

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

Gerlinde Knetsch, Marianne Rappolder¹
Eric Weihs²

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 den 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.

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“.

¹ Umweltbundesamt, Postfach 33 00 22, 14191 Berlin
E-Mail: gerlinde.Knetsch@uba.de
E-Mail: marianne.Rappolder@uba.de

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

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.

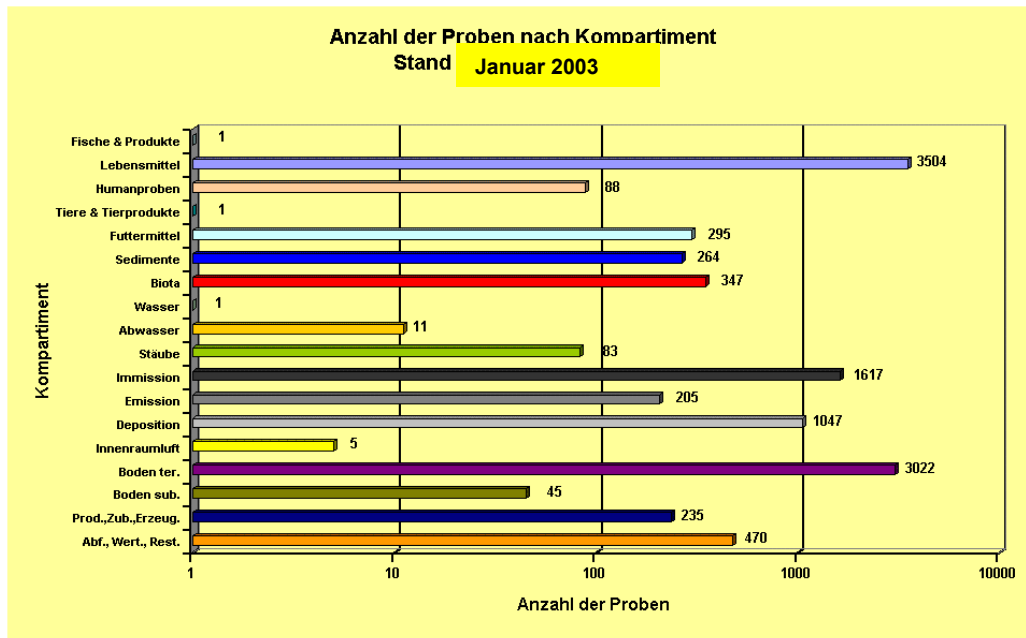


Abb. 1 Übersicht zum Datenbestand in der Datenbank Dioxine

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 Administrationsrechte 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** (Probenentnahmereinrichtung, 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, jahreszeitenspezifische und kongenere-spezifische Berichte möglich. Generierte Kreuztabellen können nach EXCEL und einem statistischen Softwareprogramm SPSS³ exportiert werden, um weitere Auswertungen und grafische Darstellungen z.B. für die Berichtserstellung vornehmen zu können.

³ Statistik-Software der Firma SAS (Superior Performing Software Systems)

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.

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 Webservices unter Berücksichtigung der Rahmenbedingungen aus der BundOnline2005 zu erfüllen

- Zugang zu den Daten über das Internet nach dem W3C⁴-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)**⁵, 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.

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. BgVV, Berichte der B/L-AG Dioxine,
- Glossar mit fachlichen Erläuterungen und Erklärung von Abkürzungen angeboten werden.

⁴ World Wide Web Consortium (W3C)

⁵ <http://www.umweltdatenkatalog.de/wwwudk/V-UDKServlet>

Mit diesem Informationsangebot für die Öffentlichkeit wird das UBA den Anforderungen aus dem Umweltinformationsgesetz 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-Umgebung
- 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)**⁶ 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.

Die Einbindung des Dioxin-Datenbank für die Öffentlichkeit in den virtuellen **Umweltdatenkatalog des Bundes und der Länder (vUDK)**⁷ findet über die

⁶ http://www.gein.de/index_de.html

⁷ <http://www.umweltdatenkatalog.de/wwwudk/V-UDKServlet>

Beschreibung des Objekts in die Objektklasse „Datensammlung/Datenbank“⁸ 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 2005⁹

Ziel 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 bereitgestellten 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:

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, Stylesheets, 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.

⁸ <http://www.umweltdatenkatalog.de/wwwudk/jsp/DataDetailBottom.jsp?uURL=5>

⁹ <http://www.bund.de/bundonline2005>

3.4 Lösungsprinzipien

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

- Dynamische Web-Seiten unter Einsatz der serverseitigen Software PHP¹⁰ und ASP¹¹ 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.

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.

