datenmodells entgegen, ohne auf die Vorteile relationaler oder hierarchischer Modelle verzichten zu müssen (Weihs 1999, 2000).

2. Zur Methode des Umweltobjektkatalogs

2.1 Rückblick

Bereits bei ersten Realisierungsversuchen einer Landschaftsdatenbank für den Geschäftsbereich wurden die Datenbestände durch ein Dokument bezogenes Metainformationssystem beschrieben, auf welches die Software während der Ausführung zugreifen konnte (Weihs 1978, 1986). Das damalige Metafile „A“ enthielt bereits einen strukturierten und maschinenlesbaren Codeplan mit Angabe der Fundstelle der Daten, Feldbeschreibung, Erläuterungen usw.
In der ad hoc Arbeitsgemeinschaft Bodeninformationssysteme der Länderarbeitsgemeinschaft Bodeninformationssystem (LABO) wurde der Gedanke bei der Spezifikation eines Kernsystems aufgegriffen und weiterentwickelt. Ergebnis war eine Empfehlung der Umweltministerkonferenz zum Aufbau von Bodeninformationssystemen (Ad Hoc AG der LABO, Weihs et al. 1994), die Kernzelle

des Umweltobjektkataloges (UOK) wurde. Mit Beginn der Arbeiten zum Kartenkatalog (Koch - Steindl et al. l992), der noch heute ein wesentlicher Bestandteil des Umweltobjektkataloges ist, entstand die Idee, im für den Anwender lesbaren Text des Kartennachweises Formatierungsregeln transparent – aber dafür mit fachlich-inhaltlicher Aussage versehen – mittels Schlüsselwörtern (den späteren XML-Tags) zu hinterlegen. Damit waren die Grundlagen des späteren XML-Datenmodells für die Programmierung des UOK unter Berücksichtigung bereits frühzeitig gelegt.
Heute basiert der UOK konsequent auf ein Datenmodell in der XML, das in einer nativen XML-Datenbank realisiert ist. Der Klassenbegriff wurde verallgemeinert, um im Sinne eines am Verwaltungsprozess orientierten Monitorings[6] den Anforderungen des künftigen UIG Rechnung tragen zu können. Mit dem Framework Architekturkonzept wird die Realisierung anspruchsvoller prozessbegleitender „Kataloge“ in eGovernment Anwendungen wie UVP, Forschungsvorhaben, Projektverwaltung etc. möglich.

2.2 Architektur

Alle im UOK verwendeten Fach- und Sicherheitsdaten, die Ein/Ausgabe bestimmenden Stylesheets, usw. basieren ausnahmslos auf der XML. Auch die Datenhaltung aller Daten erfolgt in einer nativen XML-Datenbank. Wir haben eine native XML-Datenbank[7] gewählt, da diese die Objektstruktur wie von uns vorgesehen abbildet, die Objekte einfach referenziert werden können und die Abfragemöglichkeiten des XQL (X-Query) voll ausschöpft werden. Damit entstehen auch keine Performance-Verluste. Die gewählte Datenbank selbst ermöglicht eine direkte Anbindung an den Webserver und bietet die Vorteile kommerzieller Datenbanken wie Transaktionssicherheit, Logging, Recovering usw. wie das Management großer Datenmengen (derzeit etwa 160.000 Objekte).

[6] Vg. Fußnote 2: Die beobachtende Beschreibung eines Prozessverlaufs legt die Bezeichnung Monitoring nahe.
[7] XML-Datenbank Tamino der Software AG, Darmstadt

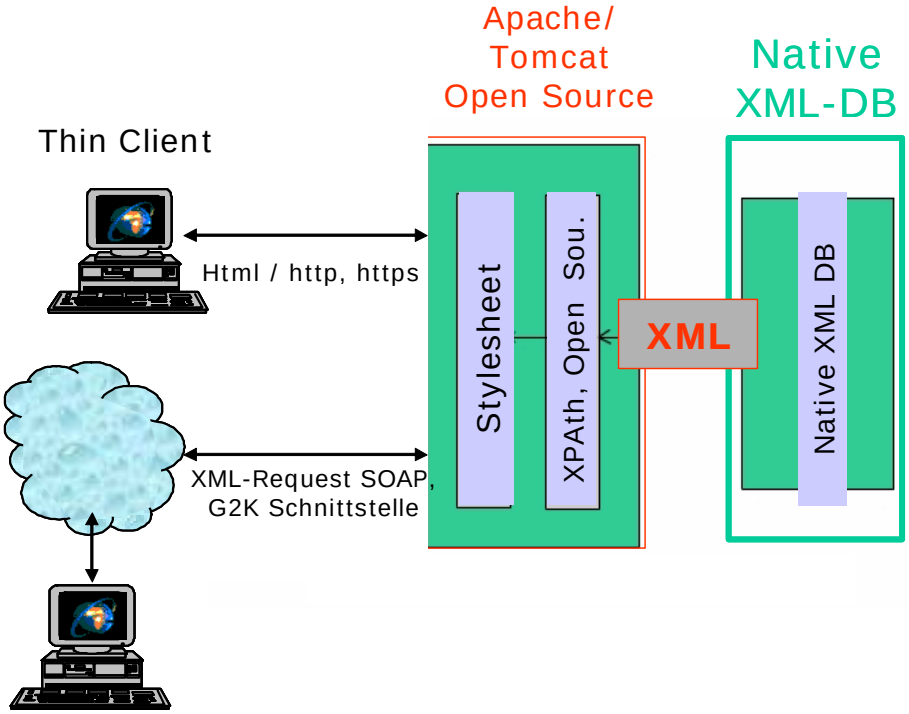


Abbildung 1: Überblick über die Systemarchitektur des Umweltobjektkatalogs

Im wesentlichen besteht der UOK nach Abbildung 1 aus den Komponenten Apache - Web-Server JAVA und Apache Tomcat-Server und der nativen XML Datenbank. Der Tomcat-Server fungiert als Webservice Container (JAVA-Servlets, Webdav) und verwaltet den Passwort geschützten Zugang. Die Logik des Systems ist im Wesentlichen in Stylesheets (XSLT) enthalten, welche die Kommunikation zwischen Client und XML Datenbank enthält. Damit wird ein hohes Maß an Flexibilität erreicht, da der Erweiterungs- und Änderungsaufwand sich auf die Stylesheets bezieht. Über die über den Stylsheet - Manager gesteuerten Stylesheets erfolgt der Call auf die XML-Datenbank. Über Stylesheets werden sowohl die SOAP- Requests gemappt (derzeit nur an den virtuellen Umweltdatenkatalog[8] der Länder) wie auch die get und post-Schnittstellen der G2K Schnittstelle für gein® (German Environmental Information Network) bedient. Für die http/s Requests der Web-Clients werden die XML- Daten in Verbindung mit den Stylesheets in html umgewandelt (XSLT–Prozessor (=Stylesheetprozessor)). Damit kann die Strategie eines schlanken Clients (thin client concept) beibehalten werden, da keine explizite Installation am Client erforderlich ist. Außer einem heute gängigen Browser ist kein weiterer Aufwand Client-seitig erforderlich.

[8] vUDK

Bis auf die Lizenz pflichtige[9] XML-Datenbank wird ausschließlich Open Source Software verwendet. Die gesamte Konfiguration ist Plattform unabhängig und auf Windows X wie UNIX oder LINUX ablauffähig. Die Entwicklung erfolgt derzeit auf Windows 2000 und Linux, die Produktionsumgebung läuft unter SUN - Solaris. Die dargestellte Konfiguration erlaubt – unter Berücksichtigung der Aufgabenstellung – eine Kosten- und Entwicklungszeitreduktion auf 20 bis 40 % zu vergleichbaren Funktionalitäten anderer Systeme.

2.3 Die Frame-Work Technologie

Da der UOK entsprechend der frühzeitig erkannten Aufgabenstellung nach 2.1 Objekte bzw. Entitäten ganz unterschiedlicher Art verwalten sollte, deren Gemeinsamkeit als die eines strukturierten Formulars *gesehen (!) werden kann,* achteten wir aus funktionalen Gründen – und um Entwicklungszeit zu sparen - auf die möglichst weitgehende Wiederverwendbarkeit der einzelnen Komponenten. Da man sich die Architektur als Rahmen vorstellen kann, dessen fachlich bestimmten „Bildinhalte“ (= Objekte; Entitäten) ausgetauscht werden können, wird diese Architektur als Frame - Work bezeichnet. Als „Bilder-Rahmen“ hat man sich Stylesheets vorzustellen, deren Funktionalität durch wieder verwendbare Servlets dargestellt wird. So werden die immer benötigten Funktionen zur Bearbeitung wie Erfassen, Aktualisieren, Löschen, Import, Export, Version[10], Rechte durch eine – wie aus Windows-Anwendungen bekannte - Funktionsleiste gesteuert, die sich abhängig vom Rechte Profil und der gewählten Sprache nach Abbildung 2.1 und 2.2 unterschiedlich darstellt.

[9] Die mögliche Eignung einer open source XML-Datenbank ist noch zu untersuchen
[10] Darstellung der Historie der Bearbeitung

Abbildung 2.1: Profil: lfw (Datenkatalog)

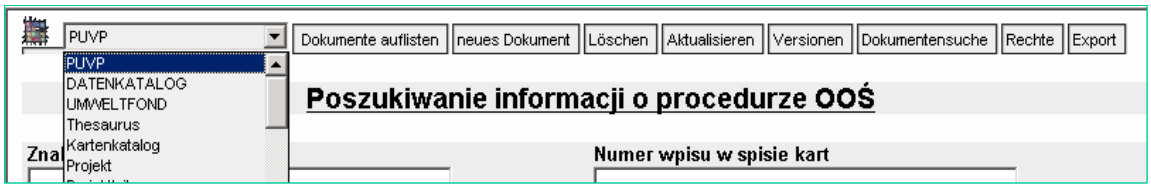


Abbildung 2.2: Profil: admin (Administration)

Nach Abbildung 2.1 ist die Funktionalität für ein typisches (hier englisches) Anwenderprofil dargestellt, in Abbildung 2.2 ein Administrationsprofil aus einem Monitoring der polnischen UVP (=SEA) (vgl. Recht, Import). In Abbildung 2.3 ist dazu der zugehörige Sourcecode des Stylesheets im Ausschnitt dargestellt.

```
<xsl:if test="$new='true'">
<a class="aButton" target="content" href="ServletEdit?new=1&amp;id=0">
 neues Dokument
</a>
<xsl:variable name="INOID">
 <xsl:value-of select="@inoId"/>
</xsl:variable>
<xsl:if test="$delete='true'">
 <a class="aButton" target="content„ ref="ServletDelete?id={$INOID}">
Löschen
 </a>
</xsl:if>
<xsl:if test="$edit='true'">
 <a class="aButton" target="content" ref="ServletEdit?id={$INOID}">
Aktualisieren
 </a>
```

Objekt ID aus Tamino

Abbildung 2.3: Ausschnitt aus der Stylesheetprogrammierung der Funktionsleiste

2.4 Sicherheitskonzept

Nach dem Architekturmodell der Abbildung 1 erfolgt der Zugriff auf die Datenbank über in den Stylesheets/Servlets enthaltenen Calls[11]. Da die Stylesheets den Anwenderprofilen (und damit auch den UOK-Klassen) zugeordnet werden, lassen sich auf dieser Ebene die Zugriffsrechte regeln. Es ist daher nicht erforderlich, einzelne Tags (in SQL-Datenbanken Felder und Tabellen) in der Datenbank mit besonderen Rechten zu belegen. Damit wird der Verwaltungsaufwand des laufenden Betriebs erheblich reduziert.

Verschieden Klassen des UOK sind in den Verwaltungsvollzug des StMLU oder anderer externer Behörden (BMU, MoE Warschau) eingebunden und müssen daher strengeren Kriterien der Qualitätskontrolle und des Datenschutzes bzw. der Datensicherheit genügen. Teilmengen dieser Daten müssen aber im Sinne der aktiven Präsentation der Daten in den Geschäftsbereichen öffentlich zugänglich sein (das neue UIG wird dies ohnehin fordern[12]), andere Klassen des UOK sind bereits vollständig öffentlich zugänglich (z. B. Daten-, und Kartenkatalog, Thesaurus etc.).

Durch die Framework-Architektur kommen die für eine bestimmte Aufgabenstellung erforderlichen Funktionen „automatisch“ allen anderen Klassen zu gute: Die für die Forschungsdokumentation unten in 2.4.1 gezeigte erforderliche chronologische Erfassung aller Änderungen (Funktion Version) war bei der Entwicklung des Monitorings der polnischen UVP erforderlich. Umgekehrt war dort die Mehrsprachigkeit (polnisch, englisch, deutsch) wie die Darstellung unterschiedlicher Zeichensätze in einer Anwendung erforderlich. So konnten nicht nur Entwicklungskosten gespart, sondern auch insgesamt die Qualität der Anwendung verbessert werden..

[11] vgl. Abbildung 2.3, z.B ServletEdit?id={INOID} zur Aktualisierung

[12] UIG (Umweltinformationsgesetz), insbes. § 4 und 12

2.4.1 Historie der Datenbearbeitung (Qualitätssicherung der Bearbeitung)

Bei jeder abgespeicherten Veränderung eines Objektes wird der vorangegangene Zustand vollständig erfasst und in der Datenbank abgespeichert. Da der Sachbearbeiter in der Regel nicht löschen darf (Regelung im Benutzerprofil) lässt sich jede Änderung nach Abbildung 3 chronologisch zurück verfolgen[13] und gegebenenfalls ein vorangegangener Zustand wieder herstellen.

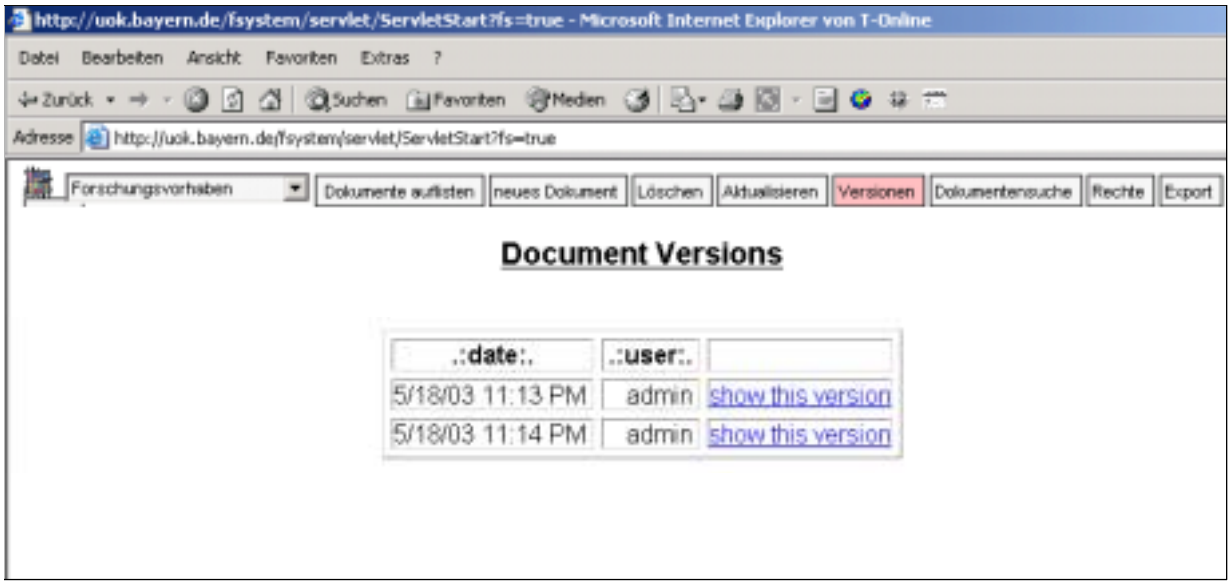


Abbildung 3: Versionierung der Objekte (hier der Forschungsvorhaben)

2.4.2 Rechte auf Klassenebene

Wegen der fachlich unterschiedlichen Bearbeitung der UOK-Klassen wurde eine klassenspezifische Rechtevergabe vorgesehen. Dazu existiert eine nur dem Administrator zugängliche Klasse Security, in der die Rechte Vergabe geregelt ist. Nach Abbildung 4 ist ersichtlich, dass die „Benutzergruppe“ SOAP[14] die Zugriffe der vUDK Schnittstelle (nur lesen, s.u.) regelt.

[13] vgl. Abb 2.1 und 2.2 Versionierung bzw, Rechtevergabe (?)
[14] der über SOAP-Schnittstelle kommunizierende Server des vUDK

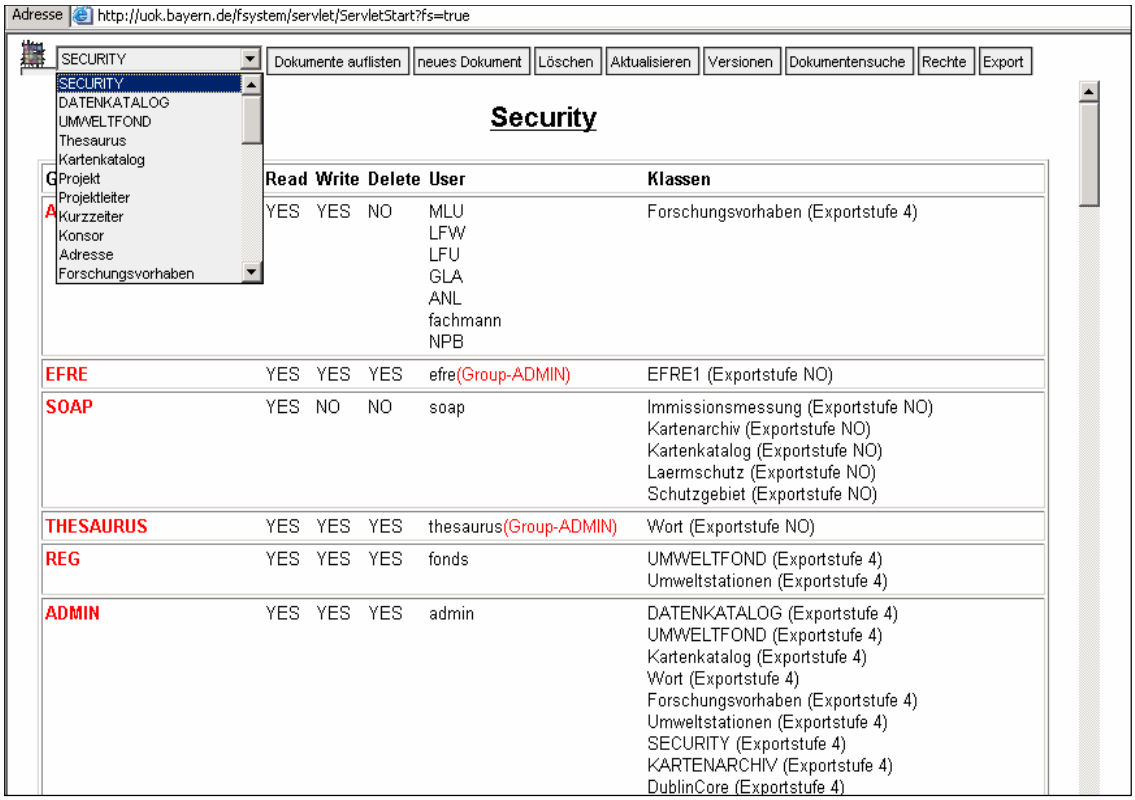


Abbildung 3: Ausschnitt aus einem Objekt der Klasse Security

Aus Abbildung 3 ist auch ersichtlich, dass die Klassen EFRE und THESAURUS über einen Gruppenadministrator verwaltet werden. Man beachte, dass auch hier die „Versionierung“ greift, also alle Aktualisierungen erfasst werden.

2.4.5 Rechte auf Objektebene

Neben den Rechten auf Klassenebene erwies es sich als erforderlich, auch Rechte auf Objektebene zu vergeben: Im Sinne größtmöglicher Transparenz[15] ist beispielsweise im Geschäftsbereich für die berechtigten Kollegen jedes in der Forschungsdokumentation angelegte „Forschungsvorhaben“ (Objekt) bereits im Status der Bearbeitung sichtbar. Änderungen kann allerdings nur der Anwender

[15] gemeint ist hier Transparenz im ursprünglichen Sinn: Jeder Anwender *sieht* die Tätigkeit der anderen, heute wird Transparenz in der IT auch im Sinne des Verbergens der Funktionalität vor dem Anwender (und Softwarekäufer) verstanden, ein bemerkenswerter Bedeutungswandel

vornehmen, der mit dem Schreibrecht seines Profils (soweit nicht weitere Rechte eingeräumt sind, etwa für den Gruppenadministrator) das Objekt angelegt hat[16].
In dem Beispiel der Abbildung 4 wurde mit dem Profil „rabus“ das Dokument angelegt, „Lautenbacher“ (Gruppenadmin) und „ANL“ können die Daten des Objekts unter „rabus“ angelegten Objektes auch löschen. Da in den Daten jedes angelegten Objektes der Forschungsvorhabens u.a. festgelegt ist, ob und wie[17] dieses im Internetangebot des Geschäftsbereichs erscheinen soll, ist sichergestellt, dass nur berechtigte Kollegen das Vorhaben „freischalten“. Die Freischaltung ergibt sich durch Eintrag in einem Datenfeld (tag) bei der Aktualisierung oder Erfassung.

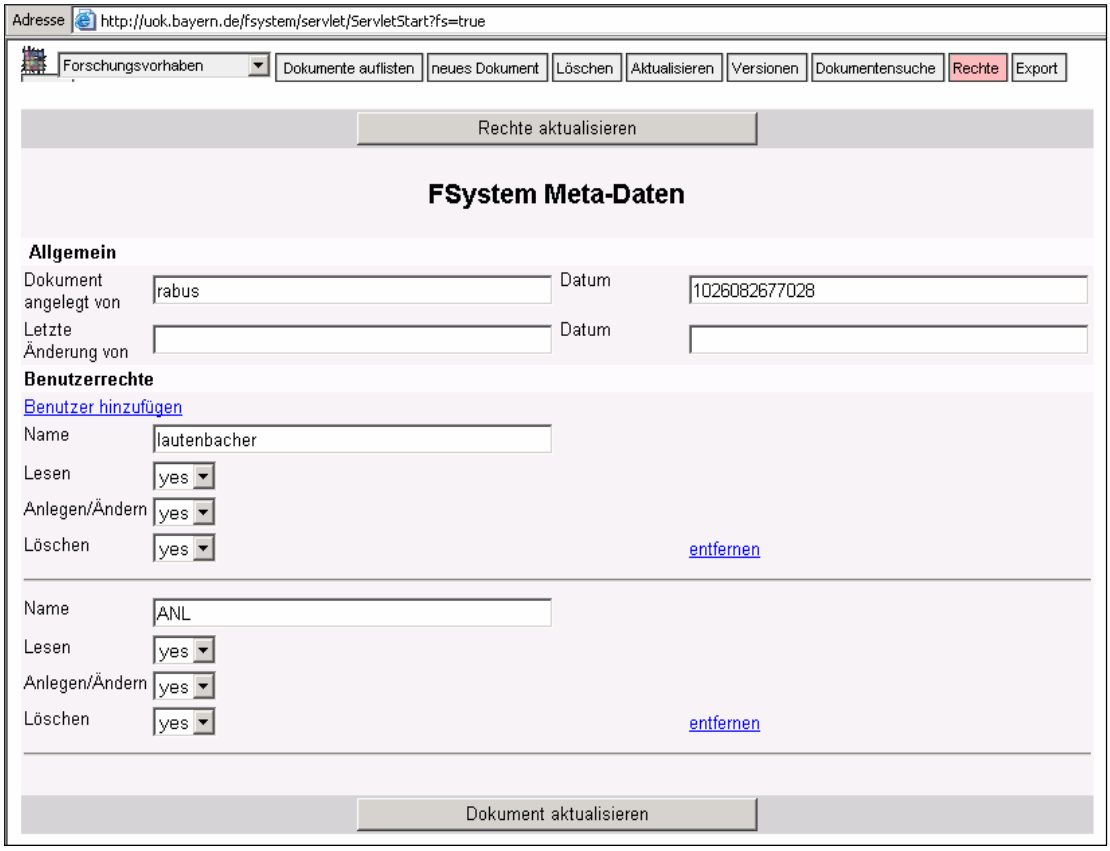


Abbildung 4: Regelung der Zugriffsrechte auf Klassenebene

[16] Pro Objekt *n* user mit *m* Rechten

[17] Die Forschungsvorhaben werden hier in der Regel öffentlich ausgeschrieben. In diesem Fall besteht die direkte Bewerbungsmöglichkeit über das Internet im Sinne des eGovernment. Soweit gibt es mehrere Möglichkeiten zur Freischaltung wie Bewerbung, Ankündigung, laufend, abgeschlossen.

2.5 Der UOK im Netz

Der UOK wurde als Webservice konzipiert und ist als Serverdienst über SOAP-Schnittstellen aufrufbar. Darüber hinaus kann der UOK, einzelne Klassen oder bestimmte Objekte über Servlets direkt aufgerufen werden. Damit ist eine Einbindung der Daten und Funktionen des UOK in andere Webdienste[18] oder CMS-Anwendungen[19] möglich: Im Web-Angebot des Geschäftsbereichs erfolgt die Darstellung der Forschungsdokumentation entsprechend der Vorgaben des Layouts des Geschäftsbereichs, der Sachbearbeitung erscheint es mit einer für die Verwaltung der Vorhaben angepassten Oberfläche und aller erforderlichen Daten.

3 Anwendungen des Frameworks

Wie in der Einführung bereits hingewiesen, führt der UOK in seinem Informationsangebot über das Katalogservice hinaus, obwohl, besonders in Verbindung mit den Länderkatalogsystemen, dort Schwerpunkte des Datenangebots liegen.
Nicht mehr in die Denkweise der klassischen Katalogsysteme passen die im UOK enthaltene Klassen wie Projektverwaltung(en), „Umweltstationen“ u.a. sowie die bereits zitierte Verwaltung der Forschungsdokumentation und das Monitoring der UVP, das an anderer Stelle behandelt werden wird.

Die Erweiterung des Klassenkonzeptes des UOK war mit vertretbarem Aufwand und für den UOK „systemkonform“ nur auf Grund der konsequenten XML-Ausrichtung möglich. Darüber hinaus war die von Anfang an festgelegte Frameworkstruktur, die auf eine die Klassen übergreifende Wiederverwendbarkeit beruht, Basis der Flexibilität des UOK für weitere Anwendungsgebiete.

[18] derzeit vUDK, gein®, Medienarchiv des StMLU, Thesaurusdienst des UOK.
[19] CMS = Contend Management Systeme, Datenbanksysteme zur Verwaltung und Standardisierung des Internetangebotes

3.1 Die Forschungsdokumentation

Die Forschungsdokumentation „monitort“ die Forschungsvorhaben des Geschäftsbereichs des Bayerischen Staatsministeriums für Landesentwicklung und Umweltfragen. Je Vorhaben werden die Daten für die Verwaltung (z.B. Budget/Jahr, Verausgabung, Forschungsnehmer, Ausschreibungsverfahren usw.) und fachliche Informationen (z.B. Kurzbeschreibung) erfasst. Die administrative Erfassung und Verwaltung der Vorhaben erfolgt über das Inter- und Intranet über gesicherte Verbindungen dezentral.

Befindet sich ein Vorhaben im Ausschreibungsstadium, wird dieses hier als eGovernment -Lösung im Web-Angebot des Geschäftsbereichs nach Abbildung 5 aufgenommen. Die Bewerbung erfolgt durch das Ausfüllen der Formulare. Es erfolgt eine Bestätigung der Eingabe durch den Bewerber. Der UOK-Server sendet dann eine e-Mail an den Sachbearbeiter zur weiteren Veranlassung.

Mit der Realisierung als UOK wurden im Framework weitere Funktionen wie die oben zitierte Versionierung und weitere Sicherheitsanforderungen (z. B. Zertifizierung) und Zugriffsrechte auf Objektebene eingeführt, die auch in den anderen Klassen Verwendung fanden. Da die Klasse Forschungsvorhaben in den UOK integriert ist, ist diese auch über die Schnittstellen zu Gin® und dem vUDK von dort erreichbar.

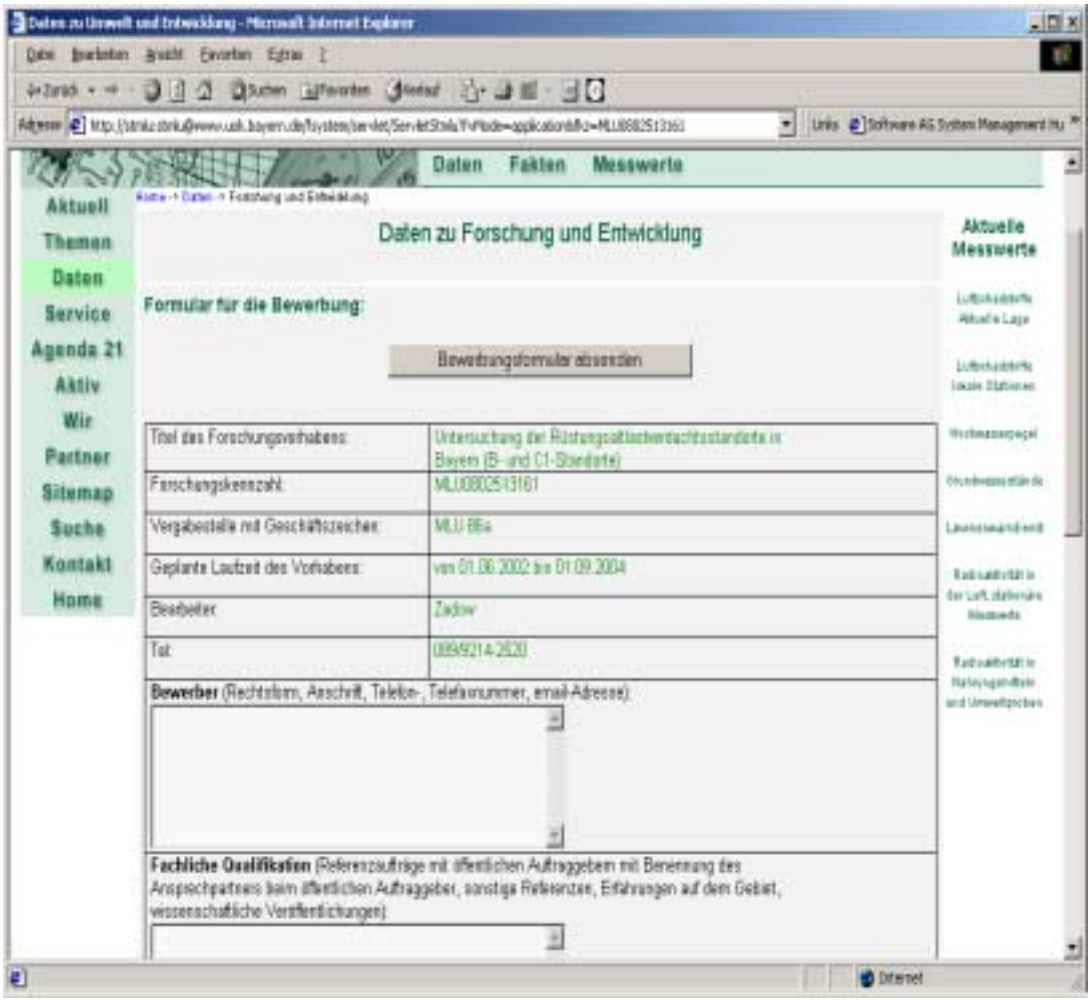


Abbildung 5: Bewerbungsformular zur Forschungsdokumentation im Internetangebot des Geschäftsbereichs

3.2 Der Thesaurus

Für die Indizierung der Objekte wie zur Recherche und Texterschließung kann ein Thesaurus hilfreich sein. Da die Praxis zeigt, dass in der Regel mit freien Zeichenketten (Texten), wie man dies von kommerziellen Suchmaschinen her kennt, gesucht wird, wird die Freitextsuche im UOK weitestgehend unterstützt.

Der Thesaurus umfasst mittlerweile ca. 50.000 raumbezogene Begriffe und ca. 100.000 fachliche Begriffe wie den GEMET und UBA Thesaurus. Auch hat sich gezeigt, dass es erforderlich ist, fachliche Mikrothesauri wie gegliederte Artenlisten oder einen Geothesaurus aufzunehmen.

Das Datenmodell entspricht dem Klassen- und Frameworkkonzept des UOK. Damit sind u.a. alle Funktionen wie Zugriffsregelung, Versionierung, Update, Recherche usw. bis auf Objektebene (Wort-/Begriffsebene) möglich. Die Freitextrecherche über den Thesaurus ermöglicht beispielsweise die Recherche auch in den Definitions“feldern“. Daneben kann entsprechend der in XML abgebildeten Wurzeln entsprechend hierarchisch gesucht werden: Jedes der derzeit ca. 150.000 Worte ist ein eigenes Objekt, welches durch Wurzelbegriffe und Synonyme mit anderen Objekten netzartig verknüpft ist. Durch die Einbeziehung mehrerer Thesauri ergibt sich eine Pyramidenstruktur nach BRITO (1990). Erst mit der Benennung eines bestimmten Thesaurus wird die bekannte hierarchische Thesaurusstruktur erzeugt. Ohne Benennung eines Thesaurus werden für den gewählten Begriff alle zutreffenden Baumteile angezeigt, in denen der Suchbegriff enthalten ist (Abbildung 8).

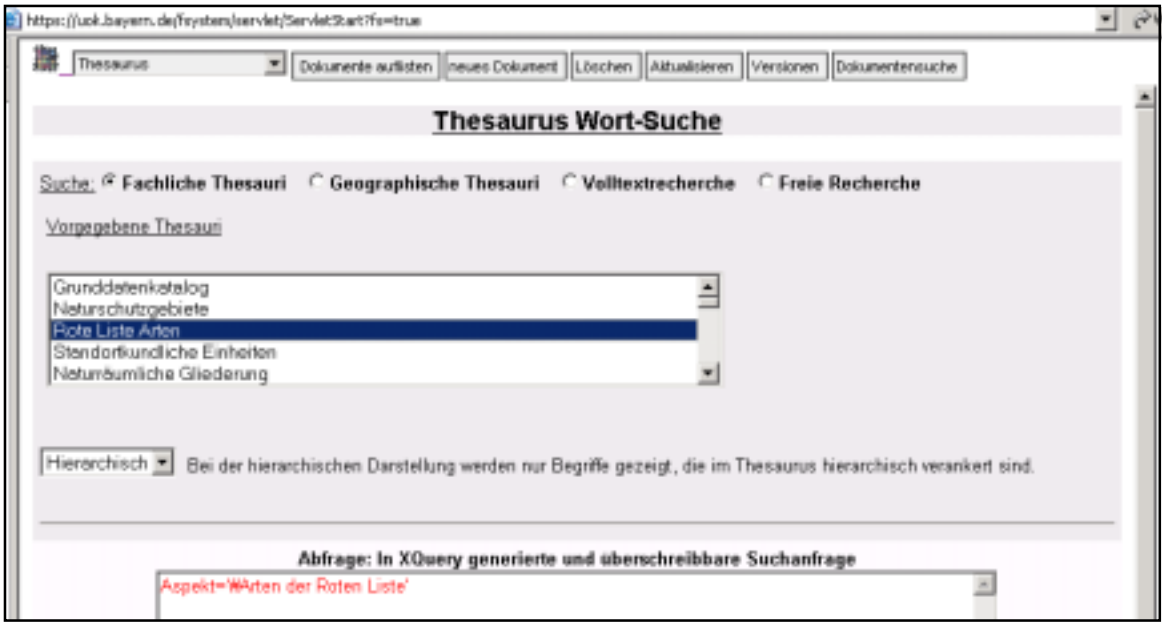


Abbildung 6: Suche im Mikrothesaurus „Rote Liste Arten“

Die Freitextsuche nach der Zeichenkette „Waller“ im Thesaurus ergibt, ohne Einschränkung auf den Mikrothesaurus „Rote Liste Arten“ mehrere Teil-Baumstrukturen. Neben der Einordnung in den Baum der Artenliste finden sich Teilstrukturen u.a. in dem raumbezogenen Thesaurus (Weihs 1993, 2000). In Abbildung 7 dargestellt ist der Teilbaum aus dem Gemeindeschlüssel GS574 Mittelfranken. Damit können Homonyme Bezeichnungen eindeutig getrennt werden, obwohl die Zeichenkette „Waller“ als Stammwort nur einmal im Thesaurus enthalten ist.

Da nach Abbildung 6 nur im Thesaurus „Rote Listen Arten“ nach dem Begriff *Waller* recherchiert wurde, stellt sich das zu erwartende Suchergebnis ein, nämlich eine hierarchische Baumstruktur (wir hätten auch eine Listendarstellung wählen können) in welcher der Begriff Waller enthalten ist.
Wie aus Abbildung 7 zu ersehen ist, steht zur Bearbeitung der Thesaurus Begriffe die volle Funktionalität des UOK wie Update, Version, Import, Export usw. zur Verfügung. Damit ist auch die Pflege des Thesaurus über Internet möglich. Jedes der zur Zeit etwa 150.000 Worte ist ein getrennt bearbeitbares Wort, das im Moment der Anwendung zum passenden Strukturbaum zusammen gefügt wird (Weihs 2002).

Abbildung 7: Ergebnisdarstellung des Freitextsuche (Netz) nach Waller

Literaturverzeichnis

AD HOC AG der LABO, Weihs, E, u.a. (1994): Aufgaben und Funktionen von Kernsystemen des Bodeninformationssystems als Teil von Umweltinformationssystemen; Hrsg: Umweltministerium Baden-Württemberg, Stuttgart.

Brito, P. Diday, E. (1990): Pyramidal representaton of symbolic objects in: Knowledge, Data Knowledge and Decision, Hamburg, Eds. Schader, Springer

Coad, P./Yourdan, E. (1994): Objektorientierte Analyse, 1. Auflage 75, Prentice Hall

Rosenkranz, D. Knetsch G., (2003): Tagung Medienübergreifende Umweltbeobachtung – Stand und Perspektiven, Landesanstalt für Umweltschutz Karlsruhe, Juni 2003

Weihs, E. (1978): Zum Stand der Entwicklungsarbeiten des bayerischen Umweltschutzinformationssystems, Natur und Landschaft 53 Heft 5, Stuttgart, p. 146-149

Weihs, E. (1993): An approach to a Space Related Thesaurus; in Information and Classification, O.Opitz, B.Lausen, R.Klar (Hrsg.), 469-476; Springer; Berlin, Heidelberg.

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. (1999): Objektorientierung bei der Speicherung und Recherche von XML-Daten in vernetzten Systemen in: Management von Umweltinformationen in vernetzten Umgebungen, Bd 21 Umweltinformatik aktuell, p. 58 - 69

Weihs, E. (2000): Anwendung des XML Thesaurus im Umweltobjektkatalog, Beitrag zur Jahrestagung der Gesellschaft für Klassifikation, Passau 2000

Weihs, E. (2002): Webservices: Der Thesaurusdienst des Umweltobjektkatalogs (UOK) im Internet als XML-Anwendung, bei: 16. Internationalen Symposium "Informatik im Umweltschutz"; Workshop; Wien 2002

Die IMPETUS-Metadatenbank

Ein Baustein interdisziplinärer Forschung

Michael Süßer[20]

Zusammenfassung

Im interdisziplinären GLOWA-Forschungsprojekt IMPETUS, das sich mit dem nachhaltigen Süßwasser-Management in Westafrika beschäftigt, arbeiten Wissenschaftler unterschiedlicher natur- und geisteswissenschaftlicher Fachdisziplinen zusammen. Um den internen Datenaustausch zu fördern wurde eine Metadatenbank entwickelt, in der Metainformationen alle im Projekt akquirierten Daten gesammelt werden. Die Daten selbst werden dezentral auf den verschiedenen Institutsservern gehalten, der Zugang ist über die Metadatenbank per Download - Link möglich.

Die Metadatenbank wurde in die vorhandene Projektverwaltungs-Datenbank in Access eingebaut. Zur Webanbindung wird die Datenbank in eine MySQL-Datenbank kopiert und diese dynamisch mittels PHP-Skripten abgefragt. Alle Datenautoren erfassen Metainformationen über ein HTML-Formular, die eigentliche Dateneingabe erfolgt zentral in der Geschäftsstelle, wodurch eine gründliche Qualitätskontrolle möglich ist.

Die Recherche in der Metadatenbank ist sowohl für Projektmitarbeiter als auch für die interessierte Öffentlichkeit zugänglich. Der öffentliche Download von Daten ist aber nur für ausgewählte Datensätze freigegeben. Als wichtige Suchfunktion zur inhaltlichen

[20] Institut für Geophysik und Meteorologie, Universität zu Köln, Kerpener Straße 13, 50923 Köln, E-Mail: suesser@meteo.uni-koeln.de

IMPETUS – Integriertes Management-Projekt für einen Effizienten und Tragfähigen Umgang mit Süßwasser in Westafrika

Einleitung

IMPETUS ist ein interdisziplinäres und anwendungsbezogenes Forschungsvorhaben, in dem die Entwicklung von Szenarien für ein nachhaltiges Management der knappen Ressource „Wasser“ im Vordergrund steht. Das Forschungsprojekt ist Teil des Förderschwerpunkts „Globaler Wandel des Wasserkreislaufs (GLOWA[21])“ des Bundesministeriums für Bildung und Forschung (BMBF). Weitere im Rahmen von GLOWA finanzierte Forschungsvorhaben beschäftigen sich mit den Einzugsgebieten von Donau, Elbe, Jordan und Volta (Westafrika).

Die Arbeiten von IMPETUS werden auf der Basis eines Kompetenznetzwerkes in zwei Flusseinzugsgebieten Westafrikas durchgeführt (Abb. 1), dem Ouémé in Benin und dem Drâa in Marokko.

Abb. 1: **Lage der Untersuchungsgebiete in Marokko und Benin.**

An IMPETUS sind fünfzehn Institute der Universitäten Köln, Bonn, Hamburg und Rostock sowie das Max-Planck-Institut für Meteorologie in Hamburg mit insgesamt ca. 120 Wissenschaftlern verschiedener Fachdisziplinen beteiligt: Meteorologen, Hydrologen, Botaniker, Vegetationskundler, Bodenkundler, Geologen, Agrarwissenschaftler, Agrarökonomen, Ethnologen, Mediziner, u.a.

Die Forschungsinitiative IMPETUS ist eingebettet in die lokale Forschungslandschaft der Staaten, in denen die Forschungsaktivitäten konzentriert werden, also überwiegend Benin und Marokko. Es wird Wert darauf gelegt, die dortigen staatlichen, traditionellen und privaten Institutionen durch eine intensive

[21] http://www.glowa.org

Hinsichtlich der Schnittstellen zwischen den erwarteten Ergebnissen des Projektes und den künftigen Konsequenzen für Ressourcenmanagement und praktischer Technologieentwicklung wird mit Partnern aus der Praxis bzw. Industrie kooperiert (IMPETUS 2003).

IMPETUS ist in mehrere Teilprojekte unterteilt, die nicht nach Fachdisziplinen sondern nach Fragestellungen und Bezugsraum (Benin, Marokko) unterschieden sind (Tab. 1):

Tab. 1: Gliederung des IMPETUS-Projekts

A	**Der hydrologische Kreislauf des Ouémé-Einzugsgebietes und sozioökonomische Implikationen**
AB1	Externe Klima-Antriebsszenarien auf der globalen und kontinentalen Skala
A1	Szenarien der raum-zeitlichen Variabilität von Niederschlag und Verdunstung auf der regionalen und lokalen Skala
A2	Bodenwasserdynamik, Oberflächenabfluss, Grundwasserneubildung und Bodendegradation auf der lokalen und regionalen Skala
A3	Funktionale Beziehungen zwischen raumzeitlicher Vegetationsdynamik und Wasserkreislauf
A4	Sozioökonomische Entwicklung im Hinblick auf die Verfügbarkeit der Ressourcen – Agrarsektormodellierung und Wasserbedarfsanalyse
A5	Verfügbarkeit, Qualität und Management von natürlichen Ressourcen: Sozialwissenschaftliche und medizinische Perspektiven
B	**Die Wasserbilanz des Drâa-Einzugsgebietes und sozioökonomische Implikationen**
AB1	Externe Klima-Antriebsszenarien auf der globalen und kontinentalen Skala
B1	Regionale und lokale Szenarien der raum-zeitlichen Variabilität von Niederschlag und Verdunstung in Marokko
B2	Wasserverfügbarkeit und Bodendegradation

B3 Steuerfunktionen der Vegetation für den Gebietswasserhaushalt des Drâa-Catchments

B4 Modellierung der Landnutzungsveränderung vor dem Hintergrund knapper Ressourcen und globalisierter Agrarmärkte

B5 Soziokultureller Wandel und Wassernutzung im Einzugsgebiet des Drâa

C	Integration, Organisation und Datenmanagement
C1	Integration und Organisation
C2	Datenmanagement und übergeordnete Arbeiten

Szenarienentwicklung

Während man sich in der ersten dreijährigen Projektphase überwiegend auf die Erfassung und Diagnose wesentlicher Aspekte des Wasserhaushaltes und des Verhalten des Menschen konzentriert hat, steht in der zweiten Phase die Analyse der zukünftigen Entwicklung im Vordergrund der Betrachtungen. Aufgrund der mit der Modellierung verbundenen großen Unsicherheiten kann die zukünftige Entwicklung dabei aber nicht direkt prognostiziert werden. Vielmehr muss mittels Szenarienanalysen versucht werden, mögliche Entwicklungen gegeneinander abzuwägen und zu bewerten, damit die Akteure Handlungsempfehlungen aus den Ergebnissen ableiten können. Hierzu werden zunächst, orientiert an den allgemeinen Entwicklungen für die untersuchten Länder und in Abstimmung mit den lokalen Stakeholdern, Szenarien entwickelt, die anschließend analysiert und bewertet werden. Betrachtet wird neben dem Klimawandel der kulturelle und soziale Wandel, der institutionelle Wandel, die Bevölkerungsentwicklung, die wirtschaftliche Entwicklung und technologische Innovationen. Die skalenabhängige, raumzeitliche Betrachtung und Bewertung der zukünftigen Entwicklung ist die unverzichtbare Grundlage für die Erstellung und Implementierung der für ein nachhaltiges Management benötigten Werkzeuge. Dies soll in der abschließenden dritten zweijährigen Phase vorgenommen werden. Eigene Messungen werden nur dann durchgeführt, wenn sie zur Realisierung des angewandten Konzeptes der Szenarienentwicklung benötigt werden und keine anderen Daten zur Verfügung stehen.

Durch die Analyse von Szenarien sollen Fragen der Art „was ist, wenn ...“ beantwortet werden. Im Gegensatz zu einem zielorientierten Szenario, bei den Optionen gegenwärtigen und zukünftigen Handelns gesucht werden, um einen gewünschten Zustand in der Zukunft zu erhalten, wird hier die Form eines extrapolativen Szenarios gewählt. Um die zu betrachtenden komplexen Systeme analysieren zu können, müssen sie nach systemtheoretischen Überlegungen in interne, eventuell rückgekoppelte Prozessketten und die sie von außen steuernde Umwelt unterteilt werden. Folgende Teilbereiche werden daher definiert (Abb. 2):

- Unabhängige Antriebskräfte (driving forces), diese müssen problemorientiert definiert werden und wirken per Definition nicht auf das betrachtete System zurück.
- Problemkomplexe, hierunter sind übergeordnete Problembereiche zu verstehen, die nur multidisziplinär zu analysieren und vorherzusagen sind. Jeder Problemkomplex setzt sich aus einer Vielzahl von einzelnen Themenkomplexen zusammen. Um die Problemkomplexe einer Bewertung zugänglich zu machen, werden Responseindikatoren definiert.
- Themenkomplexe, hier werden die Prozesse bzw. Prozessketten definiert, die direkt oder indirekt die Responseindikatoren beeinflussen

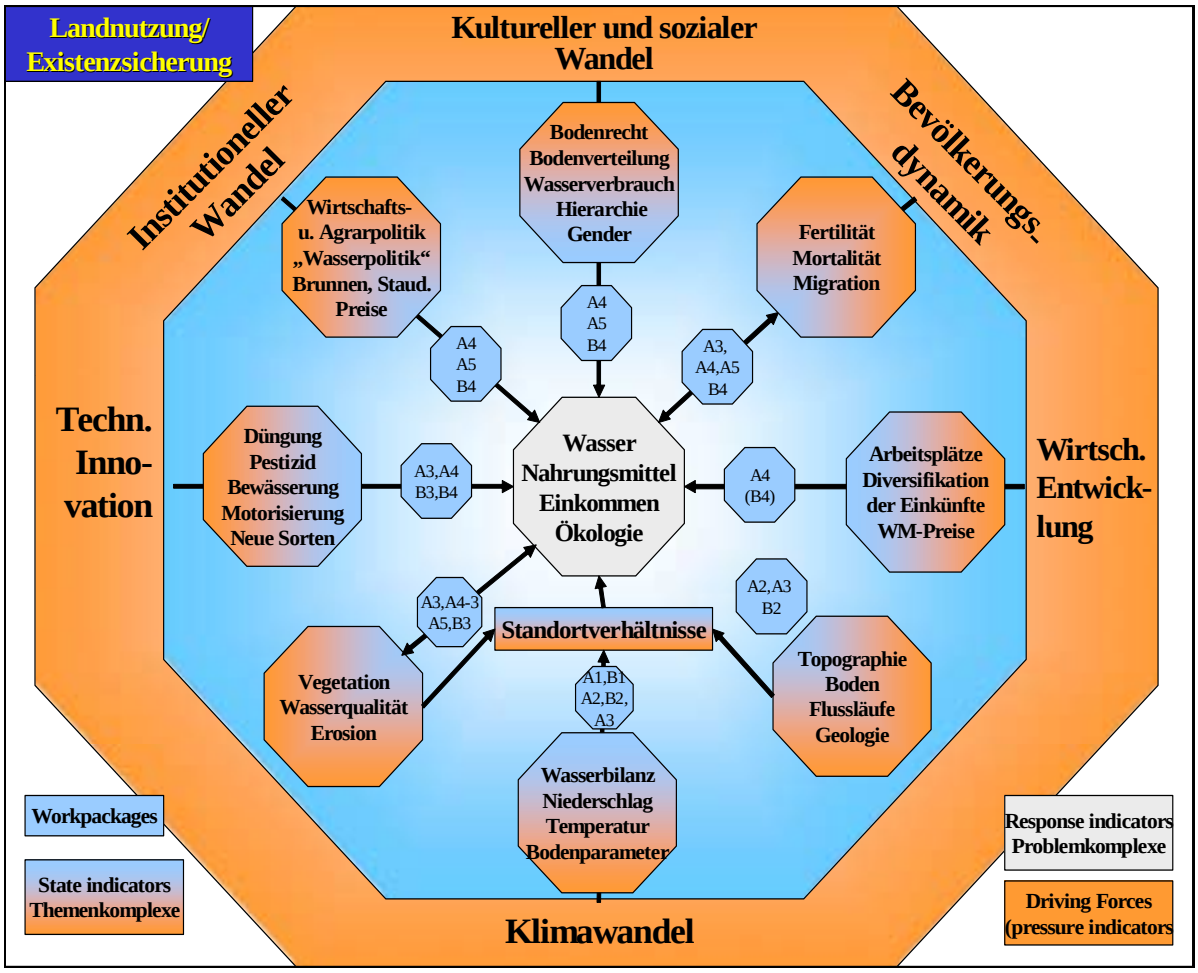


Abb. 2: **Definition der Themen- und Problemkomplexe anhand des Beispiels „Landnutzung unter dem Aspekt der Existenzsicherung**

Für die Analyse der einzelnen Themenkomplexe kann nicht ein einzelnes Modell eingesetzt werden, vielmehr müssen Teilkomplexe mit dafür geeigneten Ansätzen simuliert werden. Für die Simulation der agrarischen Produktion in einem Einzugsgebiet muss z.B. ein hydrologisches Modell mit einem Pflanzenwachstumsmodell gekoppelt werden. Als Antriebskräfte dieser Modelle dienen hierbei Informationen aus der Atmosphäre und der Anthroposphäre (Abb. 3 und 4).

