Handling Heterogeneous Data in a Federal System Using XML
Demonstration of the Web Service for the German POP-Dioxin Database

Nina Brüders¹, Gerlinde Knetsch¹, Erich Weihs²

Abstract
Since 1991, data from more than 220 monitoring programs which cover pollution with dioxins and other chlororganic substances of several environmental and human compartments are recorded and administered in an ORACLE based client-server system.
In cooperation between the Bavarian State Ministry of the Environment, Public Health and Consumer Protection (BStMUGV) and the German Environmental Agency, a web service based on XML technology has been developed to enable profile specific access to the database for three user groups: public, provider of data and special user. The web service for the German Dioxin Database is now implemented and can be visited at: www.POP-DioxinDB.de.

1. Introduction
Monitoring programs for dioxins and other POPs are carried out in Germany for different reasons, with different aims and scopes in different human and environmental compartments and in different spheres of responsibility. Therefore, data on dioxin and other POP contaminations are heterogeneously collected and stored at various Federal and Laender authorities in Germany. For effective use of this existing data in order to draw an overall picture of exposure it is vital to bring the data in a coherent form. Harmonisation of data from different sources is one of the key problems in terms of dealing with environmental data today. In Germany the federal states (Laender) as well as the federal Government have to report data about the state of the environment in various forms and with different focuses on the ground of different legal regulations.

For this purpose a standardized format as well as a collection of data at regular intervals beforehand in a central database from which the required data could easily be extracted as needed is important.

Relevant data are kept mostly in excel sheets, in access tables or in text files. One step towards standardization and harmonisation is the supply with technical solutions which are system and platform independent and highly scalable.

One example for such a centralized data management already exists with the German POP-Dioxin Database. A recently developed web service allows data retrieval and upload via the internet and it is available now at: http://www.POP-DioxinDB.de.

¹ Federal Environmental Agency, Wöllitzer Platz 1, D-06844 Dessau
email: nina.brueders@uba.de, Internet: http://www.POP-DioxinDB
² Bavarian State Ministry of the Environment, Public Health and Consumer Protection, Rosenkavalierplatz2, D-81925 Munich, Germany, email: erich.weihs@stmugv.bayern.de,
2. The German POP-Dioxin Database

In 1991, the 37th Conference of Laender Environment Ministers of Germany issued an administrative agreement to compile, document, and evaluate data from monitoring and surveillance programs at the federal and the Laender level. Substances to be included are polychlorinated dibenzo-p-dioxins (PCDD), polychlorinated dibenzofurans (PCDF), polychlorinated biphenyls (PCB) and other chlororganic substances.

Since then data of more than 220 monitoring programmes of several environmental and human compartments are recorded and administrated in an ORACLE based client-server system.

In the complex relational model a wide variety of possible information coming with analytical values can be stored and retrieved. Examples for meta information - specific for each compartment - are:
- reason and target for the investigation (e.g. incident investigation, detection of background contamination)
- location data (e.g. community indicators, easting/northing values, dioxin-relevant industrial area)
- sampling procedure (e.g. method of sampling, transport of samples)
- laboratory data (e.g. storage conditions, regulation for the preparation of samples)

Only with this additional information meaningful sharing of data is possible.

Fig. 1: Schematic representation of the dataflow. The DIOXIN-database is maintained by the Federal Environment Agency (UBA) for environmental matrices including emissions and by the Federal Office of Consumer Protection and Food Safety (BVL) and the Federal Institute for and Risk Assessment (BfR).
3. The web service for the POP-Dioxin Database

The project for development of a web service for the POP-Dioxin Database was designed and financed through cooperation of the Bavarian State Ministry of the Environment, Public Health and Consumer Protection (BStMUGV) and the German Environmental Agency. It uses XML (Extensible Markup Language) technology and provides the following features:

- public access to aggregated information
- differentiated access to the database for different user groups with different rights
- data-recall facility
- navigation to all meta data and the analytical results (concentrations of chemical substances with calculated and aggregated values).
- fast navigation (branching)
- overview: lists, details
- download of data for further evaluation
- data input and administration
- download (templates)
- upload and import into the database
- quality assurance, data release
- administration of master data
- environmental internet search engine connection (gein®)

Fig. 2: Screenshots of Examples for the service for public users (left) and for special users (right).

It is important to define roles and rules for data access. Three user groups are identified: public, data providers and special user. The profile specific retrieval is one key aspect for the acceptance of this system. To supply these functions the front-end uses XML style sheets to generate individual dynamic websites for every user. Another aspect is the down- and uploading function of data for authorised users via the internet.

The data provider is able to access and manipulate own data and to decide for whom they want to make own data readable. Data providers can use the central database to store their own data for different purposes and retrieve it at any time.

http://www.gein.de

---

3 http://www.gein.de
The datasets are listed together with a contact information, allowing restricted users to ask data owners individually for permission of data access.

The website offers an overview of existing data and contact information as well as access to analytical data. It marks the first steps on the way to decentralisation of data import.

The web service involves many components which are hidden:

- programming language Java (SDK, J2EE)
- full implementation with XML
- Apache (Webserver, https + SSL)
- Apache TOMCAT (By-pass- Server 1, ServletEngine- Server 2+ 3)
- Apache Axis (framework for web services)
- Exolab CASTOR (marshalling and unmarshalling)
- StrutsCX (XML- supplement of Apache Struts frameworks to build web-applications with Java, MVC concept)
- Xerces / Xalan (XML- Parser / XML- Prozessor, transformation)
- Log4J (Application- logging)
- ORACLE 9i (database server), JDBC
- JMeter and JUnit (performance tests, automatic tests, internal)
- CVS (version management, internal)
- Jakarta POI (create Excel- files)

A multi tier architecture separates the different layers (presentation and application layer, web service layer, database layer) and allows change of single components, e.g. replacement of the oracle database by a XML-database without changing the business logic.

The transformation of the complex relational model into an XML model and vice versa prepares the data for easy connection of exposure data with substance data as held in substance databases such as GSBL, UICLID etc.

4. Conclusions

The German POP-Dioxin Database allows the central storage of exposure data for dioxins and other POPs including a great variety of meta data. The storage of data for the relevant congeners allows flexible calculation even for older data.

The recently implemented web service provides direct access to data via the internet for three different user groups and will encourage the users to exchange data and information. Furthermore it is an instrument to harmonise heterogenic databases.

Bibliography


Copyright © Masaryk University Brno, Brno 2005. ISBN: 80-210-3780-6