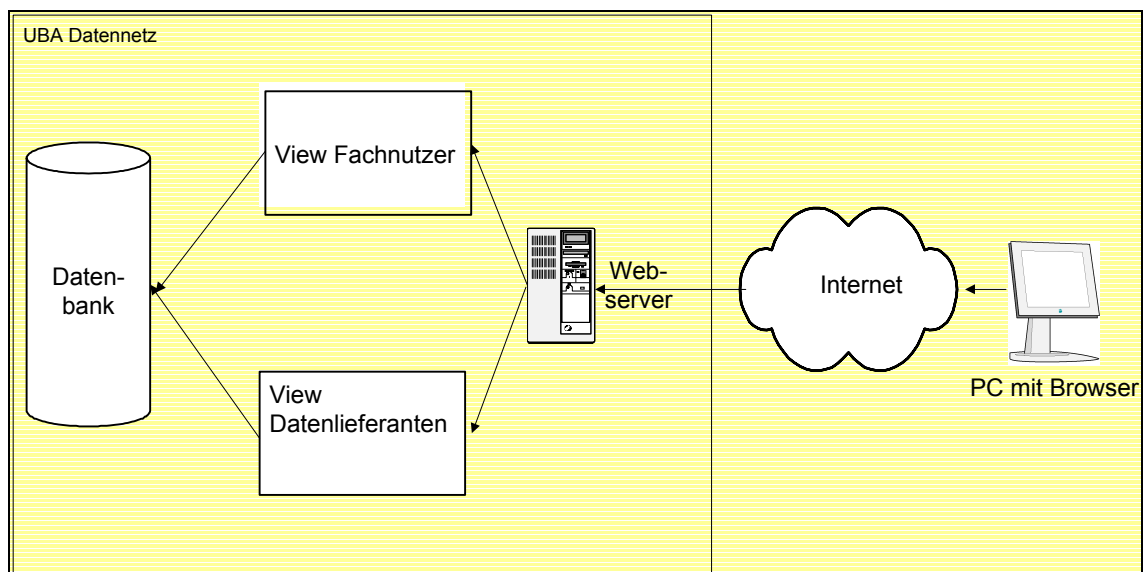


Abbildung 1: Prinzipschema für die Lösungsvariante „Dynamische Web-Seiten“, Quelle: F+E-Vorhaben des UBA, FKZ 201 62 215

¹⁰ Personal Home Page Tools, Open Source Skriptsprache zur Erstellung dynamischer Webseiten

¹¹ Active Server Pages, Technologie von Microsoft zur serverseitigen Erstellung dynamische HTML-Seiten

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**TM 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 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. Basiskomponenten des Web-Service sind Apache, Tomcat, Xerces o.ä. (Umwandlung XML in html, pdf etc über Stylesheets), Open Source Software (Java-Beans) unter Beibehaltung des W3C Standards.

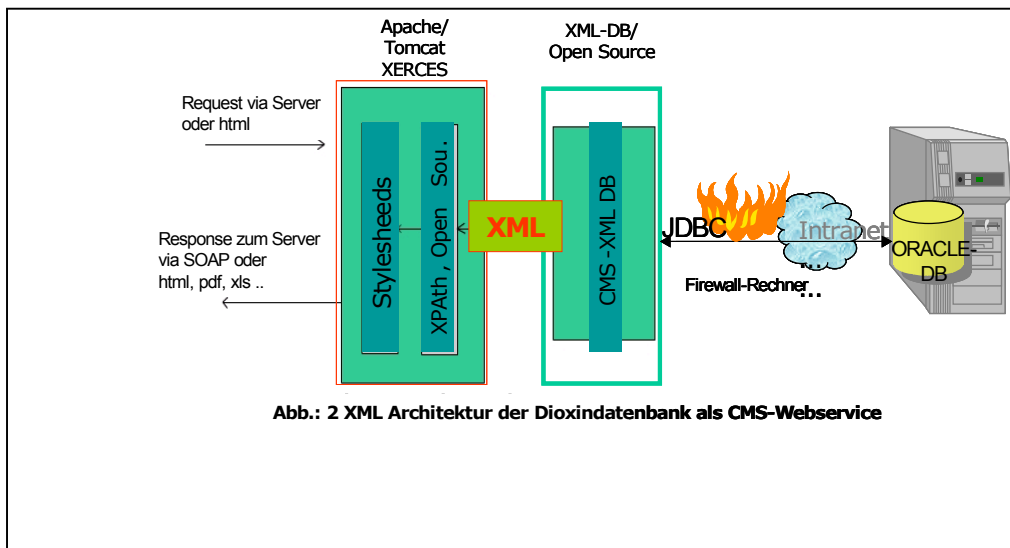


Abb.: 2 XML Architektur der Dioxindatenbank als CMS-Webservice

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 UFOPLAN 2001