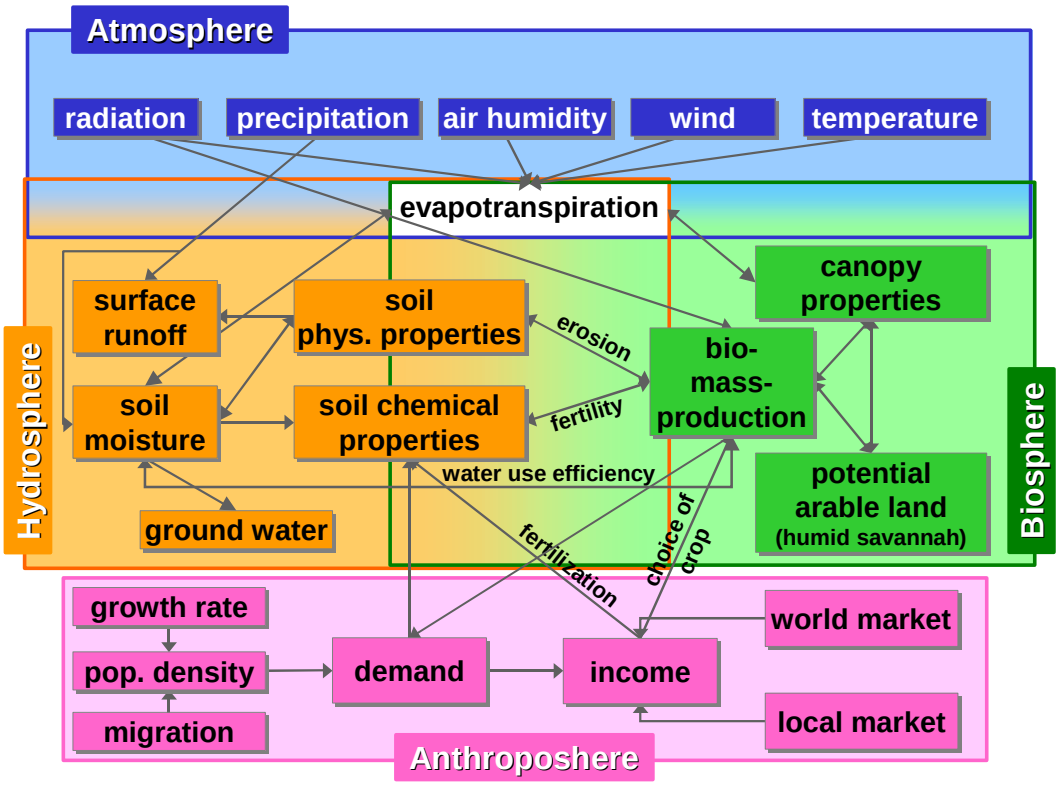


Abb. 3: **Beispiel für die Rückkopplung der Prozesse zur Beschreibung der Agrarproduktion einer Region.**

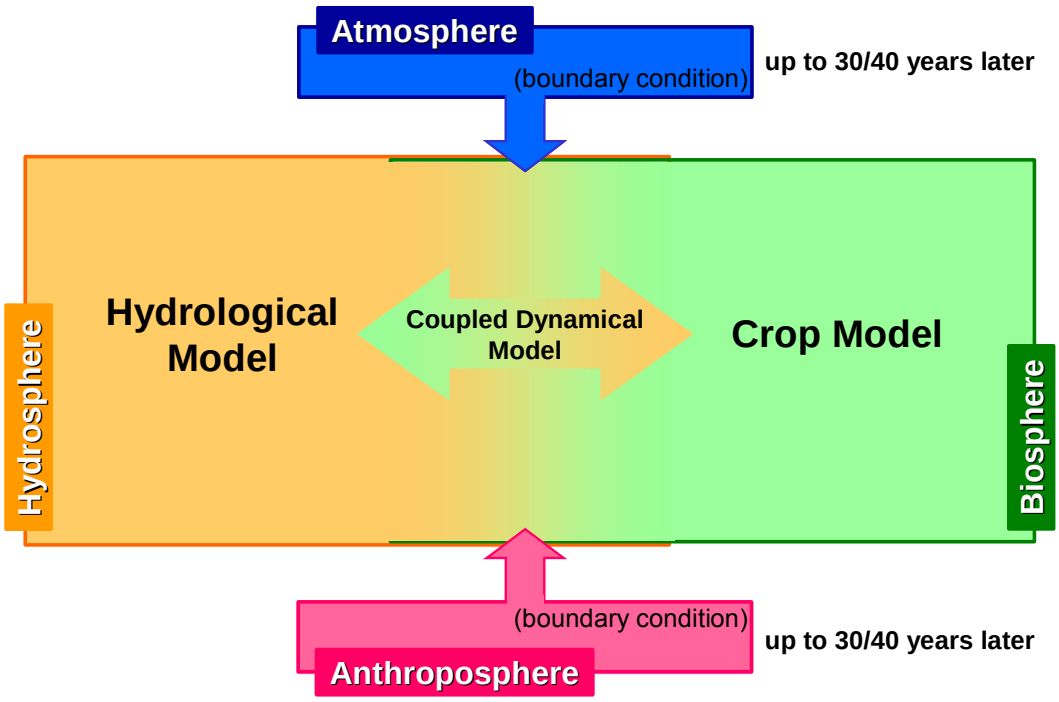


Abb. 4: **Beispiel der Simulation der Nahrungsmittelproduktion durch Kopplung eines hydrologischen Modells (inkl. eines Soil Vegetation Atmosphere Transfer-Modells) mit einem Pflanzenwachstumsmodell. Sowohl die Daten der Atmosphäre als auch der Anthroposphäre werden für die Szenarienberechnungen als Randbedingungen in dem betrachteten System berücksichtigt.**

Heterogene Datensätze

Wie aus den oben genannten Beispielen zur Szenarienanalyse ersichtlich wird, sind für die Entwicklung und Validierung der einzelnen Simulationsmodelle unterschiedliche Daten aus allen beteiligten Fachdisziplinen notwendig. Ein großer Teil dieser Daten wurde in der ersten Projektphase akquiriert – durch eigene Mess- und Feldkampagnen, durch Übernahme aus vorhergehenden Forschungsprojekten (z.B. CATCH[22]) oder durch Ankauf. In der folgenden Projektphase werden parallel zur Szenarienanalyse weiterhin Daten erworben und erhoben.

Durch die Fachdisziplinen übergreifende Modellierung der Prozesse und Prozessketten ist ein intensiver Datenaustausch zwischen den einzelnen Teilprojekten und allen beteiligten Wissenschaftlern notwendig. Bedingt durch die unterschiedlichen Vorgehensweisen, Erfassungsmethoden, Datenstrukturen und raum-zeitlichen Betrachtungsebenen der Fachdisziplinen ist es zwingend notwendig, alle verfügbaren Datensätze einheitlich zu katalogisieren.

Das Spektrum der in IMPETUS vorliegenden Datensätze umfasst

- räumliche Skalen vom globalen bis zum lokalen Maßstab,
- einmalige und mehrfache, regelmäßige und unregelmäßige Messungen,
- punktuelle, lineare, flächige und rasterbasierte Geoinformationen,
- unterschiedliche Maßstäbe, Auflösungen und Koordinatensysteme,
- Rohdaten, abgeleitete Daten, Modellergebnisse, Berichte,
- Texte, Tabellen, Karten, Bilder,
- digitale und analoge Daten,
- vielfältige Daten- und Dateiformate
-

Aufgrund dieser extremen Variabilität der Datensätze ist eine zentrale Datenhaltung in IMPETUS nicht zu verwirklichen. Daher wurde entschieden, die einzelnen Datensätze dezentral auf den Servern der beteiligten Institute zu speichern und über einen Metadatenkatalog zu erschließen und zugänglich zu machen. Ziel eines solchen Metadatenkatalogs ist es, dem Nutzer alle Informationen, aus denen er die Verwendbarkeit eines Datensatzes für eine bestimmte Fragestellung ableiten kann, schnell und leicht zugänglich zu machen.

[22] Couplage de l'atmosphère tropicale et du cycle hydrologique (Institut Français de Recherche pour le Développement IRD)

Darüber hinaus soll der Metadatenkatalog auch externen Interessenten einen Überblick über die vorhandenen Daten geben und den direkten Zugriff auf relevante Daten ermöglichen. Als Zielgruppen kommen hierbei Behörden, v.a. der beteiligten Staaten Benin und Marokko, Universitäten und Forschungseinrichtungen, Firmen und Unternehmen sowie nichtstaatliche Organisationen (NGOs) aus dem Bereich der Entwicklungszusammenarbeit in Betracht.

Metadatenbank

Anforderungen an den Metadatenkatalog

Unter den gegebenen Voraussetzungen und in Hinblick auf die projektbezogenen Ziele des Metadatenkatalogs müssen folgende Anforderungen formuliert werden:

- Schnelle und effektive Navigationsmöglichkeiten
- Fachübergreifende inhaltliche Erschließung der Datensätze
- Flexible Struktur zur Dokumentation unterschiedlichster Datentypen und Inhalte
- Keine Verwaltung der zugrunde liegenden Datensätze, aber direkter Zugang zu den einzelnen Datensätzen
- Minimaler Erfassungsaufwand für Datenautoren

Metadaten-Struktur

Vereinfacht kann man Metadaten auch als „strukturierte Daten über Daten“ bezeichnen. Während der Begriff Metadaten relativ neu ist, wird ist das zugrunde liegende Konzept z.B. in Bibliothekskatalogen bereits seit langem angewandt. Metadaten können auf unterschiedliche Weise kategorisiert werden (STROBL 1995): Formale Kategorien unterscheiden z.B. einfach

- textliche Metadaten (Beschreibung von Charakteristika mit Freitext)
- tabellarische Metadaten (Ausweisung unterschiedlicher Werte in strukturierter Form, oft mit eingeschränkten Ausprägungen)
- bildliche Metadaten (z. B. Quicklook, Thumbnail)

•

Weiterhin können Metadaten auch inhaltlich kategorisiert werden. DENZER und GÜTTLER (1994) differenzieren vier breite Kategorien:

- Semantische Metainformation (inhaltliche Beschreibung)
- Syntaktische Metainformation (Wege des Zugriffs auf die zugrundeliegenden Daten)
- Strukturelle Metainformationen (Objektstrukturierung, Hierarchien, usw.) und
- Navigatorische Metainformation (Beschreibung der Navigationsmöglichkeiten im gesamten Kontext der Daten).

Die elf Hauptklassen des ISO-Standards (ISO/TC 211 / CD 19115.3) ergeben schon ein umfassenderes Bild der Metainformationen (ARBEITSGEMEINSCHAFT DER VERMESSUNGSVERWALTUNGEN 2002):

- Identifikation
- Raumbezug
- Qualität
- Vertrieb
- Darstellungsart
- Informationsinhalt
- Signaturenkatalog
- Anwendungsschema
- Zugriffs- und Nutzungsbeschränkungen
- Datenpflege
- Erweiterte Metadaten-Informationen

Ähnlich umfangreich sind die Kategorien des FGDC Standard (FEDERAL GEOGRAPHIC DATA COMMITEE 2000), der insgesamt zehn, teils breite Kategorien unterscheidet:

- Identifikation (Identification Information)
- Datenqualität (Data Quality Information)
- Räumliche Datenorganisation (Spatial Data Organization Information)
- Definition des räumlichen Bezugs (Spatial Reference Information)
- Entitäten und Attribute (Entity and Attribute Information)
- Abgabeformate und -modalitäten (Distribution Information)
- Verantwortlicher und Aktualität der Metadaten (Metadata Reference Information)

- Zitierhinweise (Citation Information)
- Zeitliche Referenz (Time Period Information)
- Kontaktangaben (Contact Information)

Eine weitere, wichtige Unterscheidung liegt in der Differenzierung in abgeleitete (sekundäre) und primäre Metadaten. Sekundäre Metadaten können direkt aus dem Datensatz abgeleitet werden und sind insofern bis zu einem gewissen Grad auch automatisch generierbar. Im Gegensatz dazu müssen primäre Metadaten explizit vom verantwortlichen Bearbeiter eingegeben werden. Tab. 2 zeigt Beispiele für primäre und sekundäre Metadaten.

Tab. 2: **Beispiele für primäre und sekundäre Metainformationen**

primäre Metainformationen	sekundäre Metainformationen
Geodatentyp	Beschreibung
Horizontales Koordinatensystem	Schlagworte/Deskriptoren (Thesaurus)
Bounding Box	Parameter-Genauigkeit
Räumliche Auflösung	Vollständigkeit des Datensatzes
Dateigröße	Nutzungsbedingungen

Unter Berücksichtigung dieser unterschiedlichen Kategorisierungen wurde die Struktur der IMPETUS-Metadatenbank entwickelt. Folgende Aspekte mussten hierbei berücksichtigt werden:

- Nur ein Teil der in IMPETUS erhobenen Daten sind digitale Geodaten. Daneben existieren z. B. auch analoge Daten (Karten), textliche Beschreibungen und Ergebnisse von Umfragen und Interviews.
- Da für die Generierung der Metainformationen die einzelnen Bearbeiter der Datensätze zuständig sind, muss der Umfang der zu erfassenden Metainformationen so gering wie möglich gehalten werden. Umfangreiche Erfassungsformulare führen erfahrungsgemäß nur zu einer geringen Erfassungsquote.
- Um insbesondere bei textlichen Metadaten einheitliche Einträge zu erlangen, sollte jede Eingabe in die Datenbank an zentraler Stelle erfolgen, bzw. zumindest zentral kontrolliert und ggf. editiert werden.

Aus den o.a. Gründen wurden die vorhandenen Kategorisierungen der Metadatenstandards (FGDC, ISO 19115) nur als Orientierungsgrundlage herangezogen und projektspezifisch abgewandelt, d.h. in Teilen übernommen und um weitere relevante Informationen ergänzt. Da die primäre Aufgabe der IMPETUS-Metadatenbank die Erleichterung des projektinternen Datenaustauschs ist und daher projektrelevante Metainformationen an erster Stelle stehen, wurde die möglicherweise nicht vollständige Standardisierung der Metadatenstruktur in Kauf genommen. Abb. 5 sowie Tab. 3 zeigen Struktur und Inhalte der Datenbank.

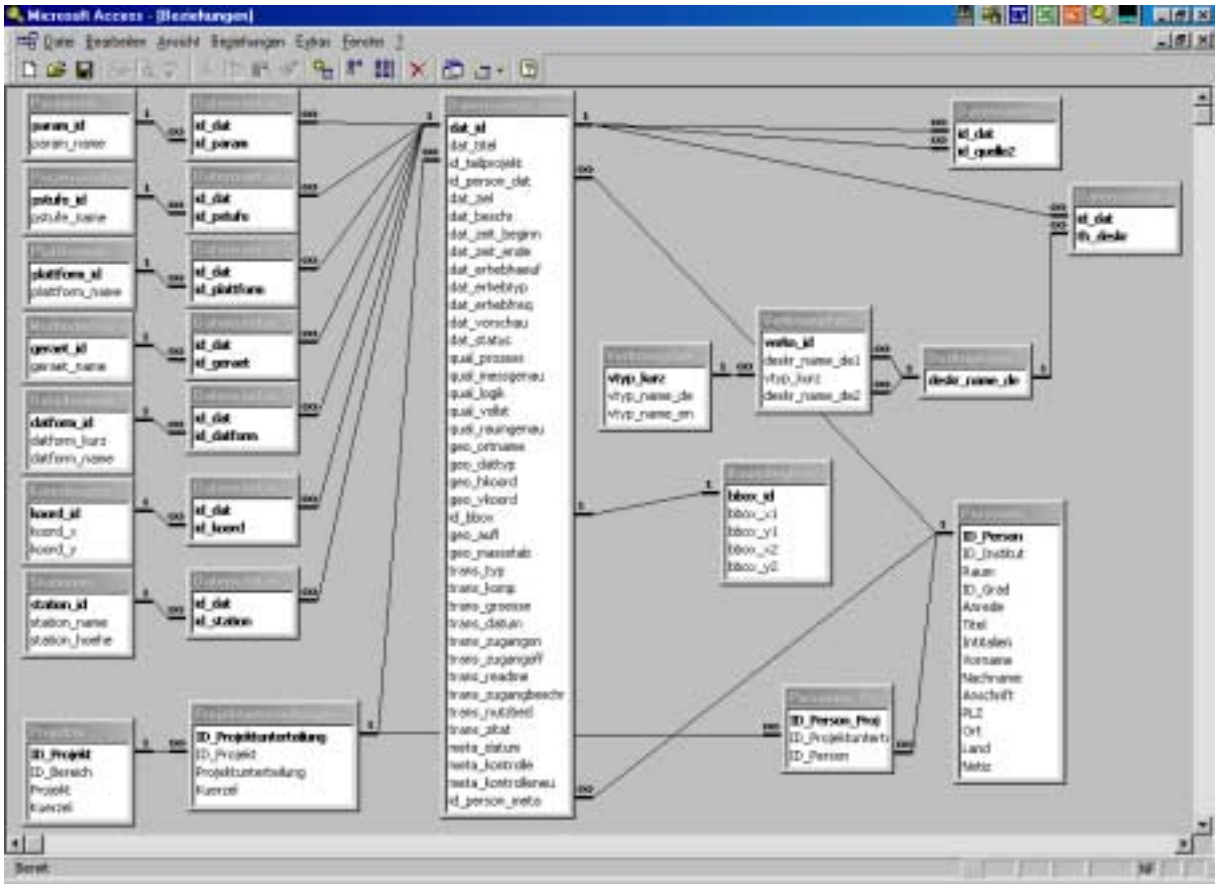


Abb. 5: **Beziehungsstruktur der Access-Datenbank (nur Metadatenbank)**

Tab. 3: **Inhalte der IMPETUS-Metadatenbank**

Metadata – Identification	
ID	
Subproject	
Contact Person	
Metadata – Content	
Aim / Purpose	
Description	
Parameter / Attribute List	
Beginning Date	
Ending Date	
Acquisition Frequency	single, multiple
Acquisition Regularity	regular, irregular
Acquisition Frequency Description	
Preview Image	
Status	finished, in progress, planned
Keyword List	
Metadata – Data Quality	
Method / Instrument List	
Processing Levels	raw data, derived/calculated data, modelling results/simulated data, report/description
Source Data / Related Data	
Data Provenance	
Platforms	satellite, airplane, balloon, tower, vehicle, person, ground, ship
Processing Description	
Parameter Accuracy Report	
Logical Consistency Report	
Completeness Report	
Spatial Accuracy Report	
Metadata – Spatial Reference	
Spatial Reference Description	
Station List	
Geodata Type	point, vector, raster
Planar Coordinate System	
Vertical Coordinate System	
Coordinate List	
Bounding Box	
Resolution	
Scale	
Metadata – Transfer Information	
Data Character	text, table, map, image
Data Form	digital, non digital
File Format List	
Compression	not compressed, ZIP, TGZ/GZ, ARJ, LZH, …
File Size uncompressed	
File Size compressed	
File Creation Date	
Online Access	
Offline Access	
Readme File	
Access Constraints	public, IMPETUS internal, no access
Use Constraints	
Citation Information	

Hierarchischer Schlagwortkatalog / Thesaurus

Eines der wichtigsten Ziele der Metadatenbank ist die fachübergreifende inhaltliche Erschließung des Datenbestands. Einerseits benötigen die beteiligten Wissenschaftler für die Szenarienentwicklung und -analyse Daten der anderen Fachdisziplinen, andererseits besitzen sie nicht die nötigen Fachkenntnisse und Fachtermini, um schnell den thematischen Kontext eines Datensatzes zu erkennen. Freie Schlagworte, wie sie häufig angewandt werden, entstammen meist der spezifischen Fachsprache und besitzen keinen thematischen Zusammenhang.

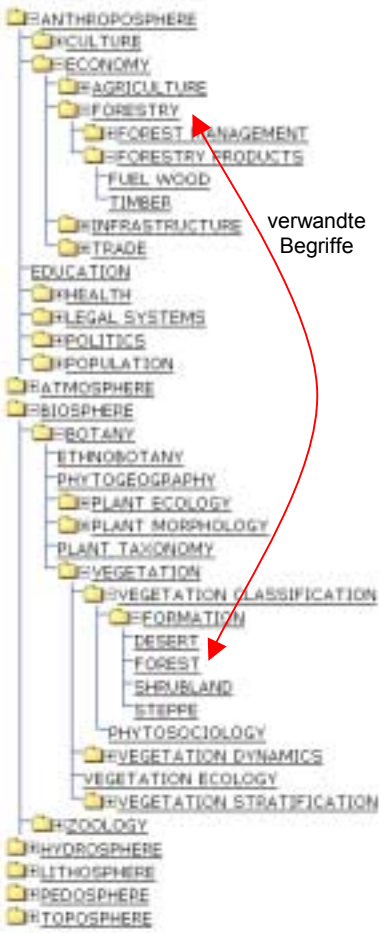


Abb. 6: **Beispielhafter Ausschitt aus dem Thesaurus**

Hier kann ein hierarchisch aufgebauter Schlagwortkatalog oder Thesaurus zum Ziel führen. Ein Thesaurus stellt eine festgelegte Menge sogenannter Deskriptoren zur Verfügung. Diese Deskriptoren sind über Verwandtschaftsbeziehungen miteinander verknüpft (Abb. 6). Zusätzlich zu über- und untergeordneten Begriffen können auch verwandte Begriffe (unabhängig von der Baumstruktur) sowie Synonyme definiert werden (WERSIG 1985). Die Recherche in der Metadatenbank mit Hilfe des Thesaurus kann dann mit unterschiedlichen Optionen erfolgen: Suche Pfad aufwärts, Suche Pfad abwärts und Einbezug verwandter Begriffe.

Die Auswahl von Deskriptoren kann sowohl alphabetisch als auch systematisch mit Hilfe einer Baumstruktur erfolgen. Ein fachfremder Nutzer kann sich im Schlagwortbaum leicht zum benötigten Terminus vortasten. Im alphabetischen Katalog können außerdem auch Synonyme aufgeführt werden.

Organisation

In der Geschäftsstelle von IMPETUS am Institut für Geophysik und Meteorologie der Universität zu Köln wird eine Projektdatenbank betrieben, in der alle verwaltungsrelevanten Informationen über Mitarbeiter, Institute, Publikationen, Veranstaltungen u.a. gespeichert sind. Zugriff auf diese Datenbank haben Geschäftsführung, Datenmanagement sowie das Sekretariat. Da mehrere – auch nicht-datenbankerfahrene – Mitarbeiter mit der Datenbank arbeiten, wodurch die Bedienungsfreundlichkeit im Vordergrund steht, wurde mit Access eine umfangreiche Datenbank mit einem Backend und mehreren Frontends entwickelt.
Um auf den Internet-Seiten von IMPETUS (http://www.impetus.uni-koeln.de) stets aktuelle Mitarbeiter- und Institutslisten mit Kontaktinformationen zu präsentieren, wird die Projektdatenbank skriptbasiert auf den Webserver der Universität Köln in eine MySQL-Datenbank kopiert und dort dynamisch mit PHP abgefragt. Sie dient damit gleichzeitig als Datenbank für das Content Management System (CMS) der IMPETUS-Internetpräsentation.
Der Metadatenkatalog ist durch den jeweiligen Ansprechpartner und die Zuordnung der Datensätze zu den einzelnen Teilprojekten direkt mit der Projektdatenbank verknüpft. Aus diesem Grund wurde auch der Metadatenkatalog in Form einer Metadatenbank in die Access-Datenbank eingefügt.

Web-Anbindung

Wie oben bereits erwähnt, werden die in einer Access-Datenbank gehaltenen Metainformationen in eine MySQL-Datenbank auf dem Webserver der Universität Köln kopiert (Abb. 7).
Die Abfrage der Metadaten erfolgt auf den Internetseiten von IMPETUS (www.impetus.uni-koeln.de) mit PHP-Skripten. Die Recherchemöglichkeiten werden unterschieden in interne und externe Nutzer.

Derzeit sind folgende Suchmöglichkeiten realisiert:

- Thesaurus
- Teilprojekt
- Bounding Box
- Kontakt-Person (nur intern)

Geplant sind weiterhin eine Suche nach Datum sowie eine Volltextsuche über ausgewählte Felder (Titel, Beschreibung, Parameter, Methoden, Instrumente, Ortsbeschreibung, usw.). Interne Nutzer erhalten über die Suchergebnisse direkt einen Download-Zugang, falls der entsprechende Datensatz auf einem Internet-Server bereit liegt. Externe Nutzer erhalten diesen Zugang nur für Datensätze, die für die öffentliche Nutzung freigegeben sind.
Für die Erfassung der Metadaten sind die Datenautoren selbst zuständig, da nur sie über alle erforderlichen Informationen verfügen. Zur Erfassung der Metainformationen ist ein Online-Formular vorhanden. Alle Einträge werden – vollständig ausgefüllte Pflichtfelder vorausgesetzt – per Email an das Datenmanagement weitergeleitet. Dort werden die übermittelten Informationen in die Access-Datenbank eingegeben. Die Erfahrung zeigt, dass diese – durchaus aufwändige – Vorgehensweise berechtigt ist. Viele Mitarbeiter sind mit dem Instrument Metadatenbank noch nicht vertraut. Daher werden z.B. öfters Informationen in falsche Felder eingetragen. Die zentrale Eingabe in die Datenbank ermöglicht darüber hinaus eine Vereinheitlichung der Einträge. Und schließlich sollte die Thesaurierung ebenfalls von eingewiesenen und erfahrenen Mitarbeitern durchgeführt werden. So kann eine gleichmäßig hohe Qualität der Metainformationen erreicht werden.

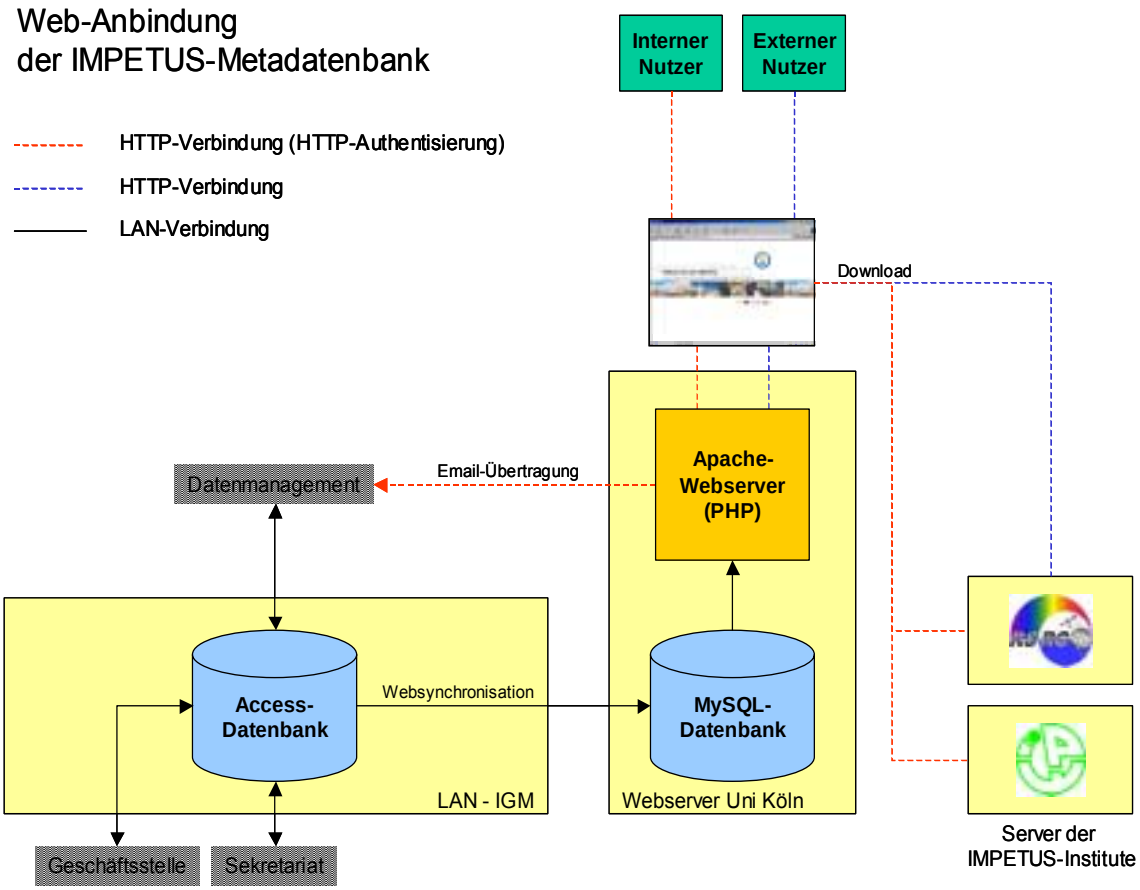


Abb. 7: **Webanbindung der IMPETUS-Metadatenbank**

Ausblick

Derzeit läuft die Metadatenbank im Testbetrieb. Bei der Erfassung der ersten Datensätze wird untersucht, ob sich die Struktur der Datenbank für die unterschiedlichen Fachdisziplinen und deren Anforderungen eignet. Gleichzeitig sollen die Datenautoren Erfahrung mit der Erfassung von Metainformationen sammeln. Eventuell lässt sich zukünftig ein direkter Eintrag in die Metadatenbank verwirklichen. Eine zentrale Kontrolle der Einträge wird aber immer erfolgen müssen. Die Metadatenbank ist bereits öffentlich zugänglich. Der direkte Zugriff auf die einzelnen Datensätze ist aber vorerst nur intern möglich. Nach und nach werden die Datensätze auch für die Öffentlichkeit freigegeben. Über die Internet-Anbindung und die englischsprachige Benutzerführung ist grundsätzlich ein weltweiter Zugang zur Metadatenbank und zu den Datensätzen möglich.

Außenstehende Interessenten müssen sich mit Namen und Email-Adresse identifizieren, den Verwendungszweck der Daten angeben und die Nutzungsbedingungen anerkennen.
Weiterhin ist vorgesehen, relevante Daten auf einem zentralen FTP-Server abzulegen, um die allgemeine Verfügbarkeit – auch über die Projektlaufzeit hinaus – sicherzustellen.

Literatur

ARBEITSGEMEINSCHAFT DER VERMESSUNGSVERWALTUNGEN ADV (2002): Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens, Kapitel 9 – Metadatenkatalog. Version 1.0, Stand 01.02.2002.

DENZER R. u. G. GÜTTLER (1994): Integration von Umweltdaten. In: PAGE, B. u. M. HILTY (Hrsg.): Umweltinformatik – Informatikmethoden für Umweltschutz und Umweltforschung. Handbuch der Informatik, Oldenbourg, München, 360 S.

FEDERAL GEOGRAPHIC DATA COMMITTEE (2000): Content Standard for Digital Geospatial Metadata Workbook Version 2.0. Federal Geographic Data Committee. Washington, D.C.

IMPETUS (2003): IMPETUS Westafrika – Integratives Management-Projekt für einen Effizienten und Tragfähigen Umgang mit Süßwasser in Westafrika, 3. Zwischenbericht, 129 S.

STROBL, J. (1995): Grundzüge der Metadatenorganisation für GIS. In: DOLLINGER, F. u. J. STROBL: Angewandte geographische Informationstechnologie VII. Beiträge zum GIS-Symposium vom 5. - 7. Juli 1995. (http://www.agit.at/papers/1995/jstrobl.pdf).

WERSIG, G. (1985): Thesaurus-Leitfaden – Eine Einführung in das Thesaurus-Prinzip in Theorie und Praxis. 2. Aufl., DGD-Schriftenreihe

GEOGRAFISCHES FACHINFORMATIONSSYSTEM ZUR VERTEILTEN VERWALTUNG VON MILITÄRISCHEN ALTLASTEN

Prof. Dr.-Ing. Uwe Rüppel
Institut für Numerische Methoden und Informatik im Bauwesen,
Technische Universität Darmstadt

Dipl.-Ing. Gerrit Seewald
Dr.-Ing. Michael Petersen
CIP Ingenieurgesellschaft mbH, Darmstadt

Kurzfassung

Sechzig Jahre nach Ende des letzten Weltkrieges stellt sich bei allen Bauvorhaben, insbesondere im Bereich Grundbau, immer noch die Frage nach möglichen Belastungen durch militärische Kampfstoffe und Rüstungsaltlasten. Für den einzelnen Bauherren besteht die Möglichkeit, eine Belastungsanfrage bei dem zuständigen Kampfmittelräumdienst zu stellen, um somit schon in der Planungsphase geeignete Maßnahmen ergreifen zu können. Die Anfrage zu einer möglichen Kampfmittelbelastung löst dabei eine Vielzahl von Arbeitsschritten aus, die momentan durch personal- und zeitaufwändige Verwaltungsvorgänge und Datenrecherchen bearbeitet werden. Die Daten liegen derzeit überwiegend noch analog bzw. in Papierform, wie beispielsweise historischen Luftbildern, Zeitzeugenberichten oder Räumberichten, vor und stellen einen umfangreichen Datenbestand dar. Um die Verwaltung der geografischen Informationen über militärische Altlasten zu verbessern und damit die Effizienz der Bearbeitung der Belastungsanfragen zu erhöhen, wurde eine detaillierte Analyse der anfallenden Datenbestände durchgeführt. Aufbauend auf dieser Analyse wurden fachspezifische Softwarekomponenten entwickelt, um durch DV-gestützte Methoden und eigenständige Fachschalen die Verwaltung der Datenbestände zu unterstützen und den digitalen Zugriff auf einzelne Datensätze zu ermöglichen.

1. Einleitung

Das Thema Kriegsaltlasten ist bei vielen Bauvorhaben in Deutschland von Bedeutung. Beginnend bei den frühen Planungsphasen und Vorerkundungen bis hin zur eigentlichen Bauausführung und dabei insbesondere den Erdbauarbeiten, muss auf die unterschiedlichen Arten von Altlasten zeitnah reagiert werden können,
Eine gezielte Erfassung der Altlasten kann jedoch nur erfolgen, wenn eine differenzierte Einteilung der unterschiedlichen Arten gegeben ist, um so fachgerechte Maßnahmen ergreifen zu können.
Im Bereich Altlasten werden unterschieden:

- Altstandorte: Ehemalige Grundstücke auf denen mit umweltgefährdenden Stoffen umgegangen wurde.
- Stillgelegte Ablagerungen: Ehemalige Abfallbeseitigungsanlagen sowie sonstige Grundstücke mit Altablagerungen.
- Altlasten-Verdachtsflächen: Alle weiteren Flächen die einen Verdacht auf das Vorliegen von Schadstoffen aufweisen.

Insbesondere für die beiden ersten Fälle existieren mittlerweile verschiedene Lösungsansätze, die sowohl auf Bundesebene, auf Initiative des Umweltbundesamtes als auch auf Länderebene gezielt vorangetrieben werden und sich mittlerweile im Einsatz bzw. in der Erprobung befinden.

Im Rahmen dieses Beitrages wird auf den Bereich der militärischen Altlasten eingegangen, die verglichen mit den übrigen Bereichen der Altlasten ein besonders hohes Gefährdungspotential und eine akute Bedrohung für Leben und Gesundheit von Menschen darstellen. Problematisch sind dabei vor allem die mögliche Selbstentzündung und Explosion sowie die Inhaltsstoffe, die durch Zersetzung bzw. Korrosion der Ummantelung entweichen können. Der Umgang mit Kampfmitteln sollte daher ausschließlich erfahrenen Fachkräften vorbehalten sein [1]. Die Verwaltung georeferenzierter Informationen über Kampmittel ist als Teil der jeweiligen Ordnungsbehörden auf der Ebene der Bezirksregierungen und Regierungspräsidien der einzelnen Bundesländer angesiedelt.

Trotz unterschiedlicher Vorgehensweisen, bedingt durch den Föderalismus der Bundesländer, sind Gemeinsamkeiten im Arbeitsablauf der einzelnen Kampfmittelräumdienste (KMRD) zu erkennen.
Im Auftrage des Landes Hessens, vertreten durch den Kampfmittelräumdienst, wurde von der Firma CIP Ingenieurgesellschaft mbH Darmstadt ein Framework verschiedener Fachapplikationen für die Unterstützung bei der Verwaltung und Bearbeitung der anfallenden Informationen bei der Kampfmittelräumung entwickelt. Diese Entwicklung verlief dabei auf der Grundlage einer umfassenden Analyse des vorhandenen Arbeitsablaufs beim Kampfmittelräumdienst. Die Implementation erfolgte in einem mehrstufigen, modularen Verfahren, das es erlaubt, die einzelnen Komponenten für die verschiedenen fachspezifischen Anwendungsfälle getrennt voneinander einzusetzen [2].
Ziel dieses geografischen Fachinformationssystems ist es, unter Verwendung moderner Informations- und Kommunikationstechnologien ein Framework an Fachapplikationen zu schaffen, um zielgerichtet eine netzwerkgerechte und bedarfsorientierte Bereitstellung georeferenzierter Fachinformationen aus einem zentralen Datenpool zu ermöglichen. Am Beispiel der Verarbeitung von Altlasten-Informationen, insbesondere aus dem Bereich der militärischen Altlasten, soll in den folgenden Kapiteln der ganzheitliche Ansatz für das Informationssystem vorgestellt und dessen Einsatz sowohl während der Vorplanung als auch während der Baudurchführung aufgezeigt werden.

2. Bedarfsanalyse

2.1. Analyse der Prozessbeteiligten

Im Hinblick auf eine Integration von Ingenieurprozessen der Planungsphasen bis hin zur Bauausführung sind zunächst die beteiligten Fachplaner und Prozessbeteiligten zu identifizieren, um darauf aufbauend die Arbeitsabläufe zu spezifizieren.

a) Bauherr, Eigentümer, Investor

Der Bauherr ist der Initiator für die nachfolgenden Prozesse. Durch rechtzeitige Anfragen bzgl. möglicher Altlasten kann er sicherstellen, dass sowohl die Durchführung der Bauarbeiten als auch eine unbeeinträchtigte Nutzung des zu erstellenden Bauwerks möglich ist.

b) *Zuständige Landesbehörde(n)*

Die zuständige Landesbehörde ist zunächst mit der Bearbeitung der gestellten Anträge betraut. Dazu ist ein effektiver Zugriff sowohl auf alle bisher bearbeiteten Anträge, auf Auswertungsergebnisse der Luftbilddetailauswertung als auch auf alle bisher durchgeführten Räum- und Sanierungsarbeiten notwendig.

c) *Fachingenieure*

Die fachliche Beurteilung von Flächen hinsichtlich möglicher Altlasten erfolgt durch spezialisierte Fachingenieure, die insbesondere für die Luftbilddetailauswertungen, weiterführende Sondierungsmaßnahmen sowie zur Definition von geeigneten Maßnahmen zum Schutz der Öffentlichkeit im Rahmen von Räummaßnahmen hinzugezogen werden. Bei der Luftbilddetailauswertung fallen dabei umfassende Datenbestände von digitalisierten Luftbildern an, die durch die alliierten Streitkräfte unmittelbar vor und nach einem Angriff systematisch erstellt wurden. Diese werden stereoskopisch ausgewertet, um so präventiv mögliche Verdachtsstellen für militärische Altlasten entdecken zu können.

d) *Sondierungs- und Räumungsfirmen*

Die Sondierungs- und Räumungsfirmen übernehmen die Dekontamination sowie die Sanierungsmaßnahmen des Baugrundes im Auftrag des Bauherrn. Es fallen dabei umfangreiche Fach- und Geodaten an, die momentan in uneinheitlicher Form dem zuständigen Kampfmittelräumdienst zur Bearbeitung übergeben werden.

e) *Externe Beteiligte*

Als externe Beteiligte werden diejenigen Behörden, Personen und Institution bezeichnet, die nicht direkt in den Prozess der Kampfmittelräumung eingebunden sind, für deren Arbeitsablauf jedoch ein stark eingeschränkter Zugriff auf die vorgehaltenen Datenbestände sinnvoll wäre. Beispielhaft sind dazu Bau- und Umweltämter zu nennen, die im Rahmen von Baugenehmigungs-, Raumordnungs- oder Planfeststellungsverfahren Informationen über Belastungen durch militärische Altlasten benötigen.

2.2. Analyse der Prozesse

Bei der Analyse der Prozesse wird zwischen folgenden Arbeitsabläufen (Workflows) unterschieden:

a) Statische (geplante) Workflows

b) ad-hoc (ungeplante) Workflows

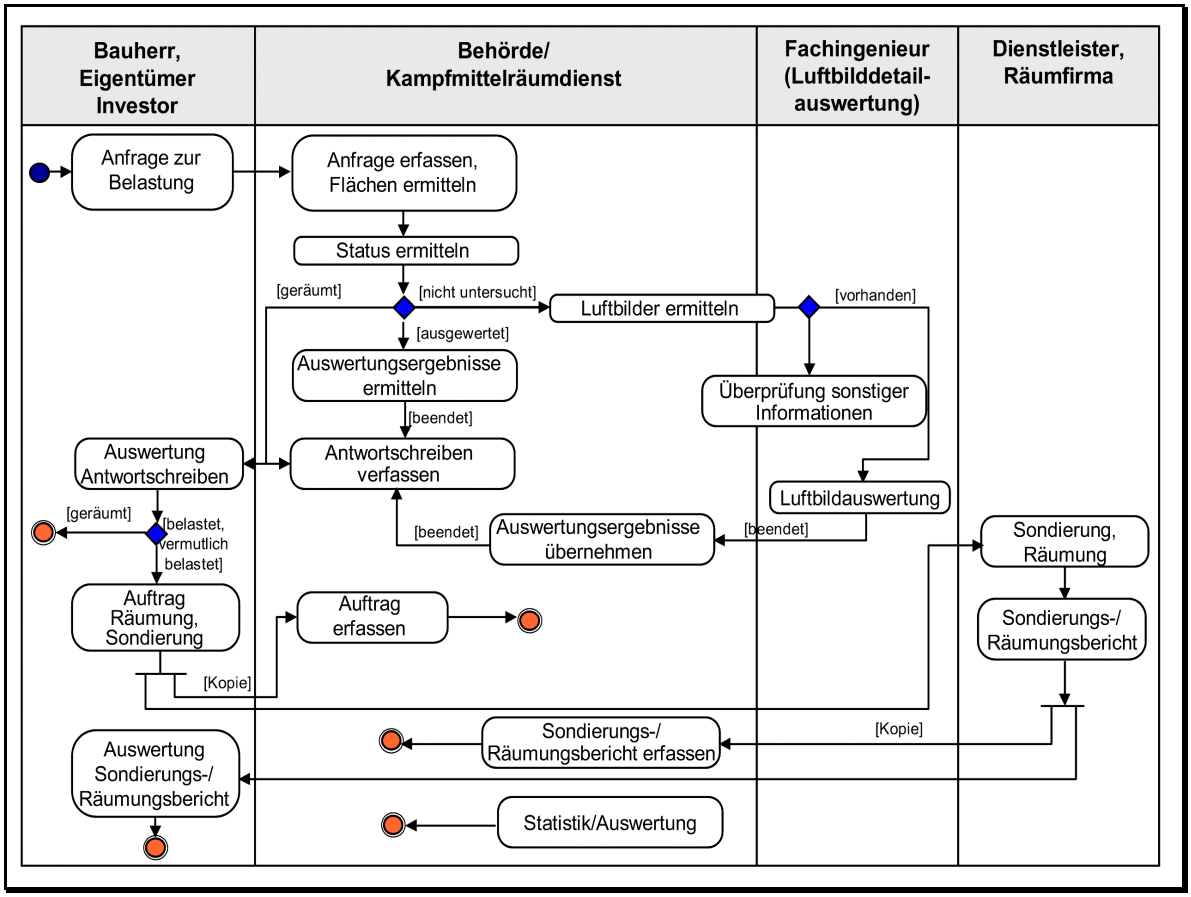


Abbildung 1: **Aktivitäten der beteiligten Fachplaner (vereinfacht) als Vorstufe der Software-Komponentenbildung [3]**

Die Einteilung nach diesen beiden Kriterien stellt eine wichtige Randbedingung für die Konzeption der Workflow-Steuerungskomponente dar. Der statische bzw. geplante Workflow stellt dabei den üblichen Prozess im Gesamtworkflow dar. Darunter fallen die normalen Anträge der Bauherren während des Bauplanungsprozesses oder während der Bauausführung. Der ad-hoc Workflow beschreibt die so genannten Einzel- und Zufallsfunde, die vor allem während des Bauablaufs auftreten und daher eine zeitnahe Untersuchung, Beurteilung und Räumung der vorhandenen militärischen Altlasten erfordern.

Entscheidend dabei ist, dass solche Situationen im Vorfeld nur schlecht planbar sind und daher den betroffenen Beteiligten eine umfassende, dynamische Unterstützung durch vorhandene Datenbestände und Programmkomponenten zur Verfügung gestellt wird.

2.3. Anforderungsanalyse

Bei den in Abbildung 1 aufgeführten Vorgängen lassen sich drei verschiedene Anforderungen identifizieren, um die vollständige, digitale Erfassung der anfallenden Daten aus allen Teilprozessen und deren Integration in ein ganzheitliches Fachinformationssystem sicherzustellen:

a) Aufbau einer Datenbank, die eine Verknüpfung von Sachdaten und geografischen Informationen (Geodaten) ermöglicht.

b) Steuerung der Zugriffe durch die Projektbeteiligten auf Sach- und Geodaten.

c) Fachgerechte Visualisierung der vorhandenen georeferenzierten Datenbestände.

d) Flexible Verwaltung der vorhandenen Arbeitsabläufe durch eine unabhängige Steuerungskomponente.

e) Berücksichtigung der unterschiedlichen Sichtweisen und Anforderungen der verschiedenen Prozessbeteiligten.

f) Implementierung einer geeigneten Schnittstelle auf Basis von Metadaten, um eine dynamische Erweiterung der Zugriffe zu ermöglichen.

3. Schnittstellendefinition

Für den eingeschränkten Zugriff auf die vorhandenen Sach- und Geodaten durch externe Beteiligte, wird eine einheitliche Schnittstelle benötigt. Sowohl Bau- und Umweltämter während ihren umfassenden Genehmigungsverfahren als auch Architektur- und Planungsbüros während der Bauplanung benötigen Informationen zu möglichen Belastungen durch Kriegsaltlasten. Um eine sinnvolle Mehrfachnutzung der vorhandenen Datenbestände zu ermöglichen, wird hierzu eine flexible und zugleich standardisierte Schnittstelle für den allgemeinen Austausch von Geodaten benötigt und in dem Fachinformationssystem abgebildet.

In der sehr heterogenen Metadatenlandschaft, die sich mit der detaillierten Beschreibung der vorhandenen Geodaten beschäftigt, existieren derzeit eine Vielzahl an amtlichen, behördlichen oder auch privaten Metadatenmodellen. Die beiden wichtigsten Normierungsansätze werden von den nachstehenden Standardisierungsgremien aufgestellt:

- ISO/TC211 19115: International Standards Organization (ISO).
- FGDC-STD-001-1998: Federal Geographic Data Comitee (FGDC)

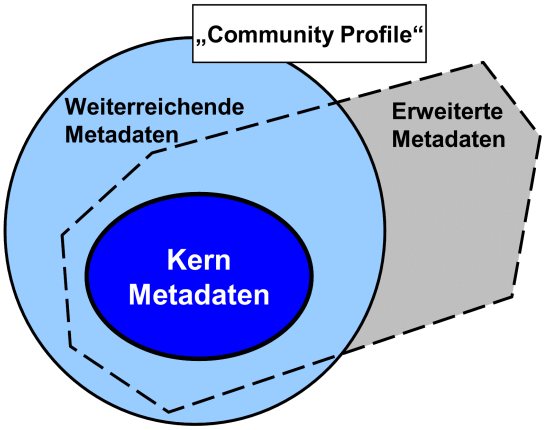


Abbildung 2: **Gliederung von Metadaten gemäß ISO 19115 [4]**

In der Beschreibung durch die ISO 19115 [4] werden dabei rund 300 Metadatenelemente aufgeführt, die sich in drei Bereiche des Metadatenmodells aufteilen (Abb. 2). Der Kernbereich umfasst die zentralen Elemente der Metadatenbeschreibung, die auf jeden Fall implementiert und durch alle Anwendungen abgedeckt werden müssen. Alle optionalen Metadatenelemente werden dem Bereich der weiterreichenden Metadaten zugeordnet. Für eine individuelle Erweiterung und Anpassung des Standards ist in der ISO bereits die Möglichkeit vorgesehen, erweiterte Metadaten zu integrieren und diese zu einem „Community Profile“ zusammenzuschließen. Dieses „Community Profile“ dient der Erweiterung von regionalen, nationalen oder fachspezifischen Anforderungen und umfasst dabei stets die Kern Metadaten sowie ausgewählte, d.h. weiterreichende Metadaten.

Die Entwicklung eines umfassenden Metadatenbankmodells (Community Profile) für militärische Altlasten und der Beschreibung der vorliegenden georeferenzierten Datenbankinhalte beschränkt sich dabei in der ersten Stufe auf den Kernbereichen (siehe Abb. 2), der im einzelnen in der ISO 19115 [4] festgelegt ist.

1. **Allgemeine Kriterien**

   a. Identifizierung der Metadaten
   Verwendeter Metadatenstandard, Urheber, Gültigkeit, Sprache, Zeichensatz *(mdStanName, mdContact, mdDateSt, mdLang, mdChar)*

   b. Beschränkungen
   Zugriffsberechtigungen, Nutzungsmöglichkeiten, Sicherheit *(accessConsts, useLimits, SecConsts)*

   c. Datenaustausch
   Austauschmöglichkeiten der Daten, Digitale Übertragung, Formatbeschreibung, Mediumbeschreibung *(MD_Distribution, MD_DigitalTransferOptions, MD_Format, MD_Medium)*

2. **Inhaltliche Kriterien**

   a. Beschreibung der Inhalte
   (MD_ContentInformation)

   b. Räumliche Ausdehnung
   (MD_CoverageDescription)

   c. Multimediale Metadaten
   Beschreibung von digitalen Bildern hinsichtlich Qualität und

3. **Informationen über die Datenart**

   a. Geografische Ausdehnung
      (EX_GeographicExtend)

   b. Zeitliche Ausdehnung
      (EX_TemporalExtent)

Die Bereitstellung dieser „Kerndaten“, die sich als Pflichtfelder aus dem Normierungsvorschlag herleiten, wurde auf Basis von XML konzipiert. Die Beschreibung der einzelnen Auszeichnungselemente, Beziehungen, Attribute, Strukturen und Bereiche erfolgt dabei in einer separaten Datei als Document Type Definition (DTD), die zunächst unabhängig von den Inhalten das Format der XML Austauschdatei beschreibt und darüber hinaus die Möglichkeit bietet, das Format hinsichtlich der Gültigkeit zu verifizieren (Abb. 3).

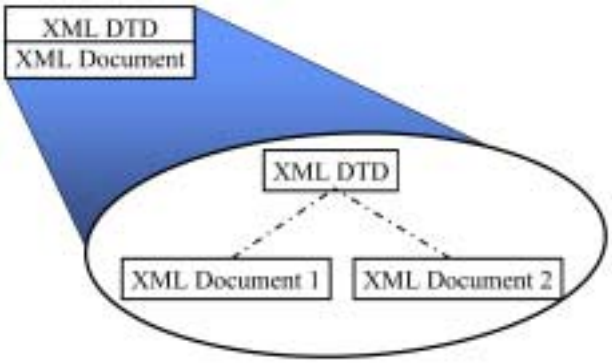


Abbildung 3: **Beschreibung von Metadaten mittels XML [4]**

4. Systementwurf

Aus den Anforderungen, die sich aus der Analyse der Anwendungsfälle und Nutzerrollen ergeben, wurde ein Framework aus einzelnen Softwarekomponenten entwickelt. Den Prozessbeteiligten wird durch die Integration von gebiets- und projektbezogenen Sichtweisen ein ganzheitliches System für die Unterstützung bei den einzelnen Arbeitsschritten der Kampfmittelräumung zur Verfügung gestellt (Abb. 4).

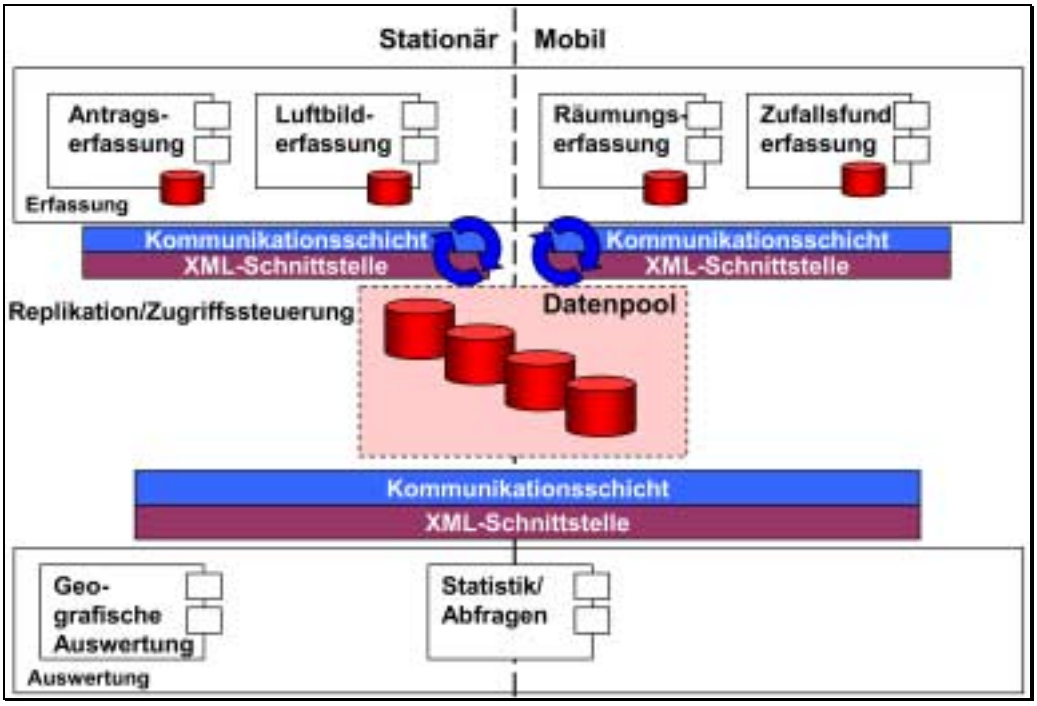


Abbildung 4: **Systementwurf**

4.1 Steuerungskomponente

Die Steuerungskomponente wurde als zentraler Serverdienst entwickelt und soll die Verwaltung und Zugriffssteuerung auf die gesammelten Daten und die Verwaltung der statischen und dynamischen Workflowbausteine sicherstellen. Als Datenbanksystem werden gegenwärtig Microsoft SQL Server/MSDE und Oracle eingesetzt. Für diese wurde ein datenbankinterner Protokollmechanismus entwickelt, der es erlaubt, auf Basis der Trigger-Technologie alle vom Nutzer durchgeführten Änderungen zu protokollieren. Auf der Basis dieser Protokolle erfolgt der XML-basierte Datenaustausch und die Replikation der verteilten Datenbanken. Weiterhin stellt diese Steuerungskomponente sicher, dass die vorhandenen Metadaten gemäß den Kernbedingungen der ISO 19115 propagiert werden.

4.2 Antragsbearbeitung

Für die Erstellung und Verwaltung der Anträge aus den Bereichen des Kampfmittelräumdienstes ist eine eigenständige Antragsverwaltung entwickelt worden. Diese erlaubt es, mittels Benutzermasken unter MS Windows, die Antragsdaten in ein eigenständig entwickeltes, relationales Datenbankschema einzuspeisen und nach Bedarf mit der zentralen Steuerungskomponente abzugleichen. Aufbauend auf der Verwaltungs-, Erfassungs- und Rechercheebene, wird durch die Steuerungskomponente auch der interaktive Austausch über die verknüpften Workflowelemente implementiert.

4.3 Geografische Komponente

Die geografische Komponente dient dem übersichtlichen, kartengestützten Zugriff auf die einheitlich im Gauß-Krüger-System vorliegenden Datenbestände. Auf Basis von Topographischen Karten in den Maßstäben von 1:200.000 bis zu 1:5.000 werden die Geodaten dynamisch dem Anwender angezeigt, so dass er direkt auf die zugrunde liegenden Sachdaten zugreifen kann.

a) Lokale Arbeitsplatzkomponente

Die Darstellung erfolgt über ein fachspezifisches Erweiterungsmodul auf Basis der ObjectARX-Schnittstelle innerhalb des CAD-Systems AutoCAD [5]. Das vorhandene Kartenmaterial sowie verwaltete Luftbilder werden hierzu als Pixelgrafiken eingebunden, auf deren Grundlage die Geodaten vektorisiert angezeigt werden.

b) Internet-Auskunftsarbeitsplatz

Für internetgerechte Darstellungen und Zugriffsmöglichkeiten auf die zugrunde liegenden Geodaten lassen sich die geografischen Informationen aus den jeweiligen Datenbanken, verknüpft mit ihren rudimentären Geodaten im Shape-Format, extrahieren. Für die fachgerechte, visuelle Darstellung im Internet dient ein sogenannter Kartenserver. Als Software-Tool für die hybride Darstellung dient der MapServer der University of Minnesota (UMN) [6].

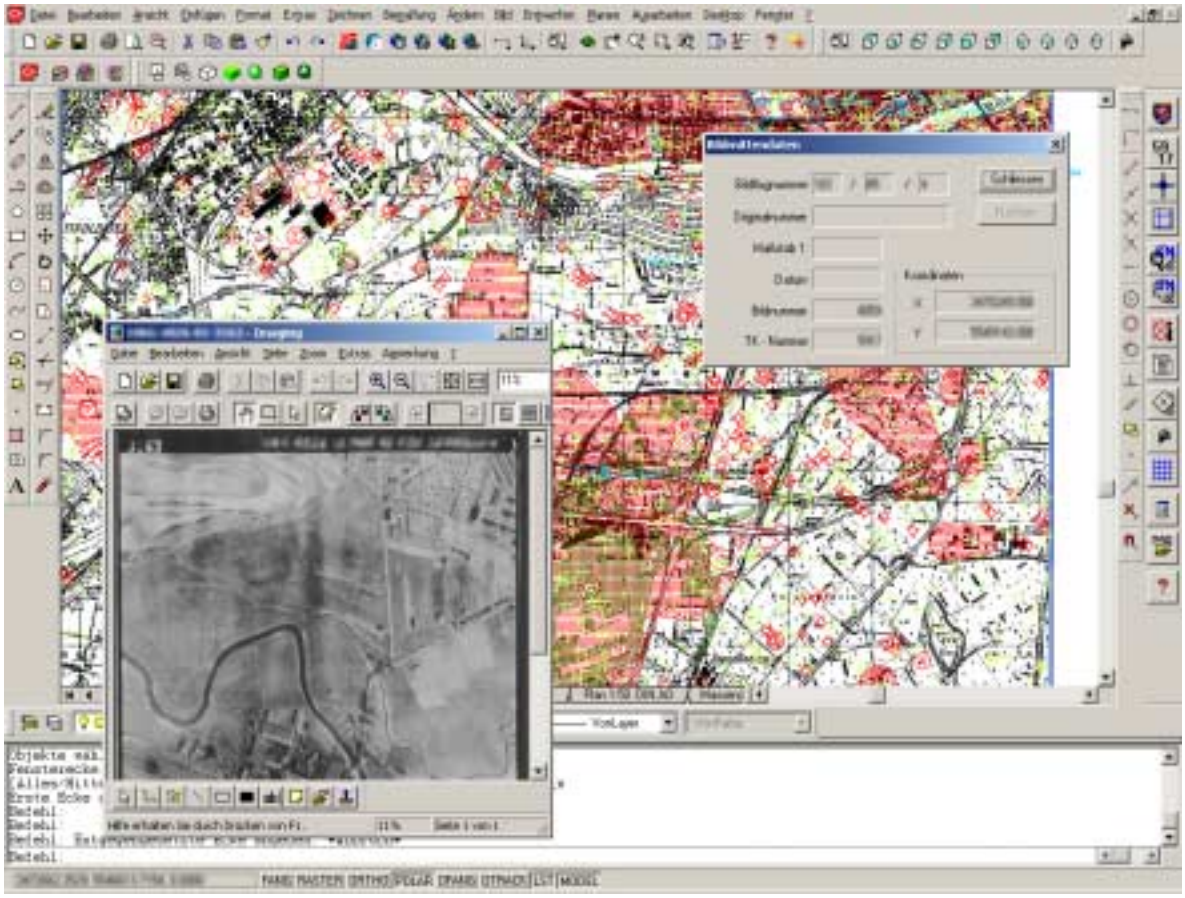


Abbildung 5: **GIS-Fachkomponente für militärische Altlasten auf Basis von AutoCAD**

4.4 Mobile Erfassung von Räuminformationen

Für eine mobile Erfassung, insbesondere bei Räumungs-, Sondierungs- und Sanierungsmaßnahmen, wurden eigenständige Softwarekomponenten entwickelt, die es den Dienstleistungsfirmen ermöglichen, die anfallenden Daten mit Hilfe von geeigneten Benutzerschnittstellen direkt vor Ort auf einem Laptop aufzunehmen. Der Einsatz von Personal Digital Assistants (PDA), mit GPS Empfängern ist vorgesehen [7]. Die erfassten Daten lassen sich nach Beendigung der Arbeiten in Berichtsform ausgeben und dienen somit sowohl der internen Verwaltung durch die Dienstleister als auch als analoger Datenaustausch in Form von Abgabebelegen für die zuständige Überwachungsbehörde. Über die entwickelten Replikationsmechanismen lässt sich darüber hinaus auch ein Austausch der erfassten Daten im XML-Format mit der zentralen Datenbank durchführen.

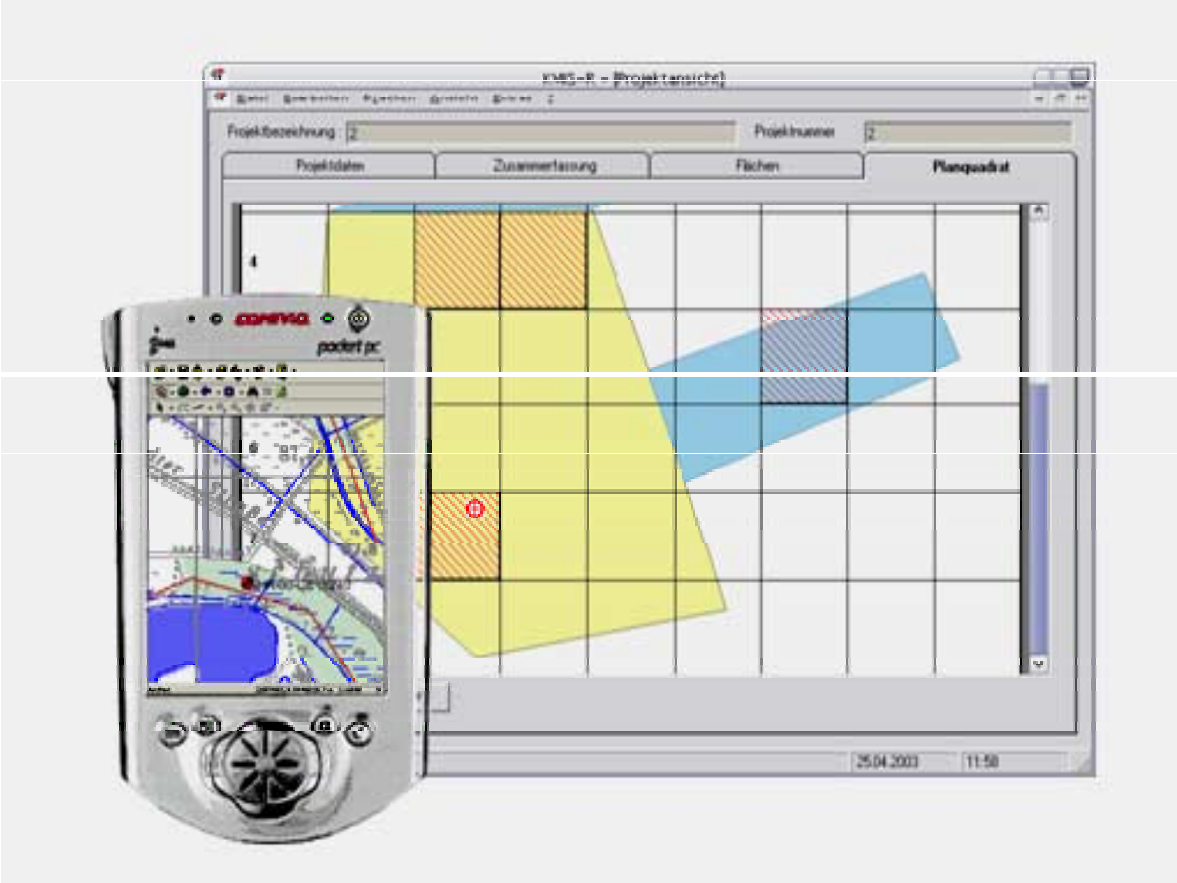


Abbildung 6: **Komponente zur mobilen Erfassung von Räumstelleninformationen**

4.5 Einbindung der Luftbilddetailauswertung

Für die Einbindung der Ergebnisse aus fachlichen Auswertungen durch externe Dienstleister bei der Luftbilddetailauswertung wird ebenfalls eine fachspezifische Softwarekomponente benötigt. Bei der Luftbilddetailauswertung handelt es sich um gezielte stereoskopische Auswertungen aus den umfangreichen Luftbildarchiven der Alliierten. Diese Archive enthalten Luftbilder, die geordnet nach Flugstreifen und Flugnummer die einzelnen Bombardierungen dokumentieren. Es wird dabei zwischen punktuellen Auswerteergebnissen, Flakstellungen und Bombenkratern sowie flächenhaften Ergebnissen, in Form von Trichterfeldern unterschieden. Bei der fachlichen Auswertung müssen die Ergebnisse der Luftbilddetailauswertung getrennt nach Geo- und Sachdaten erfasst und zentral zusammengeführt werden. Dazu wird eine GIS-gestützte Eingabeumgebung für die einzelnen Dienstleister zur Verfügung gestellt, welche die Ergebnisse fachgerecht erfassen und in einer lokalen Datenbank bzw. in einem strukturierten XML Format ablegen. Die gesammelten Daten lassen sich im Anschluss wieder für eine zentrale Auswertung über eine definierte XML Schnittstelle bereitstellen.

5. Zusammenfassung und Ausblick

Der Beitrag beschreibt eine Lösung zur vollständigen digitalen Verarbeitung der anfallenden Daten aus den Teilprozessen der Erfassung, der Verarbeitung und der Auswertung bodenbelastender Stoffe am Beispiel militärischer Altlasten. Für die unterschiedlichen Prozessbeteiligten wurde ein integriertes Informationssystem entwickelt, das alle Bearbeitungsschritte unterstützt und eine Bereitstellung der Metadaten gemäß ISO 19115 sowie einen Austausch der Daten mittels XML ermöglicht.

Für die Erweiterung der mobilen Komponente ist die Verwendung von GPS-Geräten vorgesehen, um so die Erfassung der Geodaten effizient und präzise zu gestalten. Weiterhin ist vorgesehen, Systeme zur automatisierten Luftbildauswertung und Bilderkennung in das Gesamtsystem zu integrieren. Für qualitätsbehaftete Messwerte ist eine Schnittstelle zu stationären sowie zu mobilen Messlaboratorien konzipiert, um die Ergebnisse möglichst ohne zeitliche und inhaltliche Verluste dem Anwender zur Verfügung zu stellen.

Geografisches Fachinformationssystem zur verteilten Verwaltung von militärischen Altlasten

Dipl.-Ing. Gerrit Seewald
Dr.-Ing. Michael Petersen
CIP Ingenieurgesellschaft mbH

Prof. Dr.-Ing. Uwe Rüppel
Institut für Numerische Methoden und Informatik im Bauwesen, TU Darmstadt

Berlin, 19.05.2003

Gliederung

1. **Einführung**
2. **Bedarfsanalyse und Workflowbeschreibung**
3. **Strukturierung der Fachinformationen**
4. **Systemskizze und -komponenten**
5. **Replikation, Schnittstellen und Datenaustausch**
6. **Kommunikationsschicht**
7. **Ausblick**

Gliederung

1. Einführung
2. Bedarfsanalyse und Workflowbeschreibung
3. Strukturierung der Fachinformationen
4. Systemskizze und -komponenten
5. Replikation, Schnittstellen und Datenaustausch
6. Kommunikationsschicht
7. Ausblick

Militärische Altlasten

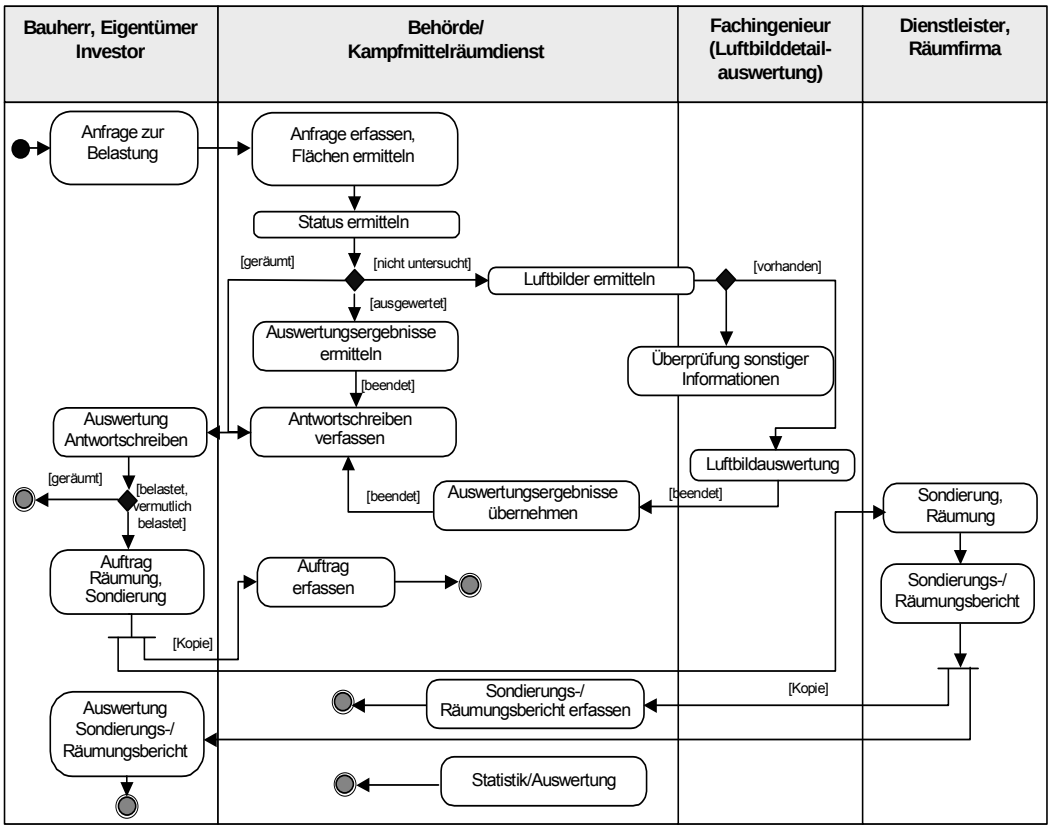
- Altstandorte
  - Sprengstoff-, Pulver- und Kampfstofffabriken
  - Füllstellen, Munitionsanstalten
  - Flug-, Übungsplätze, Kasernen
  - Lager-, Spreng- und Vernichtungsplätze
- „Ablagerungen“
  - Munitionsreste, Blingänger
  - Trichterfelder
  - Flakstellungen

➡ Hohes Gefährdungspotenzial

Aufgabenbereiche der Prozessbeteiligten

- **Bauherr, Eigentümer**
  - Antragsteller, Initiator
- **Landesbehörde**
  - Antragsbearbeitung und Auskunftspflicht
  - Verwaltung/Archivierung
- **Fachingenieure**
  - Luftbilddetailauswertung
- **Sondierungs- und Räumfirmen**
  - Sondierungs- und Räumarbeiten von Altlastenverdachtsflächen
- **Externe Beteiligte**
  - Baugenehmigungs- und Planfeststellungsverfahren
  - Ingenieur-/Architekturbüros

Workflow der Antragsverwaltung



Anforderungsanalyse

1 2 3 4 5 6 7

- Integration aller Fachinformationen in einem gemeinsamen Datenpool
- Steuerung der Zugriffe
- Fachgerechte Visualisierung
- Abbildung von ad-hoc Prozessen
- Einbindung unterschiedlicher Sichtweisen
- Implementierung geeigneter Schnittstellen

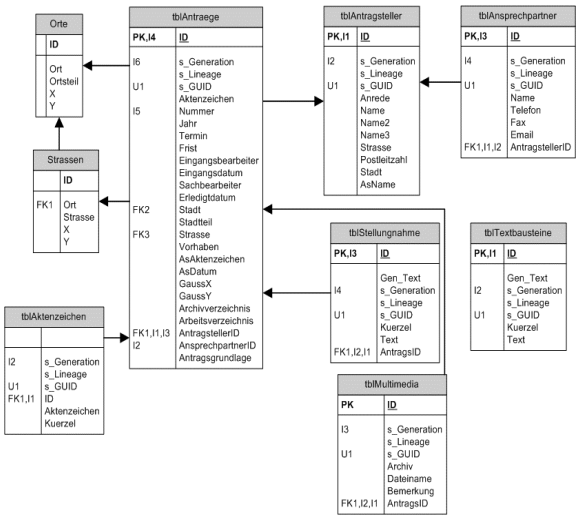
Strukturierung und Beschreibung der vorhandenen Fachinformationen

1 2 3 4 5 6 7

- Stationäre Erfassung
  - Antragsdaten
    - Personenbezogen, Sachbezogen
  - Luftbilddetaildaten
    - ausgewertete, geräumte, belastete Flächen
- Mobile Erfassung
  - Räumstelleninformationen
  - Projektablauf
  - Betriebsinformationen der Räumfirmen
- Räumliche Auswertung
  - Punktuelle Daten
    - Bildmitten von Luftbildern
    - Einzelfunde von Blindgängern
  - Flächenhafte Daten

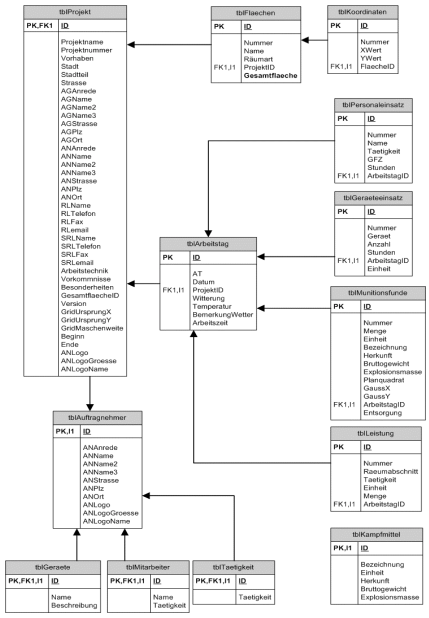
Stationär erfasste Daten/Antragserfassung

- **Allgemeine Antragsinformationen**
  - Aktenzeichen
  - Bearbeitungszeiten, Fristen
- **Personenbezogene Daten**
  - Antragssteller
  - Bearbeiter
- **Berichtswesen**
  - Textbausteine
  - Stellungnahmen
- **Multimediale Ergänzungen**
  - Antwortschreiben
- **Geografische Informationen**
  - Ort, Strasse
  - Gauß-Krüger Koordinaten



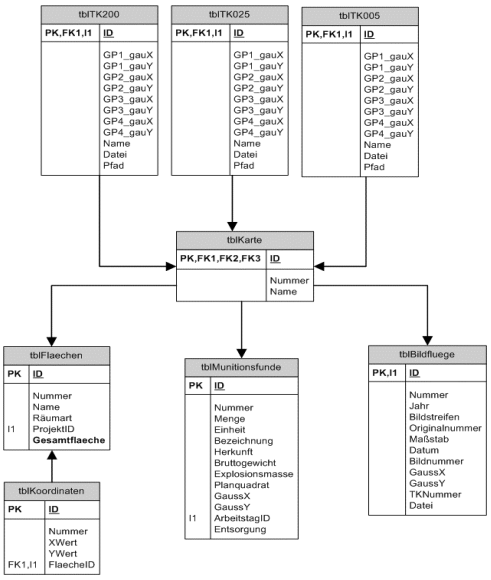
Mobil erfasste Daten/Räumstellen

- **Projektgrundlagen**
  - Detailinformationen
- **Projektablauf**
  - Personaleinsatz
  - Geräteeinsatz
  - Leistungen
  - Munitionsfunde
- **Auftragnehmer**
  - Geräteverwaltung
  - Personalverwaltung
  - Tätigkeitsbeschreibungen
  - Adressinformationen
- **Georeferenzierte Informationen**
  - Geräumte/Untersuchte Flächen
  - Punktuelle Erfassung von Munitionsfunden > 15kg
  - Erfassung von Munitionsfunden < 15kg über Planquadrate

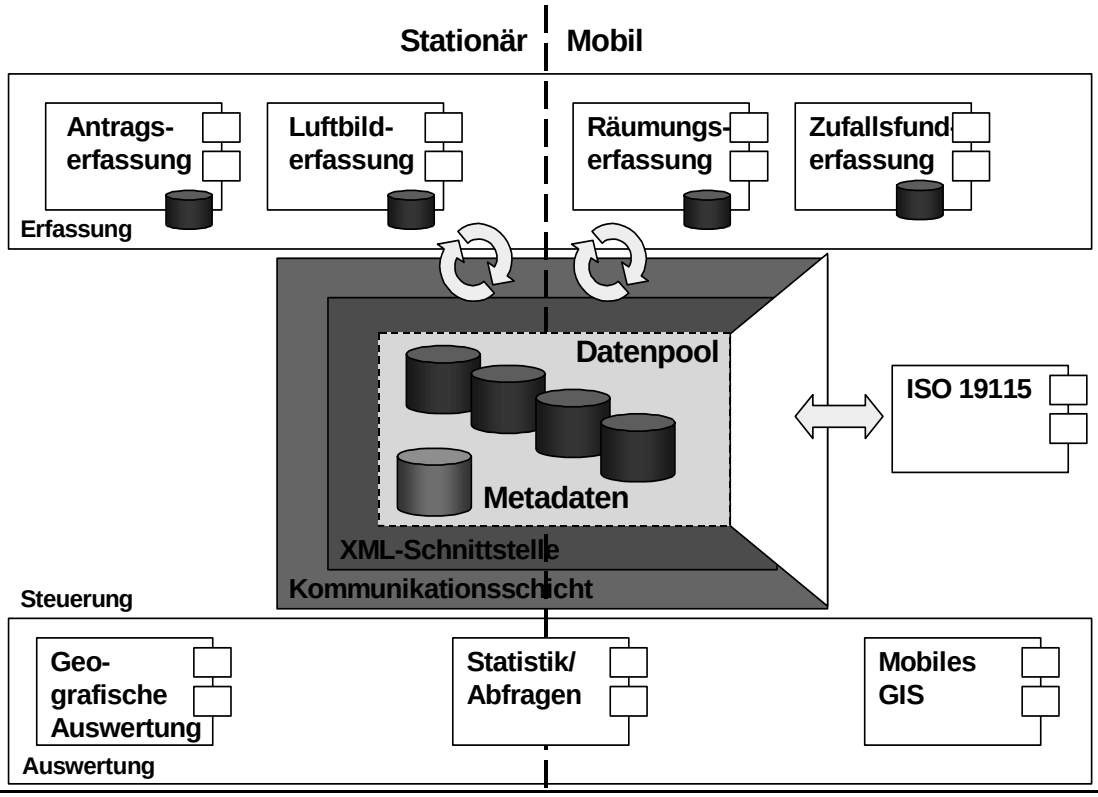


Georeferenzierte Fachinformationen/Auswertung

- **Kartengrundlage**
  - **Topographische Karten 1:5.000, 1:25.000, 1:200.000**
- **Munitionsfunde**
  - **Art, Menge, Zusammensetzung und Herkunft**
- **Bildflüge**
  - **Bildmitten**
  - **Flugdaten (Nummer, Jahr, Bildstreifen)**
- **Flächeninformationen**
  - **Art der Flächen (geräumt, belastet, ausgewertet)**



Systemarchitektur



Anforderungen an Schnittstellen und Datenaustausch

1 2 3 4 5 6 7

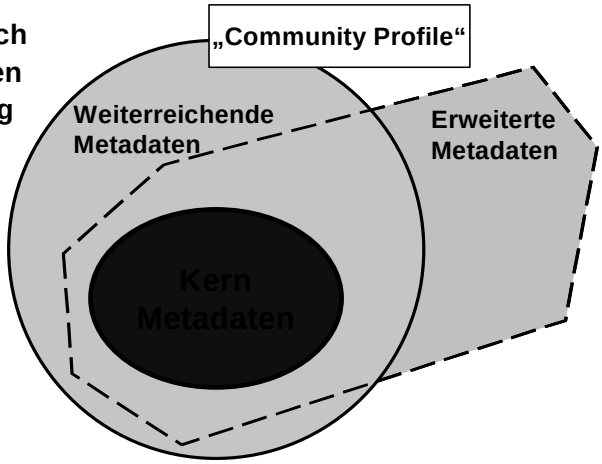
- Integration externer Geodaten (ALK, ATKIS)
- Mehrfachnutzung der Datenbestände
  - Umfassende Beschreibung der vorhandenen Geodaten
  - Nutzung standardisierter Austauschformate
  - Eingeschränkte Zugriffsmöglichkeiten durch Dritte

➡ **Entwicklung eines Metadatenmodells nach ISO/TC211 19115**

Metadatenmodell ISO/TC211 19115

1 2 3 4 5 6 7

- **Kernbereich**
  - **Identifizierung der Metadaten, Datenaustausch**
  - **Inhaltliche Beschreibungen**
  - **Geografische Ausdehnung**
  - **Zeitliche Ausdehnung**
- **Weiterreichender Bereich**
  - **optionale Erweiterungen der Kerninformationen (Beschreibungen, Zusatzinformationen)**
- **Erweiterter Bereich**
  - **Räuminformationen**
  - **Luftbilddetailauswertung**
  - **Antragsverwaltung**

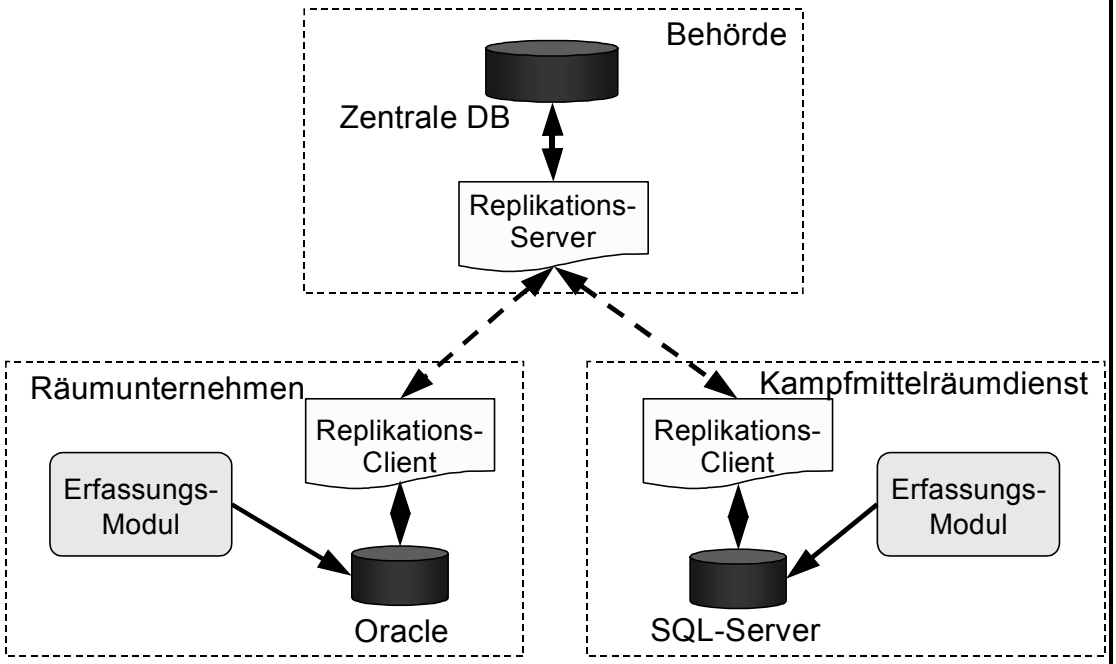


➡ **Communitiy Profile**

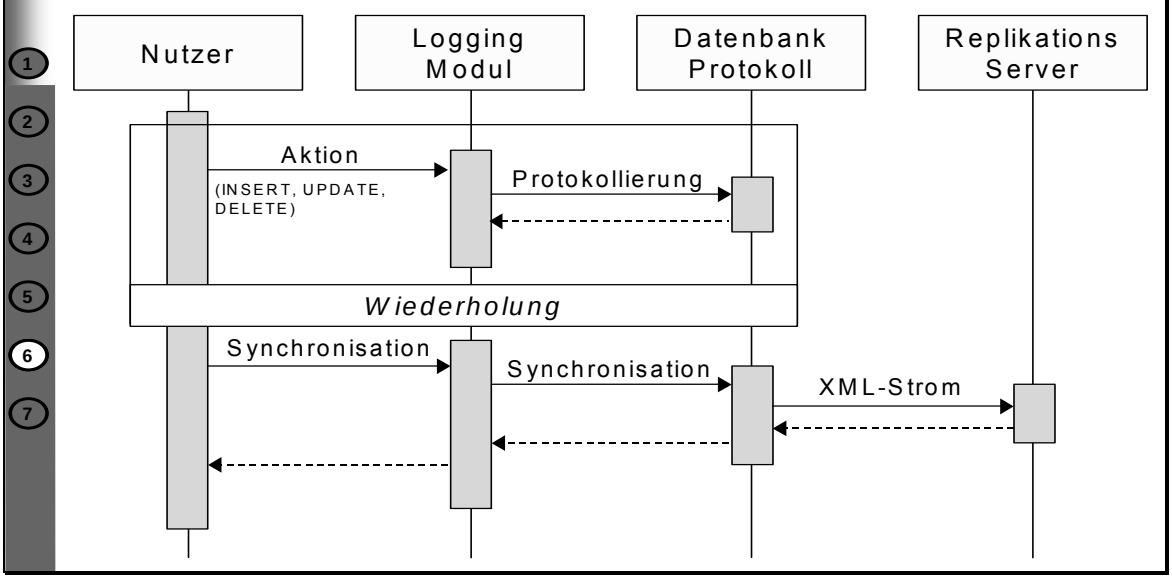
Anforderungen an die Kommunikationsschicht

- Unterschiedliche relationale Datenbanksysteme
  - Oracle, MSDE, Microsoft SQL Server
  - Heterogene Vorgaben durch beteiligte Firmen
- Frei konfigurierbarer Datenabgleich
  - Temporäre Verbindung zwischen Client und Server
  - Frei definierbarer Synchronisationszeitpunkt
  - Unterschiedliche Netzanbindung (Mobil/Stationär)
- Sicherung vertraulicher Daten
  - Verschlüsselte Datenübertragung (SSL 128-bit)
- Protokollierung aller Zugriffe
- Aktualisierung der verteilten DB-Schemata
- Datenaustausch im XML-Format

Architektur der Kommunikationsschicht

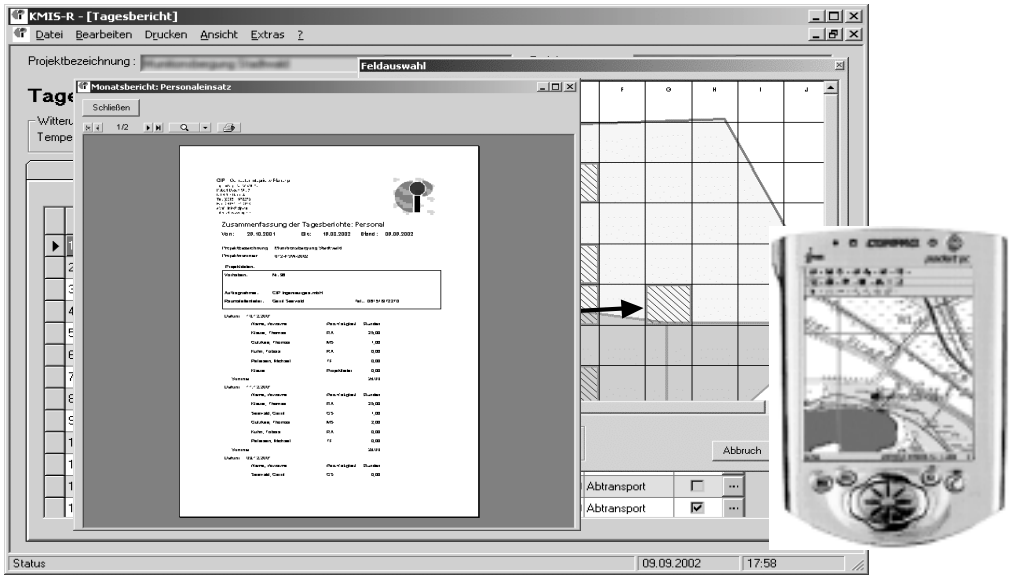


Ablauf der Replikation

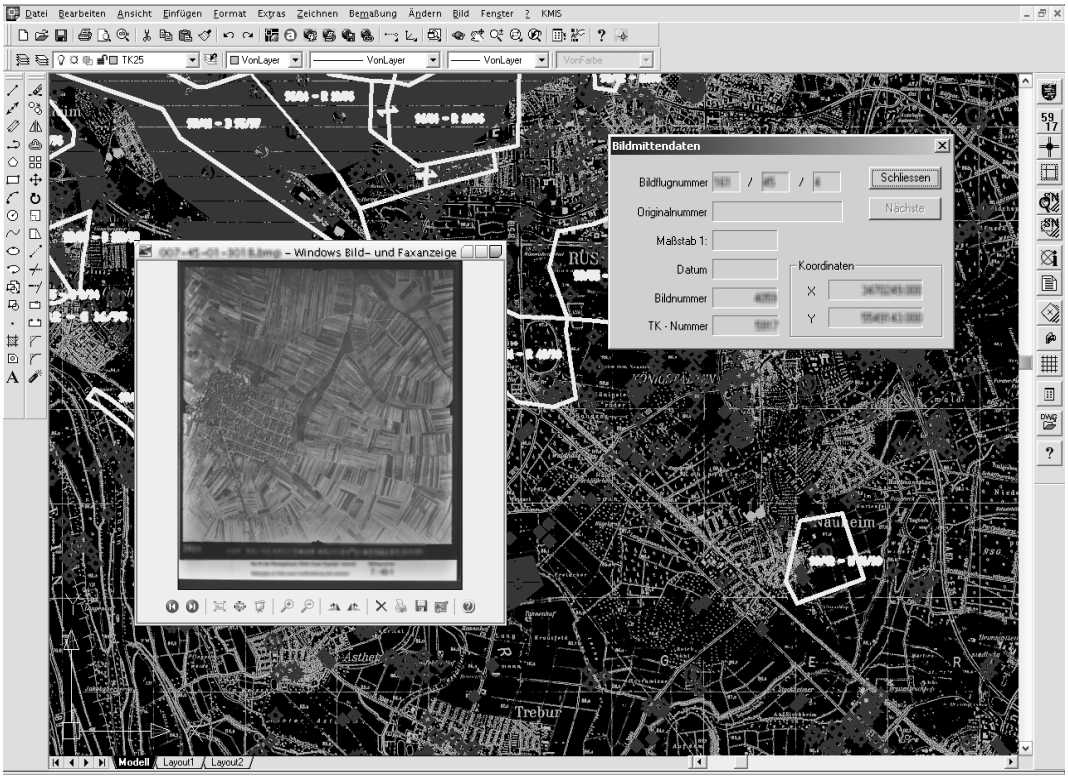


Stationäre Antragserfassung

Mobile Räumstellenerfassung



Visualisierung und Auswertung



Ausblick

1 2 3 4 5 6 7

- **Anbindung von Fachapplikationen**
  - **Stereoskopische Visualisierung**
  - **Luftbilddetailauswertung**
- **Mobile Computing**
  - **Replikationsmechanismen für mobile Endgeräte**
  - **Sprachsteuerung**
  - **GPS-Unterstützung**
- **Digitales Berichtswesen**
  - **Implementierung von rechtssicheren Methoden zur Kommunikation mit Hilfe Digitaler Signaturen**

Literatur

[1] R. Haas: Konzepte zur Untersuchung von Rüstungsaltlasten, Berlin 1992.

[2] U. Rüppel: Fachinformationssystem zur hybrid-semantischen Erfassung, Dokumentation und Auswertung von Kampfmitteln, Bauingenieur 11.2002.

[3] G. Seewald, T.Kuhn: Mobiles Informationssystem zur Verwaltung bodenbelastender Stoffe auf Basis georeferenzierter Fachinformationen.

[4] International Standards Organization (ISO): ISO/FDIS 19115

[5] McAuley, Charles: Programming AutoCAD 2000 Using ObjectARX, AutoDesk Press, New York, 2000.

[6] UMN Mapserver http://mapserver.gis.umn.edu/, http://mapserver.gis.umn.edu/mdp/

[7] U. Rüppel, I. Meissner, T. Gutzke, G. Seewald: Mobile Computing in der Grundwasserbewirtschaftung, Weimar 2003.

Entwicklung eines Web-Services für die DIOXIN-Datenbank des Bundes und der Länder

Gerlinde Knetsch, Marianne Rappolder[23]
Eric Weihs[24]

Kurzfassung

Daten über den Zustand der Umwelt werden in Fachdatenbanken bzw. Informationssystemen verwaltet. Für die Stoffgruppe der persistenten chlororganischen Stoffe (POPs) entwickelte das Umweltbundesamt gemeinsam mit den Bundesländern und dem ehemaligen Bundesinstitut für gesundheitlichen Verbraucherschutz die ***DIOXIN-Datenbank des Bundes und der Länder*** (Version 5.1), die seit 1998 als Client-Server-Anwendung im Routinebetrieb läuft. Die Datenbank enthält umfangreiche beschreibende Informationen (Metadaten) zu dem Messprogrammen, zur Probenentnahmemethodik, Analytik und zur räumlichen Einordnung der Probenentnahmestellen sowie zu den gemessenen Konzentrationen (Analysenergebnisse) in verschiedenen Umweltmedien wie z.B. Boden, Luft, Sedimente, Biota. Um den Zugriff auf diese Daten und Informationen der DIOXIN-Datenbank einem breiten Nutzerkreis zu ermöglichen, soll das System in eine web-basierende Umgebung implementiert werden. Es werden Lösungsmöglichkeiten zur Entwicklung eines Web-Services für verschiedene Nutzergruppen vorgestellt.

[23] Umweltbundesamt, Postfach 33 00 22, 14191 Berlin
E-Mail: gerlinde.Knetsch@uba.de
E-Mail: marianne.Rappolder@uba.de

[24] Bayerisches Staatsministerium für Landesentwicklung und Umweltfragen, Rosenkavalierplatz 2
81925 München E-Mail: erich.weihs@stmlu.bayern.de

1 Einführung

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 sollte mit dem Aufbau eines zentralen Datenbanksystems im UBA realisiert werden, an dem das ehemalige Bundesinstitut für gesundheitlichen Verbraucherschutz und Veterinärmedizin zu beteiligen war. 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 dar mit dem Anhang II 3. „Austausch von Daten zu polyhalogenierten Dibenzodioxinen und Dibenzofuranen sowie weitere chlororganische Stoffe“.

Die Dioxin-Datenbank des Bundes und der Länder läuft seit 1998 in einer Client-Server-Architektur im Routinebetrieb. Sie enthält Daten von über 10.000 Proben für die Umweltmedien 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 (Metadaten) in der Datenbank enthalten.

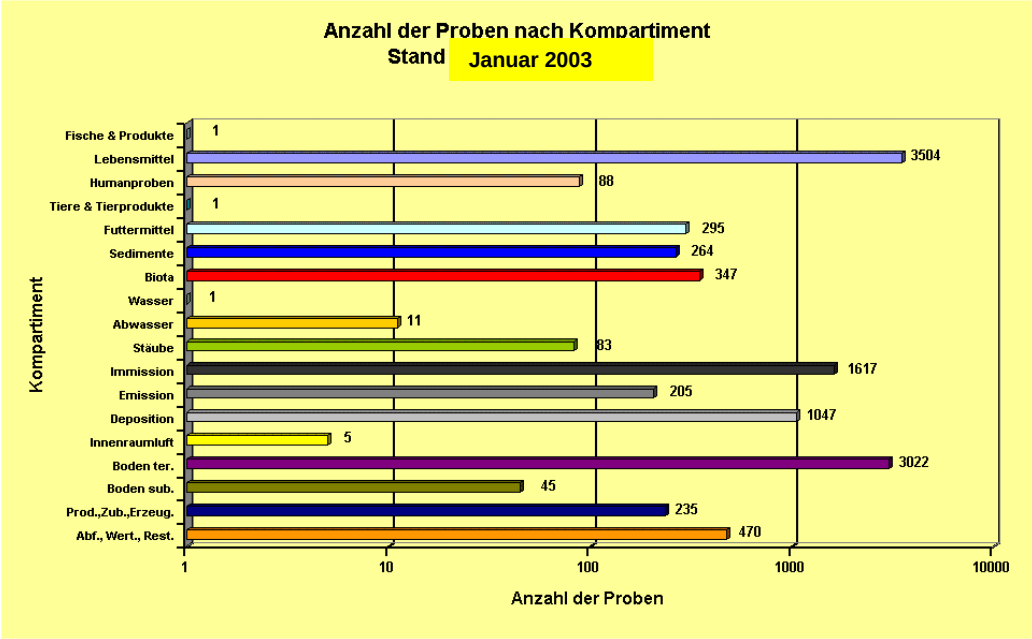


Abb. 1 **Übersicht zum Datenbestand in der Datenbank Dioxine**

2 Die bestehende Client-Server-Anwendung der DIOXIN-Datenbank

Die Dioxin-Datenbank (Version 5.1) ist als Client-Server-Anwendung im Umweltbundesamt implementiert. Der Datenserver läuft unter dem Betriebssystem UNIX mit dem Datenbanksystem Oracle Version 8.1.6.2. Die Client-Anwendung unter WINDOWS NT wurde unter Einsatz der Software ACCESS 97 entwickelt. Die Verbindung zwischen Client und Server erfolgt mit dem standardmäßig verfügbaren ODBC-Treiber von Oracle.

Die Applikation gewährleistet die Mehrbenutzerfähigkeit. Es können gleichzeitig die erfassten Daten in der Datenbank auf dem ORACLE-Server fortgeschrieben und Recherchen von anderen Benutzern durchgeführt werden. Über eine Benutzerverwaltung können jedem Anwender in Form einer Rollenzuweisung Zugriffsrechte zugewiesen werden, die lesende, schreibende und Administrierungsrechte zulassen.

2.1 Datenerfassung

Die Erfassung von Daten aus Länder- und Bundesprogrammen erfolgt über Erfassungsmasken, die für die entsprechenden Matrices wie z.B. Böden, Luft, Wasser, Sedimente typische Merkmale des Programms dokumentieren. Die Erhebungsbögen strukturieren somit umfangreiche Metadaten zu dem Messprogramm. Dies betrifft folgende Kriterien:

- **Allgemeine Angaben** (Datenliefernde Stelle, Anlass und Ziel der Untersuchung)
- **Standortdaten** (Hoch- und Rechtswert der Probenentnahmestelle, Gemeinde mit automatischer Zuordnung zum siedlungsstrukturellen Gebietstyp, raumnutzungsbezogene und juristische Besonderheiten, möglicher Emittenteneinfluss)
- **Probenentnahmedaten** (Probenentnahmeinrichtung, Zeitpunkt der Probenentnahme, Art und Menge der Proben, Beschreibung der Probenentnahmemethodik)
- **Labordaten** (Methodik der Probenvorbereitung und Analytik, Qualitätssicherung)

Die Erhebungsbögen umfassen auch die Dokumentation der analytischen Daten selbst, die sog. Einzelkongenere mit den entsprechenden Bestimmungsgrenzen. Das sind chemische Verbindungen wie z.B. das 2,3,7,8-TCDD - eine vollständige Liste der Kongenere wird in der Präsentation vorgelegt -, die Grundlage für die Berechnung von Toxizitätsäquivalenten sind. Alle dokumentierten Einzelkongenere der Dioxine und Furane sowie der Polychlorierten Biphenyle (PCB) werden unter Anwendung von mathematischen Algorithmen verrechnet und automatisch bei Auswertungen in Toxizitätsäquivalente umgerechnet. Es existieren drei Schnittstellen

- Generieren eines Vor-Ort-Erfassungsmoduls
- Import der Daten über das Erfassungsmodul
- Import von Lebensmitteldaten über individuelle Schnittstellen

die in die Anwendung integriert sind. Weitere extern gelieferte digitale Datenbeständen (vorwiegend als Excel-Dateien) werden über Importfunktionen auf der ACCESS-Ebene durchgeführt. Diese Vorgehensweise ersetzt zunehmend die Erfassung über die Erhebungsbögen.

2.2 Recherche -Funktionalitäten und Berichtswesen

Die Datenbank verfügt über Standardrecherchen getrennt für den Umwelt- und Humanbereich. Recherchekriterien lassen sich den Kategorien Raum, Zeit und Fachbezug (z.B. Kompartiment, Stoff) zuordnen, werden jedoch auch medienabhängig spezifiziert (Recherchekriterien für Boden z.B. nach Flächennutzung, Horizont). Darauf aufbauend können Berichte generiert werden, die z.B. für Datenanfragen genutzt werden. Hierbei sind standortbezogene, jahreszeiten- und kongenerspezifische Berichte möglich. Generierte Kreuztabellen können nach EXCEL und einem statistischen Softwareprogramm SPSS[25] exportiert werden, um weitere Auswertungen und grafische Darstellungen z. B. für die Berichtserstellung vornehmen zu können.

Die Recherchefunktionalitäten genügen bisher nur den grundsätzlichen Anforderungen an das Explorieren der Datenbestände der Dioxin-Datenbank und sollen mit der Weiterentwicklung zu einem Webservice ausgebaut werden.

[25] **Statistik**-Software der Firma SAS (Superior Performing Software Systems)

3 Lösungsprinzipien für einen Webservice

Um den Zugriff auf Daten und Informationen der DIOXIN-Datenbank einem breiten Nutzerkreis zu ermöglichen, hat das System den Anforderungen eines Webservice unter Berücksichtigung der Rahmenbedingungen aus der BundOnline2005 zu erfüllen

- Zugang zu den Daten über das Internet nach dem W3C[26]-Standard (*XML-Schema*)
- Nach Vorgabe des Styleguide des Umweltbundesamtes sollen **verschiedenen Nutzergruppen** die Inhalte der Dioxin-Datenbank entsprechend ihrer Benutzerprofile für Präsentation und Erfassung zugänglich gemacht werden.
- Schnittstelle zur Einbindung in das **German Environment Information Network (GEIN)** und
- Server-Schnittstellen (SOAP, XML) zum virtuellen **Umweltdatenkatalog (vUDK)[27],** um die Informationen nach Raum und Zeit verfügbar-zu machen.

3.1 Nutzergruppen

Im Konsens mit der AG DIOXINE soll für folgende Nutzergruppen der Zugriff auf allgemeine Informationen und Daten der Dioxin-Datenbank realisiert werden:

A) Öffentlichkeit/Internetangebot: Angebot zu aggregierten Daten (z.B. Jahremittelwerte der Luftbelastung) und Informationen, die dynamisch aus dem Datenbestand abgefragt werden können und in Form von Tabellen, Graphiken, Diagramme, mit erläuternden Texten dargestellt werden. Diese Daten werden durch die Länder und das UBA zertifiziert, um den Qualitätsanforderungen zu entsprechen.

[26] World Wide Web Consortium (W3C)

[27] http://www.umweltdatenkatalog.de/wwwudk/V-UDKServlet

Darüber hinaus sollen durch statische Web-Seiten:

- Allgemeine Hinweise zur DIOXIN-DB des Bundes und der Länder (z.B. Anlass, Auftrag, Zielstellung, Informationsfluss)
- Basiswissen zur Dioxin-Problematik (Einführung, Vorkommen, Entstehung, Toxizität, Bewertung),
- Ergebnisberichte von Forschungsprojekten, Veröffentlichungen des UBA und des ehem. Bundesinstituts für gesundheitlichen Verbraucherschutz und Veterinärmedizin (BgVV), Berichte der B/L-AG Dioxine,
- Glossar mit fachlichen Erläuterungen und Erklärung von Abkürzungen angeboten werden.

Mit diesem Informationsangebot für die Öffentlichkeit wird das UBA den Anforderungen aus dem Umwelti8nformationsgesetz und der Aarhus-Konvention (freier Zugang zu Umweltinformationen) gerecht.

B) Fachnutzer

Die Zugriffsberechtigung auf die Fachdaten wird durch Benutzerprofile geregelt. Dies können z.B. folgende Profile sein: Fachnutzer im UBA, in den Ländern, Forschungsnehmer. Die Rechte sind z.B. lesender und/oder schreibender Zugriff auf bestimmte Datenmerkmale nach Herkunft (z.B. länderspezifische fachliche Verantwortung). Den Fachnutzern wird entsprechend seines Benutzerprofils der Zugriff auf die Daten durch das UBA nur nach der Zustimmung der zuständigen Landesbehörden mittels Passwort erteilt.

Die von den Fachnutzern einzusehenden Daten sind im Sinne des Datenschutzgesetzes mit den Ländern abgestimmt und werden nach festgelegten Regeln – gegebenenfalls anonymisiert – dargestellt. Dieser Prozess der Anonymisierung erfolgt mit dem-downloaden der Daten über personalisierte Stylesheets, mit denen die XML in das gewünschte Exportformat (XML, pdf, xls etc.) umgewandelt werden.

Das Web-Angebot soll folgende Kern-Funktionen unter Berücksichtigung der Benutzerprofile erfüllen:

- Suchmasken, orientiert an den Recherchemasken der Dioxin-Datenbank – oder entsprechende Alternativen
- Export der Rechercheergebnisse
- 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
- Statistische Auswertungen, Graphiken und Diagramme sollte definiert werden.

Dabei muss – ebenso wie bei den Datenlieferanten - der Datentransfer gesichert erfolgen und der Export ohne zusätzliche Installationen des Nutzers möglich sein. Die übrigen Daten der Datenbank dürfen nicht einsehbar sein.

C) Datenlieferanten, Administratoren

Die datenliefernden Stellen sollen mittels eines Passworts die Einsichtnahme in die von ihnen gelieferten Daten erhalten, z.B. für die Qualitätssicherung. Es soll aus folgenden Komponenten aufgebaut werden:

- Kern-Funktionalitäten wie bei den Fachnutzern unter B)
- Zugriff / Export auf die eigenen gelieferten Daten zur Präsentation in einer eigenen Web-Umwebung
- Zugriff auf die von der Datenbank berechneten Daten (Toxizitätsäquivalente)
- Optional: Zugriff auf statistische Daten und Diagramme, wie sie entsprechend ihres gelieferten Datenumfangs aufbereitet und als Diagramme dargestellt werden können.
- Qualitätssicherung (Update, Freigabemechanismen)
- Dokumentation der Historie der Erfassung und der Fortschreibung
- Administration, Rechtevergabe

Der Datentransfer muss gesichert erfolgen und der Export sowie das Herunterladen ohne zusätzliche Installationen einer Anwendung in den allgemein verfügbaren Formaten (z.B. ASCII, Excel, PDF) möglich sein. Die Zugriffserlaubnis erfolgt selektiv über das Benutzerprofil und muss protokolliert werden.

3.2 Schnittstelle GEIN/UDK

Das Konzept des **German Environmental Information Network (GEIN)**[28] sieht vor, die für die Öffentlichkeit bestimmten Informationen bereitzustellen. Hierbei sind die unter Punkt 3.1 A) genannten Informationen vorgesehen. Um im GEIN verzeichnet zu werden, genügt es zunächst die Anmeldung beim GEIN-Broker mit Angabe der URL des Informationsangebotes sowie der Angabe einer Kontaktadresse. Bei einer spontanen Suche wird dieses Informationsangebot berücksichtigt. Zusätzlich kann der Eintrag des Informationsangebotes als sog. Portalseite beantragt werden. Dazu ist das Thema zu dem Themenkatalog zuzuordnen. Dies kann u.a. sein: Chemikalien, Boden, Luft, Wasser, Umwelt. Bei der Aufnahme der Portalseite werden diese automatisch verschlagwortet. Grundlage für die Verschlagwortung ist der UBA-Thesaurus. Um das Verhalten dieser Suchmaschine zu beeinflussen, können gezielt HTML-Meta-Informationen gesetzt werden.

Für die thematische Suche nach Raum und Zeit wird die standardisierte Schnittstelle -**G2K-Profil** des GEIN-Servers benutzt. Bereitgestellt werden entsprechend aufbereitete Statistiken. Um die Anonymität sicher zu stellen, wird eine Mindestfallzahl vorzugeben sein.

Die universelle „Daten-Sprache“ von GEIN in der auch der Zugriff über die G2K-Schnittstelle erfolgt ist XML. Obwohl die Umweltinformation selbst in der Regel in HTML vorliegt, erfolgt der Austausch zwischen dem GEIN-Server und den Informationsanbietern in XML mit einem dafür vereinbarten "Profil" (namespace). Alle Nutzdaten des GEIN-Servers (Index, Thesauri, Kalender, Portalseitenverzeichnis usw.) werden in XML gespeichert. Dies funktioniert mit der Datenbank Tamino der Software AG.

[28] http://www.gein.de/index_de.html

Die Einbindung des Dioxin-Datenbank für die Öffentlichkeit in den virtuellen **Umweltdatenkatalog des Bundes und der Länder (vUDK)**[29] findet über die Beschreibung des Objekte in die Objektklasse „Datensammlung/Datenbank“[30] statt. Von dort aus kann über die Rubrik „Verweise im WWW“ direkt auf das Informationsangebot für die Öffentlichkeit gelinkt werden. Der Nutzer erhält die Einstiegsseite zur Dioxin-Datenbank/Öffentlichkeit.

Die (eingeschränkte) selektive Suche nach Raum und Zeit erfolgt ähnlich der Vorgaben zur G2k-Schnittstelle des GEIN über eine SOAP-Schnittstelle.

3.3 Schnittstelle BundOnline-Initiative 2005[31]

Ziel ist der Initiative ist, bis zum Jahr 2005 alle internetfähigen Dienstleistungen der Bundesverwaltung online verfügbar zu machen. Künftig sollen Verwaltungsleistungen sowie andere Dienstleistungstypen wie **Informationen bereitstellen / Kommunikation und Beratung** über das Internet angeboten werden. Das betrifft Bürgerinnen und Bürger ebenso wie die Wirtschaft und die Behörden des Bundes, der Länder und der Kommunen. Informationen können somit online schneller, transparenter und kostengünstiger bereitgestellt werden.

Zu den zentral geplanten und zentral bereitgestellte IT-Basiskomponenten zählen u.a. auch das **Content Management System** (CMS). Content Management Systeme sind i.d.R Datenbank gestützte IT-Verfahren, die das Erstellen, Pflegen, Strukturieren und Auffinden von Inhalten für Web-Seiten ermöglichen. Die Dioxin-Datenbank benötigt diese Basiskomponente für die Bereitstellung der öffentlich zugänglichen Informationen. Durch die als Webservice definierte Funktionalität wird der Initiative entsprochen

[29] http://www.umweltdatenkatalog.de/wwwudk/V-UDKServlet
[30] http://www.umweltdatenkatalog.de/wwwudk/jsp/DataDetailBottom.jsp?uURL=5
[31] http: //www.bund.de/bundonline2005

Ein Prinzip von CMS ist die Trennung von Layout und Inhalt, d.h. die Speicherung der Inhalte erfolgt unabhängig von Ausgabeformat und Design. Vorlagen (Templates, Stylesheds, Active Server Pages [ASP], allgemein verwendbare Formatvorlagen) definieren die Anordnung und Darstellung der Inhaltselemente. Durch das CMS werden bei der Generierung einer Web-Seite die festgelegten Vorlagen automatisch mit den Inhalten gefüllt. Dadurch ist die konsequente Einhaltung des Layouts garantiert. Änderungen am Layout erfordern die Anpassung weniger Vorlagen, die Inhalte bleiben unverändert.
CMS entkoppeln damit die Inhaltserstellung und Programmierung der Web-Seiten. Diese Entkoppelung ermöglicht einer Vielzahl von Mitarbeitern die eigenständige Inhaltserstellung - auch ohne umfangreiche Programmierkenntnisse. Die Bearbeitung der Inhalte über eine Browsermaske ermöglicht zudem einen komplett unabhängigen Standort. Durch definierte Arbeitsabläufe und Freigabezyklen ist auch bei vielen Beteiligten die Qualität der Web-Site gewährleistet.

3.4 Lösungsprinzipien

Folgende Lösungsprinzipien zur Entwicklung eines Web-Services bieten sich an:

- Dynamische Web-Seiten unter Einsatz der serverseitigen Software PHP[32] und ASP[33] unter Beibehaltung der derzeitigen Grundfunktionalität der Dioxin-Datenbank.
- WEB-Service mit Kommunikation per XML und http als CMS-System

A) Dynamische Web-Seiten unter Einsatz der serverseitigen Software PHP und ASP
Der Zugriff auf die Daten der Dioxin-Datenbank soll über den zentralen WWW-Server des UBA erfolgen. Dazu existieren auf dem WWW-Server Java-Komponenten, die eine Anbindung der Oracle-Datenbank über JDBC ermöglichen. Der verwendete JDBC-Treiber ist ein von Oracle zur Verfügung gestellter Treiber Typ 4, der es ermöglicht durch die Firewall auf die Datenbank zuzugreifen.

[32] Personal Home Page Tools, Open Source Scriptsprache zur Erstellung dynamischer Webseiten
[33] Active Server Pages, Technologie von Microsoft zur serverseitigen Erstellung dynamische HTML-Seiten

Soweit für den Export und Schnittstellen XML-Daten benötigt werden, müssen diese aus den relationalen Daten der Oracle-Datenbank erstellt werden. Die Schnittstelle BundOnline-Initiative 2005 kann nur mit Einschränkungen erfüllt werden, da Daten und Vorlagen in getrennten Systemen gehalten werden.

Vorteil dieser Lösung ist die Nutzung länger bekannter und erprobter Technologien.

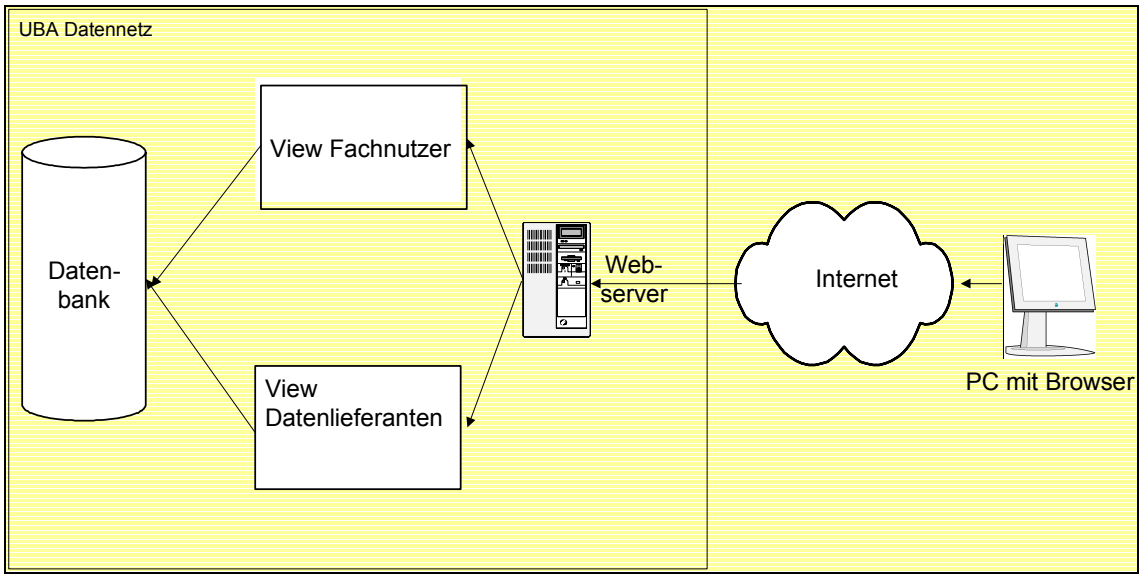


Abbildung 1: **Prinzipschema für die Lösungsvariante „Dynamische Web-Seiten“, Quelle: F+E-Vorhaben des UBA, FKZ 201 62 215**

Der WWW-Server läuft derzeit unter Windows 2000 und hat folgende Softwarekomponenten:

- Tomcat 4.0.1
- Java JDK 1.2.2 (Java Development Kit)1.3.1
- Java JSDK 2.0 (Java Servlet Development Kit).

Die Softwarekomponente **NetCharts™** von Visual Mining kann für die grafische Umsetzung der tabellarischen Rechercheergebnisse genutzt werden. Diese Software wurde schon erfolgreich in die Applikation des Webservice für das Informationssystem Umweltprobenbank eingesetzt (siehe http://193.174.169.38:8008/isupbwww/servlet/upb)

B) XML-basiertes Webservice

Funktionalitäten der Web-Services sind die Nutzung des W3C-Standards unter weitgehender Nutzung von open-source Software zur Schaffung „selbstständiger“ Komponenten, die gewünschten Funktionalitäten über das Internet bereit stellen. Die Grundfunktionalität wird durch ein CMS-System bereit gestellt. Die Funktionalitäten sind unter Punkt 3.3 *Schnittstelle BundOnline-Initiative 2005* beschrieben:

- Profilspezifische Recherchemöglichkeiten in den Dioxin-Daten
- Qualitätssicherung
- Import/Export Schnittstellen auf Basis von XML

Kennzeichnend für diese Lösung ist der Wegfall der Middleware der plattformabhängigen JDBC Treiber gegen eine reine XML-Lösung. Basiskomponeten des Web-Service sind Apache, Tomcat, Xerces o.ä. (Umwandlung XML in html, pdf etc über Stylesheeds), Open Source Software (Java-Beans) unter Beibehaltung des W3C Standards.

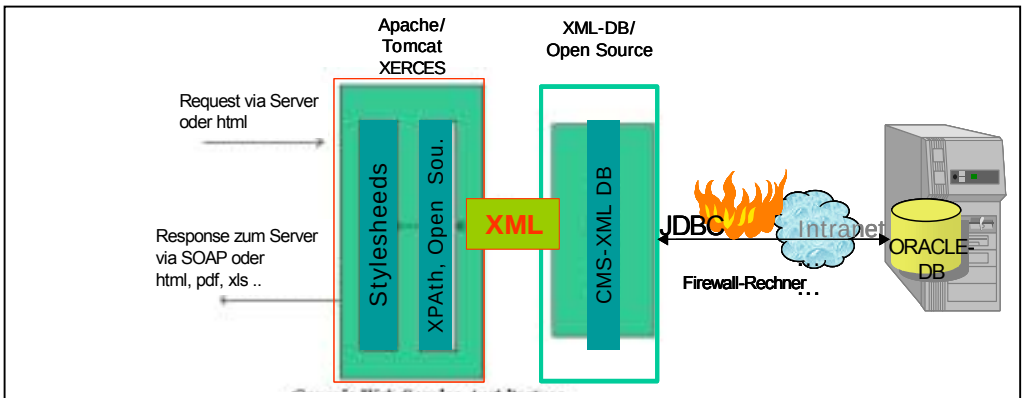


Abb.: 2 XML Architektur der Dioxindatenbank als CMS-Webservice

4 Ausblick

Die Entwicklung des Web-Services DIOXIN-Datenbank soll in enger Abstimmung mit den Nutzern der Daten und Informationen (AG DIOXINE) im Rahmen der Kooperation Umweltinformationssysteme erfolgen. Derzeit treiben der BUND/UBA und Bayern das Projekt voran. Weitere Partner sind willkommen. Die Vorgehensweise bei der Umsetzung findet auf der Grundlage des iterativen Prototypings statt. Mit der Entwicklung wird im Frühsommer 2003 begonnen.

Referenzen:

BMU (1993): 2. Bericht der Bund/Länder-Arbeitsgruppe DIOXINE; Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, Bonn 1993

BMU/UBA (2002): 3. und 4. Bericht der Bund/Länder-Arbeitsgruppe DIOXINE/Dioxin- Referenzmessprogramm, ISBN3-00-009326-5, Umweltbundesamt 2002

Knetsch, G. (1995): Dioxin-Datenbank des Bundes - Baustein eines Umweltinformationssystems in: Informatik für den Umweltschutz, Marburg 1995

Knetsch, G. (1999): Anwendung von Methoden und Verfahren zur Aus- und Bewertung von Umweltdaten in zeitlicher und räumlicher Sicht am Beispiel von Dioxindaten, Proceedings des TZI - Center for Computing Technologies, Bremen, 1999

Knetsch, G. (2000): The German Dioxin Database – Application in Trend Monitoring of Polychlorinated Dibenzo-p-Dioxins and Furans in the Environment, Organohalogen Compounds 46, 2000

Stoller et al (2001): „Entwicklung eines Konzeptes für den Zugang, Austausch und Veröffentlichung von Dioxin-Daten des Bundes und der Länder auf der Basis der Dioxin-Datenbank des Bundes und der Länder“, FKZ 201 62 215

Integrierte Meß- und Bilddatenhaltung für Umweltdatenbanken

Vortrag mit Live-Demonstration -

Peter Baumann
Rasdaman GmbH, Bunsenstr. 4, D-81735 München
E-Mail: baumann@rasdaman.com

Kurzfassung

Traditionell werden Luft/Satellitenbilder und andere Rasterdaten im Dateisystem gehalten, während die Ablage von Meta- und Vektordaten in Datenbanken heutzutage mehr oder weniger Standard ist. Erst jüngst wird Datenbankunterstützung auch für große Rasterdaten in größerem Umfang diskutiert und gefordert.
Im EU-Projekt EuroClim findet erstmalig eine datenbankgestützte Integration unterschiedlich dimensionaler Rasterobjekte statt. Als Datenhaltungsplattform für die Rasterdaten dient rasdaman, welches die klassischen Datenbankvorteile Anfragesprache, serverseitige Optimierung, Transaktions- und Speicherverwaltung für Rasterdaten aller Art und Dimension bereitstellt. Die Datenbank-Modellierung und Beispielsanfragen darauf werden in diesem Beitrag vorgestellt.

1 Motivation

Traditionell werden Luft/Satellitenbilder und andere Rasterdaten im Dateisystem gehalten, während die Ablage von Meta- und Vektordaten in Datenbanken heutzutage mehr oder weniger Standard ist.
Erst jüngst wird Datenbankunterstützung auch für große Rasterdaten in größerem Umfang diskutiert und gefordert. Es steht zu erwarten, dass sich in der nächsten Zeit Rasterdatenbanken auch kommerziell durchsetzen

So sieht Oracle nach eigener Aussage auf der GITA Annual Conference 2003 in der 2D- und 3D-Rasterunterstützung eine der beiden Kerninnovationen im Bereich Geodatenbanken in den nächsten Jahren. Derzeit verfügbar am Markt ist allerdings hauptsächlich dateibasierte Rasterdatenhaltung. Entsprechende datenbankbasierte Produkte von (z.B.) ESRI und Oracle sind in Vorbereitung.
Bereits seit mehreren Jahren verfügbar und auch im praktischen Einsatz befindlich ist dagegen der Rasterserver rasdaman. Rasdaman ist Client/Server-Middleware mit einer Anfragesprache, welche ISO-SQL92 um deklarative Raster-Operatoren erweitert. Die parallele Anfrageauswertung benutzt transparente Mechanismen zur Verbesserung der Performance und Speicherausnutzung, z. B. regelbasierte Optimierung, automatische Partitionierung (Kachelung), Geo-Indexierung, Kompression. Die Ablage erfolgt in einer relationalen Datenbank, etwa Oracle, IBM DB2 oder IBM Informix. Anfragesprache, Optimierung und Speicherverwaltung basieren auf der rasdaman Array-Algebra als mathematischem Fundament.
In mehreren EU-Projekten wurde rasdaman eingesetzt bzw. weiterentwickelt, u.a. RasDaMan (www.forwiss.tu-muenchen.de/~rasdaman), ESTEDI (www.estedi.org), FOREMMS (foremms.nr.no) EuroClim (euroclim.nr.no). Derzeit existieren rasdaman-Installationen in 12 Nationen (einschl. USA und GUS) eingesetzt, hauptsächlich für Geo-Rasterserver 1. Der holistische Ansatz, der bei der Entwicklung von rasdaman verfolgt wurde, macht das System außerdem zu einer geeigneten Basis für die Untersuchung aller Aspekte datenbankgestützter Rasterdatenhaltung. Im vorliegenden Beitrag wird der Einsatz von rasdaman in den vorgenanntenen Projekten EuroClim und FOREMMS präsentiert.
Der Schwerpunkt liegt auf konzeptueller Ebene, da sich EuroClim derzeit in der Spezifikationsphase befindet. Insbesondere wird gezeigt, wie Mess- und Bilddaten übergreifend abgefragt werden können.1 in Deutschland u.a. als offizieller Rasterserver mehrerer Landesvermessungsämter, etwa Thüringen und Bayern
Der Rest dieses Beitrags gliedert sich wie folgt. In Kapitel 2 wird ein Überblick über konzeptuelles Modell und Architektur von rasdaman gegeben. In Kapitel 3 werden Meß-, Simulations- und Bilddaten in 1D-, 2D-, 3D- und 4D diskutiert. In Kapitel 4 wird die Live-Demo des Vortrags skizziert. Kapitel 5 gibt eine Zusammenfassung.

2 Das rasdaman-System

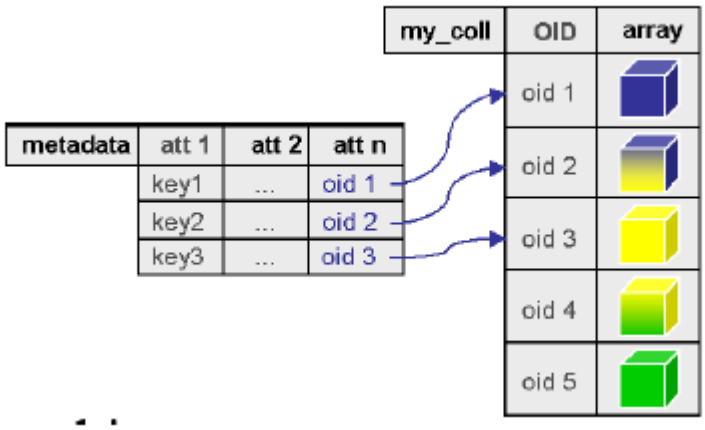
2.1 Konzeptuelles Modell

Das konzeptuelle Modell von rasdaman basiert auf n-dimensionalen Arrays (im Programmiersprachensinn, also Rasterdaten), welche von beliebiger Dimension und Ausdehnung sein können [Bau94]. Das dynamische Typsystem [Wid00] unterstützt alle C/C++-Typen als Raster-Zellentyp („Pixeltyp“), einschließlich geschachtelter structs und ausschließlich Zeigertypen (welche in der Datenbank keinen Sinn machen). Die Ober- und Untergrenze jeder Dimension kann zur Typdefinitionszeit festgelegt oder variabel gelassen werden. Die Typdefinition geschieht durch die Definitionssprache rasdl, welche auf der ODMG ODL (Object Definition Language) basiert [Cat96]. Ein Rastertyp wird definiert mittels eines Templates

marray<b,d>

welches mit einem Zellentyp b und der Raster-Ausdehnung *(spatial domai*n) d, instantiiert wird. Die Ausdehnung wird angegeben durch die Unter- und Obergrenze pro Dimension, wobei jeder Wert durch “*” ersetzt werden kann, um dynamische Feldgrenzen fest zu legen. Beispielsweise kann ein blattschnittfreies Rasterbild unbegrenzter Ausdehnung definiert werden durch

typedef marray
< struct{ char red, green, blue; },
*[*:*, *:*] >*
ColorOrthoImg;



;

Abb. 1: Verweise von relationalen Metadaten auf Rasterobjekte in rasdaman

Durch Angabe von Punktkoordinaten, Begrenzungsboxen oder –polygonen lassen sich Regionen auch innerhalb eines Rasterobjekts referenzieren. Selbstverständlich ist aber auch inhaltsbezogene Referenzierung möglich, in dem mit dem Geoobjekt eine Anfrage abgelegt wird, die beim Zugriff entsprechende Informationsextraktion bewirkt. Beispielsweise könnte dies sein *„Alle Gebiete im DEM mit Höhen zwischen 0 und 100 m über N*N“.

2.2 Raster-Anfragen

Die rasdaman-Anfragesprache rasql bietet Rasterprimitive eingebettet in SQL92; wie üblich gibt eine select-Anfrage eine homogene Resultatmenge zurück. Raster-wertige Ausdrücke können sowohl in der select- als auch in der where-Klausel verwendet werden. *Trimming* liefert einen rechteckigen Ausschnitt.
Beispiel: “Ein Ausschnitt mit Eckpunkten (1000,1000) und (2000,2000) aus den DOPs”:

select OrthoCollection[1000:2000,1000:2000]
from OrthoCollection

Für jede Operation auf dem Zelltyp wird vom Server automatisch eine entsprechende *induzierte Operation* bereitgestellt, welche diese Operation simultan auf alle Zellen eines Objekts anwendet. Sowohl unäre (z.B. Komponentenzugriff) als auch binäre Operationen (z.B. Jmaskeirung und Überlagerung) lassen sich induzieren.

Beispiel: "Topographical map bit layer 3 oveölaid with the (grayscale) ortho image":

*select OrthoCollection overlay bit(Map, 3) * 255c*
from Map, OrthoCollection

Im allgemeinen können Rasterausdrücke im select-Teil einer Anfrage auftreten und, wenn der Ergebnistyp der äußersten Operation vom Typ Bool ist, auch im where-Teil. Die Mächtigkeit von rasql umfasst viele Signalverarbeitungs-, Bildverarbeitungs- und Statistikoperationen, bis hin zu beispielsweise Fourier-Transformationen; sie endet bei rekursiven Operationen, so dass jede Anfrage terminiert, mithin rasql sicher in der Auswertung ist. Details der unterliegenden rasdaman Array-Algebra sind in [Bau99] beschrieben.

2.3 Anfrageverarbeitung

In rasdaman stehen etwa 150 Ersetzungsregeln für heuristische Optimierung zur Verfügung [Rit99]. Beispiele sind *"pull out disjunctions while aggregating cell values of a raster object using logical o*r" oder *"push down geometric operations to expression leave*s". Die letztere Regel erreicht, dass nur die minimale Datenmenge gelesen wird, welche für die Berechnung des Anfrageunterbaums erforderlich ist. Das Prinzip der *Common Subexpression Elimination* wurde um räumliche Überlappung erweitert, um Gebiete jeweils nur einmal zu laden. Die Kachel-basierte Abarbeitung des Anfragebaums bewirkt, dass pro Zeitpunkt nur die minimale erforderliche Kachelanzahl im Hauptspeicher gehalten werden muss.
Der rasdaman-Server unterstützt bereits Inter-Objekt-Parallelität, in dem eine Anfrage an einen Server im Pool zugewiesen wird. Derzeit erfolgt die Erweiterung um Intra-Objekt-Parallelität, so dass auch eine einzelne Anfrage auf mehrere CPUs bzw. Rechner verteilt werden kann [Hah02].

2.4 Speicherverwaltung

Raster-Objekte werden intern partitioniert in sog. Kacheln. Die Partitionierung ist nicht auf reguläre Gitter beschränkt. Eine Reihe von Kachelungsstrategien steht für den Datenbankprogrammierer zur Verfügung, um eine für bestimmte Anfragesets optimale Kachelung einzurichten. Nähere Beschreibungen und Benchmarks finden sich in [Fur99]. Jede Kachel wird in ein relationales BLOB gespeichert. Geoindexverfahren (u.a. R+-Baum) sorgen für die schnelle Selektion der für eine Anfrage benötigten Kacheln.

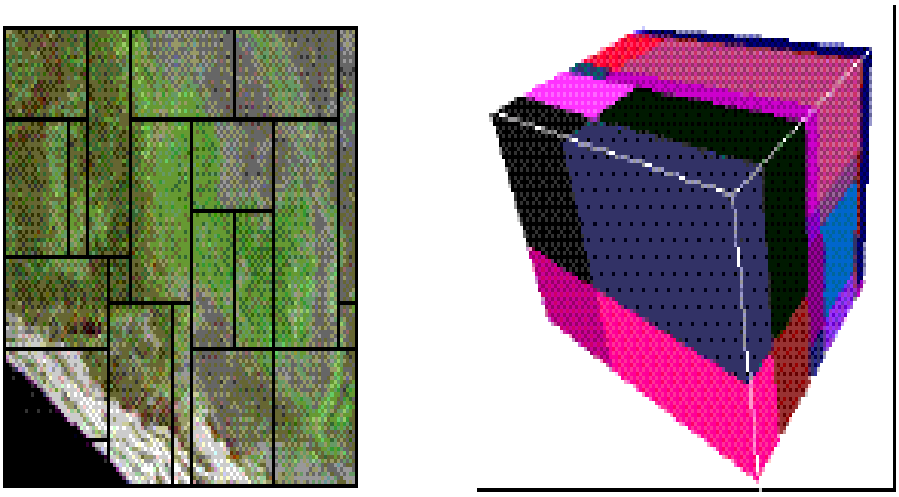


Abb. 2: irreguläre 2D- und 3D-Kachelungen

Heutige Plattensysteme erlauben bereits Terabyte-Archive; der Bedarf entwickelt sich jedoch rasant weiter, so dass für großvolumige Daten wie Satellitenarchive auch weiterhin Bandspeicher zum Einsatz kommen werden. Daher wurde für rasdaman eine Tertiärspeicher-Anbindung entwickelt [Rei02].
Die Ein- und Auslagerung erfolgt wahlweise automatisch oder manuell, wobei der Speicherverwalter räumliche Clusterung auf dem Tertiärspeicher berücksichtigt.

2.5 Architektur

Der rasdaman-Server ist als komplettes Client/Server-Datenbanksystem implementiert, mit eigenen Komponenten für Client/Server-Kommunikation, Anfrageanalyse und –optimierung, Katalogverwaltung etc. Die unterste Schicht nimmt eine Speicherabbildung auf relationale Blobs vor; Adapter existieren derzeit für Oracle, IBM DB2 und IBM Informix.
Ein verteilbarer Server-Manager teilt den Clients jeweils einen Serverprozess zu. Jeder Server kann unterschiedliche relationale Datenbanken ansprechen. Damit ist eine hohe Skalierbarkeit gegeben während Datenbanken bereits auf PC-Servern mit 256 MB Hauptspeicher Zugriffe unter einer Sekunde erlauben, lassen sich ebenso große Serverfarmen aufbauen (Abb. 3).

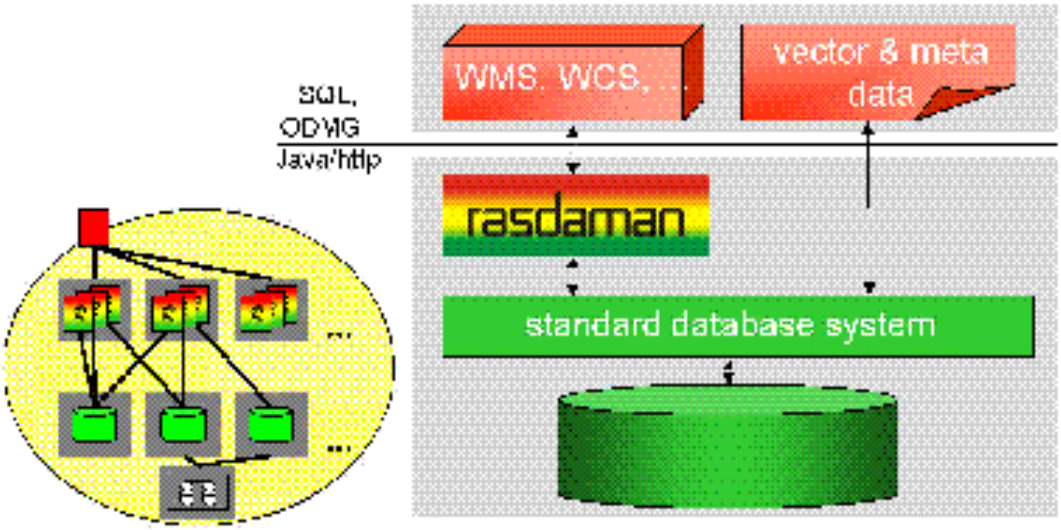


Abb. 3: rasdaman-Architektur (rechts: Blockschaltbild, links: hochparallele Serverfarm mit zentralem Einsprungspunkt)

3 EuroClim

Im Rahmen des EU-IST-Projekts EuroClim (euroclim.nr.no) wird ein europaweites kryosphärisches Klimainformationssystem, also mit Schwerpunkt auf Eis-, Schnee- und Gletscherbedeckung, entwickelt. Unter der Leitung des Norwegian Computing Institute in Oslo arbeiten hauptsächlich ,nordeuropäische Teilnehmer aus den Bereichen Klimaforschung, Erdbeobachtung und Geo-IT zusammen. Derzeit befindet sich das Projekt in der Spezifikationsphase, i.w. ist das Schema für Satellitendaten entwickelt. Trotzdem lassen sich bereits die prinzipiellen Schemata darstellen. Vorausgeschickt sei, dass die Datenaufbereitung (L1 bis L3) von eigener Software erledigt wird. Zum Zeitpunkt des Einfügens liegen die Daten radiometrisch korrigiert und georeferenziert vor. Das Einfügeformat bei Messreihen

3.1 Messreihen (1D)

Eindimensionale Messreihen lassen sich folgendermaßen modellieren. Der Zellentyp besteht aus einem Messwert und einem Indikator zur Validität; darüber wird ein eindimensionales Array mit offener Obergrenze definiert, entsprechend einer zeitlich unbegrenzten Laufzeit. Die rasdl-Definition lautet

typedef marray< struct { float value; char flag; }, [0:] > TimeSeries;*

Entsprechende Metadatenstrukturen sind im konventionellen Katalog zu definieren, um die zeitliche Auflösung, den Anfangszeitpunkt der Messungen, Sensorspezifikationen u.a. zu bestimmen.
Für value werden Gleitpunktzahlen entgegengenommen, die z.B. per Telemetrie automatisch in die Datenbank eingefügt werden können. Das entsprechende rasql-Update-Statement wäre etwa wie folgt.

update Sensor_X
set SensorX[sdom(Sensor_X)[0].hi+1].value assign $1,
SensorX[sdom(Sensor_X)[0].hi+1].flag assign $2

Hierbei sind die Positionsparameter $i durch die jeweiligen Eingangswerte zu substituieren. Die Standardfunktion sdom(X) liefert die aktuelle Begrenzungsbox zu Objekt X; hiervon wird Dimension 0 abgegriffen und davon wiederum die Obergrenze; durch Inkrementierung um 1 wird das Rasterobjekt erweitert, der übergebene Wert wird sodann der neuen Position zugewiesen.
Aus einem solcherart aufgebauten Sensorobjekt lassen sich Kenngrößen wie z.B. der bisher aufgetretene Maximalwert abfragen, wobei nur gültige Werte berücksichtigt werden (gekennzeichnet durch leeres Flag).

*select max_cells(Sensor_X.value * Sensor_X.flag = ‘ ‘)*
from SensorX

Der Einfachheit halber wurde hier angenommen, dass nur positive Messwerte auftreten können. Damit lassen sich mit der Multiplikation über den Booleschen flag-Wert alle ungültigen Werte auf 0 setzen.

3.2 Satellitenbilder (2D)

Die im Projekt ausgewählten MODIS-Daten sind hyperspektraler Natur, d.h. ein aufgenommener Punkt hat eine große Zahl (hier: mehrere Dutzend) von spektralen Kanälen. Unter Ausnutzung räumlicher Clusterung lässt sich dies in rasdl definieren als:

typedef marray
< struct{ char band1, …, bandn; },
[xmin:xmax,ymin:ymax] >
ModisImage;

mit entsprechend gewählten Begrenzungskoordinaten für Europa. Ziel ist, eine blattschnittfreie Datenbank zu erhalten, auf der beliebiges Zoom und Pan möglich ist. Auf einer solchen Datenbank lässt sich beispielsweise der normalisierte Vegetationsindex (NDVI) folgendermaßen gewinnen.

select abs(m.near_infrared – m.red) / (m.near_infrared + m.red)
from ModisCollection as m

Dabei wird angenommen, dass die Kanalbezeichnungen für nahes Infrarot bzw. Rot gegeben seien durch near_infrared und red.

3.3 Satellitenbild-Zeitreihen (3D)

Nach dem Schema der oben definierten Modis-Kollektion lässt sich mit einer kleinen Erweiterung ein blattschnittfreier 3D-Würfel aufbauen, dessen dritte Koordinate die Zeit bildet:

typedef marray
< struct{ char band1, …, bandn; },
[xmin:xmax,ymin:ymax,tmin:] >*
ModisTimeseries;

Auf diesem 3D-Würfel lassen sich Schnitte entlang aller Dimensionen legen, beispielsweise ein Kartenausschnitt zum Zeitpunkt T:

select m[x0:x1,y0:y1,T]
from ModisTimeseriesCollection as m

Aber auch ein Zugriff entlang der Zeitachse zu einer bestimmten Position x/y ist möglich; in diesem Fall ist das Resultat eindimensional; abgegriffen wird der gesamte verfügbare Zeitbereich:

select m[x,y,:*]*
from ModisTimeseriesCollection as m

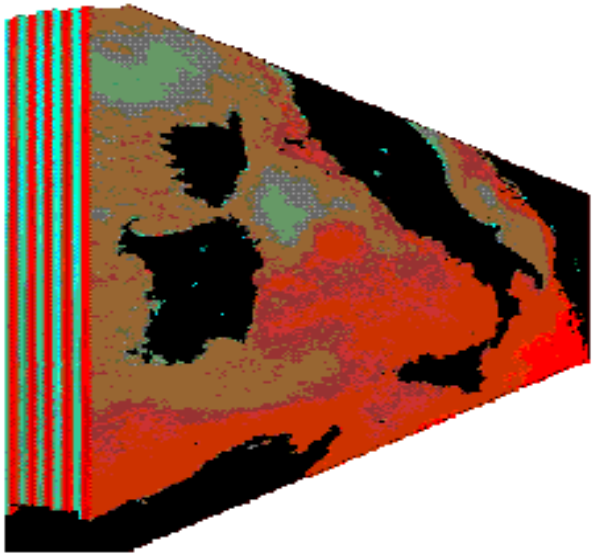


Abb. 4: blattschnittfreie Satellitenbild-Zeitreihe als 3D-Würfel
(mit frdl. Genehmigung DFD/DLR)

3.4 Dimensionsübergreifende Anfragen

Aufbauend auf den vorangehenden Definitionen lassen sich Rasterobjekte in Anfragen verknüpfen. Beispiel: *„den maximalen NDVI für alle Werte, an denen der Sensor an Position P=(x,y) eine Überschreitung des Schwellwerts T gemeldet hat*“. Dazu wird zuerst aus den (relational abgelegten) Metadaten über eine Standard-SQL-Anfrage der Sensor zu Position P selektiert:

select rasdaman_id
from Sensors
where posx = x and posy = y

Das Ergebnis ist der Identifikator des gewünschten Sensors in der rasdaman-Kollektion. Damit kann im zweiten Schritt die rasdaman-Anfrage formuliert werden:

select max_cells(
abs(m.near_infrared – m.red) / (m.near_infrared + m.red)

```
*
( s.flag = ‘ ‘ and s.value > T )
)
from ModisTimeseriesCollection as m, Sensors as s

where s.id = $1
```

4 Demonstration

Die Demonstration von rasdaman erfolgt über das graphisch-interaktive Frontend rView und über Web-Browser. Die C++-Applikation rView erlaubt das Absetzen von rasql-Anfragen an eine Datenbank und die Visualisierung von 1D- bis 4D-Anfrageergebnissen. Das Java-Servlet rasgeo generiert puren HTML-Zugang für Standard-Webbrowser, um auf 2D-Karten zu navigieren, sie zu überlagern und einzufärben. Die Beispielsdatensätze umfassen mehrere Dutzend GB große 2D-Georaster (Orthophotos, TK, DHM), 3D-Zeitreihen und eine 4D-Klimasimulation.

5 Zusammenfassung

Rasterdaten bilden einen der letzten „weißen Flecken“ auf der Landkarte der Internet-Geodienste. Die Integration der Rasterdatenhaltung in eine allgemeine, datenbankgestützte Verwaltung sämtlicher relevanter Daten bietet eine Reihe von Vorteilen: zuvorderst die Integration der Datenhaltung, welche die Konsistenz erhöht und die Verwaltung erleichtert. Weiterhin lassen sich sämtliche bereits vorhandenen Datenbank-Werkzeuge und Mechanismen nutzbringend einsetzen, u.a. Views, Trigger/Con-straints, Stored Procedures, Optimierung/Tuning, Transaktionskonzept, Zugriffsschutz, Replikation, Online-Backup, Report-Generatoren, 4GL-Werkzeuge. Schließlich gewinnt man mit in hohem Maße Flexibilität durch die Fähigkeit der Datenbanksysteme zur ad-hoc Definition neuer Anfragen und Tabellenstrukturen. Erlaubt ein solcher Rasterserver die gleichzeitige Haltung von Daten unterschiedlicher Dimension (von 1D-Zeitreihen bis 4D-Klimamodellen), dann ergibt sich ein besonders hoher Integrationsgrad.

Anfragen werden damit möglich, welche eine Verknüpfung der diversen Sensordaten erlauben (z.B. Ergebnisse von Messbojen korreliert mit Satellitenbildern Klimasimulationsdaten) und damit neuartige Dienstsqualitäten hervorbringen.

Rasdaman ist das erste System seiner Art, welches ein mathematisch fundiertes Modell zur Definition und Bearbeitung von multidimensionalen Rasterdaten anbietet und gleichzeitig mit seiner gesamten Architektur diese speziellen Strukturen unterstützt. Eine große Zahl untersuchter Anwendungen hat gezeigt, dass dieser Ansatz tragfähig ist – was sich nicht zuletzt darin zeigt, dass große GIS- und Datenbankhersteller mittlerweile an eigenen Varianten der datenbankgestützten Rasterverwaltung arbeiten. Aktuelle Themen der Weiterentwicklung von rasdaman konzentrieren sich auf den effizienten Zugriff mit besonders komplexen Anfragen bzw. auf extrem große Datenvolumina. FORWISS München (http://www.wibas.forwiss.tu-muenchen.de) arbeitet an Intra-Query-Parallelität zur Verteilung von Anfragen auf mehrere CPUs bzw. Rechner in Clustern [Hah02] sowie weiteren Anfrageoptimierungen sowie an der Einbindung von Bandrobotern mit besonderem Schwerpunkt auf räumliche Clusterung der Daten im Bandarchiv [Rei02]. Der Autor arbeitet aktiv im OGC an der Definition der Web Coverage Services (WCS) mit.

Danksagung

Für wertvolle Beiträge, Diskussionen und Implementierungen gilt der besondere Dank des Autors Karl Hahn, Bernd Reiner, Steve Pester, Asgeir Finnseth, Jason Baragry, Rune Solberg, Bernhard Buckl, Erhard Diedrich, Clemens Glock, Bruno Roder.

Referenzen

[Bau99] P. Baumann: *A Database Array Algebra for Spatio-Temporal Data and Beyond.* Next Gen-eration Information Technology and Systems (NGITS) '99, Zikhron Yaakov, Israel, 1999.

[Bau94] P. Baumann: *On the Management of Multidimensional Discrete Dat*a. VLDB Journal 4(3)1994, Special Issue on Spatial Database Systems, pp. 401-444.

[Cat96] R. Cattell: *The Object Database Standard: ODMG-9*3. Morgan Kaufmann Publishers, 1996.

[Fur99] P. Furtado, P. Baumann: *Storage of Multidimensional Arrays Based on Arbitrary Tiling.ICDE'9*9, Sidney - Australia 1999.

[Hah02] Karl Hahn, Bernd Reiner, Gabriele Höfling, Peter Baumann: *Parallel Query* Support for Multidimensional Data: Inter-object Parallelism. 13th International Conference on Database and Expert Systems Applications (DEXA), September 2-6, 2002, Aix enProvence, France.

[Rei02] Bernd Reiner, Karl Hahn, Gabriele Höfling, Peter Baumann: *Hierarchical Storage Supportand Management for Large-Scale Multidimensional Array Database Management System*s.3th International Conference on Database and Expert Systems Applications (DEXA),September 2-6, 2002, Aix en Provence, France.

[Rit99] R. Ritsch: Optimization and Evaluation of Array Queries in Database Management Systems. PhD Thesis, Technische Universität München, 1999.

[Wid00] N. Widmann: *Efficient Operation Execution on Multidimensional Array Data.* PhD Thesis, Technische Universität München, 2000.

Wie bewältigt man Stationaritätsannahmen in der Geostatistik?

Alexander Brenning[34] & K. Gerald van den Boogaart[35]

Zusammenfassung

In der Geostatistik ist die Stationarität des Zufallsfeldes eine zentrale Annahme. Die räumliche Variabilität vieler Phänomene in unserer Umwelt hängt stark von den Verhältnissen in einer lokalen Umgebung ab, die meist aber instationär sind. Um damit umgehen zu können, wird das Konzept der Stationarität des Zufallsfeldes ersetzt durch eine Stationarität des Einflusses der lokalen Umweltverhältnisse, wie sie in einem GIS gespeichert sind, auf das lokale Kovariogramm. Es wird eine Konstruktionsmethode benutzt, die auf sinnvolle Art räumliche Informationen aus dem GIS in Kovariogramm-Modelle einbinden kann, etwa im Untersuchungsgebiet variierende Umweltverhältnisse, sich räumlich verändernde Anisotropie im Gebirgsrelief oder geologische Störungen, welche die Kontinuität unterbrechen.

1 Einführung

Die in Geo-Informationssystemen vorhandene Menge räumlicher Daten hat in den vergangenen Jahrzehnten nicht zuletzt durch die Nutzung moderner Fernerkundungsmethoden sehr schnell zugenommen. Diese Daten eignen sich häufig für die Modellierung lokaler Anisotropien, die beispielsweise tektonischen Strukturen oder dem Relief folgen.

[34] Humboldt-Universität zu Berlin, Geographisches Institut, Unter den Linden 6, 10099 Berlin, E-Mail: alexander.brenning@rz.hu-berlin.de
[35] TUFreiberg, Institut für Geologie, B.-v.-Cotta-Str.2, 09596 Freiberg, E-Mail: boogaart@iname.com

Infolgedessen entsteht ein Bedarf an instationären geostatistischen Methoden, die diese Informationen auf sinnvolle Weise berücksichtigen, so dass komplexe Umweltkonfigurationen modelliert werden können. Zu diesen Umweltbedingungen zählen etwa tektonische Verwerfungen, welche zu Diskontinuitäten führen, oder hydrographische Einzugsgebiete, die hierarchisch verschachtelte Anisotropiestrukturen nach sich ziehen. Methoden der Geostatistik ohne Stationaritätsannahmen müssen mit Geo-Informationssystemen verknüpft werden, um entsprechende Umweltdaten nutzen zu können.

2 Grundlagen

Die Geostatistik stellt Methoden zur Analyse (skalarer) räumlicher Daten zur Verfügung. Diese Daten (Beobachtungen) werden als Realisierungen eines räumlichen stochastischen Prozesses oder Zufallsfeldes $(Z(s))_{s \in D}$, hier mit $D \subset \mathbb{R}^2$, angesehen, also als Menge von Zufallsvariablen auf einer räumlichen Domäne.

Räumliche Abhängigkeiten werden stochastisch durch Kovarianzstrukturen, sogenannte Kovariogramme, beschrieben:

$$C(s, t) = \mathrm{Cov}(Z(s), Z(t)) = E[(Z(s) - m(s)) \cdot (Z(t) - m(t))],$$

wobei $m(s) = E(Z(s))$ den Erwartungswert einer Zufallsvariable bezeichnet – vgl. z.B. GOOVAERTS (1997).

2.1 Stationarität

In Theorie und Praxis werden in aller Regel für Erwartungswerte und Kovariogramme Stationaritätsannahmen getroffen; dabei handelt es sich um räumliche Invarianzforderungen bezüglich Mittelwert und Kovariogramm. Die meist verwendete Stationarität 2. Ordnung fordert:

i) Der Erwartungswert m ist konstant auf *D*;

ii) das Kovariogramm *C* hängt nur von der Differenz *t – s* der Beobachtungsstellen ab (und nicht von ihrer tatsächlichen Lage).

(Zur technischen Vereinfachung gehen wir hier und im folgenden nicht auf Semivariogramme und intrinsische Stationarität ein; vgl. hierzu BRENNING 2001 oder GOOVAERTS 1997.)

2.2 Zwei Fälle von Instationarität

In folgenden für die Praxis wichtigen Fällen erfüllen Umweltvariablen die Stationaritätsbedingungen nicht – vgl. BRENNING & VAN DEN BOOGAART (2001), CRESSIE (1993):

i) Vorhandensein eines Trends: Es existiert eine an räumliche Kovariablen geknüpfte systematische „Tendenz“ – beispielsweise in einem Gebirge die lineare Abnahme der Temperatur mit zunehmender Höhe über dem Meer. In diesem Fall ist der Erwartungswert nicht auf *D* konstant, sondern eine Funktion von räumlichen Kovariablen.

ii) Lokale Anisotropien und Diskontinuitäten: Das Streichen geologischer Schichten kann dazu führen, dass Korrelationen quer zur Vorzugsrichtung bei gleicher Entfernung kleiner sind als parallel zu ihr. Ferner können Beobachtungen derselben Umweltvariable in Gebieten, die durch eine geologische Störung getrennt sind, unkorreliert sein.

iii) In diesen Fällen hängt das Kovariogramm nicht nur von der Distanz zwischen Beobachtungen ab, sondern von lokalen Veränderungen der Kovariablen.

Zwar sind in beiden genannten Fällen Stationaritätsannahmen verletzt. Dennoch entstehen die jeweiligen Instationaritäten auf der ganzen Domäne auf dieselbe Weise: Lediglich die Kenntnis bestimmter Kovariablen ist nötig, um diese Instationaritäten zu erklären. Daher wird im folgenden das Konzept der Stationarität des Zufallsfeldes ersetzt durch eine Stationarität des Einflusses der lokalen Umweltverhältnisse auf das lokale Kovariogramm (VAN DEN BOOGAART 1999).

Fortan wird lediglich der Fall instationärer Kovariogramme (bei Fehlen eines Trends) betrachtet. Geostatistische Schätzung mit linearem Trend wird von GOOVAERTS (1997), VAN DEN BOOGAART & BRENNING (2001) sowie BRENNING (2001) behandelt.

3 Zur Konstruktion instationärer Kovariogramme

Das Konzept der Stationarität des Einflusses der lokalen Umweltverhältnisse auf das lokale Kovariogramm wird nun in Form eines generischen Stationaritätsbegriffes eingeführt und anschließend mit Hilfe einer Konstruktionsmethode operationalisiert. Relevante Kovariablen (die „lokalen Umweltverhältnisse“) werden im folgenden durch eine Abbildung g auf der Domäne D repräsentiert.

3.1 Generische Stationarität

Ein Kovariogramm C heißt *generisch stationär* bezüglich g, falls es eine Funktion C_g gibt, so dass für alle s,t,h gilt:

$$C(s,t) = C_g(s,t;\, g),$$

Die Kovarianz von Umweltvariablen an den Stellen s und t hängt somit von den lokalen Umweltverhältnissen ab; und wenn an den Stellen $s+h$, $t+h$ dieselben lokalen Umweltverhältnisse vorliegen wie bei s und t, werden auch dieselben Kovarianzen auftreten. An jedem Punktepaar ($s+h$, $t+h$) wirkt daher das gleiche Einflussgesetz $C_g(s,t; \cdot)$, jedoch bezüglich „verschobener" Umweltverhältnisse $g(\cdot + h)$. Auf diese Weise kommt eine Stationarität des Einflusses der lokalen Umweltverhältnisse auf das Kovariogramm zustande.

Je nachdem, welche Kovariablen g und welches Einflussgesetz C_g zum Modellieren verwendet werden, ist das resultierende Kovariogramm mehr oder weniger stationär oder instationär – es handelt sich sozusagen um eine „skalierbare" Instationarität.

3.2 Eine Konstruktionsmethode

Allgemein lässt sich zeigen (VAN DEN BOOGAART 1999, BRENNING 2001), dass sich eine hinreichend große Klasse von Kovariogrammen konstruktiv durch eine Faltung erzeugt werden kann:

$$C_w(s,t) := \int_E w(s,p)w(t,p)\,\mathrm{d}p,$$

wobei $w : D \times E \to \mathbb{R}$ eine geeignete Gewichtsfunktion auf einer Domäne $D \times E$ sei.

Werden radialsymmetrische Gewichtsfunktionen verwendet, so können isotrope Kovariogramme erzeugt werden. Dies ist etwa für die Indikatorfunktion auf einer Kreisscheibe der Fall, während die Indikatorfunktion einer Ellipse mit lokal variierender Hauptachsenorientierung ein lokal anisotropes Kovariogramm induziert.

Die Unabhängigkeit der beobachteten Umweltvariable auf verschiedenen Teilgebieten der Domäne lässt sich durch heranmultiplizieren weiterer Indikatorfunktionen erzielen. Durch geeignete Wahl von Gewichtsfunktionen $w(s,t; g)$ können somit die lokalen Umweltverhältnisse g auf ein sie berücksichtigendes generisch stationäres Kovariogramm abgebildet werden.

Das hier vorgestellte Konstruktionsprinzip repräsentiert daher einen Modellierungsansatz, dessen Ziel darin besteht, Kovarianzmodelle baukastenartig zusammenzusetzen, wobei Kenntnisse über Strukturen und Prozessmuster in unserer Umwelt abgebildet werden. Das folgende Beispiel soll dies verdeutlichen.

3.3 Eine hydrologisch motivierte Anwendung

Bei der Untersuchung bestimmter hydrologischer oder hydrogeologischer Variablen kann angenommen werden, dass Paare von Punkten, die in unterschiedlichen Einzugsgebieten liegen, unkorreliert sind; dagegen sind Punkte mit sich überschneidenden Teileinzugsgebieten umso stärker korreliert, je größer deren Schnittfläche ist.

Bezeichnet nun $A(s)$ das zu einem Punkt s gehörige hangaufwärts gelegene Teileinzugsgebiet, so lässt sich vermittels einer Gewichtsfunktion mit Träger $A(s)$ – z.B. einer Indikatorfunktion – ein Kovariogramm induzieren, welches unser qualitatives Prozesswissen repräsentiert. Dieses Kovariogramm ist generisch stationär, jedoch in hohem Maße instationär im klassischen Sinne.

Das Teileinzugsgebiet zu einem Punkt kann mit Hilfe von digitalen Geländemodellen ermittelt werden.

3.4 Implementationsansätze

Ausgewählte generisch stationäre Modelle wurden von BRENNING (2001) implementiert. Dabei kam die open-source Datenanalyse-Umgebung und Programmiersprache R (www.r-project.org) zum Einsatz. Die für die Berechnung von Kovariogrammen (genauer: Semivariogrammen) aufgrund des beschriebenen Konstruktionsschemas benötigte numerische Integration wurde als Quasi-Monte-Carlo-Verfahren in C implementiert. Ferner wurde eine einfache GIS-Schnittstelle zur kommerziellen Software ArcView 3.1 in der Programmiersprache AVENUE realisiert, um den Geodatenaustausch zu erleichtern.

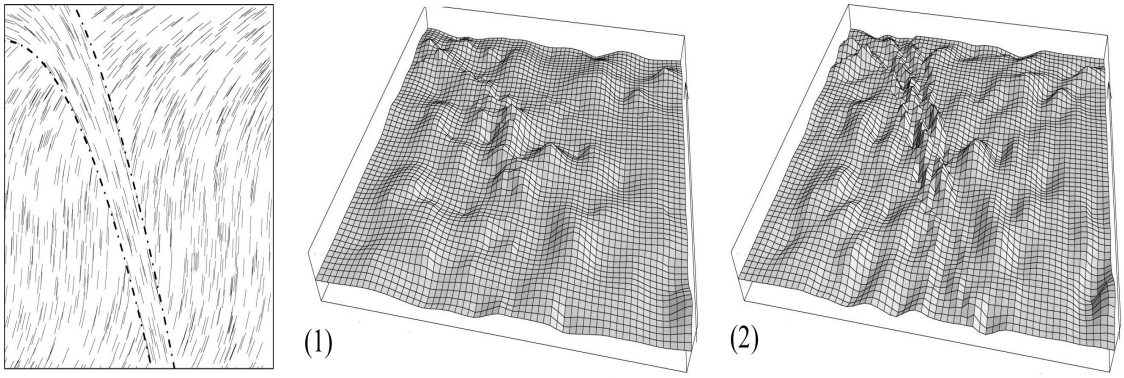


Abbildung 1: Links: Die für die Simulation eines generisch stationären Datensatzes angenommenen Umweltverhältnisse: Es werden zwei unabhängige Teil-Zufallsfelder mit lokal variierenden Anisotropierichtungen (durch Linien angedeutet) betrachtet. Mitte und rechts: Kriging-Oberflächen, die unter Verwendung geeigneter stationärer (Mitte) bzw. generisch stationärer Modelle (rechts) berechnet wurden.

4 Ein simulierter Beispieldatensatz in komplexen Umweltverhältnissen

Als Beispiel für die Anwendung generisch stationärer Kovarianzstrukturen wurde ein zu einem instationären (generisch stationären) Gaußschen Zufallsfeld gehörender Datensatz aus 259 unregelmäßig verteilten Werten mit Hilfe von Pseudozufallszahlen simuliert. Hierbei wurden lokale Instationaritäten auf voneinander unabhängigen Teilgebieten angenommen, wie sie durch häufig auftretende Umweltverhältnisse induziert werden können (Abb. 1). An diesen Datensatz wurden geostatistische Modelle angepasst, welche (1) stationär bzw. (2) generisch stationär unter Berücksichtigung der zugrunde liegenden (bekannten) Ahnisotropiestruktur sind.

Die Ergebnisse der geostatistischen Schätzung mit Hilfe von Kriging sind in Abbildung 1 dargestellt. Wie deutlich zu erkennen ist, werden die durch die anisotropen Kovariablen vorgegebenen Richtungsabhängigkeiten und Unstetigkeiten unter Stationaritätsannahmen nicht reproduziert. Lediglich in der Mitte des dargestellten Gebietes ist eine etwas ausgeprägtere Variabilität der interpolierten Werte zu beobachten, welche jedoch in alle Richtungen unscharf abnimmt. Dagegen zeichnet die mit einem generisch stationären Modell erzeugte Interpolation relativ scharf die Grenze zwischen den beiden unabhängigen Gebieten sowie die Richtung der lokalen Anisotropien (langgestreckte „Hügel“) nach. Das Kriging-Ergebnis unter

der Annahme generischer Stationarität kommt somit den in unserer meist instationären Umwelt näher.

5 Herausforderungen für die Zukunft

Bei den vorgestellten generisch stationären Modellen handelt es sich um eine Auswahl, welche das angegebene Konstruktionsprinzip veranschaulichen und wichtige Anwendungsfelder aufweisen soll. Eine Herausforderung für zukünftige Arbeiten stellt nun die Bereitstellung einer Reihe von weiteren einfachen, generisch stationären Kovarianzmodellen dar, welche wichtige Umweltprozesse und -strukturen – etwa in Abhängigkeit von Relief oder Hydrographie – berücksichtigen. Aufbauend auf diesen einfachen Modellen können dann nach dem Baukastenprinzip komplexere Umweltverhältnisse wiedergegeben werden. Empirische Analysewerkzeuge müssen geschaffen werden, um in der Praxis die Anwendbarkeit solcher Modelle auf einen konkreten Datensatz zu prüfen. In Verbindung mit dem GIS-gestützten Einsatz sollte ein Expertensystem dem Anwender die Verwendung generisch stationärer Modelle erleichtern.

Integrierte Meß- und Bilddatenhaltung für Umweltdatenbanken

Workshop GI-AK Umweltdatenbanken
Berlin, 19. Mai 2003

Peter Baumann

rasdaman GmbH,

FORWISS (Bayerisches Forschungszentrum für Wissensbasierte Systeme)

©2003 rasdaman GmbH

rasdaman

Gliederung

- EuroClim
- rasdaman
- Live-Demo
- Zusammenfassung

©2003 rasdaman GmbH

rasdaman

Datenmaterial

- European Climate Change Monitoring and Prediction System
- Datenmaterial
  - Schwerpunkt: Schnee/Eis-Bedeckung, -Temperatur
  - AVHRR Polar Pathfinder, MODIS etc.
  - Polar stereographic; Eingabeformate: HDF 5, GRIB (DKRZ)

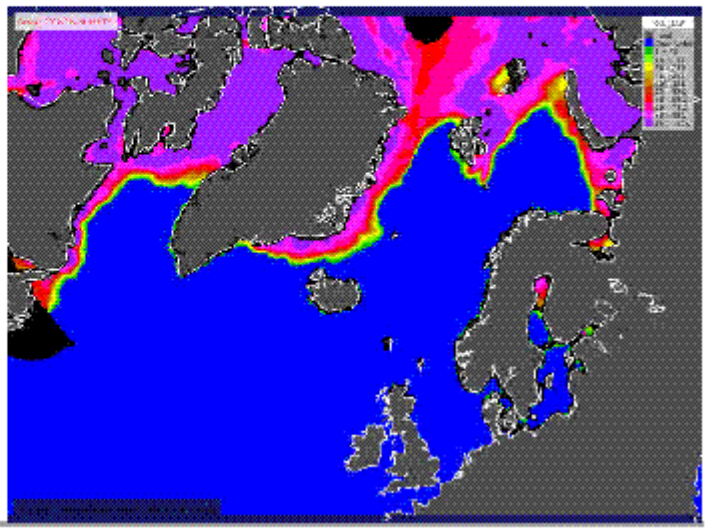
Mode values for cryosperic parameter characteristics across all organisations								
	Snow-on-land variables				Sea ice variables		Glacier vars	
	Albedo	*Coverage*	*Temp*	*Wetness*	*Concen-tration*	*Tick-ness*	*Firn line altitude*	*Volume change*
Spatial Resol'n	**1km 10km** (3)	**1km** (8)	**1km 10km** (3)	**1km** (4)	**1km 100km** (2)	**1km 100km** (2)	**1km** (3)	**1km** (2)
Update Freq.	**7day** (5)	**1/day 5day** (5)	**1/day** (3)	**7day** (3)	**1, 7, 30dy** (2)	**1, 7, 30dy** (2)	**30dy** (4)	**30dy** (3)

©2003 rasdaman GmbH

rasdaman

Datenmaterial

- Sea ice concentration
- O&SI SAF (EUMETSAT)
  - SSM/I



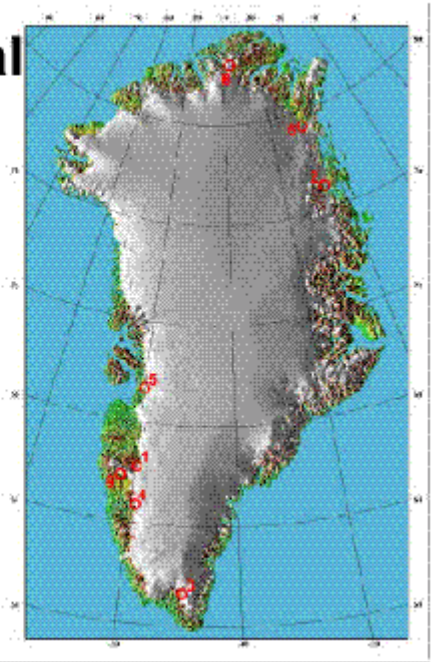
©2003 rasdaman GmbH

rasdaman

EuroClim

Datenmaterial

- Temporal coverage: 1982 - now
- Variables
  - T, humidity, precipitation cumulative, wind (speed and direction), sunshine hours, evaporation, global radiation



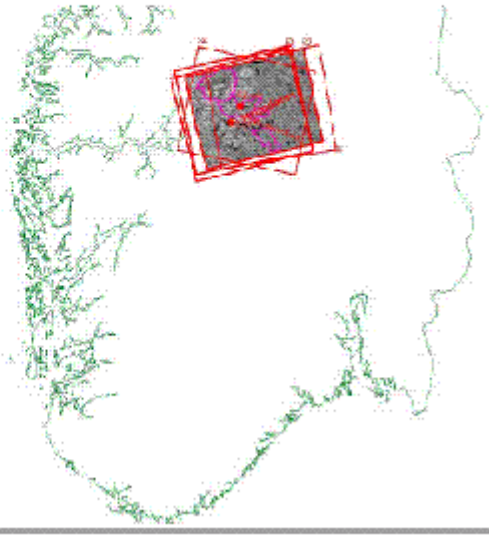
©2003 rasdaman GmbH

rasdaman

EuroClim

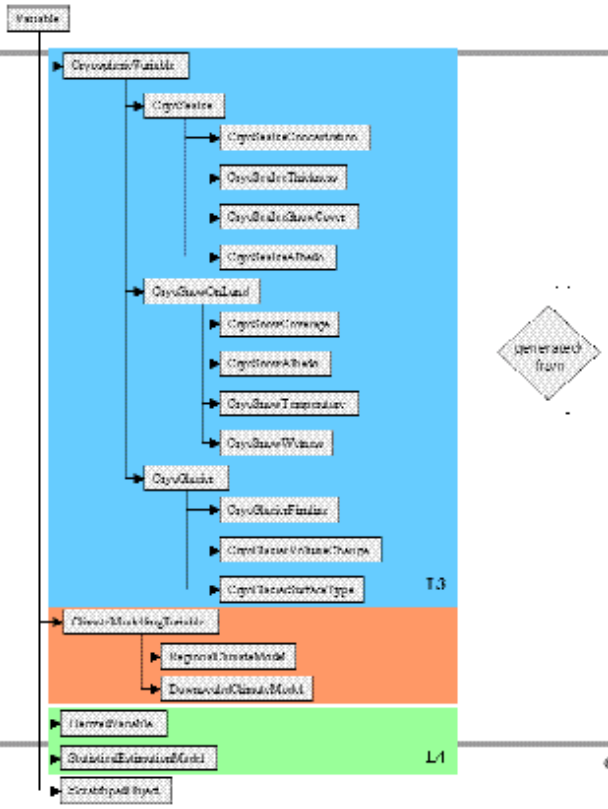
Datenmaterial

- RADARSAT/Envisat, ERS-2
  - 100km



©2003 rasdaman GmbH

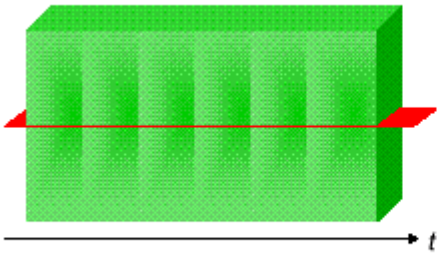
2D-Schema (vorläufig)



©2003 rasdaman GmbH

4D-Klimamodelle

- Beispiel: ECHAM T42-Modell
  - 128 x 64 x 17 x 2,190,000 Voxel
  - 2.5 TB pro Variable (24min)
  - 50 oder mehr Variablen pro Modelllauf



©2003 rasdaman GmbH

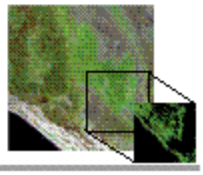
rasdaman

- Middleware-Toolkit für großvolumige Rasterdaten
  - unbeschränkte Größe, Dimension, Zellentypen
  - multidimensionales SQL, Java, C++
  - intelligente Speicher- und Anfrageoptimierung
  - interoperabel: DB-Systeme, GIS (OGC, ESRI)

metadata	att 1	att 2	att n
	key1	...	oid 1
	key2	...	oid 2
	key3	...	oid 3

my_coll	OID	array
	oid 1	
	oid 2	
	oid 3	
	oid 4	
	oid 5	

- rasql = multidimensionale Ausdrücke in SQL
  - typedef marray
    < struct{ char red, green, blue, …; },
    [x0:x1,y0:y1]
    > LandsatImage;
  - select img.green[x0:x1,y0:y1] > 130
    from LandsatArchive as img



©2003 rasdaman GmbH

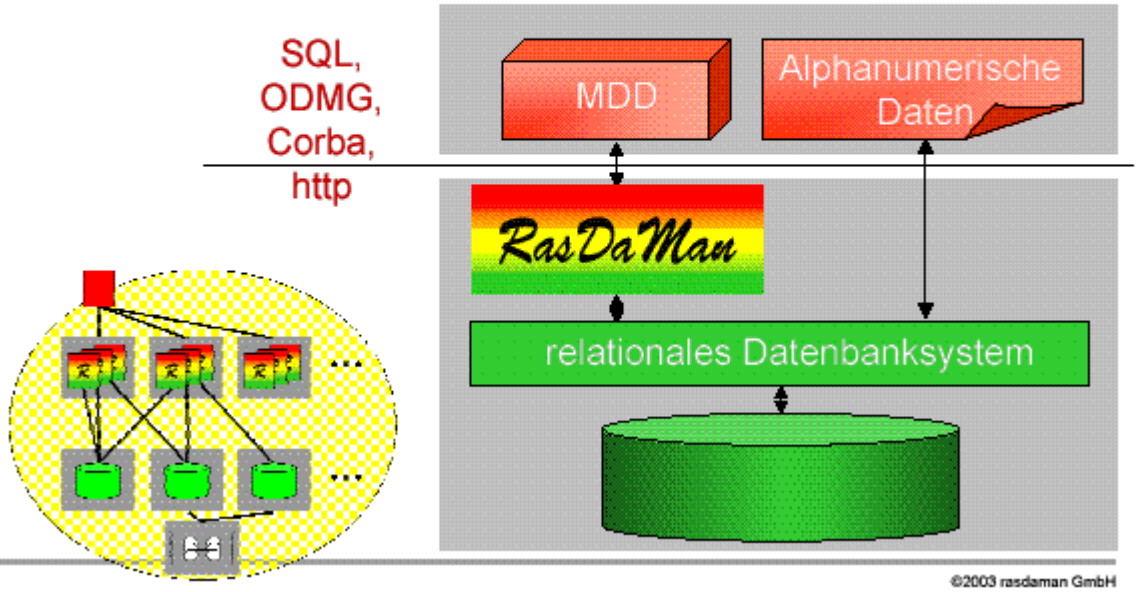
Multidimensionale Definition & Abfrage

```
typedef marray< struct{ char red, green, …; }, [x0:x1,y0:y1] > LandsatImage;


SELECT jpeg(
                scale(img0[X00:X01,Y00:Y01],[1:300,1:300])          * { 1c, 1c, 1c}
       overlay (scale(img1[X10:X11,Y10:Y11],[1:300,1:300]) < 71.0) * {51c, 153c, 255c }
       overlay bit(scale(img2[X20:X21,Y20:Y21],[1:300,1:300]), 2)  * {230c, 230c, 204c}
       overlay bit(scale(img4[X20:X21,Y20:Y21],[1:300,1:300]), 7)  * {102c, 102c, 102c}
       overlay bit(scale(img5[X20:X21,Y20:Y21],[1:300,1:300]), 6)  * {255c, 255c, 0c}
       overlay bit(scale(img6[X20:X21,Y20:Y21],[1:300,1:300]), 3)  * {191c, 242c, 128c}
       overlay bit(scale(img7[X20:X21,Y20:Y21],[1:300,1:300]), 4)  * {191c, 255c, 255c}
       overlay bit(scale(img8[X20:X21,Y20:Y21],[1:300,1:300]), 1)  * {0c, 255c, 255c}
       overlay bit(scale(img9[X20:X21,Y20:Y21],[1:300,1:300]), 0)  * {102c, 102c, 102c}
       )
FROM …
```

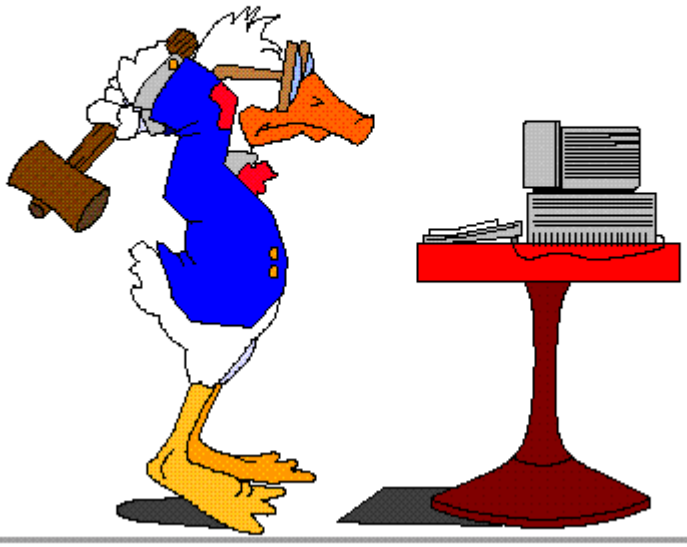
©2003 rasdaman GmbH

Architektur



©2003 rasdaman GmbH

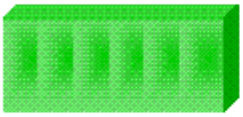
Live Demo



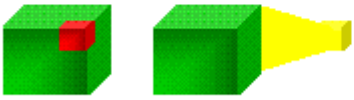
©2003 rasdaman GmbH

Zusammenfassung

- Aufgabe der Datenhaltung:
  - Import system-getrieben



  - Abgabe nutzergetrieben



- Beispiel: EuroClim
  - Dimensionsübergreifende Datenhaltung
- Rasterdatenbanken „heisses Thema"
  - „Kerninnovation" (Oracle)
- Flexibilität der Anfragesprache!

©2003 rasdaman GmbH

Literatur

BRENNING, A. (2001): Geostatistics without stationarity assumptions within Geographical Information Systems, Freiberg Online Geosciences 6.

BRENNING, A. & VAN DEN BOOGAART, K. G. (2001): Geostatistics without stationarity assumptions within GIS, Proc. 7th Annual Conference of IAMG, Cancún.

CRESSIE, N. A. C. (1993): Statistics for spatial data, Wiley, New York.

GOOVAERTS, P. (1997): Geostatistics for natural resources evaluation, Oxford University Press, New York et al.

VAN DEN BOOGAART, K. G. (1999): A new possibility for modelling variograms in complex geology, Proc. of StatGIS Klagenfurt 1999.

VAN DEN BOOGAART, K. G. & BRENNING, A. (2001): Why is Universal Kriging better than IRFk-Kriging: Estimation of variograms in the presence of trend, Proc. 7th Annual Conference of IAMG, Cancún.

Anwendung von Metadaten im Küstenzonenmanagement

Carsten Heidmann[36], Rainer Lehfeldt[37], Wassilios Kazakos[38] und Frank Simmering[39]

Zusammenfassung

In NOKIS wird eine Auswahl von Metadatenelementen aus ISO 19115 und Dublin Core verwendet, um einen handhabbaren Kern von Metadaten für die Küstenzone zu definieren. Zur Erfassung der Metadaten wird ein Editor verwendet, der durch Änderungen an dem zu Grunde liegenden XML-Schema an örtliche Besonderheiten angepasst werden kann. Die mit diesem Werkzeug lokal erfassten Metadaten werden auf einem zentralen Server repliziert und über ein Webportal präsentiert.
Stichworte: Nord- und Ostsee-Küsten-Informationssystem, Metadateneditor, ISO 19115, Dublin Core, NOKIS

Einleitung

Das Kuratorium für Forschung im Küsteningenieurwesen (KFKI) [7] hat ein dreijähriges Forschungs- und Entwicklungsprojekt initiiert, um ein Nord- und Ostsee-Küsteninformationssystem (NOKIS) [16] zu entwickeln. Die Website des Projektes

[36] Bundesanstalt für Wasserbau – Dienststelle Hamburg, Wedeler Landstraße 157, 22559 Hamburg, Tel. 040.81908.345, Email: heidmann@hamburg.baw.de

[37] Bundesanstalt für Wasserbau – Dienststelle Hamburg, Wedeler Landstraße 157, 22559 Hamburg, Tel. 040.81908.312, Email: lehfeldt@hamburg.baw.de

[38] Forschungszentrum Informatik, Haid-und-Neu-Straße 10-14, 76131 Karlsruhe, Tel. 0721.9654.712, Email: kazakos@fzi.de

[39] plan-GIS GmbH, Gesellschaft für digitale Planung, Gustav-Elster-Str. 1, 26789 Leer, Tel. 0491.979901.17, Email: f.simmering@plan-gis.de

mit allen dazugehörigen Daten wird dauerhaft von der Bundesanstalt für Wasserbau in Hamburg unterhalten (http://nokis.baw.de). Die Aktivitäten in dem seit 2001 laufenden Projekt bilden einen wichtigen Baustein innerhalb von integrierten Managementsystemen, die derzeit in Deutschland und Europa [3] mit dem Ziel initiiert werden, Netzwerke für den Austausch von Informationen und Erfahrungen zwischen den in diesem Bereich Tätigen aufzubauen.

NOKIS zielt darauf ab, aktuelle, relevante, verlässliche und zielgruppenorientierte Informationen für Bundes- und Landesdienststellen bereitzustellen, die mit Küsteningenieurs- und Umweltfragen beschäftigt sind. Das Hauptaugenmerk richtet sich dabei auf Informationen zum Auffinden von Daten und zugehörigen Publikationen aus dem Küstenbereich wie Forschungsberichte, Veröffentlichungen in Fachzeitschriften oder technische Berichte. In der gegenwärtigen Projektphase setzen sich die Datenlieferanten aus Nationalparkämtern, für den Küstenschutz zuständigen Bundes- und Landesdienststellen sowie aus einer Wasser- und Schifffahrtsdirektion zusammen. Für den Projektträger Jülich wird im Rahmen des Projektes eine Datenbank erstellt, die Beschreibungen aller gegenwärtigen und abgeschlossenen Projekte aus dem Bereich des Küsteningenieurwesens enthält. Eine weitere Quelle für das Auffinden von Forschungsberichten und Gutachten wird mit der Datenbank für die Zeitschrift "Die Küste", die durch das KFKI herausgegeben wird, aufgebaut. Diese Datenbank enthält ebenfalls die in den enthaltenen Artikeln zititerte Literatur, und bildet damit eine wichtige Referenz für die vielfältige "graue Literatur" der sektoralen und regionalen Forschungsprogramme. Die auf das Küsteningenieurwesen spezialisierte Bücherei des KFKI enthält die meisten dieser Ressourcen, die ab Sommer 2003 auf dem NOKIS und dem KFKI Webportal zugänglichen webOPAC verfügbar sein werden.

Diese (Meta-)Datenquellen bieten eine breite und solide Basis für die vom Integrierten Küstenzonenmanagement (IKZM) benötigten sektoralen Informationen, damit dieses die Anforderungen erfüllen kann, einen holistischen Ansatz zu bieten, in dem das Ökosystem als ganzes mit allen biotischen und abiotischen Komponenten sowie mit allen Arten küstenspezifischer Nutzungen inklusive der Beziehungen zwischen verschiedenen Nutzungsansprüchen und der Konflikte zwischen Nutzungsansprüchen und der Umwelt enthalten ist [20].

Auch Untersuchungen der Beziehungen zwischen sozioökonomischen und ökologischen Aspekten hängen von Informationen aus verschiedenartigen Quellen ab
Die Bereitstellung strukturierter Metadaten ist eine zentrale Anforderung an Informationssysteme, um effiziente Suchmechanismen für das Auffinden von Daten Dokumenten, Karten, Bildern, Forschungsberichten und ähnlichem zur Verfügung stellen zu können. Metadaten zur räumlichen und zeitlichen Ausdehnung sowie zum thematischen Inhalt jeder einzelnen Ressource erlauben genaue Suchabfragen für Aufgabenstellungen des Küsteningenieurwesens.

NOKIS stellt dazu eine Kombination von karten- und textbasierten Auswahltools zur Verfügung, wobei die textbasierten Werkzeuge entweder von den vorhandenen Stichwörtern oder von dem Index einer Freitextsuche über den gesamten verfügbaren Metadatenbestand Gebrauch machen. Der nächste logische Schritt in dieser Reihe ist der direkte Zugang zu ausgewählten Daten und Webservices für Visualisierungs- und Analysezwecke [15].

Die neuen Aspekte von NOKIS sind ein standardisierter Satz von Metadaten für die Aufgabenstellungen der Küstenzone und ein Metadateneditor, der sowohl als Eingabetool für die Erfassung der Metadaten beim Datenanbieter, als auch als Darstellungsinterface für die Anzeige der Suchergebnisse verwendet werden kann. Der verwendete Map Server und die Datenbank sind Open Source Software (OSS) und der neu entwickelte Metadateneditor steht allen Institutionen, die sich an NOKIS beteiligen wollen, frei zur Verfügung.

Informationen aus der Küstenzone

Die anfänglichen sechs NOKIS Partner repräsentieren typische Informationsanbieter und steuern typische küstenspezifische Daten bei. Zur gleichen Zeit stellen sie auch die typischen Nutzer dar, die nach Daten und Dokumenten suche.
In den Nationalparkämtern sind Daten zu Biologie, Geomorphologie, Flora und Fauna, rechtliche Parameter, sozioökonomische und administrative Informationen von vordringlicher Bedeutung.

Diese Daten werden zum größten Teil in Geografischen Informationssystemen (GIS) der Firma ESRI (ArcView 3.x, ArcInfo, ArcGIS 8.x) vorgehalten, die relativ detaillierte Metadaten in einem proprietären Format mitführen.
Standardisierte Metadaten erleichtern die Suche, den Austausch und das Verschneiden von Daten und kartographischen Informationen aus dem Bereich der Küstenzone.
Innerhalb der Behörden, die sich mit dem Küstenschutz beschäftigen, zählen hydrologische und meteorologische Daten von Beobachtungsstationen, Höhenprofile und Topografien des Wattenmeeres, Fernerkundungsdaten, Katasterdaten zu Deichen und Küstenschutzbauwerken, Luftbilder sowie digitale Geländemodelle (DGM) zu den wichtigsten Arbeitsgrundlagen. Von der BAW werden im Rahmen von NOKIS Metadaten aus dem WSV-eigenen System WaGIS (Wasserstraßen-Geoinformationssystem) für die Peildatenbank Küste (PDBK) erstellt. Zu diesem Zweck wurde in der BAW Ilmenau ein Mapping von proprietären WaGIS-Metadaten zu NOKIS (ISO 19115) Metadaten erarbeitet. In einem weiteren Schritt ist die Erfassung von Pegelstammdaten aus dem Deutschen Gewässerkundlichen Jahrbuch (DGJ) der Bundesanstalt für Gewässerkunde geplant.
Die unterschiedlichen Archivierungsmethoden für die verschiedenen Datentypen bieten zumeist keine Möglichkeit für eine integrative Betrachtungsweise des verfügbaren Datenbestandes. So gibt es zum Beispiel keine Möglichkeit, ein Luftbild automatisch mit einem Höhenprofil zu verknüpfen. Die existierenden Beschreibungen der Datensätze konzentrieren sich meist auf technische Details und folgen dabei einer Vielzahl von Dokumentationsrichtlinien die in den einzelnen Bundesländern gültig sind. Zudem sind sie oft in inkompatiblen Formaten gespeichert, so dass ein erfolgreiches Auffinden von Daten verhindert wird.
Auf Bundesebene unterhält die Wasser- und Schifffahrtsverwaltung eine Reihe von lokalen Dienststellen, die sich mit Tiefenvermessungen, Kartenerstellung, Daten zu Schifffahrt und Hydrologie, Modellergebnissen usw. beschäftigen. Diese einzelnen Dienststellen sind untereinander per Intranet verbunden, um auf gemeinsame Datenbestände zugreifen zu können. Spezielle Webservices im Intranet stellen den Inhalt spezialisierter Archive dar und bieten Methoden für Datenzugang und -analyse an.

Zum momentanen Zeitpunkt gibt es so gut wie keine Möglichkeit einer interdisziplinären Datenbetrachtung über Organisationsgrenzen hinweg. Daher sind einige der wichtigsten Voraussetzungen eines IKZM [1] schwer zu erfüllen:

- Entscheidungen sollen aufgrund einer fundierten Datengrundlage und aufgrund von zuverlässigen Informationen getroffen werden.
- Entscheidungsvorgänge sollen alle Interessenvertreter und alle relevanten Behörden miteinbeziehen.
-

Entscheidungsunterstützungssysteme (Decision Support Systems, DSS) benötigen in diesem Zusammenhang Informationssysteme, die alle genannten Datenquellen gleichermaßen abdecken. Für einen schnellen Zugang zu den entsprechenden Ressourcen werden daher standardisierte Metadaten benötigt, die sowohl die technischen Informationen zur Verarbeitung der Daten als auch Informationen zur Beurteilung des Dateninhaltes enthalten.

Metadaten in NOKIS

In den letzten Jahren haben sich zahlreiche Metadatenstandards entwickelt. Sie alle dienen der formalisierten Datenbeschreibung zum Zweck der Dokumentation von Datenbeständen und zur Unterstützung von formalisierten Suchalgorithmen für ein systematisches Auffinden von Informationen.
Der Dublin Core Standard [2] ist der kleinste standardisierte Satz von Metadatenelementen, die Informationen über das "Was, Wann, Wo und Wer?" der Daten und über das "Wie?" des Zugangs und der Weiterverarbeitung dieser Daten enthalten. Die "International Organization for Standardization" (ISO) [9] ist mit der Entwicklung eines Standards ISO 19115 beschäftigt, der mehrere hundert Metadatenelemente definiert. Diese enthalten detaillierte Informationen über die zu beschreibenden Daten. Das US-amerikanische "Federal Geographic Data Commitee" (FGDC), von dem der etablierte Metadatenstandard "Content Standard for Digital Geospatial Matadata" (CSDGM) stammt, hat angekündigt den ISO-Standard in naher Zukunft zu übernehmen und damit den eigenen Standard abzulösen, der seit seiner Verabschiedung 1994 für die Beschreibung aller US-amerikanischen Geodatensätze verwendet wurde.

Innerhalb von NOKIS gibt es verschiedene Anwendungen von Metadaten, die mit den verschiedenen Datenquellen verbunden sind:

- bibliographische Informationen (z.B. KFKI-Zeitschrift "Die Küste"), sowie die Dokumentation von Forschungsprojekten,
- Dokumentation von GIS Ressourcen,
- Dokumentation von Datenbeständen, die in verschiedenen Datenbanken abgelegt sind (z. B. Karten, Photographien, Rohdaten),
- Dokumentation von numerischen Modellergebnissen.

Viele Veröffentlichen existieren auch in einer PDF-Version, die auf Servern im Internet gespeichert sind. Zusätzlich zu der Verwendung der standardisierten bibliographischen Informationen bietet die durch NOKIS durchgeführte Volltextindexierung dieser Dokumente eine weitere Möglichkeit zum Auffinden von Informationen anhand von beliebigen Begriffen.

Anwendungen für Dublin Core Metadaten

Der NOKIS Datenbestand enthält einen Katalog von Forschungsprojekten aus dem Küstenbereich, die über das Webportal dargestellt werden können. Abbildung 1 zeigt die HTML-Darstellung der in einer PostgreSQL-Datenbank gespeicherten Dublin Core Metadatenelemente. Die Erzeugung der Seiten erfolgt über ein Servlet [6], das die entsprechenden Seiten bei einer Anfrage aus der Datenbank erstellt.

NOKIS Küstenforschung KFKI Projekt Nr. 72 Pt KFKI

Titel	NOKIS. Metadatenbank über Forschungsaktivitäten und Forschungsergebnisse des Küsteningenieurwesens
Laufzeit	2001 - 2004
Fördernummer	03 KIS 027
Koordinator	Dr.-Ing. Hanno Heyer [heyer@hamburg.baw.de]
Institution	BAW [http://www.hamburg.baw.de/]
Identificator	http://nokis.baw.de
Beschreibung	Entwicklung und Einführung eines zentralen Systems für Metadaten-Verwaltung und -Vermittlung einschl. Einrichtung lokaler Metadatenbanken bei Küstendienststellen als Basis-Informationssystem für die Küstenforschung und das Küstenzonenmanagement. Das System hilft Forschung zu optimieren und den Informationsaustausch - auch zwischen Ingenieuren und dem Umweltbereich - zu fördern. Es wird zunächst für wenige Dienststellen als Pilotversion eingerichtet und danach dauerhaft - beliebig erweiterbar - fortgeführt.
Stichwörter	NOKIS, Ostsee, Nordsee, Küstenforschung, IKZM, Ingenieurwesen, Metadaten Informationssystem, Gewässerkunde, Meteorologie, Sedimentologie, Ozeanographie, Meereskunde, Topographie, Bathymetrie, GIS, Modellierung, Modelle NOKIS, Baltic Sea, North Sea, coastal research, iczm, engineering, metadata information system, hydrology, meteorology, sedimentology, oceanography, topography, bathymetry, GIS, modeling, models
Räumliche Ausdehnung	
Veröffentlichungen	1. Zwischenbericht. Lehfeldt, R., Heidmann (2002) NOKIS - Erstellung eines Metadateninformationssystems für die Küstenforschung und das Küsteningenieurwesen 2. Zwischenbericht. Lehfeldt, R., Heidmann (2003) NOKIS - Erstellung eines Metadateninformationssystems für die Küstenforschung und das Küsteningenieurwesen

	Geographische Koordinaten
Nord	57,00
Süd	51,00
West	3,50
Ost	15,50

Abbildung 1: Auflistung von Projektdaten aus DC Metadaten generiert

Die erste und zweite Spalte in Tabelle 1 zeigt das Mapping zwischen den Begriffen, die in der in Abbildung 1 gezeigten Projektbeschreibung verwendet werden, und den entsprechenden Dublin Core Qualifiers/Refinements. Die dritte Spalte führt das Mapping weiter fort auf die entsprechenden ISO 19115 Elemente, die dann auch eigene Felder für Links für Email, von Websites der beteiligten Institutionen sowie der Projekthomepage besitzen.

Der jeweilige Projektkoordinator erstellt für das eigene Projekt eine Kurzbeschreibung und Stichwörter, vorzugsweise aus einem Thesaurus, in Deutsch und Englisch.

Projekt-beschreibung	Dublin Core Element	ISO 19115 Element
Koordinator	DC.Creator	Metadata.mdContact.RespParty (role=principal investig.)
Koordinator	DC.Contributor	Metadata.mdContact.RespParty (role=resource provider)
Koordinator	DC.Publisher	Metadata.mdContact.RespParty (role=publisher)
Laufzeit	DC.Date	Metadata.Ident.Citation.resRefDate
	DC.Date.*created*	Metadata.mdDateSt
Titel	DC.Title	Metadata.Ident.Citation.resTitle
	DC.Title. *alternative*	Metadata.Ident.Citation.resAltTitle
Beschreibung	DC.Description.*abstract*	Metadata.Ident.idAbs
	DC.Description.*tableOfContents*	
Stichwörter	DC.Subject	Metadata.Ident.descKeys
	DC.Type	Metadata.Ident.Citation.citIdType
	DC.Format	Metadata.Ident.Format.formatName
	DC.Format.*extent*	Metadata.Ident.distTransOps.transSize
	DC.Format.*medium*	Metadata.Ident.distTransOps.medName
Identificator	DC.Identifier	Metadata.Ident.Citation.citId
	DC.Source	Metadata.Ident.Citation.citRespParty
Veröffentli-chungen	DC.Relation	Metadata.Ident.mdParentId
	DC.Language	Metadata.contInfo.catLang
Räumliche Ausdehnung	DC.Coverage.*spatial (DCMI Box)*	Metadata.dataIdInfo.geoBox
	DC.Coverage.*spatial (TGN)*	Metadata.dataIdInfo.geoDesc.geoId
	DC.Coverage.*spatial (polygon)*	Metadata.dataIdInfo.datExt.geoEle.polygon
	DC.Coverage.*spatial (vertical)*	Metadata.dataIdInfo.datExt.vertExtent
	DC.Coverage. *temporal*	Metadata.dataIdInfo.datExt.exTemp
	DC.Rights	Metadata.Ident.mdConst.useConsts

Tabelle 1: Mapping von Dublin Core zu ISO 19115 für Projektbeschreibungen

Zur Zeit wird bei den im Rahmen von NOKIS verwendeten Stichworten mit zweisprachigen Wortlisten technischer Ausdrücke gearbeitet, da die Terminologie des Küsteningenieurwesens nur sehr unzureichend in bestehenden Thesauri abgebildet wird. Die mit einem Projekt assoziierte räumliche Ausdehnung erzeugt eine "Bounding Box" des Projektgebietes in dem in Abbildung 7 gezeigten Mapserver.

Die für die Projektbeschreibungen ausgewählten Metadatenelemente erlauben zwei verschiedene Arten von Links in einem Publikationseintrag: der Erste referenziert mit dieser Ressource verbundene und online verfügbare PDF-Dokumente, der Zweite erzeugt automatisch eine Email an die KFKI Bücherei, die bereits alle notwendigen Daten wie Signatur und Titel des Dokuments enthält.

ISO19115 metadata entity set information	Metadaten Elemente	ISO core	NOKIS
MD_Metadata	Metadata file identifier	o	m
	Metadata standard name	o	o
	Metadata standard version	o	o
	Metadata language	c	m
	Metadata character set	c	m
	Metadata point of contact	m	m
	Metadata date stamp	m	m
	Scope to which the metadata applies	(o)	m
MD_Identification	Dataset title	m	m
	Dataset reference date	m	m
	Dataset responsible party	o	m
	Abstract describing the dataset	m	m
	Descriptive keywords	(o)	m
	Status	(o)	m
MD_DataIdentification	Spatial representation type	o	o
	Spatial resolution of the dataset	o	o
	Dataset language	m	m
	Dataset character set	c	m
	Dataset topic category	m	m
	Geographic location bounding box or geographic identifier	c	m
	Vertical and temporal extent of dataset	o	m
MD_Constraints	Name of handling restriction	(o)	m
DQ_DataQuality	Scope of data	(o)	m
	Lineage statement	o	o
MD_MaintenanceInformation	Maintenance frequency	(o)	o
MD_SpatialRepresentation	Grid spatial representation	(o)	o
	Vector spatial representation	(o)	o
MD_ReferenceSystem	Reference system	o	o
MD_ContentInformation	External feature catalogue	(o)	o
	Coverage content	(o)	o
	Image description	(o)	o
MD_PortrayalCatalogue-Reference	Portrayal catalogue citation	(o)	o
MD_Distribution	Distribution format	o	o
	On-line resource	o	o
MD_MetadataExtension-Information	Information on metadata extensions	(o)	o
MD_ApplicationSchema-Information	Application schema information	(o)	o

Tabelle 2: Der NOKIS-Metadatenstandard als Erweiterung des ISO 19115

m = mandatory (Pflichtelement)
c = conditional (konditionales Element)
o = optionales Element, (o) = nicht Teil des ISO recommended core

Für die Zeitschrift "Die Küste" gibt es eine ähnlich aufgebaute Datenbank. Auch hier werden über ein Servlet die entsprechenden Seiten aufgebaut, in denen separate Angaben zum gesamten Band oder zu den einzelnen Artikeln enthalten sind. Ebenso besteht die Möglichkeit, eine Kopie des entsprechenden Artikels über eine automatisch generierte Email von der KFKI Bücherei zu bestellen.

Eine Volltextindexierung dieser beiden Datenbanken bietet eine ergänzende Methode zum Auffinden einzelner Publikationen. Dies ist ein sehr pragmatischer Ansatz, der zum Teil auch vom Erfolg der entsprechenden Internet-Suchmaschinen motiviert ist. Diese Art der Suche ergibt gute Resultate, wenn sie auf den gesamten Kontext eines Datensatzes mit allen damit verbundenen Metadaten angewandt wird, da die bei solchen Suchanfragen verwendeten, nicht standardisierten Begriffe oft irgendwo in den Metadaten der zugehörigen Dokumente auftauchen.

Anwendungen von ISO 19115 Metadaten

Das technische Komitee 211 der ISO (TC 211) hat eine Serie von Standards und technischen Empfehlungen für geographische Informationen entwickelt (ISO 191xx). In dieser Reihe wird durch den ISO 19115 [9] ein umfassender Satz von Metadatenelementen zur Beschreibung räumlicher Daten entwickelt, der mehr als 400 Einzelelemente umfasst.
Der empfohlene Kernsatz von Metadaten enthält 22 essentielle Elemente aus den in Tabelle 2 aufgelisteten Gruppen. Wenn die optionalen und die empfohlenen konditionalen Elemente außer Acht gelassen werden, verbleiben sieben Pflichtelemente (grau markiert), die für eine ISO 19115 konforme Beschreibung eines Datensatzes benötigt werden. Diese identifizieren den Datensatz (Titel, Referenzdatum, Zusammenfassung, Thema und Sprache) und den Metadatensatz (Ansprechpartner, Datum).

Viele der Einträge werden aus vordefinierten Codelisten ausgewählt, so dass die Verwendung eines standardisierten Vokabulars sichergestellt ist, das auch im internationalen Zusammenhang verwendet wird. Wo immer dies möglich ist, besteht die Möglichkeit, Referenzen zu nennen, um so ein korrektes gemeinsames Verständnis der Begriffe für die Datenbeschreibung zu erlangen. Die verwendeten Stichwörter sollten aus Thesauri stammen, die auch bei der Eingabe der Stichwörter referenziert werden sollen. Zusammenfassungen und Beschreibungen zur Datenqualität werden als Freitext eingegeben.

Das Metadatenmodell des ISO 19115 berücksichtigt die Möglichkeit zur Eingabe mehrsprachiger Einträge in allen Freitextfeldern, ebenso können zusätzliche Metadatenelemente standardkonform definiert oder bestehende Elemente abgeändert werden, um so auf die Anforderungen spezieller Nutzergemeinden einzugehen.
Der durch die NOKIS-Partner verabschiedete Metadatenstandard spiegelt die Anforderungen verschiedener Nutzergemeinschaften des Küstenbereichs wieder.

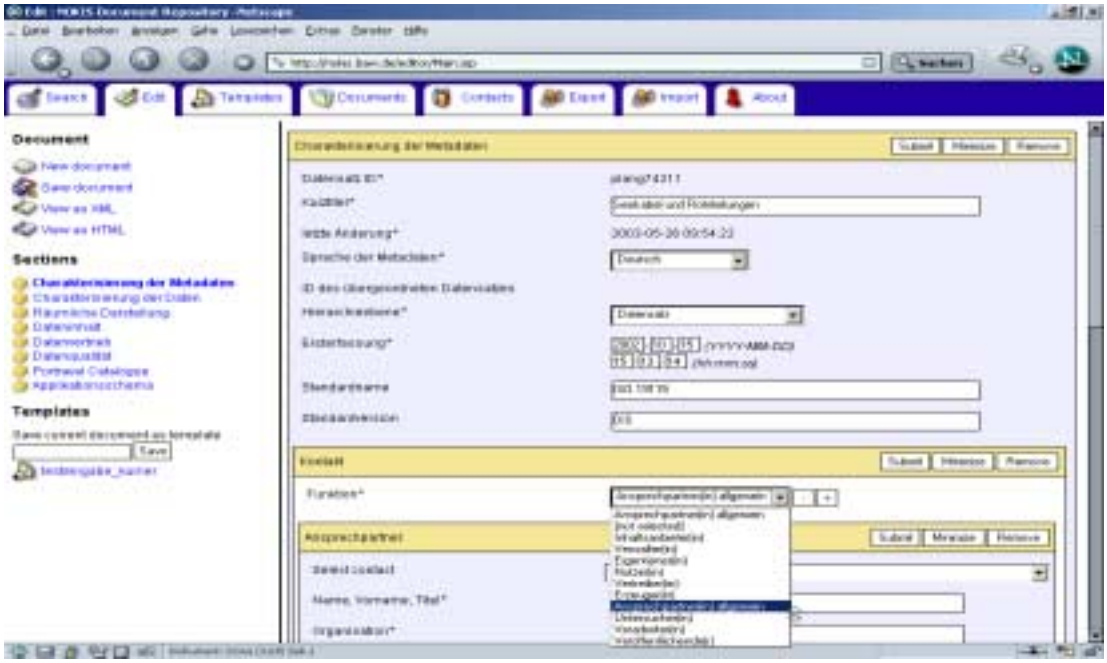


Abbildung 2: NOKIS Metadateneditor

Die teilnehmenden Dienststellen erzeugen oder erweitern ihre proprietären Metadaten und harmonisieren diese mit dem NOKIS-Metadatenstandard, dessen Kernsatz an Elementen in Tabelle 2 dargestellt ist und der eine substantielle Erweiterung gegenüber dem von der ISO vorgeschlagenen Satz von Metadatenelementen darstellt. Auf diese Weise ist es möglich, eine umfassende bibliographische und technische Dokumentation für die lokalen Datenbestände mit ihren Daten aus den Bereichen GIS, Luftbilder, Literatur oder gewässerkundlichen Datenreihen aufzubauen. NOKIS verlangt die Angabe von 19 Pflichtelementen, die vor allem die Bereiche Datensatzcharakterisierung, Nutzungseinschränkungen und Datenqualität genauer beschreiben. Zusätzlich wurde eine Zahl von neuen optionalen Feldern, die nicht Teil des ISO sind, definiert, um damit Richtlinien für die sachgerechte Verwendung der beschriebenen Daten zu geben.

Die räumliche Ausdehnung der Daten in Form von Eckkoordinaten einer Bounding Box ist eine wichtige Metainformation, die in kartenbasierten Suchmechanismen verwendet wird. Zusätzlich kann die zeitliche Ausdehnung als Filterkriterium dienen.

Stichwörter sowie eine thematische Kategorisierung der Datensätze sind von besonderer Bedeutung für die bibliographische Indexierung der Datenquellen und der dazugehörigen Dokumente. NOKIS erfordert zusätzlich zu Stichwörtern in der ersten Sprache (in diesem Fall meist deutsch) die Angabe der entsprechenden englischen Begriffe.

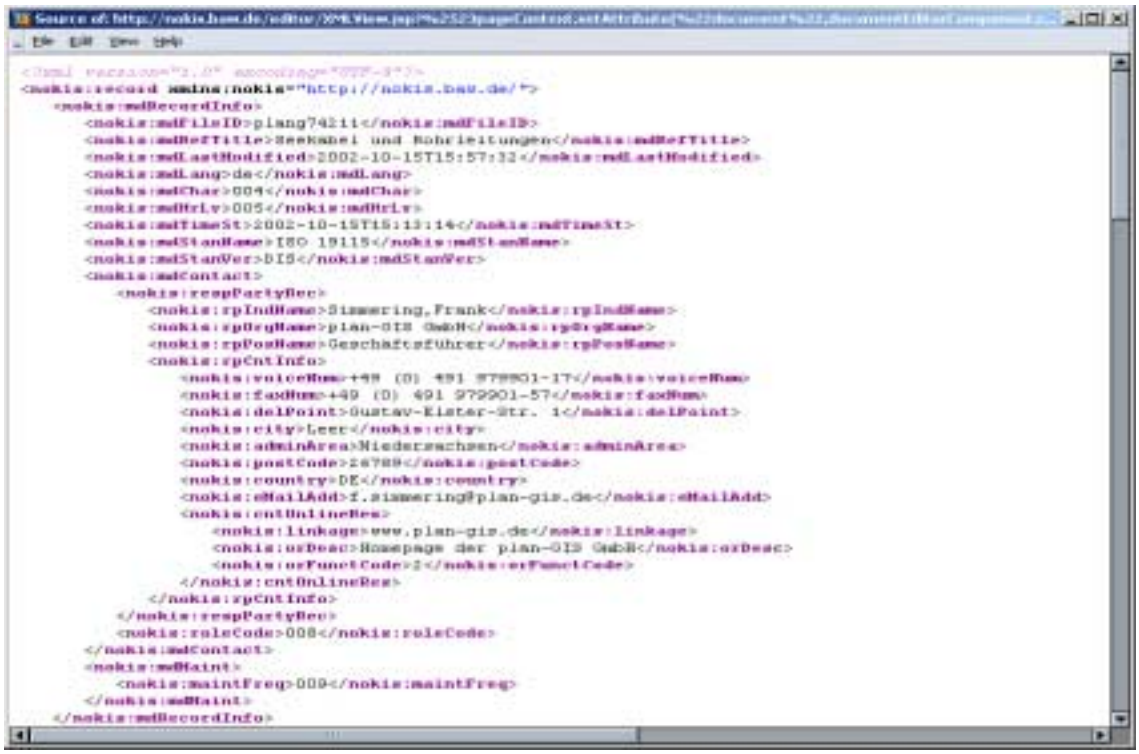


Abbildung 3: XML Sicht des Metadatensatzes

Es besteht ein Mangel an konsistent definierten geographischen Namen, für die Bereiche der Nordsee mit ihren Wattbereichen ebenso wie für die Boddengewässer in der südlichen Ostsee. Der "Getty Thesaurus of Geographic Names" (TGN) [17] enthält nur teilweise Bezeichner für den aquatischen Bereich der Küstenzone und andere verlässliche Gazeteer-Services für die Küstenzone müssen erst noch entstehen.

Referenzsysteme für zeitliche und räumliche Bezeichner sowie für Koordinatensysteme stellen eine wichtige Komponente der Metainformationen dar. Für die Angabe von Koordinatensystemen referenziert NOKIS explizit die Bezeichner der "European Petrol Survey Group" (EPSG) [4].
Durch die im ISO 19115 definierten "Metadata Extensions" war es möglich, das durch das FGDC definierte "Shoreline Profile" [5] für die speziellen Anforderungen der Küstenregion standardkonform miteinzubeziehen.

Volltextindexierung

Die zur Indexierung der Seiten verwendete Software ht://Dig [8] führt drei Schritte aus, um eine Umgebung für eine erfolgreiche Volltextsuche zu erzeugen.
Zunächst erfolgt eine Erfassung aller verlinkten Inhalte innerhalb einer zu spezifizierenden Domain, die in speziellen Suchdatenbanken abgelegt werden. Mit Hilfe dieser Methode ist es möglich, vorhandene Webseiten zu indexieren, wie z.B. die nicht formalisierten Metainformationen zu Forschungsprojekten, die von der GKSS durchgeführt wurden [13]. Weitere Suchdatenbanken wurden für einzelne Forschungsprojekte, wie z.B. das Projekt MorWin [14] und für Datenbanken aus Umweltverträglichkeitsuntersuchungen (UVU Elbe [19]) angelegt.

Nachdem dem Anlegen dieser individuellen Suchdatenbanken erfolgt eine Verschneidung der Ergebnismengen, um eine gemeinsame Suchdatenbank aufzubauen, die bei der textbasierten Suche in NOKIS verwendet wird. Regelmäßige Updates der Suchdatenbanken stellen sicher, dass auch aktuelle Ergebnisse gefunden werden. Analog zu den externen Quellen werden die auf dem NOKIS Server abgelegten standardisierten Metadaten durchsucht und mit den anderen Datenbanken verschnitten, so dass es möglich ist, die dokumentierten Daten über eine Freitextsuche zu erreichen.
Auch wenn die hier beschriebene Volltextindexierung noch nicht die für komplexere Suchmechanismen notwendigen strukturierten Metadaten verwendet, so unterstützt sie doch die von vielen Nutzern verwendete Methode, Datensätze zunächst über die Angabe einiger Stichworte oder Akronyme zu suchen, die den Inhalt oder die räumliche Ausdehnung der Daten einschränken.

Erzeugung von Metadaten

Um dem Anwender beim Erstellen der notwendigen Metadaten zu helfen und um eine gleichbleibende hohe Qualität der Metadaten zu gewährleisten, sind Eingabewerkzeuge erforderlich. Das NOKIS Metadatenmodell ist in XML Schema [22] implementiert und ermöglicht damit die Interoperabilität von Anwendungen und einen verlässlichen Austausch der generierten Informationen.

Automatisch generierte Benutzeroberfläche

Anders als viele andere zur Zeit erhältliche Metadateneditoren ist dieses in Java erstellte Tool eine exakte Abbildung des zugrundeliegenden Metadatenmodells. Es interpretiert direkt das XML Schema, das zur Beschreibung des Metadatenmodells verwendet wird, und erstellt daraus die Benutzeroberfläche. Veränderungen im XML-Schema wirken sich nach einem automatischen Update-Mechanismus unmittelbar auf die Benutzeroberfläche aus.
Diese automatisierte Methode der Oberflächengenerierung [10] vermeidet den zeitraubenden Prozess der Anpassung von Menüs und Formularen bei einer Änderung des Metadatenmodells. Bei jeder lokalen Installation des Editors kann eine beliebige Anzahl einfacher und komplexer proprietärer Metadatenelemente hinzugefügt werden, um so das Tool auch zu einem vollwertigen Dokumentations- und Recherchetool für alle lokalen Anwender zu machen.

Die lokale Bereitstellung des NOKIS Metadateneditors erfolgt über eine Installations-CD, die bei der Projektleitung angefordert werden kann. Die rein in Java implementierte Client/Server-Applikation beruht auf den OSS-Komponeneten Jakarta Tomcat, Saxon und Xerces.
Das in Abbildung 2 gezeigte Menü bildet in der linken Navigationsleiste direkt die in Tabelle 2 genannten einzelnen Abschnitte des ISO 19115 ab (Metadateninformationen, Dateninformationen, usw.). Das dargestellte Formular enthält Freitextfelder, als Dropdown-Listen dargestellte Codelisten und andere formatierte Elemente, die gemäß andern internationalen Standards aufgebaut sind.

Für die Übernahme bereits vorhandener Metadaten wurden Tools entwickelt, mit denen sich aus den ESRI-Produkten (ArcView 3.x, ArcINFO, ArcGIS 8.x) heraus NOKIS-konforme Metadaten exportieren lassen.

Die erfassten Metadaten werden in einem relationalen Datenbanksystem gespeichert (PostgreSQL), das für alle gängigen Betriebssysteme als OSS zur Verfügung steht. Aus diesen Metadatenbeständen lassen sich mit Hilfe von XML Stylesheet-Transformations (XSLT [23]) beliebige Darstellungen der Daten erzeugen. Abbildung 3 zeigt das Fragment einer XML-Datei, die gleiche Daten enthält, welche auch im Editorinterface in Abbildung 2 gezeigt werden.

Metadatenreplikation

Der zentrale Server mit dem Metadatenbestand von NOKIS wird bei der Bundesanstalt für Wasserbau in Hamburg betrieben. Wann immer bei den teilnehmenden Partnern neue Metainformationen erzeugt worden sind, werden dort mit dem Exporttool des Editors XML-Dateien erzeugt, die nur die Standardelemente des zentralen Metadatenschemas enthalten. Dadurch werden alle lokalen Erweiterungen des Schemas ausgefiltert und es wird vermieden, dass diese öffentlich sichtbar werden.

Die komprimierten Datenbestände können auf verschiedenste Weise zum zentralen Server gebracht werden, standardmäßig ist FTP vorgesehen. Sobald die Daten auf dem Server eingespielt sind, werden sie automatisch in die zentrale Metadatenbank aufgenommen und sind danach sofort per Internet recherchierbar. Für die lokalen Installationen ist ein Administrationstool in Entwicklung, welcher den Prozess der Metadatenerzeugung und -veröffentlichung administrativ und qualitätssichernd begleiten soll.

Export von Metadaten

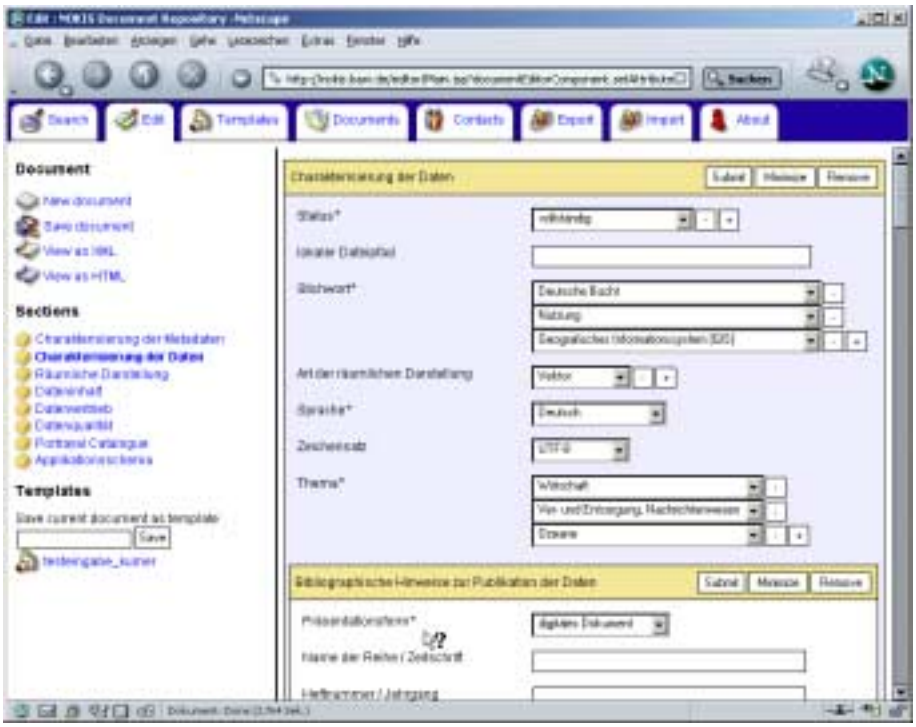


Abbildung 4: Kontextsensitives Hilfesystem

In Koordination mit der Koordinierungsstelle Umweltdatenkatalog/GEIN im niedersächsischen Umweltministerium wurde ein Exportfilter entwickelt, der es erlaubt Metadaten zu exportieren, die der semantischen XML-Schnittstelle des Umweltdatenkataloges entsprechen. Auf diese Weise wird den Bedürfnissen des überwiegenden Teils der NOKIS Partner entsprochen, die dem UDK gegenüber berichtspflichtig sind.

Kontextsensitive Hilfe

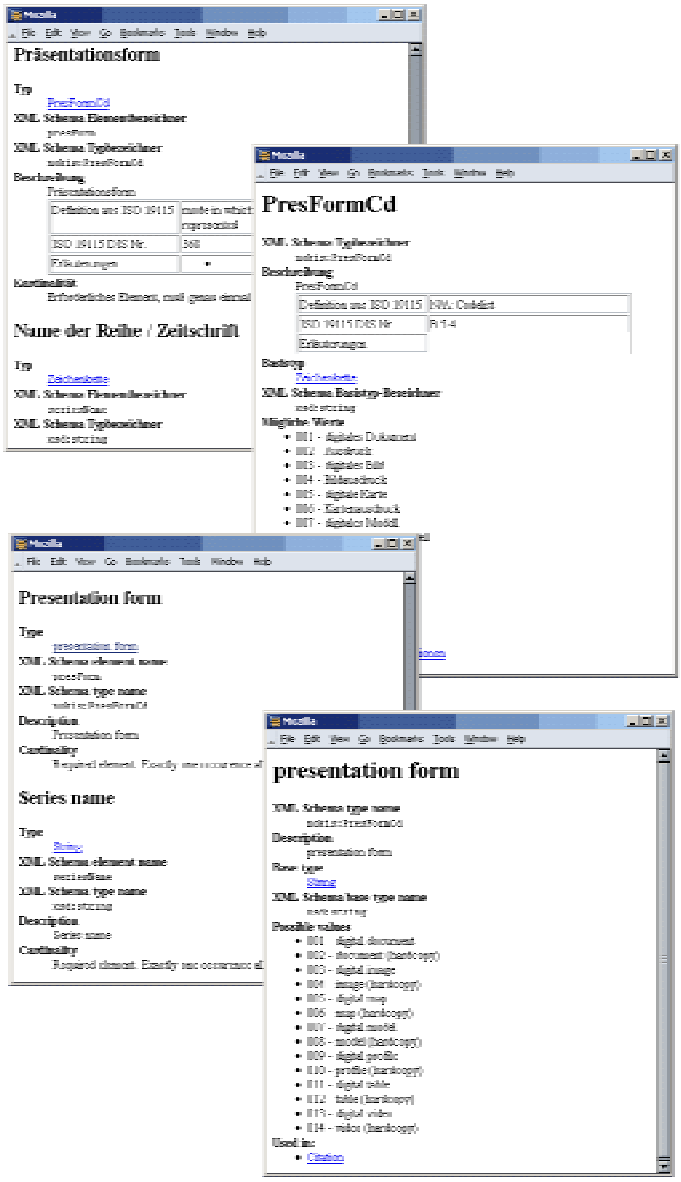


Abbildung 5: Elemente des Hilfesystems

Ein auf die Erfahrungen der bisherigen Prototypnutzer aufbauendes umfassendes Hilfesystem ist Teil des Metadateneditors. Zu jedem einzelnen Metadatenelement kann eine kontextsensitive Hilfe aufgerufen werden (siehe Abbildung 4), die neben allgemeinen Informationen zu diesem Element auch Hilfestellungen zur konkreten Belegung dieses Elements anbietet.

Der ISO 19115 beruht auf einer Vielzahl von Codelisten, deren Werte ebenfalls wie in Abbildung 5 dargestellt in der Hilfe angezeigt werden können.
Basierend auf der eingestellten Sprachpräferenz des Webbrowsers wird die gesamte Oberfläche in der entsprechenden Sprache angeboten, sofern nötigen Sprachpakete vorhanden sind. Während der Projektphase werden lediglich Deutsch und Englisch implementiert; ein Hinzufügen weiterer Sprachen ist jederzeit durch das Bereitstellen einer entsprechenden XML-Datei mit fremdsprachlichen Einträgen zu allen Elementen des Standards und der Benutzeroberfläche möglich.

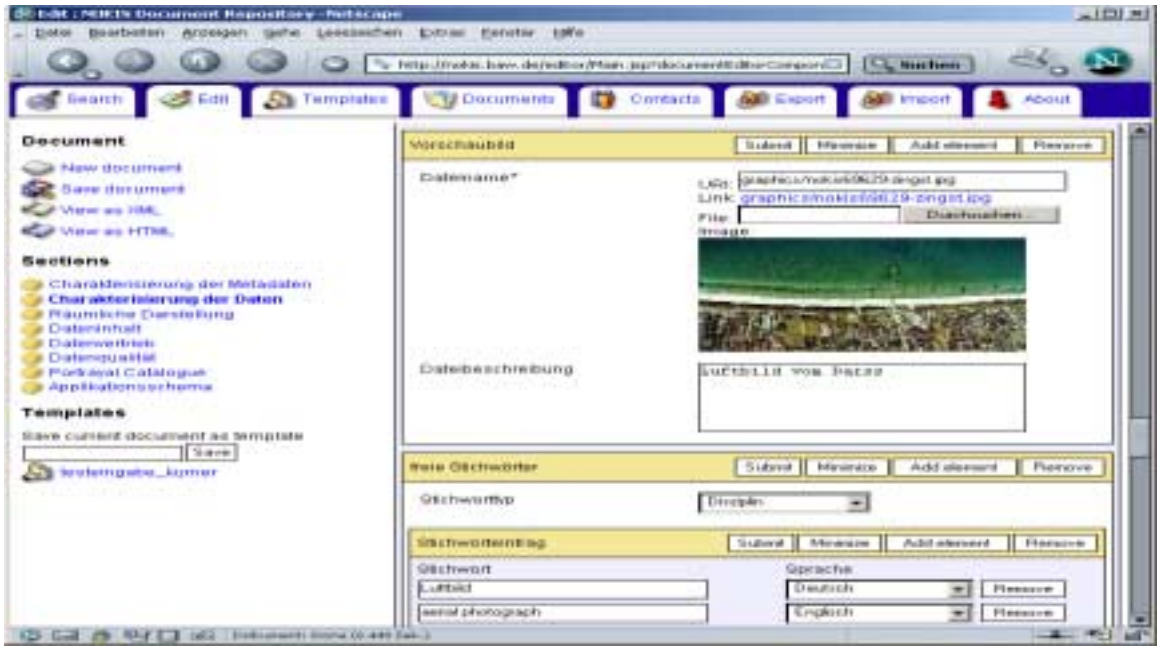


Abbildung 6: Vorschaubilder im Editorinterface

Nutzung von Metadaten

Die in den Partnerinstitutionen erstellten und gepflegten Metadaten dienen als lokale Kataloge zur Dokumentation der verschiedenen Datentypen und -archive.
Die Ansicht der kompletten Metadaten ist mit dem Editor möglich und in Kombination mit einem Mapserver können, die geographischen Elemente als Vorauswahl benutzt werden.

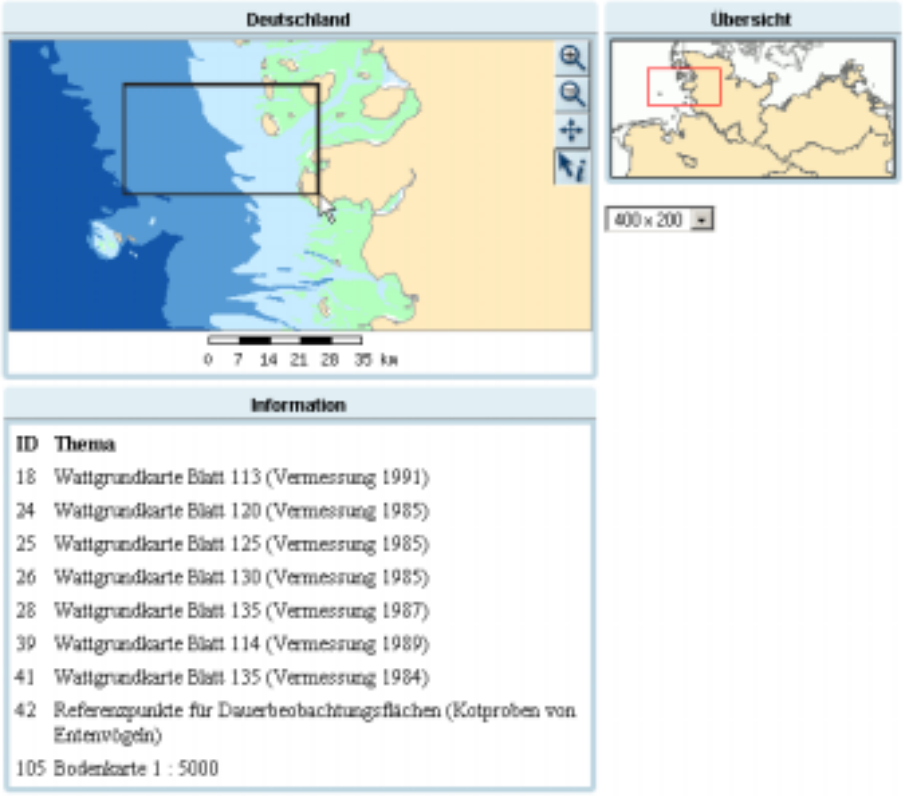


Abbildung 7: NOKIS Mapserver

Vorschaubilder

Abbildung 6 zeigt die eingegebenen Metadaten zur Beschreibung eines Luftbildes. Als Metainformation ist dort auch ein Vorschaubild enthalten. Es gibt eine große Anzahl von Datentypen, die sich gut mit einem Vorschaubild beschreiben lassen, um so bereits einen schnellen Eindruck über die Daten zu vermitteln.

Beispielsweise lässt sich die Qualität von Zeitreihen hydrologischer oder meteorologischer Parameter (z.B. Wind, Wasserstand, Strömungsgeschwindigkeit, usw.) rasch durch eine Betrachtung einer Visualisierung der Daten abschätzen. Auch das von einer flächenhaften Untersuchung des Meeres erfasste Gebiet lässt sich durch eine schematische Darstellung des Gebietes in einer Karte gut einschätzen.

Mapserver

Abbildung 7 zeigt die kartenbasierte Auswahl von Datensätzen aus der Datenbank. Entsprechend der in den Datensätzen angegebenen Bounding Box oder dem geographischen Namen werden alle innerhalb des aufgezogenen Rechtecks liegenden Datensätze ausgegeben und können dann selektiert werden. In NOKIS wird eine Implementierung des OSS Mapservers der University of Minnesota [18] verwendet, der über die Editor-Software mit der Datenbank kommuniziert. Suchabfragen nutzen eine Kombination von kartenbasierter Suche und weiteren Einschränkungen über das textbasierte Interface.

Dieser dynamische Mapserver ermöglicht ein stufenloses Zoomen in der Karte, was es dem Benutzer erlaubt, eine beliebige Maßstabsstufe zur Eingabe seiner Bounding Box zu wählen. Als hauptsächliche Kartengrundlage wird hier eine detaillierte Tiefenkarte verwendet werden, die aus dem Bereich der numerischen Modellierung stammt.

Schlussfolgerungen

Der Metadatenstandard ISO 19115 stellt eine breite Basis an Metadatenelementen aus den Bereichen der bibliographischen und technischen Information zur Verfügung, aus denen ein handhabbarer Kern von 19 Pflichtelementen zur Beschreibung von Daten aus dem Küstenbereich ausgewählt wurde.
Dieser "NOKIS-Standard" bietet Informationen zu räumlicher und zeitlicher Ausdehnung und charakterisiert die Datensätze durch Stichwörter und eine Kurzbeschreibung als Grundlagen für eine systematische Suche.

Die Zusammenstellung dieser Metainformationen wird durch einen Metadateneditor erleichtert, der als lokales Dokumentations- und Suchwerkzeug verwendet werden kann und der dazu karten- wie textbasierte Suchmechanismen anbietet.
Die mehrsprachige Editoroberfläche wird automatisch aus dem zugrundeliegenden XML-Schema generiert und bietet eine an die Spracheinstellungen des Editors angepasste Oberfläche.

Beim Umgang mit dem umfangreichen Metadatenstandard wird dem Nutzer durch ein kontextsensitives Hilfesystem Unterstützung angeboten.
In der momentanen Pilotphase werden bei den Partnern aus Bundes- und Landesdienststellen Metadatenbanken für den lokalen Gebrauch erstellt. Diese werden durch die Replikation auf dem zentralen Server, http://nokis.baw.de, auf dem auch Dienste für das Auffinden von und den Zugang zu Daten und zugehörigen Publikationen angeboten werden, im Internet recherchierbar.

Danksagung

Die Förderung des Projektes NOKIS erfolgt für drei Jahre durch das Bundesministerium für Bildung und Forschung, vertreten durch den Projektträger Jülich unter der Fördernummer 03KIS027 NOKIS

Erfahrungen mit ISO 19115
Vorführung und Stand NOKIS

Carsten Heidmann, BAW Hamburg
Rainer Lehfeldt, BAW Hamburg
Wassilios Kazakos, FZI Karlsruhe

NOKIS@AK Umweltdatenbanken

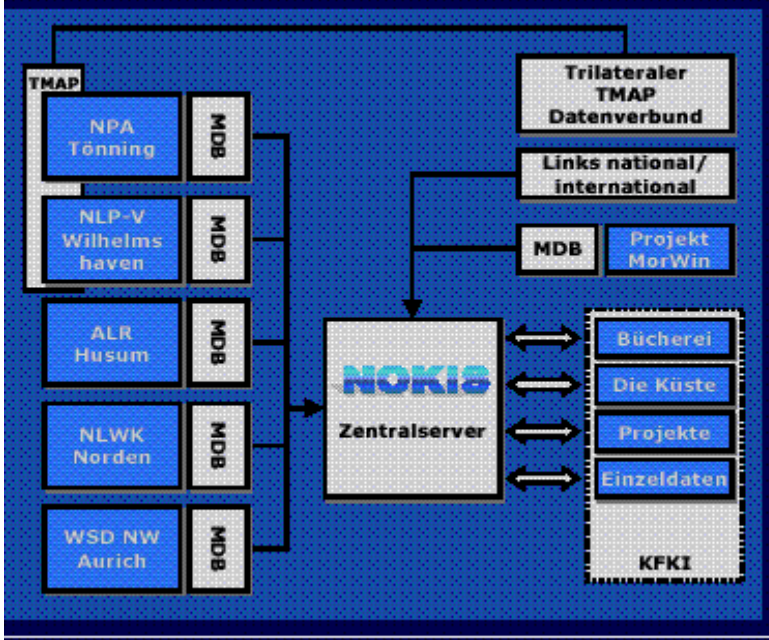
- Stand des Projektes
- Projektpartner und Datentypen / Datenaufkommen
- Struktur des Projektes
- Softwarekonzept
- Stand der Editorentwicklung
- Map Server
- Vorführung des Editors

Stand des Projektes

- Projektlaufzeit 2001 - 2004
- 5 Partner bei Projektbeginn
  + BAW (PDBK)
  + GKSS (LOTSE)
- Softwareerstellung für MD-Editor nahezu abgeschlossen
- spätestens Ende des Jahres voller Probebetrieb auf Zentralserver

NOKIS AK Umweltdatenbanken 2003 Berlin

NOKIS Struktur



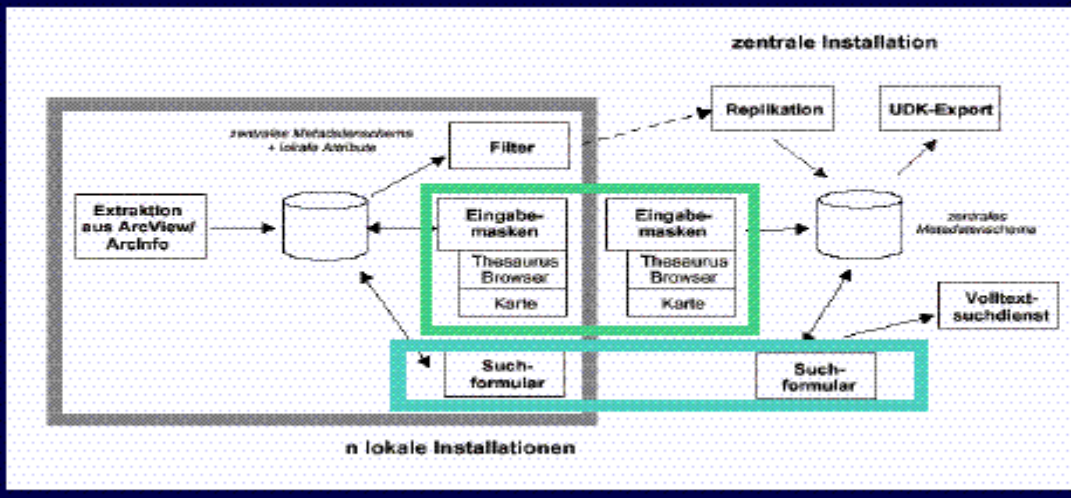
Metadaten zu:

- GIS-Daten
- Zeitreihen
- KFKI-Projekten
- "Die Küste"
- Einzeldaten

NOKIS AK Umweltdatenbanken 2003 Berlin Folie 4

Software- und Systemkonzept:

⇨ identisches Erscheinungsbild lokal und global



NOKIS AK Umweltdatenbanken 2003 Berlin

Metadata Standards

ISO 19115 metadata entity set information	Metadata Elements
MD_Metadata	Metadata file identifier
	Metadata standard name
	Metadata standard version
	Metadata language
	Metadata character set
	Metadata point of contact
	Metadata time stamp
	Scope to which the metadata applies
MD_Identification	**Dataset title**
	Dataset reference date
	Dataset responsible party
	Abstract describing the dataset
	Descriptive keywords
	Status
MD_DataIdentification	Spatial representation type
	Spatial resolution of the dataset
	Dataset language
	Dataset character set
	Dataset topic category
	Geographic location
	bounding box or geographic identifier
	Vertical and temporal extent of dataset
MD_Constraints	Name of handling restriction
DQ_DataQuality	Scope of data
	Lineage Statement
MD_MaintenanceInformation	Maintenance frequency
MD_SpatialRepresentation	Grid spatial representation
	Vector spatial representation
MD_ReferenceSystem	Reference system
MD_ContentInformation	External feature catalogue
	Coverage content
	Image description
MD_PortrayalCatalogueReferenc	Portrayal catalogue citation
MD_Distribution	Distribution format
	On-line ressource
MD_MetadataExtension	Information on metadata extensions
MD_ApplicationSchemaInformatic	Application schema information

- **ISO mandatory**
- **ISO recommended core**
- **ISO NOKIS**

n 2003 Berlin Folie 7

Metadata Standards

set information	
MD_Metadata	**Metadata file identifier**
	Metadata standard name
	Metadata standard version
	Metadata language
	Metadata character set
	Metadata point of contact
	Metadata time stamp
	Scope to which the metadata applies
MD_Identification	**Dataset title**
	Dataset reference date
	Dataset responsible party
	Abstract describing the dataset
	Descriptive keywords
	Status
MD_DataIdentification	**Spatial representation type**
	Spatial resolution of the dataset
	Dataset language
	Dataset character set
	Dataset topic category
	Geographic location
	bounding box or geographic identifier
	Vertical and temporal extent of datase
MD_Constraints	Name of handling restriction
DQ_DataQuality	Scope of data
	Lineage Statement
MD_MaintenanceInformation	Maintenance frequency
MD_SpatialRepresentation	Grid spatial representation
	Vector spatial representation
MD_ReferenceSystem	**Reference system**
MD_ContentInformation	External feature catalogue
	Coverage content
	Image descrption
MD_PortrayalCatalogueReferenc	Portrayal catalogue citation
MD_Distribution	**Distribution format**
	On-line ressource
MD_MetadataExtension	Information on metadata extensions
MD_ApplicationSchemaInformatio	Application schema information

- **ISO mandatory**
- **ISO recommended core**
- **ISO NOKIS**

2003 Berlin

Folie

Metadata Standards

set information	
MD_Metadata	**Metadata file identifier**
	Metadata standard name
	Metadata standard version
	Metadata language
	Metadata character set
	Metadata point of contact
	Metadata time stamp
	Scope to which the metadata applies
MD_Identification	**Dataset title**
	Dataset reference date
	Dataset responsible party
	Abstract describing the dataset
	Descriptive keywords
	Status
MD_DataIdentification	**Spatial representation type**
	Spatial resolution of the dataset
	Dataset language
	Dataset character set
	Dataset topic category
	Geographic location
	bounding box or geographic identifier
	Vertical and temporal extent of datase
MD_Constraints	Name of handling restriction
DQ_DataQuality	Scope of data
	Lineage Statement
MD_MaintenanceInformation	Maintenance frequency
MD_SpatialRepresentation	Grid spatial representation
	Vector spatial representation
MD_ReferenceSystem	**Reference system**
MD_ContentInformation	External feature catalogue
	Coverage content
	Image descrption
MD_PortrayalCatalogueReferenc	Portrayal catalogue citation
MD_Distribution	**Distribution format**
	On-line ressource
MD_MetadataExtension	Information on metadata extensions
MD_ApplicationSchemaInformatio	Application schema information

- **ISO mandatory**
- **ISO recommended core**
- **ISO NOKIS**

2003 Berlin

Folie

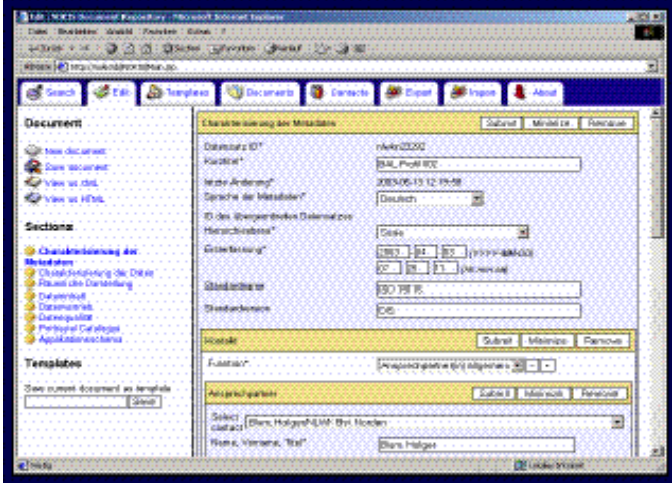
Metadata Standards

ISO 19115 metadata entity set information	Metadata Elements
MD_Metadata	Metadata file identifier
	Metadata standard name
	Metadata standard version
	Metadata language
	Metadata character set
	Metadata point of contact
	Metadata time stamp
	Scope to which the metadata applies
MD_Identification	Dataset title
	Dataset reference date
	Dataset responsible party
	Abstract describing the dataset
	Descriptive keywords
	Status
MD_DataIdentification	Spatial representation type
	Spatial resolution of the dataset
	Dataset language
	Dataset character set
	Dataset topic category
	Geographic location
	bounding box or geographic identifier
	Vertical and temporal extent of datase
MD_Constraints	Name of handling restriction
DQ_DataQuality	Scope of data
	Lineage Statement
MD_MaintenanceInformation	Maintenance frequency
MD_SpatialRepresentation	Grid spatial representation
	Vector spatial representation
MD_ReferenceSystem	Reference system
MD_ContentInformation	External feature catalogue
	Coverage content
	Image descrption
MD_PortrayalCatalogueReferenc	Portrayal catalogue citation
MD_Distribution	Distribution format
	On-line ressource
MD_MetadataExtension	Information on metadata extensions
MD_ApplicationSchemaInformatio	Application schema information

- ISO mandatory
- ISO recommended core
- ISO NOKIS

NOKIS 2003 Berlin Folie 9

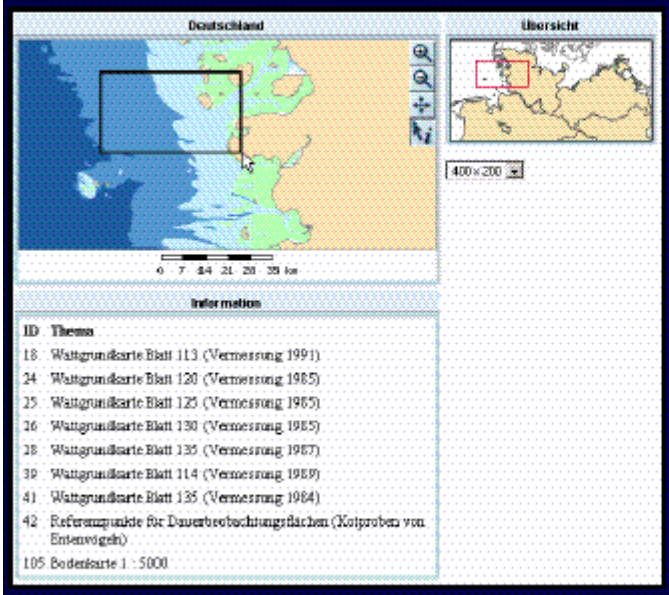
Stand des Editors



- komplett in Java
- XML-basiert
- automatisch generiert
- Mehrsprachigkeit
- PostgreSQL als DB-Backend

NOKIS AK Umweltdatenbanken 2003 Berlin Folie 10

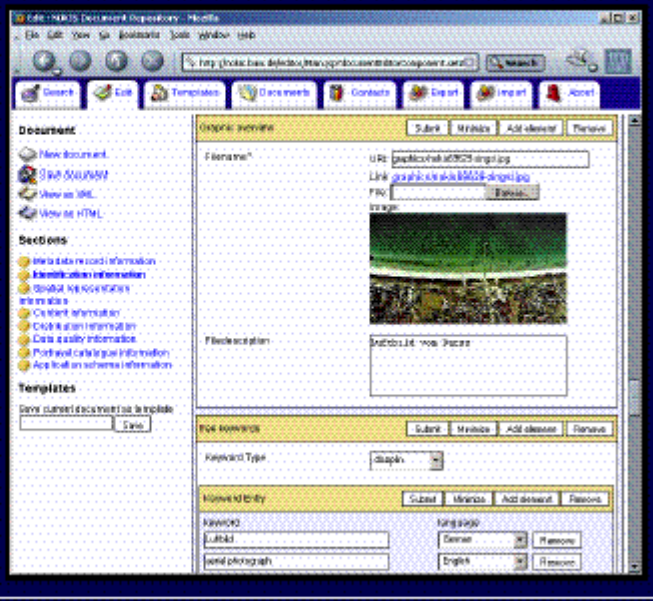
Editor und Mapserver



- Mapserver für Metadaten-erfassung und Metadateneingabe
- UMN Mapserver
- OSS mit grosser Entwickler- und Anwendergemeinde

Editor und Mapserver

- Verwendung von Thumbnails zur besseren Charakterisierung der Daten



Vorführung des Editors

- Abbildung von ISO 19115
- XML Basis
- Funktionalität des Editors
- Views/Export
- Integration Mapserver
- Erweiterbarkeit

NOKIS AK Umweltdatenbanken 2003 Berlin

Schluss

nokis.baw.de

NOKIS AK Umweltdatenbanken 2003 Berlin

Referenzen

[1] **Burbridge, P.R. (1999)** The Guiding Principles for a European ICZM Strategy. Retrieved May 12, 2003, von der zentralen Website des “Integrated Coastal Zone Management (ICZM)”: http://europa.eu.int/comm/environment/iczm/burbridge.pdf.

[2] **Dublin Core Metadata Initiative (2002).** Dublin Core Metadata Element Set, Version 1.1: Reference Description. zuletzt besucht am 10. Mai 2003, von der DCMI Website: http://dublincore.org.

[3] **Environment Directorate-General of the EU (2003).** Introduction to the new EU Water Frame-work Directive. zuletzt besucht am 10. Mai 2003, URL: http://europa.eu.int/comm/enviroment/water/waterframework/overview.html.

[4] **European Petroleum Survey Group (2002).** zuletzt besucht am 10. Mai 2003. URL: http://www.epsg.org/.

[5] **Federal Geographic Data Committee (2001)** Shoreline Metadata Profile. zuletzt besucht am 10. Mai 2003, von der FGDC Website. URL: http://www.csc.noaa.gov/metadata/sprofile.pdf.

[6] **Java Servlet Technology von der Website von Sun Microsystems.** zuletzt besucht am 10. Mai 2003, URL: http://java.sun.com/products/servlet/

[7] **KFKI (2000).** Die Aufgabe des KFKI. zuletzt besucht am 10. Mai 2003, URL: http://kfki.baw.de/general/aufgabe.de.jsp

[8] **The ht://Dig Group (2002).** How it works. zuletzt besucht am 10. Mai 2003, URL: http://www.htdig.org/.

[9] **ISO 2002. ISO 19115:** Geographic information – Metadata. Draft International Standard. zuletzt besucht am 10. Mai 2003, von der International Organization for Standardization. URL: http://www.iso.ch/iso/.

[10] **Kazakos W., Valikov A., Schmidt A., Lehfeldt R. (2002)** Automation of Metadata Repository Development with XML Schema. In: Pilmann, W. and Tochtermann, K. (Hrsg.), Environmental Communication in the Information Society. 16th International Conference Informatics for Environmental Protection, Volume I, Seite 400 - 407, Wien, Oktober 2002. ISEP.

[11] **Lehfeldt, R., Heidmann, C. (2002),** NOKIS – a metadata information system for German coastal regions of the North Sea and Baltic Sea, in RA Falconer, B Lin, EL Harris, CAME Wilson, ID Cluckie, D Han, JP Davis, S Heslop (Hrsg.) Hydroinformatics 2002. Proc. 5th Intl. Conf., Cardiff. IWA Publishing, London, Seite 1507-1512.

[12] **Lehfeldt , R., Heidmann, C., Piasecki, M. (2002),** Metadata in Coastal Information Systems, in Holz, KP, Kawahara, M, Wang, SY (eds) Advances in Hydro-Science and –Engineering Volume 5. Proceedings of the 5th International Conference on Hydro-Science and -Engineering, Warsaw. Abstract Volume, Seite 181.

[13] **LOTSE Land Ocean Thematic Search Engine (2000).** zuletzt besucht am 10. Mai 2003, von der GKSS Website. URL: http://w3k.gkss.de/lotse/.

[14] **MorWin: Morphodynamic Modeling of Wind Influenced Flats -** Internet Based Collaborative Project Handling in Coastal Engineering (2000) zuletzt besucht am 10. Mai 2003, von der MorWin Projekt Website. URL: http://morwin.baw.de/.

[15] **Nebert, D.D. (2001)** Developing Spatial Data Infrastructures: The SDI Cookbook. Global Spatial Data Infrastructure Technical Working Group. zuletzt besucht am 10. Mai 2003, URL: http://www.gsdi.org/pubs/cookbook/cookbook0515.pdf.

[16] **Nord- und Ostsee-Küsteninformationssystem (2001).** zuletzt besucht am 10. Mai 2003, von der NOKIS Website. URL: http://nokis.baw.de/abstract/index-en.jsp.

[17] **Getty Thesaurus of Geographic Names (2000).** zuletzt besucht am 10. Mai 2003. URL: http://www.getty.edu/re-search/tools/vocabulary/tgn/index.html.

[18] University of Minnesota Mapserver. zuletzt besucht am 27. Mai 2003, URL: http://mapserver.gis.umn.edu/

[19] **UVU- und Beweissicherungsdatenbank** (2002) Anpassung der Unter- und Außenelbe an die Containerschiffahrt. zuletzt besucht am 10. Mai 2003, von der Wasser- und Schifffahrtsamt Cuxhaven Website. URL: http://www.cux.wsd-nord.de/htm/Start.asp.

[20] **Vallega, A. (1999).** Fundamentals of Integrated Coastal Management. Kluwer Academic Publishers, Dordrecht.

[21] **Web-OPAC. zuletzt besucht am 10. Mai 2003,** URL: http://opc4.tib.uni-hannover.de:8080/

[22]**World Wide Web Consortium (W3C):** XML-Schema. von der W3C Website. zuletzt besucht am 23. Mai 2003, URL: http://www.w3.org/XML/Schema

[23]**World Wide Web Consortium (W3C)**: XML Stylesheet Language. von der W3C Website. zuletzt besucht am 23. Mai 2003, URL: http://www.w3.org/Style/XSL/

Triggerbasierte Replikationsmechanismen zum Austausch von Grundwasserinformationen in heterogenen Datenbankumgebungen

Prof. Dr.-Ing. Uwe Rüppel

Dipl.-Ing. Thomas Gutzke

Institut für Numerische Methoden und Informatik im Bauwesen,

Technische Universität Darmstadt

Dr.-Ing. Michael Petersen

Dipl.-Ing. Gerrit Seewald

CIP - Computer Integrierte Planung –
Ingenieurgesellschaft mbH, Darmstadt

Abstract

Das Pilotprojekt “Grundwasser-Online“ verfolgt das Ziel einer zeitnahen, unternehmensübergreifenden Grundwassersteuerung. Im Projektgebiet – dem Hessischen Ried – kooperieren zu diesem Zweck sechs große ansässige Wasserversorgungsunternehmen (WVU) gemeinsam mit dem Regierungspräsidium in Darmstadt. Das von der TU Darmstadt und den Büros CIP Ingenieurges. mbH und BGS Umweltplanung GmbH initiierte und geleitete Projekt umfasst dabei Module zur vernetzten Datenerfassung innerhalb der einzelnen Unternehmen sowie den Abgleich dieser lokalen Datenbestände mit einer zentralen Serverdatenbank über trigger- und protokollbasierte Replikationsmechanismen. Über feingranulare Zugriffsrechte sind berechtigte Personen in der Lage, über das Internet und ohne räumliche oder zeitliche Einschränkungen auf diesen unternehmensübergreifenden Datenbestand zuzugreifen.

Zur Auswertung dieser Daten werden internetbasierte Aufbereitungs- und Visualisierungsmodule zur Verfügung gestellt, die den Anwender bei der Bewertung und Steuerung der Grundwassersituation fachgerecht unterstützen.

Einführung in das Pilotprojekt "Grundwasser-Online"

Beschreibung des Projektgebiets

Das Hessische Ried ist eines von Deutschlands größten und zugleich hydrogeologisch kritischsten Grundwasserreservoirs. Bei dem Projektgebiet handelt es sich um den nördlichen Teil des Rheingrabens, dem Oberrheingraben. In dieser Region leben ca. 800.000 Menschen auf einer Fläche von 1238 km² [Abb. 1]. Das Gebiet liegt zwischen den dichten Besiedlungsräumen der Rhein-Main- und der Rhein-Neckar-Region und wird durch die hydrogeologischen Grenzen des Mains im Norden, des Neckars im Süden, des Rheins im Westen und des Odenwalds, der als natürliche Wasserscheide fungiert, im Osten begrenzt.

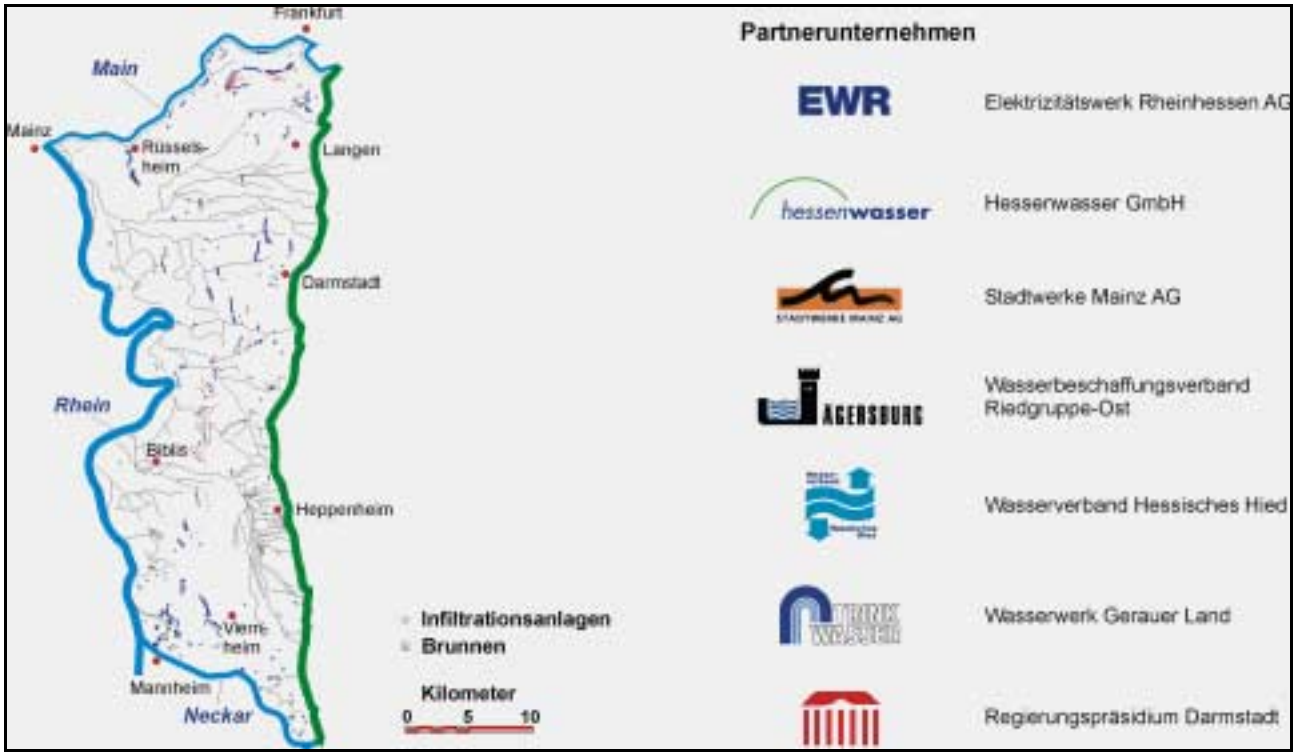


Abb. 1: Projektgebiet "Hessisches Ried"

Ausgangssituation

In Folge stark schwankender Grundwasserstände in den vergangenen Jahrzehnten [Abb. 2] kam es in vielen Bereichen des Hessischen Rieds zu Setzungs- aber auch zu Vernässungsschäden. Die Möglichkeiten der Behörden und der Wasserversorger, auf die Grundwassersituation steuernd Einfluss zu nehmen, waren und sind auf Grund des komplexen hydrogeologischen und wasserwirtschaftlichen Zusammenspiels, der großen Anzahl an beteiligten Unternehmen und nicht zuletzt der inhomogenen und inkonsistenten Datengrundlage und der zeitverzögerten Datenweiterleitung stark eingeschränkt [5].

Der derzeitige Informationsfluss zwischen Wasserwerken und Behörden umfasst vom Zeitpunkt der Erfassung von Grundwassermessständen im Feld, über die Eingabe der Daten in die lokalen Verwaltungsapplikationen der einzelnen WVU, bis zur internen Auswertung und Weiterleitung an die zuständigen Überwachungsbehörden im Schnitt eine Dauer von drei bis sechs Monate. Das Regierungspräsidium als Genehmigungs- und Überwachungsbehörde ist bestrebt, frühzeitig auf aktuelle Entwicklungen reagieren zu können. Die Einflussnahme auf die Grundwasserstandsentwicklung ist von der Natur der Sache her allerdings eingeschränkt und erfordert langwierige Maßnahmen. Im Hessischen Ried erfolgt dies z.B. über die Erhöhung oder Absenkung der Einleitungsmengen in den Grundwasserleiter über die vorhandenen und zum Teil noch in Planung befindlichen Rheinwasseraufbereitungs- und Infiltrationsanlagen.

Die Datengrundlage des Regierungspräsidiums besteht aus den in regelmäßigen Intervallen von den WVU zu erstellenden Monitoringberichten in Papierform. Eine raum- und zeitübergreifende Suche sowie die Aus- und Bewertung der Grundwassersituation auf dieser Datengrundlage ist zeitaufwendig und komplex.

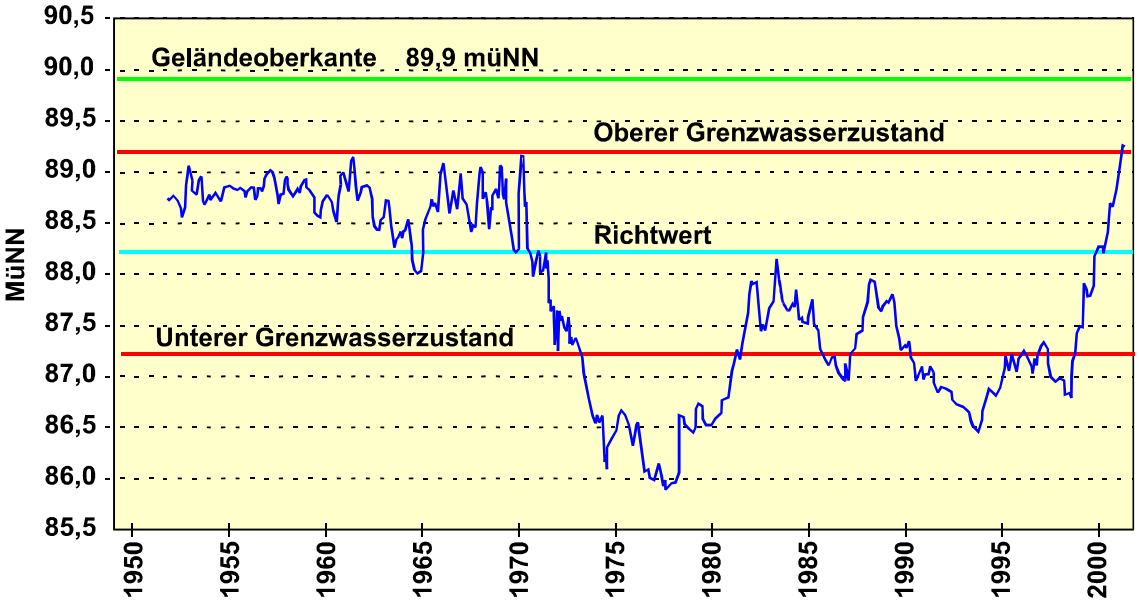


Abb. 2: Charakteristische Grundwasserentwicklung (1950 bis 2003)

Projektziele

Ein zentrales Ziel des Pilotprojekts Grundwasser-Online ist es, die anfallenden Rohdaten bei den WVU aber auch den Behörden vor Ort fachgerecht und in einem einheitlichen Format zu erfassen, um in einem zweiten Schritt diese dezentralen Datenbestände in einen gemeinsamen, zentralen Datenpool zu überführen. Um diese Informationen wiederum unternehmensübergreifend den berechtigten Fachplanern zeitnah und ortsunabhängig zur Verfügung zu stellen, werden geeignete internetbasierte Technologien eingesetzt. Zur Analyse und Bewertung der vorliegenden Grundwasserinformationen werden den dazu berechtigten Anwendern geeignete Werkzeuge zur Verfügung gestellt.

Neben der Verbesserung der Workflowprozesse zwischen den beteiligten Unternehmen und Behörden werden auch die internen Firmenstrukturen sowie die Arbeitsabläufe neu strukturiert und somit deren Effizienz gesteigert. Dies wird durch den Wegfall von manuellen Datenaufbereitungen und -übergaben, sowohl firmenintern als auch unternehmensübergreifend, erreicht. Hierdurch lassen sich Arbeitszeitaufwand und damit Kosten für alle beteiligten Unternehmen - bei gleichzeitig qualitativ veredelter Datengrundlage - reduzieren.

Auf Grund der seit Jahrzehnten kritischen Grundwassersituation hat die betroffene Öffentlichkeit großes Interesse an aktuellen Grundwasserständen und spezifischen Hintergrundinformationen. Der Öffentlichkeit soll über ein frei zugängliches Internetportal ein benutzerfreundlicher Zugang bereitgestellt werden, um durch ein weitreichendes Informationsangebot ein umfassenderes Verständnis der einzelnen Zusammenhänge zu erhalten [1].

Systemarchitektur

Aufbauend auf den zu Beginn des Pilotprojekts analysierten, heterogenen Firmenstrukturen und den durch geografische Randbedingungen sehr unterschiedlichen Internetanbindungen, ergab sich die Notwendigkeit, ein kombiniertes Online-/Offline-System zu konzipieren [Abb. 3].

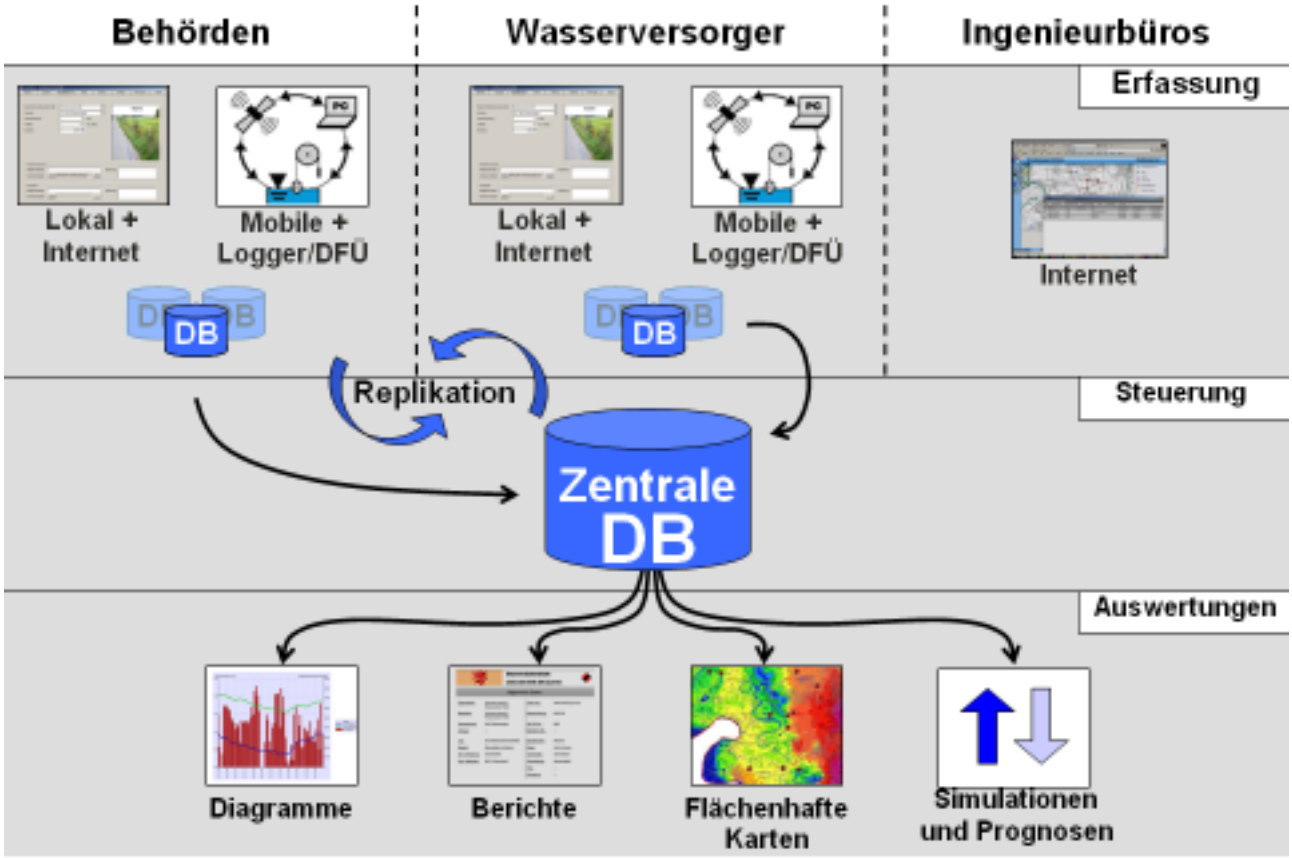


Abb. 3: Systemarchitektur von Grundwasser-Online

Lokale, mobile und internetbasierte Datenerfassung

Die Erfassung der Grundwasserdaten sowie die firmeninterne Weiterverarbeitung erfolgt dabei in der Regel über bereitgestellte, lokale Eingabemodule. Dabei werden die Daten in eine im Intranet des jeweiligen Unternehmens aufgesetzte, lokale Datenbank abgelegt. Die Datenerfassung kann bei Bedarf auch über Mobile Endgeräte erfolgen, über die Grundwasserstandsdaten direkt im Feld erfasst, auf Plausibilität überprüft und in die lokale Datenbank eingespeist werden können. Die Integration von Datenloggern mit eingebauter DFÜ-Schnittstelle wird derzeit geprüft. Um bei der Dateneingabe nicht auf einen Arbeitsplatzrechner mit der installierten Software angewiesen zu sein, werden zusätzlich Active Server Pages (ASP) bereitgestellt, die eine zeitnahe und räumlich flexible Erfassung sicherstellen. Insbesondere bei den ehrenamtlich tätigen Ablesern des Landesgrundwasserdienstes kann dies zu deutlichen Erleichterungen bei der Datenweitergabe führen. Diese Daten werden dann direkt in der zentralen Serverdatenbank gespeichert.

Zugangsberechtigungen

Bei den Grundwasserständen, Förder- und Infiltrationsmengen handelt es sich zum Teil um Inhalte, die vor Zugriffen Dritter zu schützen sind. Über feingranulare definierbare Zugriffsberechtigungen können administrativ tätige Personen eines Wasserwerks einerseits die firmeninternen aber auch die Berechtigungen an externe Unternehmen erteilen. Die Vergabe der Zugriffsrechte basiert auf einem Mehrstufensystem:

Auf der obersten Stufe wird durch den Systemadministrator von Grundwasser-Online für jedes beteiligte Unternehmen eine eigene Betreiber-Administratorgruppe angelegt. Diese Betreiber-Administratorgruppe ist wiederum berechtigt, die Abbildung der firmeninternen Strukturen vorzunehmen, indem sie alle erforderlichen, internen Benutzergruppen anlegt und die entsprechenden Mitarbeiter diesen Gruppen zuordnet. Beim Anlegen der einzelnen Gruppen werden diesen Gruppen sogenannte Basisrechte zugewiesen. Basisrechte definieren beispielsweise, welche Funktionen die Personen einer Gruppe besitzen, d.h. ob sie nur Daten eingeben können, die verschiedenen Auswertungsmodule aufrufen dürfen.

Weiterhin wird über diese Basisrechte definiert, welche Rechte eine Gruppe auf die einzelnen Objekte (Brunnen, Grundwassermessstellen etc.) erhält, die ein Unternehmen besitzt. Getrennt für Stamm- und Verlaufsdaten können Rechte gänzlich verwehrt, nur lesende Rechte eingeräumt oder aber volle Zugriffsrechte, d.h. Bearbeitungsrechte erteilt werden.

In einem dritten Schritt können die erteilten Objektrechte individuell für jede Gruppe angepasst werden. Hierdurch wird erreicht, dass für jede Gruppe definiert werden kann, auf welches Objekt welche Rechte bestehen.

Ein Betreiber-Administrator kann die bereitgestellten Funktionalitäten der individuellen Rechtzuweisung auch nutzen, um eigene Objekte gezielt an externe Unternehmen freizugeben. Hierfür selektiert er die für ein externes Unternehmen relevanten und aus seiner Sicht unkritischen Objekte und erteilt die entsprechenden Zugriffsberechtigungen (kein Zugriff, nur Lesezugriff, Vollzugriff) getrennt für die Stamm- und Verlaufsdaten. Auf diese Weise können beispielsweise der Überwachungsbehörde ausgewählte Brunnen, Infiltrationsanlagen und Grundwassermessstellen freigegeben werden, so dass diese sich jederzeit über die aktuelle Grundwassersituation informieren kann.

Auswertungsmodule

Gekoppelt an die Basis- und Objektrechte werden sowohl über das lokale als auch über die bereitgestellten Internetmodule verschiedene Aufbereitungs-, Auswertungs- und Visualisierungsmodule zur Verfügung gestellt.

Dynamische Berichtsgenerierung mittels XML und XSL

Die Berichtsgenerierung ermöglicht es unabhängig vom jeweiligen Betreiber, die vorhandenen Datenbestände für die jeweils zuständigen Sachbearbeiter zeitnah und somit aktuell auszugeben. Die Aufbereitung der Daten erfolgt dabei unter Verwendung der Extensible Markup Language (XML), die eine systematische und anwendungsunabhängigen Strukturierung ermöglicht und mit Hilfe entsprechender Layout-Schablonen (Extended Stylsheets) diese Daten in der jeweils gewünschten Form darstellen kann. Die Ausgabe von Berichten erfolgt im PDF-Format (Portabel Document Format) von Adobe, einem offenen Standard für die Verbreitung von elektronischen Dokumenten, der einen einheitlichen Ausdruck ermöglicht.

Grafische Auswertungen

Das Diagrammmodul ermöglicht dem Anwender, die zeitlichen Verläufe von punktuellen Grundwasserinformationen zu visualisieren. Dabei können beispielsweise Ganglinien von Grundwasserständen oder auch Fördermengen von Brunnen oder Infiltrationsanlagen dynamisch aus der Datenbank abgefragt und dargestellt werden. Dem Anwender werden im Programm unterstützende Assistenten angeboten, über die er seine situationsbezogenen Anforderungen definieren kann. Hierzu zählen auch statistische Größen, wie Maxima und Minima, Trends in Form von Regressionsgraden oder auch Standardabweichungen, die eine fachgerechte Bewertungsgrundlage darstellen.

Internetbasierte, flächenhafte Auswertungen

Um räumliche Zusammenhänge zu visualisieren, werden in der Praxis flächenhaft ausgewertete Karten als Grundwassergleichenpläne (GGP), Differenzenpläne (DP) oder Flurabstandspläne (FAP) erstellt [Abb. 4]. Über einen internettauglichen Kartenserver-Dienst ist es möglich, das relevante Gebiet frei zu selektieren sowie die Darstellungsart der Karte und das bei der Berechnung zu berücksichtigende Zeitintervall anzugeben.

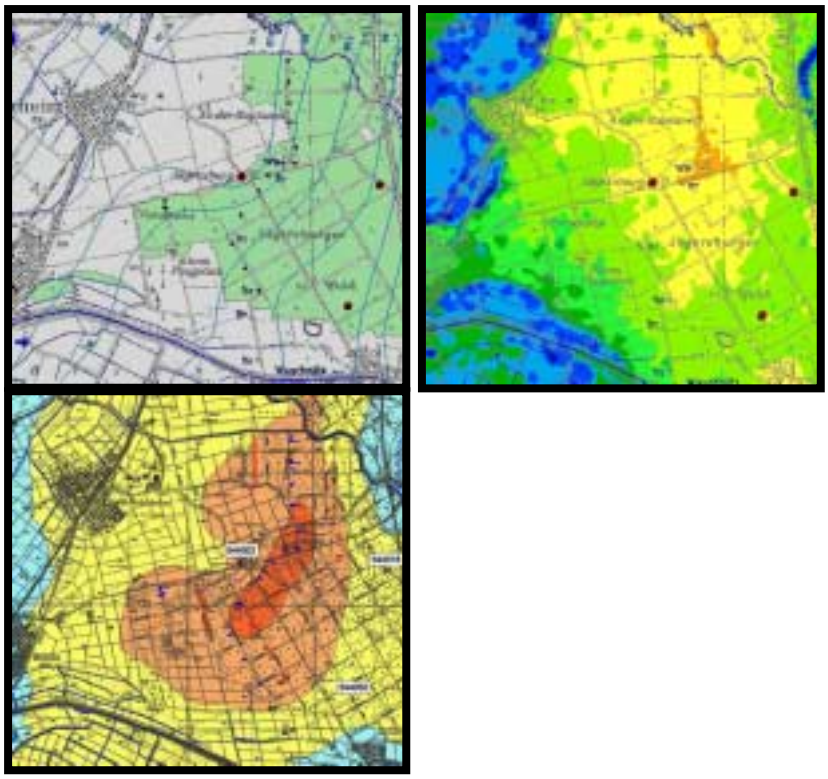


Abb. 4: Grundwassergleichen-, Flurabstands- und Differenzenplan

Replikation der Datenbestände

Relationales Datenbankmodell

Das zugrunde liegende relationale Datenbankmodell basiert im Wesentlichen auf der Schnittstellendefinition [3] des Hessischen Landesamtes für Umwelt und Geologie (HLUG). Für die Erfassung der Bohrprofile wurde die SEP3-Schnittstelle [4] des Niedersächsischen Landesamtes für Bodenforschung (NLFB) verwendet. Diese beinhaltet alle Informationen zu Bohrprofilen und Schichtverzeichnissen nach DIN 4022 und DIN 4023. Auf Grund der unterschiedlichen Anforderungen, die durch die beteiligten Unternehmen vorgegeben wurden, sind diese Datenbankschnittstellen sukzessive erweitert worden.

Das existierende Datenbankmodell unterteilt sich in unterschiedliche Teilbereiche:

1. Unternehmensinterne Strukturen
   - Betreiber, Wasserwerke, Anlagen
2. Objekttypen mit dazugehörigen Stammdaten
   - Brunnen, Infiltrationsorgan, Quellen, Bohrprofile, Grundwasser-, Gewässer- und Klimamessstellen
   - Gauß-Krüger-/Lage-Koordinaten, Gelände- / Messpunkthöhen, Detaildaten
3. Bohrprofile (SEP3-Schnittstelle)
   - Verfilterung, Ausbau, Schichtenverzeichnis
4. Multimediale Datenverwaltung
   - Binäre Dateiablage (z.B. Lagepläne, Pumpversuche etc.)
5. Verlaufsdaten/Messwerte
   - Mess- und Mengenwerte (quantitative Werte)
6. Zugriffsrechte/Benutzersichten
   - Gruppen- und Benutzerverwaltung, Zugriffsrechte, Sichten

Für den Datenaustausch mit anderen Behörden und externen Unternehmen sind XML-Schnittstellen vorgesehen, über die die gesamten Datenbestände bei Bedarf ausgetauscht werden sollen.

Triggerbasierte Replikationsmechanismen

Die erfassten Datenbestände werden derzeit in heterogenen Datenbanksystemen verwaltet. Dazu werden derzeit Oracle sowie der SQL-Server 2000 von Microsoft eingesetzt. Die Anforderung, verschiedene Datenbanksysteme zusammenzuführen, wurde erforderlich, da einerseits einige Betreiber über vorhandene Oracle-Systeme verfügten. Andererseits existiert mit der kostenlosen MSDE-Lösung des SQL-Servers ein System, das insbesondere für kleine und mittlere Wasserwerke sehr reizvoll und in seinem Funktionsumfang ausreichend ist.
Um die Datenbestände der dezentral vorliegenden, lokalen Datenbanken mit der zentralen Serverdatenbank abzugleichen, müssen geeignete Replikationsmechanismen eingesetzt werden. Da bei den einzelnen Wasserversorgern und Behörden aus den bereits erwähnten Randbedingungen unterschiedliche Datenbanksysteme installiert werden mussten, könnten keine datenbankspezifischen Mechanismen genutzt werden [7].

Replikationsmechanismen

Als Replikation wird ein Prozess bezeichnet, der es ermöglicht, eine definierte Menge von Grundwasserinformationen an mehr als einem Ort vorzuhalten. Dabei werden Änderungen an einem Ort (z. B Wasserwerk) an die anderen Orte (Serverfarm) propagiert. Die Synchronisation der Daten findet hierbei in beide Richtungen statt [6].

Bei dem Replikationszeitpunkt unterscheidet man zwischen synchronen (eager replication) und asynchronen Replikationen (lazy replication). Bei der synchronen Replikation [Abb. 5] werden die Änderungen sofort an die Replikate propagiert, wodurch sich alle Replikate innerhalb einer Transaktion aktualisieren. Diese Art der Replikation verursacht eine hohe Netzwerkauslastung. Im Gegensatz dazu werden bei einer asynchronen Replikation [Abb. 6] die Änderungen erst später an die Replikate propagiert. Die Aktualisierung der Replikate erfolgt dann in separaten Transaktionen, die von jedem Wasserwerk individuell angestoßen werden können. Die Netzwerkauslastung ist dementsprechend gering.

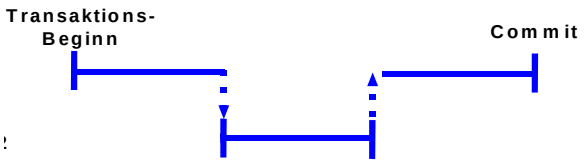


Abb. 5: Synchrone Replikation

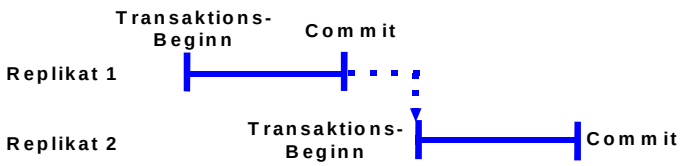


Abb. 6: Asynchrone Replikation

Bei der Auswahl des Konsistenzerhaltungsverfahrens musste zwischen einem optimistischen und einem pessimistischen Verfahren ausgewählt werden. Das optimistische Verfahren ermöglicht einen Vollzugriff auf die lokalen Replikate. Eine Netzwerkverbindung wird nur während der Replikation aufgebaut. Dabei werden temporäre Inkonsistenzen in Kauf genommen, um die Verfügbarkeit der Daten innerhalb der einzelnen Unternehmen zu erhöhen. Bei einem pessimistischen Verfahren wird mit verteilten Sperren gearbeitet, die Konsistenz der lokalen Kopie (single-copy-consistence) gewährleisten. Auf Grund der starken Einschränkungen beim Datenzugriff, scheidet dieses Verfahren für ein verteiltes, paralleles Arbeiten aus.

Das gewählte optimistische Verfahren kann vergleichsbasiert oder protokollbasiert umgesetzt werden. Beim vergleichsbasierten Verfahren werden Zeitstempel in der jeweiligen Datenbank gesetzt, die zur Identifikation der getätigten Änderungen die Replikate inhaltlich vergleichen. Bei dem gewählten, protokollbasierten Verfahren werden alle Änderungen in einem Protokoll (XML-basiert) festgehalten [Abb. 7]. Dieses Protokoll wird an die Zieldatenbank propagiert und dort abgearbeitet [Abb. 8].

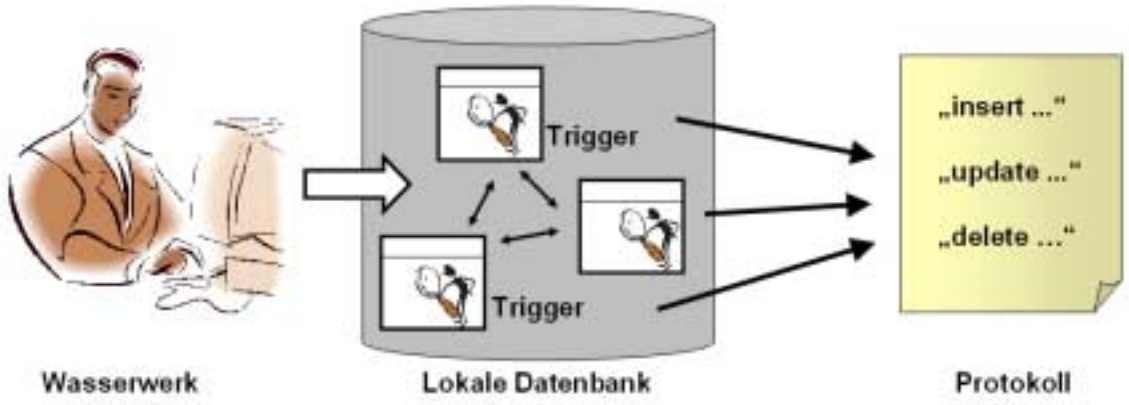


Abb. 7: Stufe 1 – Protokollierung aller Datenbankänderungen

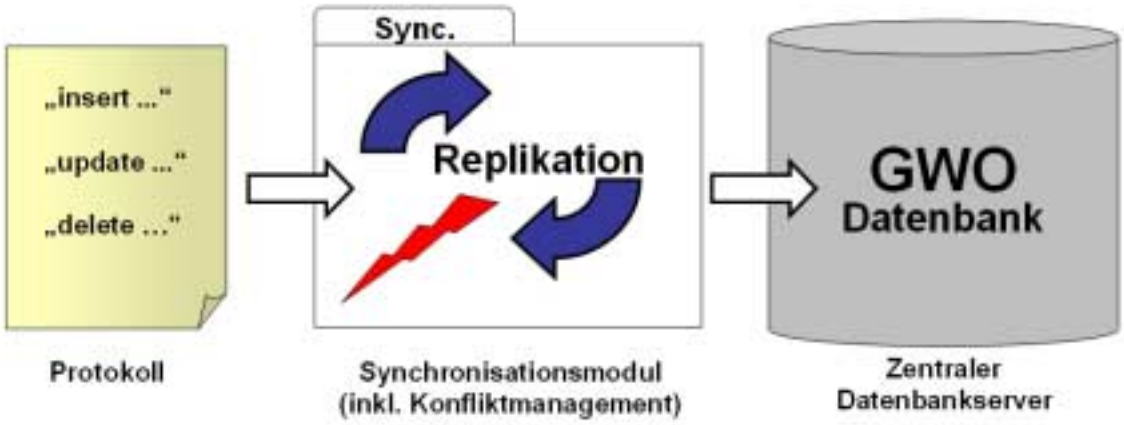


Abb. 8: Stufe 2 – Synchronisation der Replikate durch Auswertung des Protokolls

Eventuell auftretende Konflikte durch veraltete oder in beiden Datenbanken vorgenommene Änderungen der Datenbestände, werden erkannt und an einen Konfliktassistenten übergeben und können dort wahlweise manuell oder automatisch (z.B. neu überschreibt alt oder zentrale Datenbank überschreibt lokale) aufgelöst werden. Für die konkrete Umsetzung im Rahmen des Grundwasser-Online-Projekts heißt dies, das ein Client/Server-Modell mit der zentralen Datenbank als Primärkopie gewählt wurde, das über asynchrone, optimistische Replikationsmechanismen realisiert wird, um einen lokalen Vollzugriff auf die Daten, ohne eine permanente Netzwerkverbindung, zu ermöglichen.

Protokollmechanismen und Inhalte

Für die Aufzeichnung aller vorgenommenen Änderungen werden Trigger auf alle potenziell veränderlichen Tabellen und Attribute der Datenbank installiert. Trigger sind in den ANSI SQL99 Standards (Paket 008) definiert. Ein Trigger kann bei Änderungen auf Tabellen automatisch vordefinierte Folgeaktivitäten auslösen [2]. Durch den Aufruf einer Stored Procedure (einer in der Datenbank gespeicherten Prozedur mit vorkompilierten SQL-Anweisungen) können beliebig umfangreiche Verarbeitungsvorgänge angestoßen werden.

Im Änderungsprotokoll des GWO-Systems selbst werden sämtliche Änderungs-operationen eines Nutzers im Wasserwerk auf Tupelebene mitprotokolliert. Da die Zeitintervalle zwischen zwei Synchronisationsvorgängen variabel sind, d.h. zum Teil mehrere Wochen umfassen können, ist es möglich, dass Tupel gegebenenfalls mehrfach verändert wurden. Hierfür sind Regeln für die Änderungsanweisungen aufzustellen, da immer nur eine Anweisung pro Tupel übertragen und abgearbeitet werden kann.

Als Datenformat für den Austausch des Änderungsprotokolls wird XML verwendet, das sich wiederum aus einen Header- und einen Body-Teil zusammensetzt. Im Header-Teil werden die gültige Sitzungs-ID, der Synchronisationszeitpunkt, Information über die Art der übertragenen Daten sowie Informationen, die das Replikat beschreiben, vorgehalten. Im Body werden alle Tupel und die dazugehörigen Tabellennamen, Primärschlüssel, die Änderungsart sowie der entsprechende Wert und für jedes Feld der entsprechende Datentyp übertragen.

Sicherheitsaspekte bei der Protokollübertragung

Übertragen wird bei der Replikation das angelegte XML-Änderungsprotokoll. Da die übertragenen Daten als sensibel eingestuft werden, muss der Zugriff durch Unbefugte ausgeschlossen werden. Daher werden bei der Protokollübertragung starke kryptografische Verschlüsselungsverfahren eingesetzt. Zum Einsatz kommt dabei das mit 128-bit verschlüsseltes SSL-Verfahren (Secure Sockets Layer).Anwendungsszenario für die Replikation von Grundwasserdaten
In der folgenden Grafik [Abb. 9] wird der vorgestellte Replikationsprozess am Beispiel der Erstellung und Vergabe von Zugriffsberechtigungen zweier Objekte veranschaulicht:

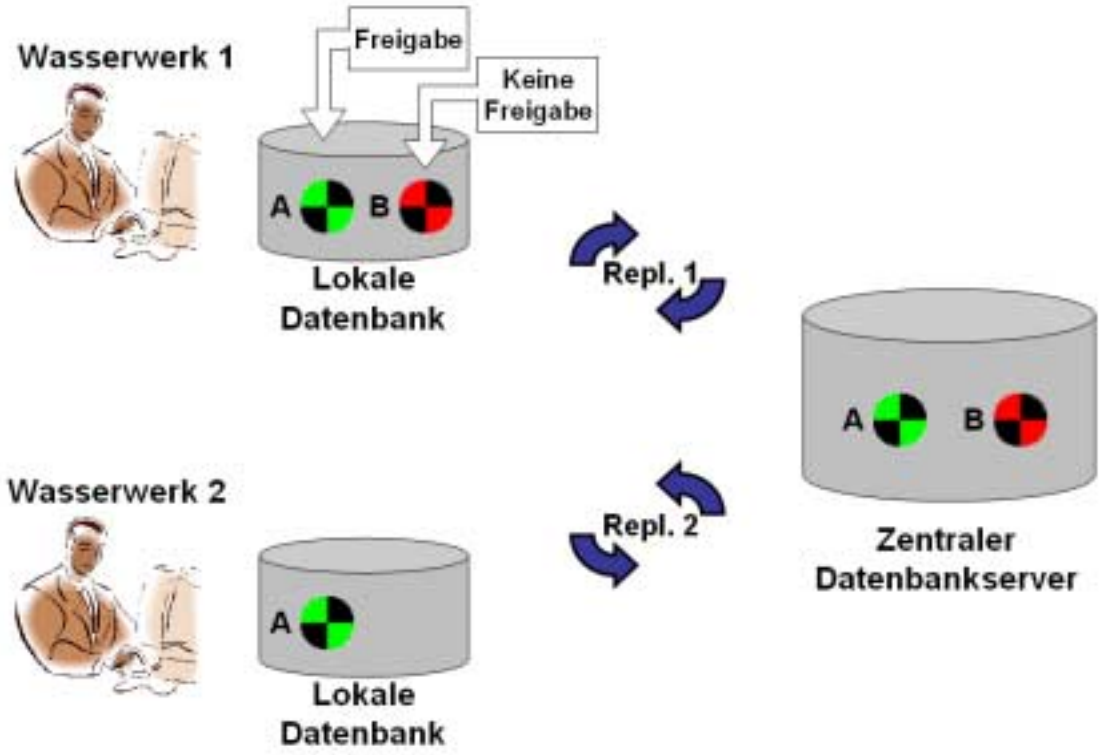


Abb. 9: Szenario des Replikationsvorgangs

Der Prozess gliedert sich in die folgenden Phasen:

1. Das Wasserwerk 1 legt in seiner Datenbank zwei neue Objekte an.
2. Das Objekt A wird für das Wasserwerk 2 freigegeben, d.h. Wasserwerk 2 erhält mindestens lesende oder schreibende Rechte auf die Stamm- oder Verlaufsdaten dieses Objekts. Für Objekt B werden hingegen keine Zugriffsrechte (Freigaben) erteilt.
3. Ein Mitarbeiter des Wasserwerks 1 startet den Replikationsprozess (1).
4. Die beiden Objekte A und B werden mit allen neuen bzw. geänderten Informationen in der zentralen Serverdatenbank abgelegt.
5. Ein Mitarbeiter des Wasserwerks 2 startet den Replikationsprozess (2).
6. Zusätzlich zum Abgleich der eigenen Daten, werden die Informationen des vom Wasserwerk 1 freigegebenen Objekts A auf die lokale Datenbank des Wasserwerks 2 übertragen, so dass die berechtigten Gruppen, entsprechend den vom Wasserwerk 1 definierten Rechten, mit dem Objekt A arbeiten können.

Zusammenfassung

Durch den gewählten Ansatz einer verteilten Datenerfassung und einer Zusammenführung dieser dezentralen Datenbestände über protokollbasierte Replikationsmechanismen in eine zentrale Serverdatenbank, werden die technischen Voraussetzungen für eine zeitnahe und unternehmensübergreifende Grundwassersteuerung geschaffen und gleichzeitig eine größtmögliche Freiheit bei der Bearbeitung der Grundwasserdaten in den einzelnen Wasserversorgungsunternehmen ermöglicht. Die unternehmensspezifischen Anforderungen im Hinblick auf die internen Strukturen und Zuständigkeitsressorts kann über die feingranularen Zugriffsrechte, die bis auf Objektebene individuell angepasst werden, realisiert werden.

Eine nachträgliche Datenbankerweiterung ist auf Grund des gewählten Ansatzes mit Triggern leicht möglich. Eine inhaltliche wie strukturelle Erweiterung des Systems ist durch den modularen Aufbau und die Skalierbarkeit jederzeit möglich.

TECHNISCHE UNIVERSITÄT DARMSTADT

Triggerbasierte Replikationsmechanismen zum Austausch von Grundwasserinformationen in heterogenen Datenbankumgebungen

Prof. Dr.-Ing. Uwe Rüppel
Dipl.-Ing. Thomas Gutzke

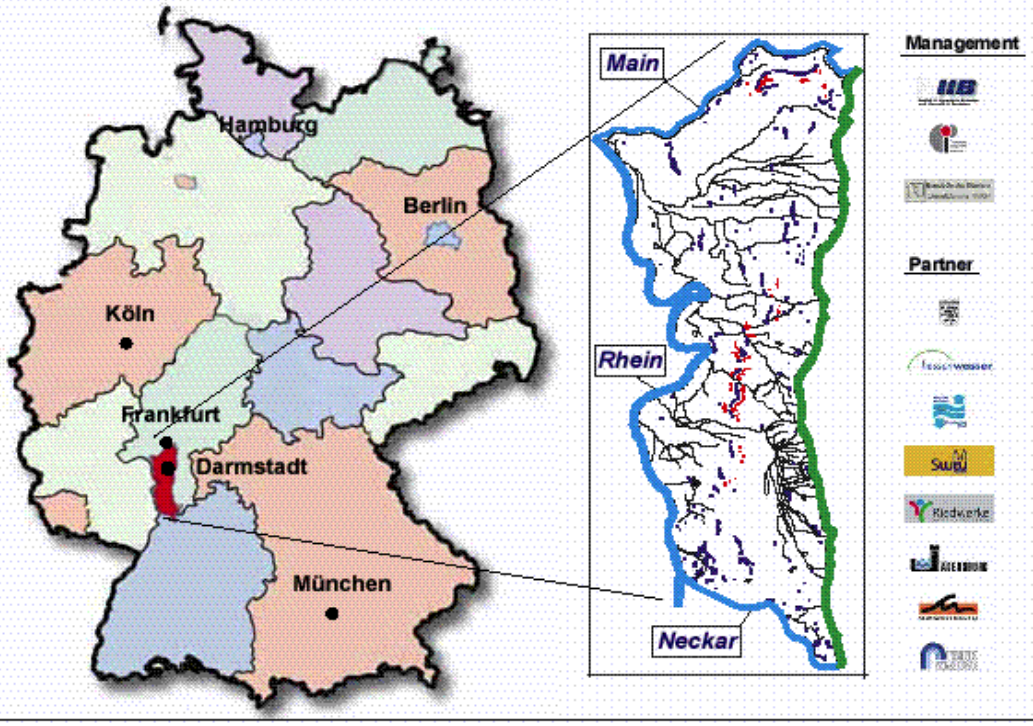
Institut für Numerische Methoden und Informatik im Bauwesen
Technische Universität Darmstadt

Dr.-Ing. Michael Petersen
Dipl.-Ing. Gerrit Seewald

CIP Ingenieurgesellschaft mbH, Darmstadt

TECHNISCHE UNIVERSITÄT DARMSTADT

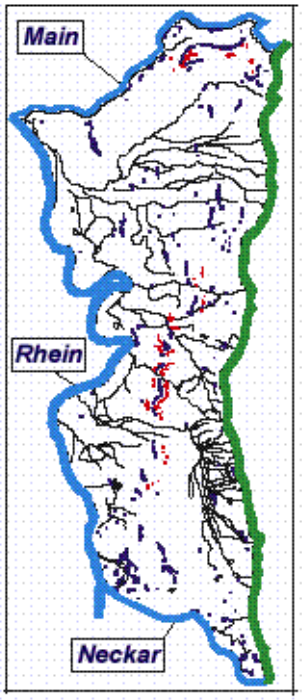
Projektgebiet



TECHNISCHE UNIVERSITÄT DARMSTADT

Projektgebiet

- Einzugsgebiet: 1238 km²
- Ca. 800.000 Einwohner in 36 Städten und Gemeinden
- Großer natürlicher GW-Speicher
- Entnahme von Trink- und Brauchwasser
- Versorgung der Ballungsräume Rhein-Main und Rhein-Neckar
- Steuerungsmöglichkeiten durch Infiltration von Rheinwasser



Management

Partner

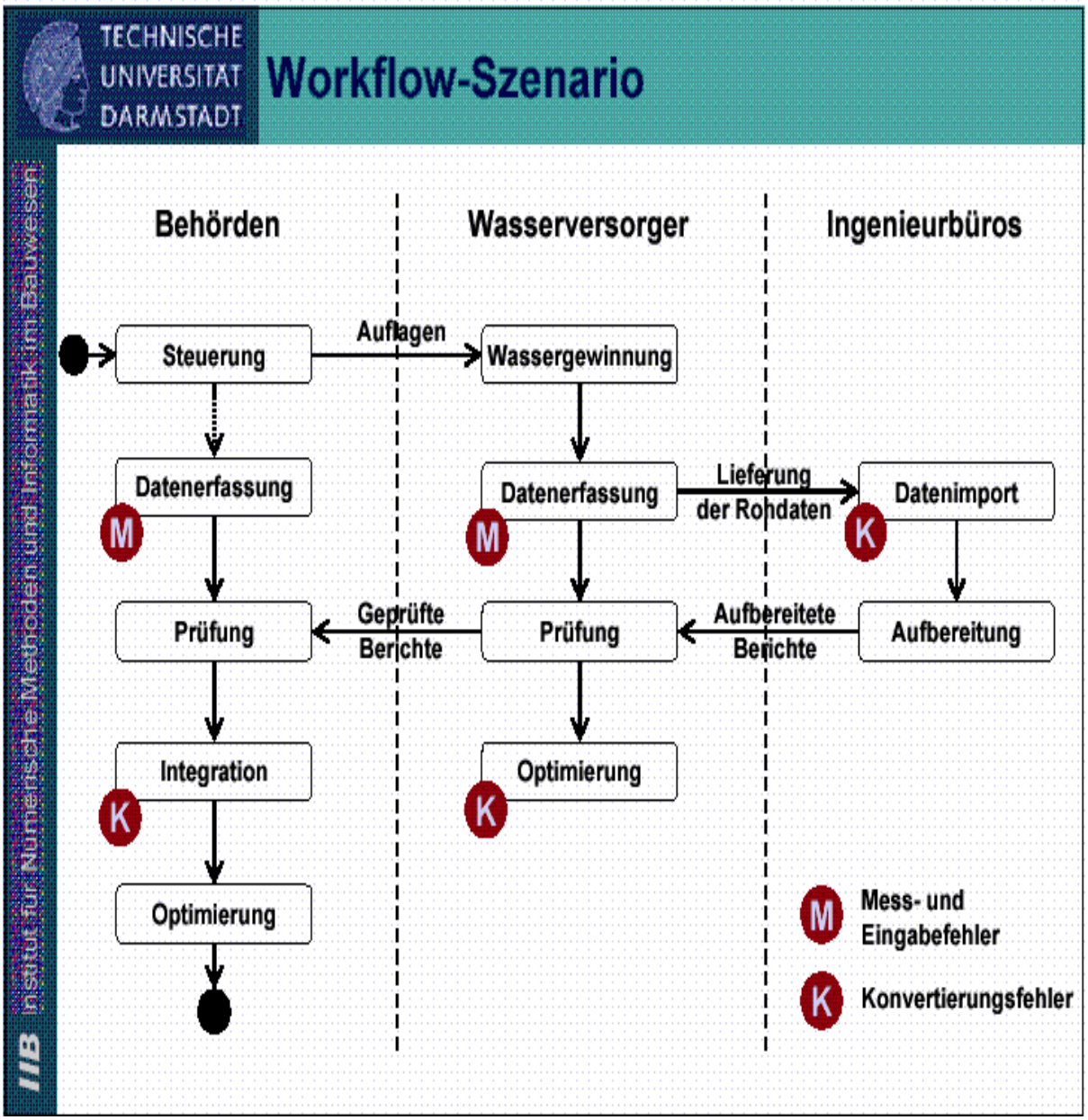
IIB Institut für Numerische Methoden und Informatik im Bauwesen

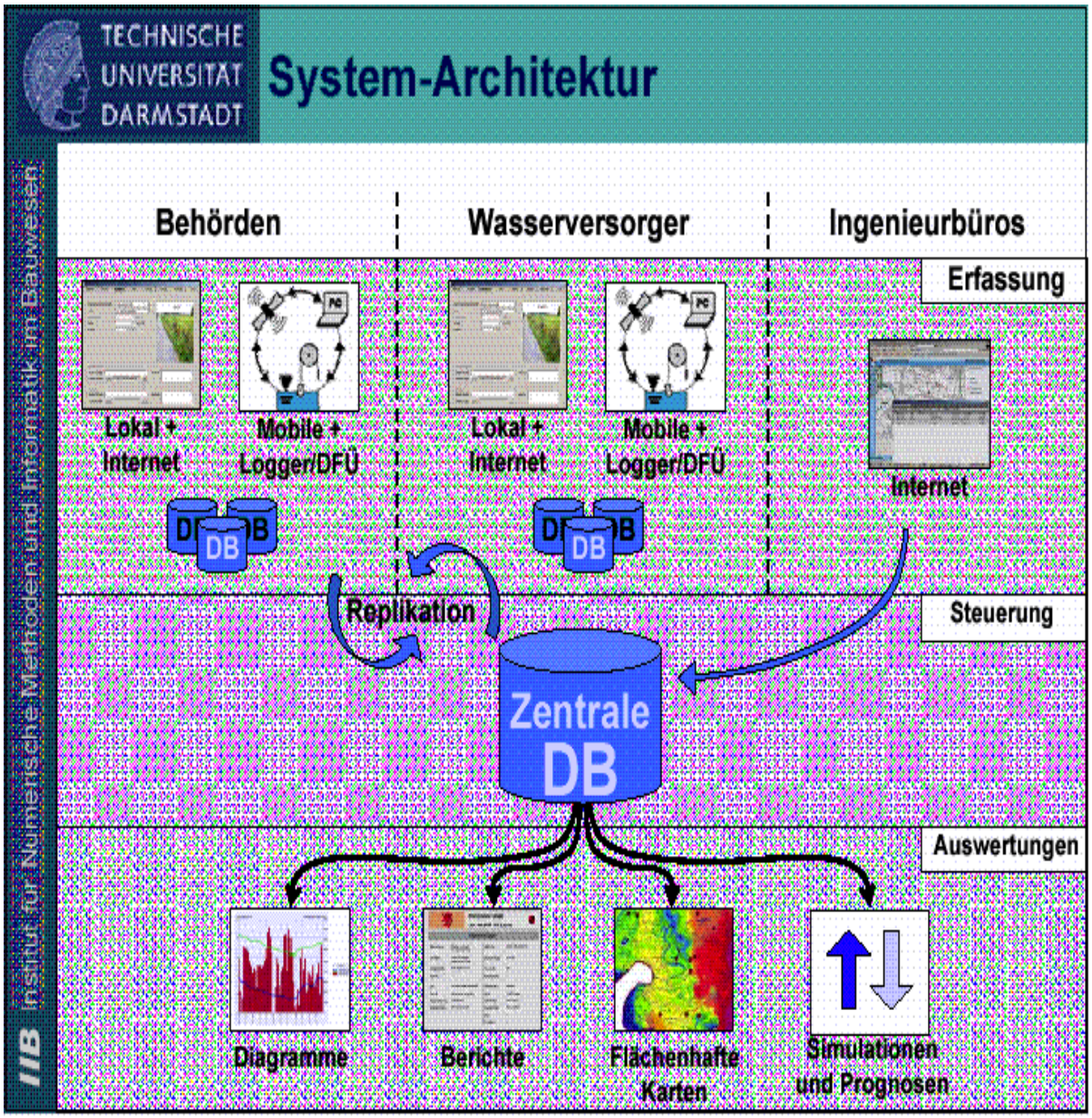
TECHNISCHE UNIVERSITÄT DARMSTADT

Ausgangssituation

Grundwasserentwicklung im Hessischen Ried
(1950-2003)

IIB Institut für Numerische Methoden und Informatik im Bauwesen


TECHNISCHE UNIVERSITÄT DARMSTADT
Workflow-Szenario
IIB Institut für Numerische Methoden und Informatik im Bauwesen
Behörden
Wasserversorger
Ingenieurbüros
Steuerung
Auflagen
Wassergewinnung
Datenerfassung
Datenerfassung
Lieferung der Rohdaten
Datenimport
Prüfung
Geprüfte Berichte
Prüfung
Aufbereitete Berichte
Aufbereitung
Integration
Optimierung
Optimierung
Mess- und Eingabefehler
Konvertierungsfehler

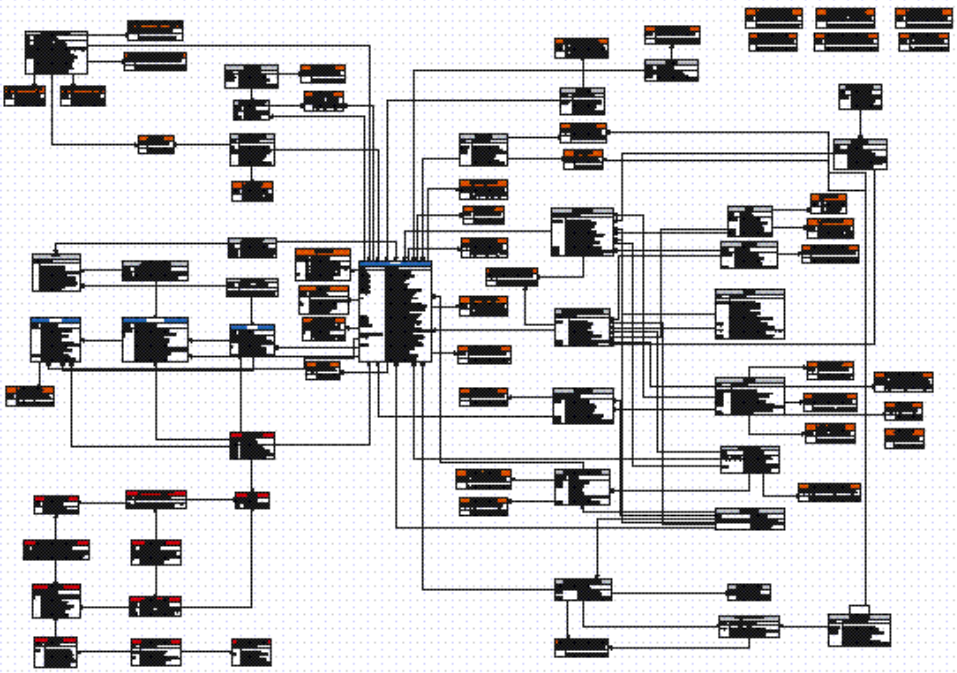

TECHNISCHE UNIVERSITÄT DARMSTADT
System-Architektur
Institut für Numerische Methoden und Informatik im Bauwesen
IIB
Behörden
Wasserversorger
Ingenieurbüros
Erfassung
Lokal + Internet
Mobile + Logger/DFÜ
Lokal + Internet
Mobile + Logger/DFÜ
Internet
DB
Replikation
Steuerung
Zentrale DB
Auswertungen
Diagramme
Berichte
Flächenhafte Karten
Simulationen und Prognosen

TECHNISCHE UNIVERSITÄT DARMSTADT

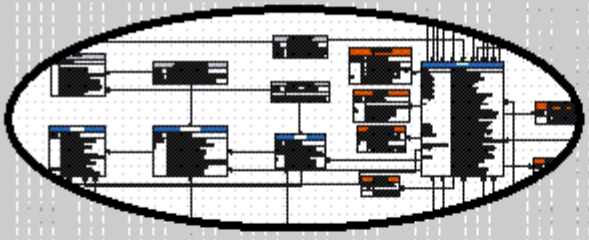
Relationales Datenbank Modell

Institut für Numerische Methoden und Informatik im Bauwesen

Unterstützte Systeme: Oracle 9i und MSSQL-Server 2000



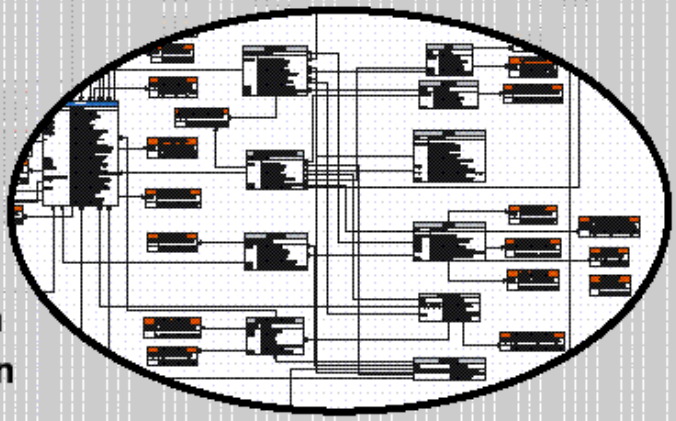
- **Struktur-Informationen**



- Betreiber
- Wasserwerke
- Anlagen

Brunnen, Infiltrationsanlagen, Quellen, Bohrprofile, Gewässerpegel, Grundwasser- und Klimastationen

- Struktur-Informationen
- Stammdaten
  - Geografie
  - Geologie
  - Bohrungen
  - Filtereigenschaften
  - Rechtliche Auflagen
  - Multimediale Dateien



TECHNISCHE UNIVERSITÄT DARMSTADT

Relationales Datenbank Modell

IIB Institut für Numerische Methoden und Informatik im Bauwesen

- Struktur-Informationen
- Stammdaten
- Verlaufsdaten

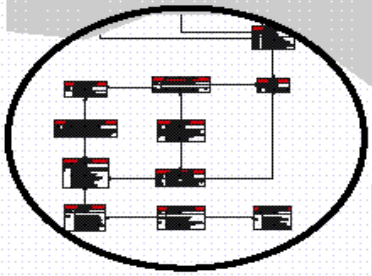
  - Messdaten
    (z.B. Grundwasserstand, Wassertemperatur)
  - Mengendaten
    (z.B. Förder- und Infiltrationsmengen)
  - Messtypen
  - Messintervalle

TECHNISCHE UNIVERSITÄT DARMSTADT

Relationales Datenbank Modell

IIB Institut für Numerische Methoden und Informatik im Bauwesen

- Struktur-Informationen
- Stammdaten
- Verlaufsdaten
- Zugriffsrechte

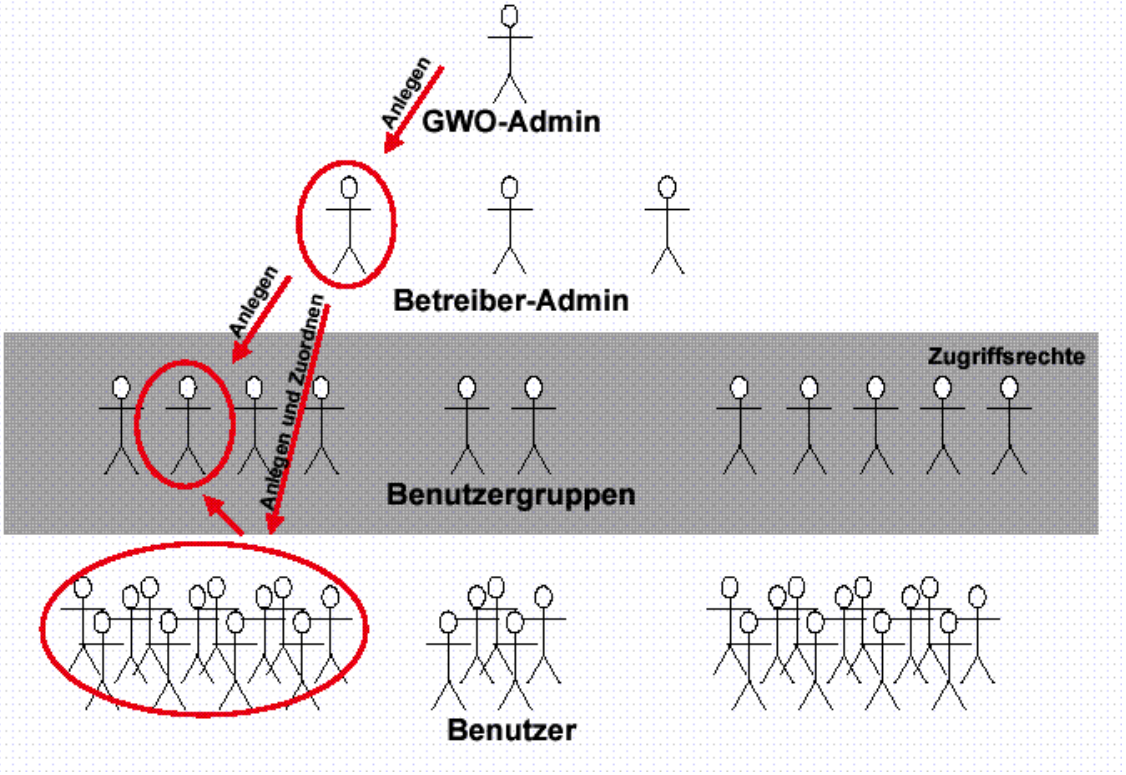


- Feingranulares Rechtesystem (Benutzer- und Objektverwaltung)
- Lese- und Schreibrechte auf Objektebene

TECHNISCHE UNIVERSITÄT DARMSTADT

Benutzergruppenverwaltung

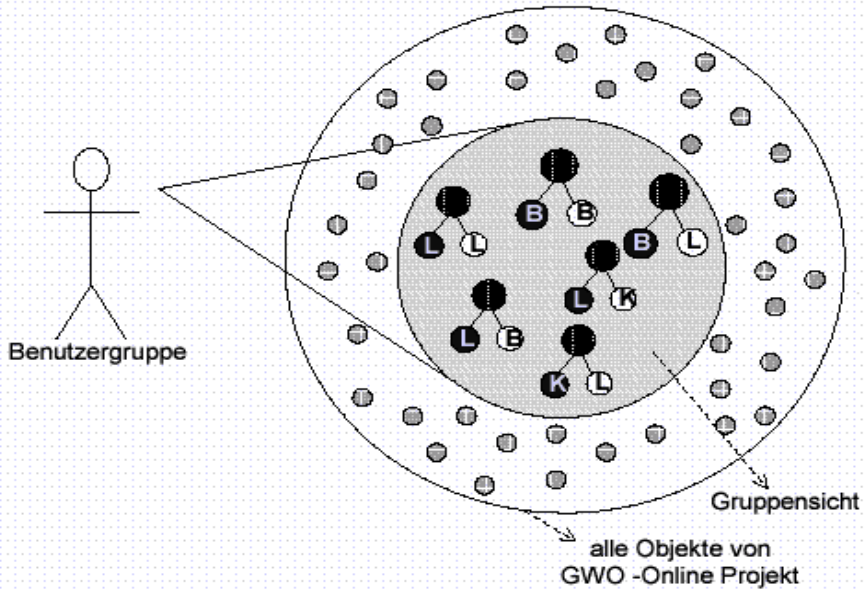
IIB Institut für Numerische Methoden und Informatik im Bauwesen

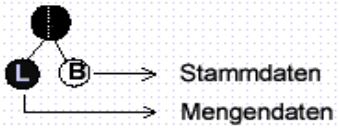


TECHNISCHE UNIVERSITÄT DARMSTADT

IIB Institut für Numerische Methoden und Informatik im Bauwesen

Zugriffsrechte und Sichtenkonzept





Zugriffsrechte:

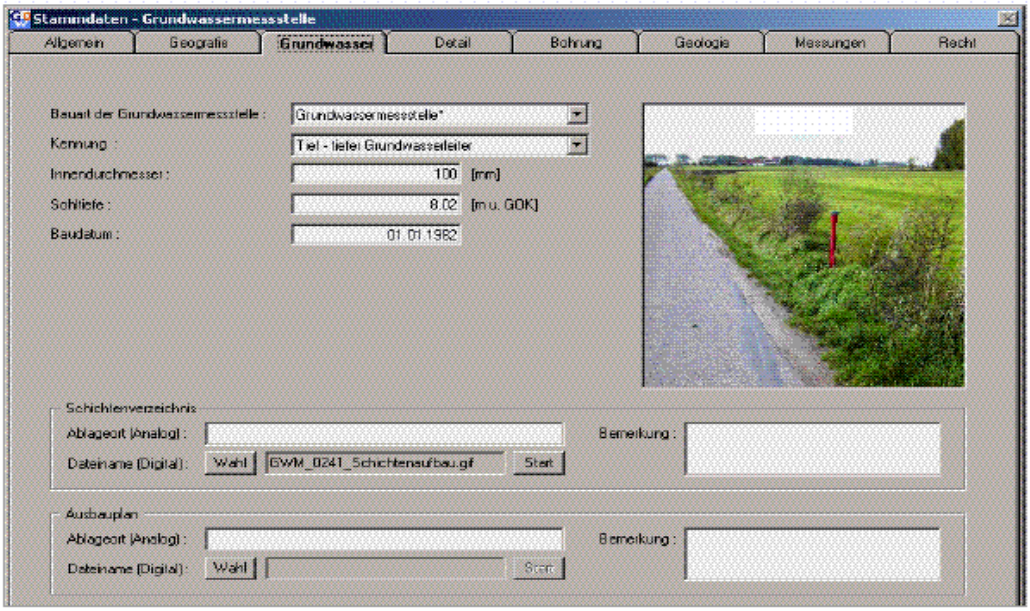
L: Lesen
B: Bearbeiten
K: Kein Zugriff

TECHNISCHE UNIVERSITÄT DARMSTADT

IIB Institut für Numerische Methoden und Informatik im Bauwesen

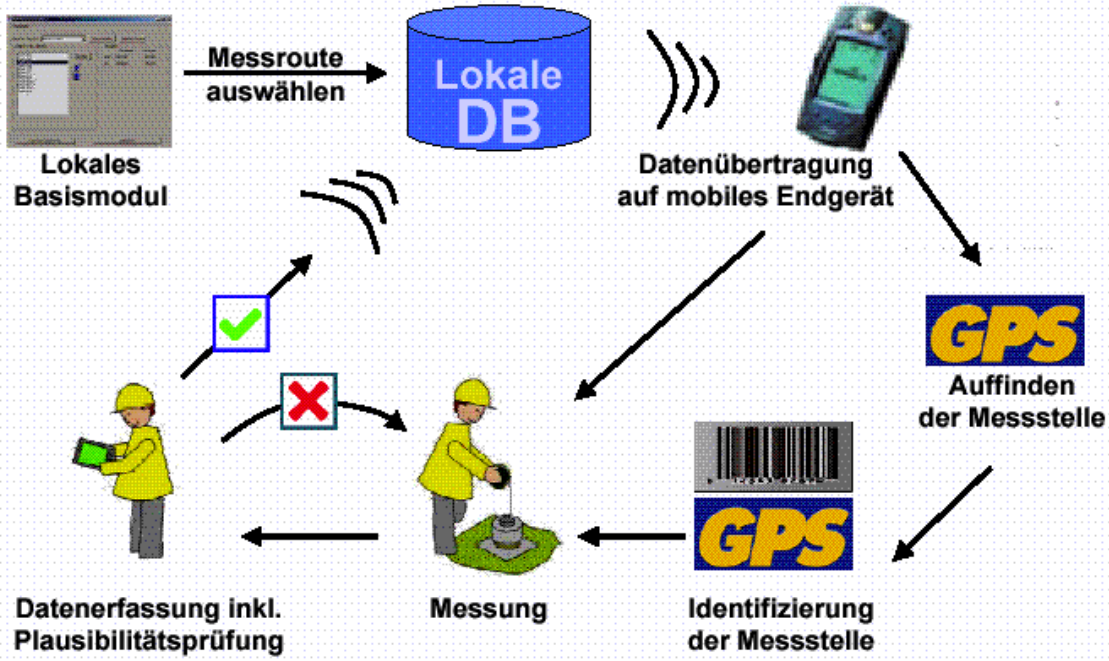
Lokale Datenerfassung

- Eingabe der Stammdaten über fachgerechte Frontends



TECHNISCHE UNIVERSITÄT DARMSTADT

Mobile Datenerfassung



IIB Institut für Numerische Methoden und Informatik im Bauwesen

TECHNISCHE UNIVERSITÄT DARMSTADT

Anforderungen an die Replikation

- Heterogene Umgebungen
  - **Unterschiedliche RDBMS (Oracle, SQL-Server…)**
  - **Geringe Internet-Bandbreiten**
  - **Unterschiedliches Datenvolumina**
- „Partly-Online“ Verbindung zwischen Client und Server
  - **Synchronisation nach Benutzeranforderung**
- Kontrolle von Zugriffsrechten
- Verschlüsselte Datenübertragung (SSL 128-bit)
- Protokollierung der Synchronisation
- Möglichkeiten zum Aktualisieren der DB-Schemata

IIB Institut für Numerische Methoden und Informatik im Bauwesen

Literaturverzeichnis

[1] **Grundwasser-Online**: Zukunftsweisende Überwachung und Steuerung der Grundwasserbewirtschaftung im Hessischen Ried. http://www.grundwasser-online.de.

[2] **Härder, T.; Rahm, E.:** Datenbanksysteme: Konzepte und Techniken der Implementierung. Springer-Verlag, Berlin 1999.

[3] **Hessisches Landesamt für Umwelt und Geologie (HLUG):** Schnittstellendefinition zum GruWa-Projekt "FIS-GW" – Fachinformationssystem Grundwasserschutz – Wasserversorgung. Wiesbaden 2000.

[4] **Niedersächsisches Landesamt für Bodenforschung (NLFB):** Aufschlusstyp SEP 3 (Beschreibung der Schnittstelle). http://www.nlfb.de.

[5] **Rüppel, U.; Meißner, U.; Gutzke, T.; Diaz, J.; Seewald, G.:** Vernetzt-kooperative Grundwasserbewirtschaftung im Hessischen Ried. In: Tagungsband zu: Bauen mit Computern – Kooperation in IT-Netzwerken, Bonn 2002, VDI-Berichte 1668, Düsseldorf 2002.

[6] **Saito, Y.: Optimistic Replication Algorithms.** In: International Symposium on Distributed Computing, S. 297-314, 2000.

[7] **Stockinger, H.:** Data Replication in Distributed Database Systems. CMS Note 1999/046, 1999.

Umweltforschungsplan
des Bundesministers für Umwelt, Naturschutz und Reaktorsicherheit
Grundlagen der Umweltpolitik, Umweltstrategien

Forschungsbericht 201 11 612

Erstellung eines semantischen Netzwerkservice (SNS) für das Umweltinformationsnetz Deutschland - German Environmental Information Network (*gein*®) -

Kurzfassung

Thomas Bandholtz

SchlumbergerSema **GmbH, Köln**

IM AUFTRAG
DES UMWELTBUNDESAMTES
Mai 2003

Einleitung

Seit dem Jahr 2000 hat das Umweltinformationsnetz Deutschland (German Environmental Information Network, *gein*®, http://www.gein.de) eine Textanalyse mit automatischer, thesaurus-basierter Verschlagwortung als ersten Schritt zu jeder weiteren Inhalts-Integration der verschiedenen Informationsquellen im Internet implementiert. In der Folge dieser ermutigenden Erfahrung wurde 2001 das Forschungs- und Entwicklungsprojekt "Erstellung eines semantischen Netzwerkservice (SNS) für das Umweltinformationsnetz Deutschland - German Environmental Information Network (gein®) -" begonnen. Das Ziel war einige Schwachstellen diese ersten Version von Ontologie-Management und automatischer Indexierung zu überwinden, vor allem durch Verbesserungen wie:

- semantische Integration von Thesaurus, geographischem Namensgut und Chronologie;
- Auflösung von Mehrdeutigkeiten durch Homonyme mit Hilfe einer Kontext-Analyse;
- differenziertere Kriterien für die Gewichtung der Schlagworte hinsichtlich ihrer Signifikanz für das Dokument;
- einfacher, verteilter Zugriff auf diese semantischen Methoden durch Web Services.
-

Integration der gein® Taxonomie

Die *gein*® Taxonomie ist seit 1999 durch Integration und Erweiterung der grundlegenden terminologischen Quellen der Umweltinformation entstanden. *gein*® benutzt drei semantische Strukturen:

1. einen Thesaurus mit derzeit 39 143 umweltrelevanten Begriffen (UmThes®),
2. einen Geo-Thesaurus, der die Lagebeziehungen zwischen 48 213 geographischen Objekten verschiedenster Typen beinhaltet,
3. eine Chronologie („Umweltkalender") von historischen und zeitgenössischen Umweltereignissen.

UmThes® ist ein vollständiger Thesaurus, der alle von ISO 2788/5964 geforderten Relationen (Ober-/Unterbegriff; Synonym; verwandter Begriff; Zusammensetzung) unterstützt, und er beinhaltet die komplette Wortmorphologie.UmThes® wird ebenfalls von vielen deutschsprachigen Informationssystemen benutzt, wie z .B. dem Umweltdatenkatalog, und er ist die deutsche Quelle des GEneral Multilingual Environmental Thesaurus (GEMET, Batschi).
Der gein® Geo-Thesaurus basiert auf dem geographischen Namensgut GN250 (Bundesamt für Kartographie und Geodäsie), welches durch verschiedene umweltrelevante Objektklassen ergänzt wurde, und er enthält alle räumlichen Lagebeziehungen als explizite, vorausberechnete Relationen, um performante Recherchen zu ermöglichen.
Die gein® Chronologie wurde von *gein*® aufgebaut, um so etwas wie einen "Zeit-Thesaurus" wichtiger Umweltereignisse zu schaffen – dies ist aber nicht als Terminkalender für kommende Konferenzen usw. zu missverstehen.

Topic Maps

Im Forschungsprojekt sind diese drei Strukturen in ein gemeinsames semantisches Modell integriert worden, um eine konsistente Ontologie zu entwickeln. Hierzu wurde eine Topic Map verwendet. Der Begriff "Topic" bezieht sich daher auf jedes Objekt dieser Ontologie unabhängig davon, ob es aus dem Thesaurus, dem Geo-Thesaurus oder der Chronologie stammt. Vom Standpunkt der Metadaten aus kann "Topic" als equivalent für "Schlagwort" aufgefasst werden.
Das Topic Map Modell erlaubt nicht nur die Integration von so unterschiedlichen Strukturen. Weiterhin wurde eine semantische Vernetzung durch Quer-Assoziationen zwischen Thesaurus Deskriptoren, Ereignissen und Orten vorgenommen.
Topic Maps (ISO 13250) haben sich als bevorzugtes Modell bewährt, wenn Thesauri, Gazetteers, Klassifikationen oder Schlagwortlisten miteinander vernetzt werden sollen. Topic Maps sind ein recht neuer Standard in der Welt der Taxonomie und Ontologie, und sie haben sich bereits zu einem der meistdiskutierten Beiträge entwickelt. Auf der XML Europe 2002 (XMLe) gab es 10 Vorträge über Topic Maps, und der Knowledge Technology Track war davon dominiert.

Topic Maps haben ihre Ursprünge in SGML und sie sind gleichzeitig mit und in ständiger Auseinandersetzung mit dem Semantic Web (Bandholtz, 1999) standardisiert worden. Heute erscheint es wahrscheinlich, dass beide Bewegungen ihrer verdienten Integration entgegengehen.

Grundsätzlich bestehen Topic Maps aus drei Komponenten: Topics, Assoziationen und Vorkommen (Occurrences).

Topics können jegliche Art von Thema (*subject*) repräsentieren:
"In the most generic sense, a subject is anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever." (ISO 13250)
Daher kann ein Topic ebenso gut ein Thesaurus Deskriptor oder Synonym sein wie ein geographisches Objekt, ein Ereignis, eine Person, eine Organisation, *whatsoever*. Unterschiedliche Arten von Topics werden als Topic Typen in einer Topic Map Anwendung definiert.
Assoziationen können Topics in eine semantische Beziehung bringen, z. B. zwischen Deskriptoren und Synonymen, Ereignissen und Schauplätzen, Deskriptoren und Ereignissen, *whatsoever*. Verschiedene Arten von Assoziationen sind an bestimmte Topic Types als Mitglieder gebunden und werden in einer Topic Map Anwendung als *Association Templates* definiert.
Vorkommen (*occurrences*) sind Informationen die zur Definition eines Topic beitragen. Generell kann jegliche existierende Information über ein Topic als *Occurrence* aufgefasst werden. Aber, da *Occurrences* als "groupings of addressable information objects around topics" (ISO 13250) definiert sind, sollten sie nicht mit dem allgemeinen Dokumentenindex einer Wissensdomäne wie *gein*® verwechselt werden. In SNS wird der Dokumentenindex außerhalb der Topic Map gesehen. Topic-Referenzen werden als klassifizierende Eigenschaften in den Metadaten zum Dokument verwendet, also eher: "groupings of topics around addressable information objects" (Bandholtz, 2002).

Das XML Modell der SNS Topic Map basiert auf ISO/IEC 13250, aber es verwendet nicht die XTM 1.0 DTD. XTM wurde als Austauschformat entworfen, nicht als Produktionsformat (Bandholtz, 2002). Wir haben es vorgezogen, ein eigenes XML Schema für Topic Maps zu entwickeln, nachdem wir die ursprüngliche ISO SGML DTD, die XTM XML DTD, und Martin Bryan's frühen XML Schema Vorschlag (Bryan) verglichen hatten.

Das SNS Topic Map Model

SNS hat seine Topic Typen und *Association Templates* so definiert, dass die drei Komponenten der *gein*® Taxonomie verlustfrei aufgenommen werden, wie in Abbildung 7 gezeigt. Der *Thesaurus* Typ und seine Subtypen bilden die klassische Thesaurus Struktur der ISO 2788/5964 nach. Der *Location* Typ ist der (abstracte) Ursprung aller spezifischen Orts-Typen wie z.B. Gemeinden, Wassereinzugsgebiete oder Nationalparks (nicht in der Abbildung gezeigt). Ebenso ist der *Event* Typ die Grundlage von Konferenzen, Störfällen, usw.. Bis hierhin wurde die übernommene Taxonomie lediglich ohne jede semantische Änderung in einer gemeinsamen Topic Map abgebildet.

Darüber hinaus wurden die drei Komponenten durch zwei neue Assoziationstypen (*where* und *what*) untereinander vernetzt. Beide verwenden die Ereignisse (*Events*) als Integrationsbezug. Die *Where*-Assoziation verbindet Ereignisse (*Event)* und Raumbezug (*Location*), indem sie aufzeigen, an welchem Ort das Ereignis geschah. Die *What*-Assoziation verbindet Ereignisse (*Event)* und Deskriptoren (*Descriptor*) und beschreibt, welche Themen das Ereignis betrifft.

Thesaurus, *Location*, *Event* sind abstrakte Oberklassen, welche die drei Dimensionen Thema-Raum-Zeit verkörpern. Da sie abstrakt definiert sind, können sie selbst nicht als Typen für konkrete Topics verwendet werden. Aber in *Association Templates* können sie dazu dienen, die jeweils zulässigen Topic Typen auf die jeweiligen Unterklassen zu beschränken. In Abbildung 7 wird dies z.B. benutzt um auszudrücken, dass in der Assoziation „*what*" jeder von *Event* abgeleitete Typ mit Deskriptoren, aber nicht mit Synonymen verknüpft werden kann.

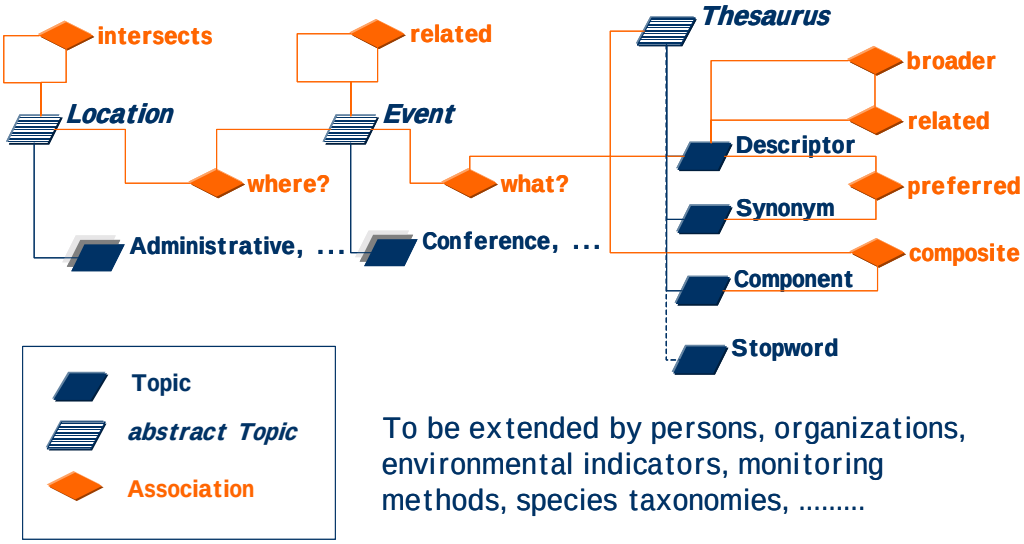


Abbildung 7 Die SNS Topic Map Typologie

Diese Struktur ist von Grund auf erweiterbar. Es wurde z. B. erwogen, Topic Typen wie „Person“, „Organisation“, Umweltindikator“, „Beobachtungsverfahren“ oder Artenverzeichnisse zu ergänzen, jedoch konnte dies aufgrund der beschränkten Projektmittel nicht realisiert werden.

Allgemein kann die heute in SNS implementierte Typologie als die Kern-Typology der Umweltterminologie aufgefasst werden, die durch die genannten oder weitere Typen ergänzt werden kann. Dazu gehört weiter die Einrichtung zusätzlicher Assoziationstypen z.B. zwischen Ereignissen und Personen, ohne dass die Kern-Typologie deswegen in irgendeiner Weise überarbeitet werden müsste.

Funktionale Dienste

Das bisher beschriebene Informationsmodell ist auf eine generischen Topic Map Engine der SchlumbergerSema GmbH implementiert worden.
Das SNS Projekt hat auf dieser Grundlage eine sogenannte “eXtensible Topic Map Engine Architecture” (Abbildung 8) realisiert. Diese Architektur beruht auf XML Formaten für Zugriffe (*requests*) und die daraus resultierenden Topic Map Auszüge (*responses*). Damit werden grundlegende Dienste zugänglich gemacht, die von Prozessen genutzt werden können, die ihrerseits außerhalb der Betrachtung des SNS Projekts liegen, also z. B. heute noch gar nicht existieren oder noch nicht als Nutzer betrachtet wurden. Der primäre Anwendungsfall für das Projekt ist *gein*®: in der letzten Projektphase wurde der *gein*® Broker mit SNS beispielhaft und produktiv integriert. Jedoch ist darin nicht das abschließende Projektergebnis zu sehen. Vielmehr wird erwartet, dass die SNS Dienste Voraussetzungen für künftige Anwendungen des Semantic Web schaffen, selbst wenn diese Anwendungen heute noch nicht im Einzelnen vorhersehbar sind.

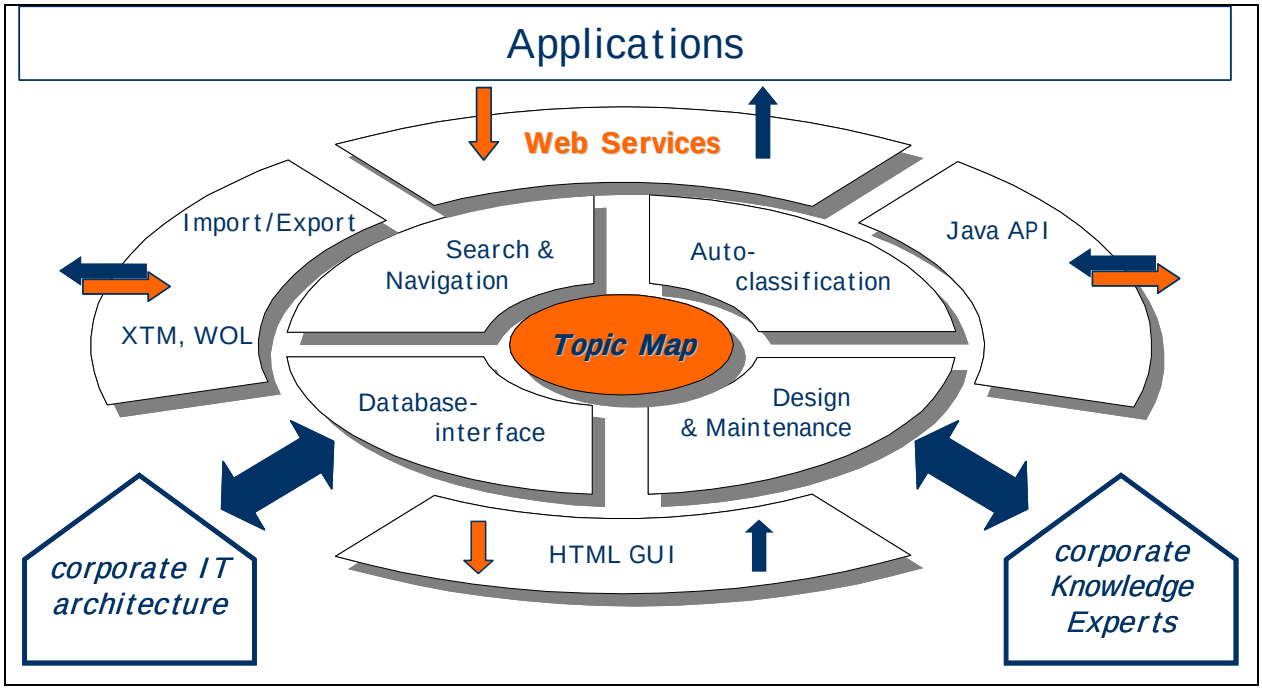


Abbildung 8: eXtensible Topic Map Engine Architecture

Innerhalb des SNS Projekts wurden 4 funktionale Dienste implementiert:

- Topic Map Browser and Navigator
- Redaktionsdienste
- Verbesserte automatische Verschlagwortung
- Semantic Web Services

Topic Map Browser und Navigator

Auch wenn dies nicht im Zentrum der Projektentwicklungen stand, so beinhaltet SNS doch eine eigene Weboberfläche, die von jeder nutzenden Applikation unabhängig ist. Damit wird die Suche in der Topic Map, allgemeine Visualisierungsmethoden und interaktive Navigation geboten.

Dies beinhaltet die Darstellung von Topics and Assoziationen in HTML Seiten, Formulare für die Eingabe von Suchbedingungen, sowie die Anzeige von Suchergebnissen oder alphabetischen Listen. Alle Topics und Assoziationen werden hier als Hyperlink dargestellt, so dass eine einfache, intuitive Navigation von Topic zu Topic geboten wird, indem man jeweiligen Assoziationen folgt. All dies erfordert beim Anwender lediglich einen Webbrowser (ohne weitere Client-Software) und Internet/Intranet-Anbindung.

Um auch semantische Netze angemessen visualisieren zu können, wurde eine *interaktive Graphik* (Abbildung 9) implementiert, die eine intuitive Darstellung von „Knoten und Kanten“ (nodes and arcs) ermöglicht. Jeder Knoten repräsentiert ein Topic. Per Mausklick können die Eigenschaften einzelner Topics eingeblendet werden, und jedes Topic kann zum Zentrum des dargestellten Netzausschnitts gewählt werden, um so in einem graphischen Fenster durch die gesamte Topic Map zu navigieren.

Diese interaktive Grafik ist mit Scalable Vector Graphics (SVG), einem XML Standard, implementiert. Dafür wird derzeit noch ein Plug-In benötigt, welches kostenfrei von den größeren Anbietern graphischer Software (Adobe, Corel, usw.) bezogen werden kann.

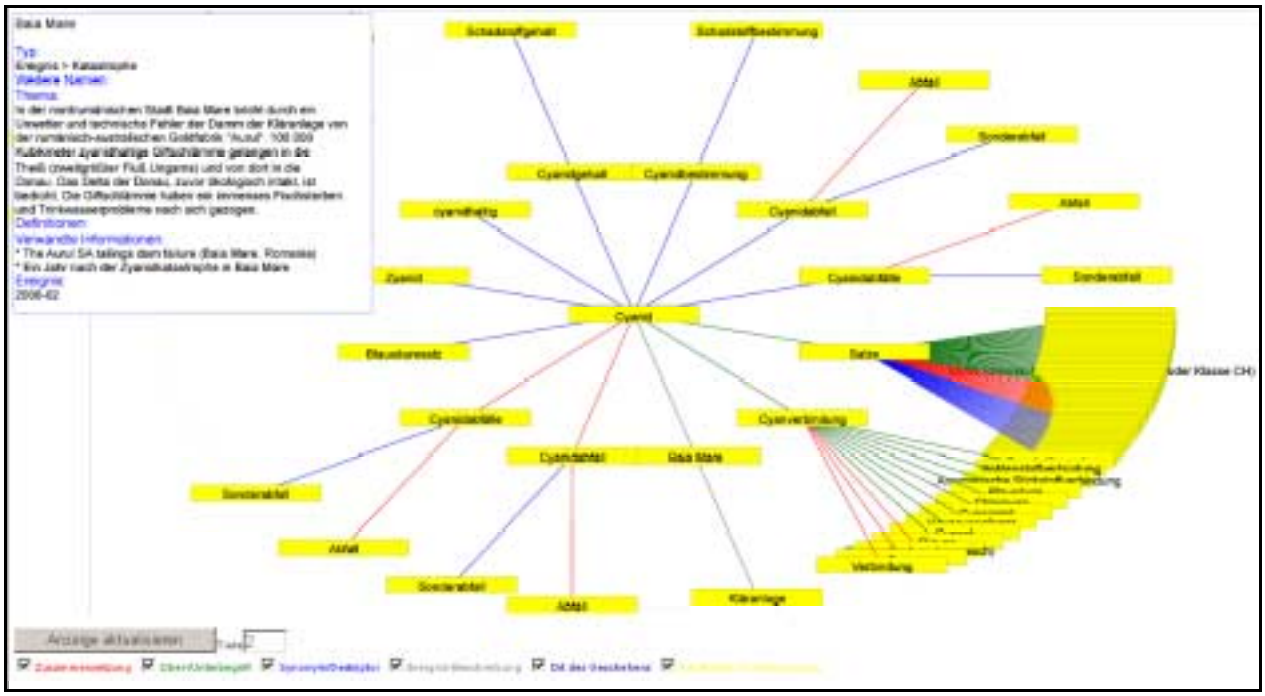


Abbildung 9: Interaktive Graphik

Redaktionsdienste

Für eine „Wortgutredaktion“ bietet SNS eine editierbare Version der HTML-Oberfläche. Hier können Topics und Assoziationen geändert, ergänzt oder entfernt werden. Im gegenwärtigen Einsatz wird diese Oberfläche dazu verwendet, den *gein*®-Umweltkalender laufend nachzuführen. Die Darstellung des Kalenders in *gein*® wird unmittelbar aus den Ereignis-Topics generiert (Zugriff über Web Service). Der größte Teil der integrierten Terminologie (UmThes® und GN250) wird heute außerhalb von SNS gepflegt, daher wurden grundlegende Importschnittstellen für beide Formate implementiert. Es ist jedoch geplant, den Geo-Thesaurus künftig direkt im XML-Topic Map Format zu pflegen, was die dauerhaft aktuelle Integration erleichtern würde.

Für SNS Redakteure sind grundlegende statistische Dienste über die Verwendung von Topics und über die in Freitexten nicht erkannten Zeichenketten konzipiert (zum Teil bereits durch *gein*® implementiert). Beides hilft bei der laufenden Aktualisierung der Terminologie. Diese Ansätze können künftig weitergehend automatisiert werden.

Verbesserte Automatische Indexierung

Eines der wichtigsten Ziele war es, die semantische Zuverlässigkeit der bisher in *gein*® implementierte automatischen Indexierung (Verschlagwortung) zu verbessern.
Dies wurde in einer Kombination mehrerer Ansätze erreicht:

- Vermeidung der häufigen Missverständnisse im Zusammenhang mit geographischen Namen,
- Generell eine bessere Auflösung von Mehrdeutigkeiten im Fall von Homonymen,
- genauere Erkennung von zusammengesetzten Begriffen,
- verbesserte Gewichtung der Signifikanz jedes ermittelten Topics für die Gesamtbedeutung des Dokuments.

Abgesehen von einigen kleineren Optimierungen beim Parsen von HTML und PDF Formaten wurde diese Verbesserung durch die Integration der *gein*® Terminologie in eine gemeinsame Struktur (die Topic Map) und durch ihre weitere Quervernetzung ermöglicht, wie in Kap. 0 näher dargestellt wird.
Die Analyse und Gewichtung der Assoziationen zwischen den Topics, die im ersten Schritt im Dokument erkannt (oder vermutet) werden, hat sich als das wichtigste Werkzeug für die Klärung der tatsächlich korrekten und signifikanten Topics herausgestellt. Obwohl hierbei theoretisch kaskadierende Missverständnisse auftreten können, weil missverstandene Topics Assoziationen mit anderen missverstandenen Topics aufweisen mögen, kann dieses Risiko nach den bisherigen Erfahrungen in der Praxis vernachlässigt werden. Die Assoziationsanalyse kann in der Mehrzahl der Fälle Mehrdeutigkeiten zwischen geographischen Namen und Deskriptoren korrekt auflösen, und dies sogar auf zwischen gleichlautenden (homographen) geographischen Namen.

Semantic Web Services

Seit 1999 war das gesamte Wortgut auf dem *gein*® Broker installiert. Es wurde u.a. für die Verschlagwortung von derzeit rund 200.000 Webseiten von 89 Informationsanbietern genutzt. Seit dieser Zeit hat es immer wieder Anfragen von Informationsanbietern gegeben, ob sie dieses Wortgut und das automatische Verschlagwortungsverfahren auch für weitere, eigene Zwecke nutzen können.
Die Überlegungen über den Aufwand, ein kompaktes Modul zu isolieren und für bis zu 89 verschiedene technische Umgebungen zu portieren, wurde einem Online Zugriff auf einen zentralen Service der Vorzug gegeben. Dieser Service sollte die grundlegenden Verfahren automatischer Inhaltsklassifizierung bereitstellen, wie z. B.:

1. Topics finden – die Suche nach Topics, die mit einer gegebenen Zeichenkette wahrscheinlich in Zusammenhang stehen,

2. Topics navigieren – von einem gegebenen Topic aus entlang seiner Assoziationen zu weiteren Topics fortschreiten,

3. Automatische Verschlagwortung – mit eine neuen, auf die Topic Map aufsetzenden linguistischen Analyse die charakteristischsten Topics für ein gegebenes Dokument ermitteln.

gein® blickt auf sehr positive Erfahrungen mit einer verteilten Recherche zurück, bei der XML Formate in HTTP-Requests kommuniziert werden. Aus heutiger Sicht war diese in 1999 konzipierte verteilte Recherche eine Vorwegnahme des Simple Access Protocol (SOAP), welches heute in der W3C Web Services Activity (http://www.w3.org/2002/ws/) aufgegangen ist.
Mit Bezug auf die dort erarbeiteten Web Service Description Usage Scenarios (W3C, 2002), repräsentiert ein “UC0015 Request-Response” Scenario, definiert als: *“... a web service that responds with an output message or a fault based on at least one or more input messages received.”*

Beispiele für Nachrichten und Antworten können z. B. sein::

SNS-Client: “Mein Anwender sucht nach Informationen mit dem Suchbegriff ‚*xyz*’ – gibt es irgendein Topic, das damit zu tun haben könnte?“

*SNS: “*Ich finde *X* Topics, hier sind ihre Definitionen, in dem XML Format, das wir vereinbart haben.”

SNS-Client: “Ich verstehe. Bitte gib mir alle Assoziationen von Topic *xyz*.”

SNS: “Topic *xyz* ist Unterbegriff von Topic *abc*, es ist verknüpft mit dem Ereignis Topic *def*, *etc, etc.*”

SNS-Client: “Gib mir bitte die Ergebnisse der automatischen Verschlagwortung für das Dokument an der Adresse http://www.abc.xyz/any.foo. Bitte ermittle die 10 signifikantesten Topics!”

SNS: “Die 10 signifikantesten Topics sind ....”

Darüber hinaus könnten Nachrichten relevant werden, die mit dem Anwendungsfall “UC0025 Event notification” übereinstimmen *(bisher nicht implementiert).* Dieser wird beschrieben als: *“An application subscribes to notifications of certain named events from an event source. When such events occur, notifications are sent back to the originating application (first party notification) or to another application (third party notification).”* (W3C, 2002)

In diesem Scenario stellt SNS die “event source” dar, die “named events” (z.B. dass eine neue Version der Topic Map aktiviert wurde) an die SNS-Clients sendet, die diesen Dienst zuvor abonniert haben.

Die XML Struktur und Syntax dieser Kommunikation sind mit XML Schema und Web Service Description (W3C, 2001) wohl definiert und testweise in einer Universal Description, Discovery, and Integration (UDDI) Registry registriert worden. Auf Details dieser Dokumente kann in der Kurzfassung nicht eingegangen werden.

Indem SNS mit diesen Standardisierungsprozessen konform entwickelt wurde, konnte eine grundlegende Interoperabilität des Systems verankert werden, die über den unmittelbaren Anwendungsfall *gein*® weit hinausgeht. Nicht nur die *gein*® Informationsanbieter, sondern jede möglicherweise interessierte Anwendung weltweit wird in die Lage versetzt, die technische Integration von SNS auf Anhieb zu verstehen, sofern es selbst globale Standards beachtet.

Integration

SNS ist nicht als alleinstehendes System konzipiert worden. Sein semantisches Modell und die funktionalen Dienste wurden entwickelt, damit sie in grundsätzlich jedes von Umweltbehörden betriebenes Informationssystem eingebunden werden können.

Als ein *vorläufiges* Ergebnis des Projekts wurden die semantischen Funktionen von *gein*® (entwickelt von 1999-2000) vollständig durch die SNS Dienste ersetzt.

Dies geht bereits weit über einen prototypischen Machbarkeitsnachweis hinaus, da *gein*® das produktive öffentliche Portal der deutschen Umweltbehörden von Bund und Ländern darstellt.

Darüber hinaus ist die Integration in *gein*® eine Demonstration der Qualität, Performanz und des Nutzens der SNS Dienste beim Einsatz mit Umweltinformationssystemen.

Auf diesem Gebiet gibt es zahlreiche Integrationsoptionen mit verschiedenen Anwendungsfällen.

1. UDK (Umweltdatenkatalog): Die (VV UDK/GEIN), eine Verwaltungsvereinbarung von Bund und Ländern über die künftig integrierte Pflege und Weiterentwicklung von UDK und *gein*®, ist am 1.1.2003 in Kraft getreten. Als eines der wichtigsten Themen wird dabei diskutiert, wie die Wortgutunterstützung beider Systeme zusammengeführt werden kann. Zur Zeit verwenden beide UmThes®, aber in unterschiedlichen Implementierungen. Es ist beabsichtigt, SNS als gemeinsame Basis für beide Systeme zu verwenden.

2. *gein*® Informationsanbieter: die mitwirkenden Organisationen (derzeit 89) sind eingeladen, SNS über die Web Services für jede Art von Umweltinformation zu nutzen. Auch wurde Interesse an lokalen Versionen von SNS für landesspezifische Aktivitäten geäußert. Langfristig kann daraus ein Netzwerk von kaskadierenden Topic Maps entstehen, von dem unterschiedliche Teilsichten verwendet werden können, je nach dem jeweiligen räumlichen und thematischen Fokus der Anwendung.

3. GeoMIS.Bund: das " Metainformationssystem für Geodaten des Bundes" wird vom (IMAGI) innerhalb der "nationalen Geodatenbasis" entwickelt. Ein erster Prototyp ist von SchlumbergerSema Anfang 2002 realisiert worden. Dabei wurde der SNS-Raumbezug („Geo-Thesaurus") in einer sehr frühen prototypischen Version eingebunden. Das Bundesamt für Kartographie und Geodäsie wird diesen Geo-Thesaurus künftig inhaltlich pflegen, und eine intensivere Integration mit Attributen des ISO 19115 Datensatz liegt nahe.

4. Europa: Die eEIONET Teilnehmer befassen sich mit "*environmental web services e.g. Reportnet, country networks, and metadata, as well as terminology/ontology issues*" auf Europäischer Ebene (EIONET). Da die Beziehung zwischen GEMET und UmThes® sehr eng ist, und da SNS bereits zweisprachig arbeitet (Deutsch/Englisch), empfiehlt sich eine Erweiterung auf den europäischen Raumbezug und auf die volle Mehrsprachigkeit (derzeit 19 GEMET Sprachen). Dies ist bereits in (Bandholtz, 2001) aus der EIONET work conference in Thun (CH) vorgeschlagen worden und führte zu einer Expression of Interest (EoI) im 6. Rahmenprogramm der Europäischen Kommission.

Zusammenfassung

SNS hat die bestehenden *gein*® -Thesauri erfolgreich miteinander integriert und zu einem Service-orientierten, aktiven Ontologie-System weiterentwickelt, um von der Umweltinformations-Domäne über Standard-Internet-Schnittstellen genutzt zu werden. Die erste anwendende Applikation ist der *gein*® Broker, welcher durch das erweiterte Terminologiemanagement und die verbesserte automatische Verschlagwortung aufgewertet wird. SNS ist produktionsreif implementiert und kann von nationalen und regionalen Systemen wie z. B. dem UDK eingebunden werden. SNS beinhaltet auch grundlegende Dienste für eine Wortgutredaktion.
SNS hat die Eignung des semantischen Modells der Topic Map ebenso nachgewiesen wie den Realismus der Web Service Architektur, mit der funktionale Dienste von den anwendenden Prozessen modular separiert werden. Aufgrund dieser Architektur kann SNS eine tragende Rolle in der weiteren Entwicklung der öffentlichen Umweltinformation der nahen Zukunft spielen.

Literatur

Angrick M., Bös R., Rüther M., Bandholtz T.: Semantic Network Services (SNS). EnviroInfo 2002. http://enviroinfo.isep.at

Bandholtz, T. (1999): GEIN 2000 and beyond: Information about the environment in the „semantic web“. In: Environmental Markup Language (EML). Proceedings of Workshop 1, Berlin 1999. Metropolis. Marburg, Germany 2000

Bandholtz, T. / Bös, R. / Rüther, M. (2000): The German Environmental Information Network (GEIN). In Cremers, A.; Greve, K. (eds.): Computer Science for Environmental Protection '00. 2000.

Bandholtz, T. : Semantic Network Services (SNS) - a webservice for environmental taxonomy. CDS and e-EIONET Work Conference 2001: Building environmental web services. http://www.eionet.eu.int/events/cdsittag2001/abstracts/band.htm

Bandholtz, T. : A Taxi in Knowledge Land. XMLeurope 2002, Barcelona. http://62.231.133.220/idea-eks-nav/papers/03-05-03/03-05-03.html

Bandholtz, T: Taxonomie und Topic Maps. Vom Stichwortverzeichnis zur Wissensnavigation. In: XML & Web Services Magazin, Heft 2, 2002. http://www.xmlmagazin.de/itr/ausgaben/show.php3?id=103&nodeid=67

Bandholtz, T,_ Semantic Network Services (SNS). Open Forum 2003 on Metadata Registries. January 20-24, 2003. Santa Fe, New Mexico, USA. http://metadata-stds.org/OpenForum2003/

Bandholtz, T,_ Semantic Network Services (SNS). Web Services for the Integrated Enterprise. OMG's 2nd Workshop On Web Services. Modeling, Architectures, Infrastructures And Standards. February 10-13, 2003. München/Philadelphia. http://www.omg.org/news/meetings/webservices2003europe/index.htm

Batschi, W.D. Felluga, B., Legat, R., Plini, P., Stallbaumer, H., Zirm, K.: „SuperThes“: A New Software for Construction, Maintenance and Visualisation of Multilingual Thesauri. In: EnviroInfo 2002, Vienna 2002.

Bryan, Martin : XML Schema for ISO 13250 Topic Maps. Proposed Syntax 29th January 2001. http://www.diffuse.org/TopicMaps/schema.html

EIONET: European Environment Information and Observation Network. eEIONET Work Conference from 26-28 September 2002 in Vienna. Released: 2002/07/03. http://eea.eionet.eu.int/Best_Practice/eEIONET2002

EoI: European Environmental Topic Map Engine with Multilingual Auto-Classification (EETM). Expression of Interest to the 6th Framework Programme of the European Commission. June 2002. http://www.jiscmail.ac.uk/files/DC-ENVIRONMENT/EoI_Bandholtz.doc

FP6KT: Knowledge Technologies Workshop. Luxembourg, 15 May 2002. Summary report. (ISTweb | KA3 | IAF) http://www.cordis.lu/ist/ka3/iaf/iaf_workshop.htm.

IMAGI: Interministerial Committee for Geoinformation (IMAGI). http://www.imagi.de/welcome_eng.htm

ISO 13250: Topic Maps (ISO/IEC FCD 13250:2000). Prepared by: ISO/IEC JTC1/SC34 - Document Description and Processing Languages. http://www.ornl.gov/sgml/sc34/document/0058.htm

IST: Knowledge Technologies within the IST 2002-2006 programme. Some reflections. Internal report on Knowledge Technologies (aka IRG Report). Luxembourg 2002.

SNS: Federal Environmental Agency of Germany, SchlumbergerSema: Semantic Network Services. Research Project UFOPLAN-Ref. No. 20111612, promoted by BMU/UBA, Germany

UDDI: UDDI Version 3.0 Specification. Open Draft, Dated 03 July 2002. http://www.uddi.org/pubs/UDDI-V3.00-Open-Draft-20020703.htm

VV UDK/GEIN: Verwaltungsvereinbarung zwischen Bund und Ländern über den gemeinsamen Betrieb und die gemeinsame Entwicklung und Pflege des Metainformationssystems Umwelt -Datenkatalog UDK und des Umweltinformationsnetzes Deutschland GEIN.

W3C (2001): Web Services Activity: Web Services Description Language (WSDL) 1.1. W3C Note 15 March 2001. http://www.w3.org/TR/wsdl

W3C (2002): Web Service Description Usage Scenarios. W3C Working Draft 4 June 2002. http://www.w3.org/TR/2002/WD-ws-desc-usecases-20020604/

XMLe: XML Europe 2002. Down to business: Getting serious about XML. Barcelona, 20-23 May 2002. Proceedings. http://www.idealliance.org/papers/xmle02/

XTM: Members of the TopicMaps.Org Authoring Group: XML Topic Maps (XTM) 1.0. TopicMaps.Org Specification. http://www.topicmaps.org/xtm/index.html

Semantic Network Services

Maria Rüther (Umweltbundesamt) Thomas Bandholtz (SchlumbergerSema)

sns

Arbeitskreis "Umweltdatenbanken„ der GI-Fachgruppe 4.6.1
Workshop 2003 in Berlin (19./20. Mai 2003)

Research project UFOPLAN-Ref. No. 20111612, promoted by BMU/UBA, Germany

UFOPLAN 201 11 612

- **Erstellung eines semantischen Netzworkservice (SNS) für das Umweltinformationsnetz Deutschland - German Environmental Information Network (*gein*®) –**
- Mai 2001 – Dezember 2002

Ziele

- Integration der in *gein*® verwendeten Thesauri
- Quervernetzung zwischen Thema-Raum-Zeit
- Automatische Verschlagwortung verbessern
- Funktionalität als Web Service bereit stellen

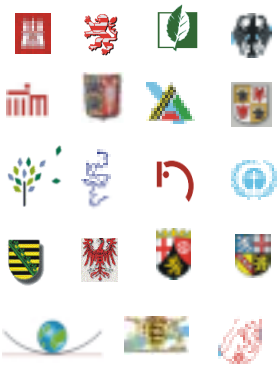
Inhaltliche Komponenten

- **UmThes®** (Themen)
  1:1 + neue Rechtschreibung + Phonetik
- **GeoThesaurus** (Orte)
  1:1
- **Umweltkalender** (Ereignisse)
  1:1 + Verweise im Internet
- **plus** Assoziationen von Ereignissen zu Orten und Themen

Integration in gein®

- UmThes® und GeoThesaurus über Web Service
- Automatische Verschlagwortung über Web Service
- Pflege des Umweltkalenders in SNS – bisher ohne spezielle Anwendungsoberfläche.

80 Behörden - 1 gemeinsames Vokabular



> 80 Informationsanbieter

> 200.000 einzelne Webseiten.

> 500.000 Informationen aus angeschlossenen Datenbanken

gein® + UDK SNS

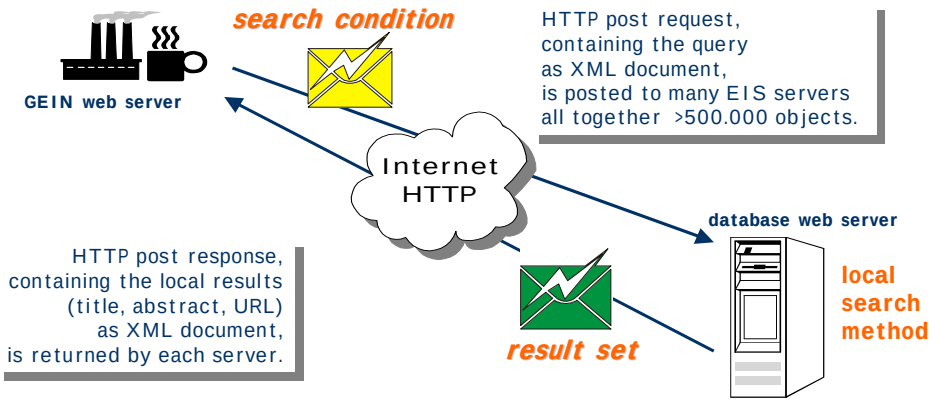
German Environmental Information Network (gein®):
Web crawler and automated search engine features. Distributed query (SOAP) to include database objects.

Umweltdatenkatalog (UDK):
Manually edited metadata, includes pointers to off-line information.

Semantic Network Services (SNS):
Web Services access to shared ontology and automated indexing.

Verteilte Recherche (seit 1999)

„gein®-POST-Schnittstelle“, „G2K-Profil“

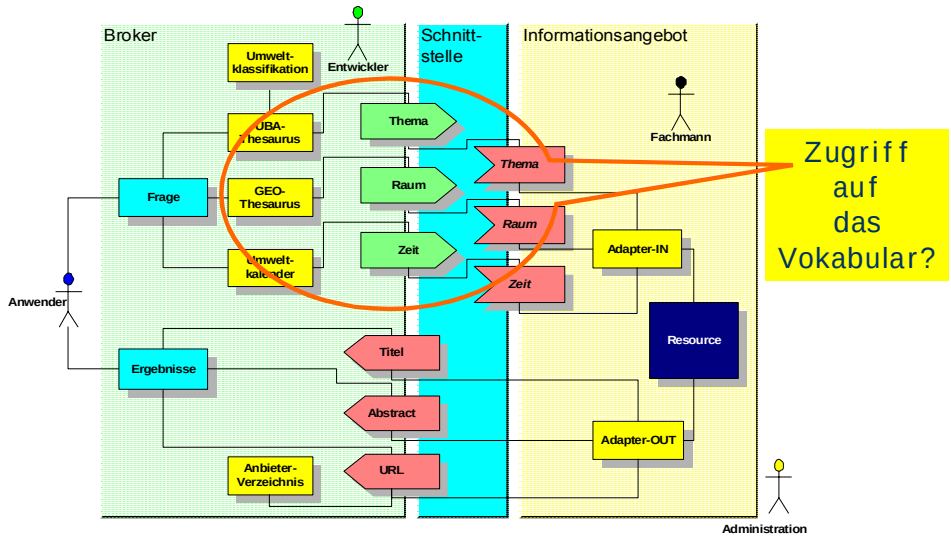


Currently discussed to use Web Services here also.

Angeschlossene Datenbanken

FloraWeb	Daten und Infos zu Wildpflanzen und zur Vegetation Deutschlands
GISU	Geographisches Informationssystem Umwelt
GSBL	Gemeinsamer Stoffdatenpool Bund/Länder
UDK	Umweltdatenkatalog des Bundes und der Länder
UFORDAT	Umweltforschungsdatenbank
ULIDAT	Umweltliteraturdatenbank
UOK	Umweltobjektkatalog Bayern
UPB	Umweltprobenbank des Bundes
WebDIM	Umweltdaten online des MUNLV NRW

Integration über das „G2K-Profil“ (1999)



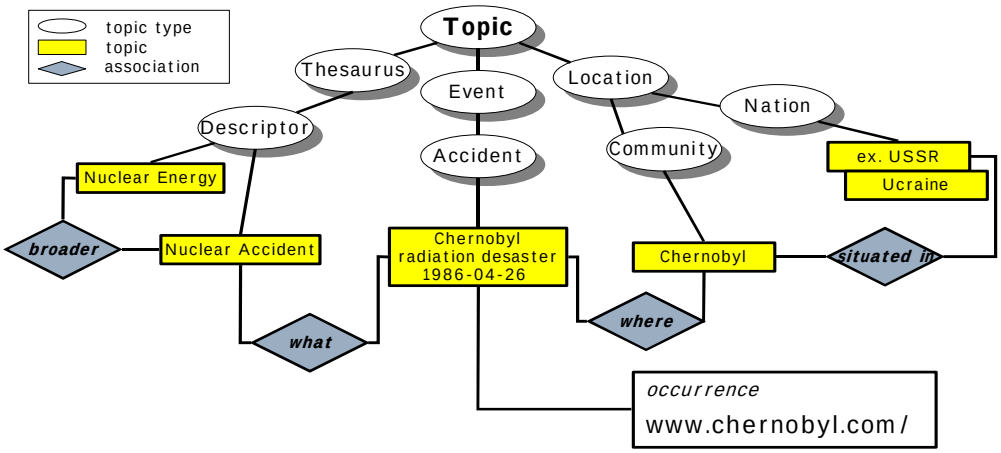
The gein® Ontology

Bi-lingual German/English

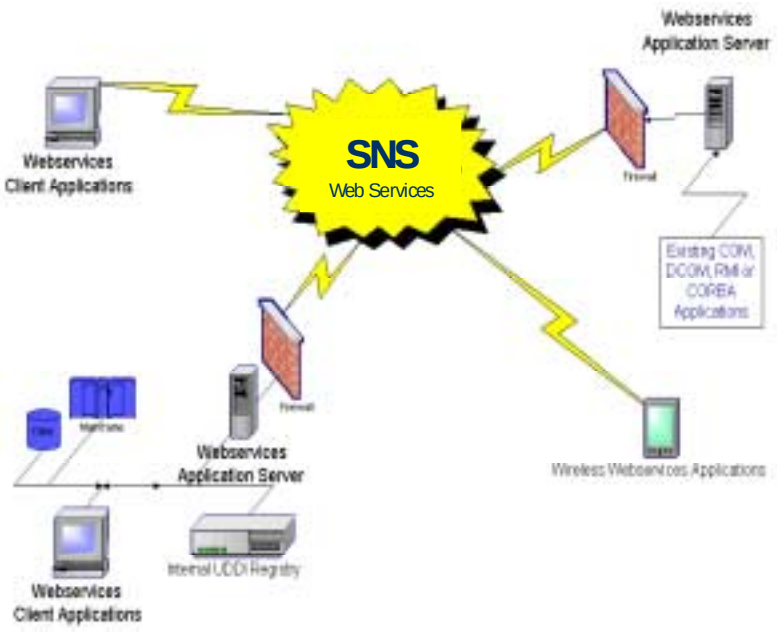
1. a **thesaurus** of currently 39,143 environmental terms (UMTHES®, German source of GEMET)
2. a **gazetteer** including the intersections between 48,213 geographical objects of all kinds;
3. a **chronology** of historical and contemporary events that have affected the environment, currently 544 events.

till 2003 these have been three distinct collections.

Integration in a Topic Map



SNS Web Services



(Original picture by webservices.org)

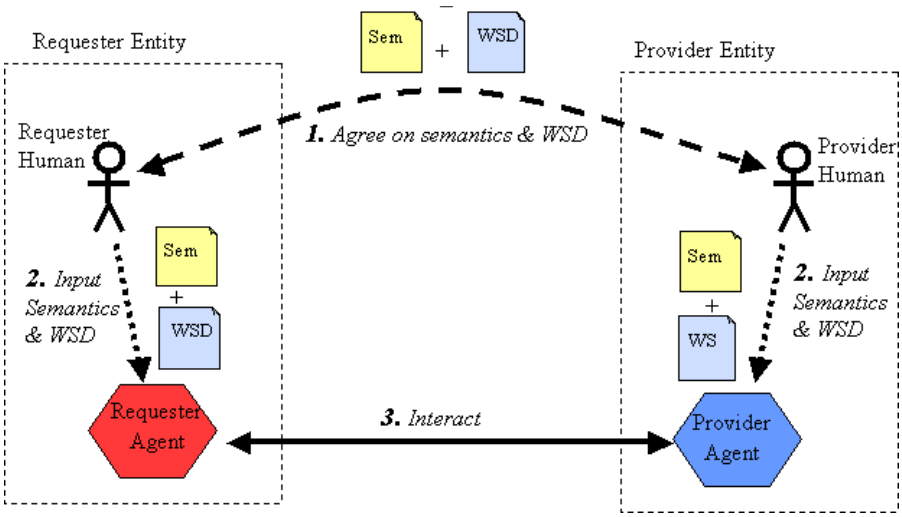
Die drei SNS Services 2003

- **findTopics** sucht nach Topics mit den übergebenen Parametern. Die Rückgabe ist ein Topic Map Fragment, welches die gefundenen Topics enthält, aber noch keine Assoziationen.

- **getPSI** (PSI= „Published Subject Identifier“) übergibt die vollständigen Eigenschaften eines gegebenen Topic. Assoziierte Topics können in wählbarer Tiefe mit ausgegeben werden.

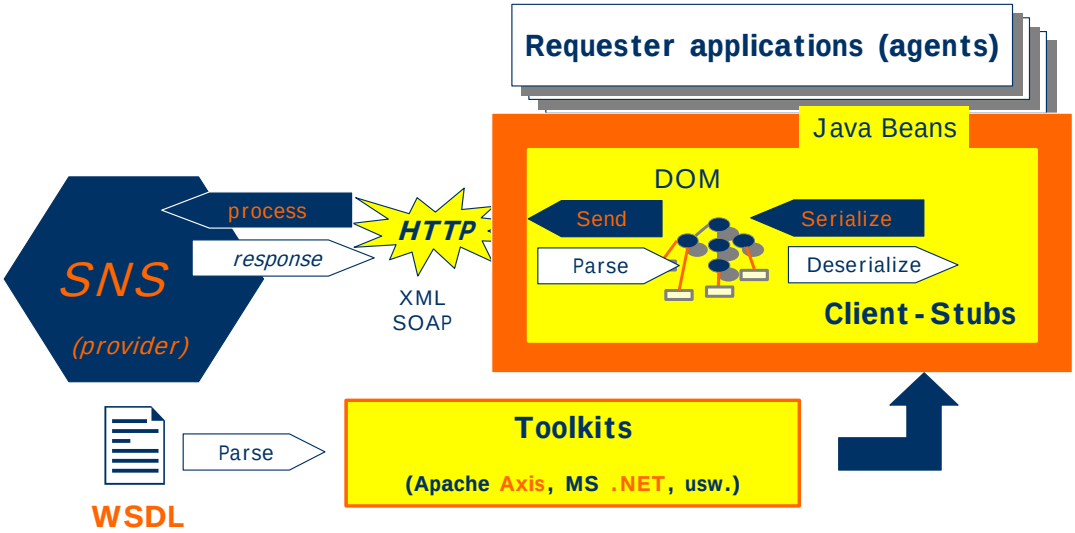
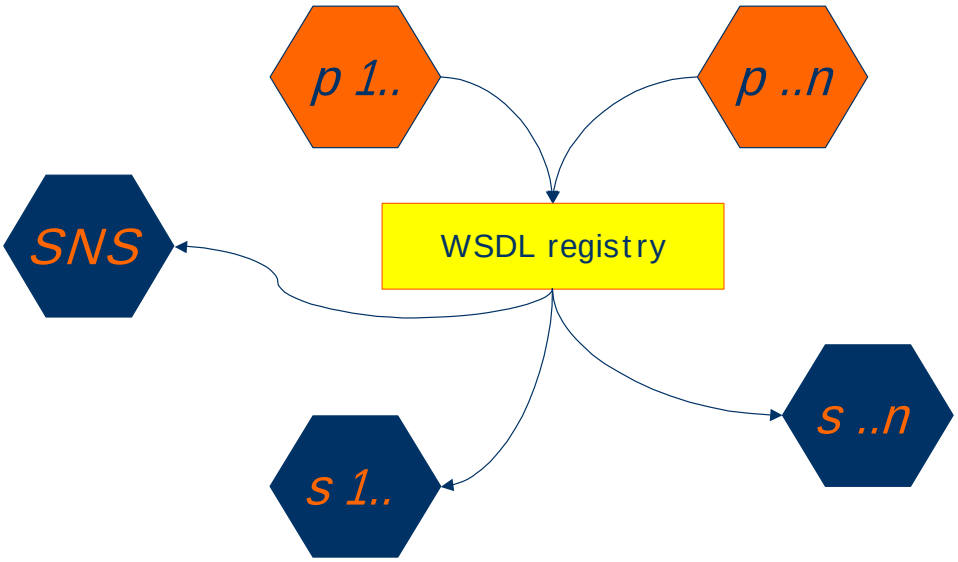
- **autoClassify** stellt die automatische Indexierung zur Verfügung. Der zu indexierende Text kann als Konstante (*document*) oder als Webadresse (*url*) übergeben werden. Das Ergebnis ist eine gewichtete Liste von signifikanten Topics.

Web Services: Basic Architectural Roles



from http://www.w3.org/TR/ws-arch/

Einfache Integration durch Standards



Trennung von Prozessen und Services