

The Design of a Multi-catalog System for Distributed Environmental Information Resources

Wolf-Fritz Riekert¹, Gerlinde Wiest², Oliver Günther³, Jan Röttgers³,
Jo van Nouhuys⁴, Birgit Mohaupt-Jahr⁵, and Jürgen Seggelke⁵

Abstract

A multi-catalog approach is presented that simplifies the utilization of the Web technology for the information transfer between suppliers and users of distributed environmental information resources. A Web-based facility, also referred to as *Locator and Communication Service Environment* (German acronym: VKSU), was designed that allows users in the German Federal Environment Agency to search for and directly access environmental information resources in the intranet of the agency. The core of VKSU is a *Broker* that integrates a number of specialized digital catalogs rather than build one single ‘universal catalog’ which records as many resources as possible. This allows the VKSU to make use of the existing information infrastructure in the UBA that already comprises a variety of information catalogs and meta-information systems.

1. Starting Point

In the last decade, a large number of information systems, databases, and data collections has been compiled in the German Federal Environment Agency UBA (Umweltbundesamt). These information resources are not only important for their owners but also for users in other departments of the UBA. Particularly the combination of information originating from multiple departments yields an added value to the existing information. There is an increasing tendency of the owners of the information to make their ‘information treasures’ available to a wider community of users, thus raising the benefit derived from the information tremendously.

¹ Dr. W.-F. Riekert, FAW Ulm, PO Box 2060, D-89010 Ulm, riekert@faw.uni-ulm.de

² G. Wiest, ISB GmbH, Karlstr. 52-54, D-76133 Karlsruhe, wiest@isb-ka.de

³ Prof. Dr. O. Günther and J. Röttgers, Humboldt-Universität zu Berlin, Spandauer Str. 1, D-10178 Berlin, guenther@wiwi.hu-berlin.de, jroettg@wiwi.hu-berlin.de

⁴ J. van Nouhuys, Condat GmbH, Alt Moabit 91d, D-10559 Berlin, nouhuys@condat.de

⁵ B. Mohaupt-Jahr and Dir. u. Prof. Dr. J. Seggelke, Umweltbundesamt, PO Box 330022, D-14191 Berlin, Birgit.Jahr@uba.de, Juergen.Seggelke@uba.de

From the users, an urgent need for information from foreign sources has shown up in many cases, particularly, if the users have to compile studies and expert opinions under time pressure. This need for information can be filled by using the traditional communication channels such as personal contacts or publications provided by a library. However, the publications used are often generated from older versions of databases and data collections and do not reflect their latest revision states.

Since the mid-nineties, the upcoming Web technology has proved its practical value in many Internet applications. Moreover, this technology can also be successfully applied in the Intranet of organizations such as the UBA. Particularly, the Web can be used as a powerful communication channel between suppliers and users of environmental information resources in the UBA.

This idea underlies the project VKSU ('Verweis- und Kommunikationsservice Umwelt', i.e., 'Locator and Communication Service Environment') which was started by the UBA after various preparatory activities, particularly the development of a prototype for the German Environmental Information Network GEIN (Riekert et al. 1997, Tochtermann et al. 1997). The main purpose of VKSU is to simplify the utilization of the Web technology for the information transfer between suppliers and users of environmental information. In a first phase, the UBA commissioned a team from FAW Ulm, CADMAP GmbH Berlin / Dipl.-pol. Jo van Nouhuys, and Enterprise Consulting Prof. Günther, Berlin to elaborate a conception for VKSU and to start exploratory steps for its realization (Seggelke/Mohaupt-Jahr 1997, Mohaupt-Jahr et al. 1998).

2. The Locator and Communication Service Environment

The main objective of the project was the specification of two facilities referred to as *Locator* and *Communication* services.

The Locator service is to provide *metadata* on environmental information resources which have been released by their owners for use in the Intranet of the UBA. Beyond metadata, i.e., descriptions of the information resources and their owners, users demand facilities for the direct access to the original information resources. For this purpose, the Communication service establishes telecommunication connections to many (in the ideal case: *all*) of the information resources described in the catalog, particularly various UBA-internal databases on topics such as toxic substances, technologies, the state of the environment, and bibliographic information, as well as external systems such as the meta-information system for climatic data ZUDIS.

The user requirements were ascertained and documented thoroughly in a number of interviews and workshops with members of various departments in the UBA (Röttgers/Günther 1998). As a first approach to fulfill these requirements, a demonstrator for the intended system was developed and put up for discussion. For this purpose, we made use of concepts and solutions from existing developments such as

the Environment and Traffic Information Service Baden-Württemberg UVIS⁶ (Henning et al. 1998), the GEIN prototype (Riekert et al. 1997, Tochtermann et al. 1997), the Environmental Data Catalog UDK⁷ (Günther et al. 1996), and the Geographic Information System Environment GISU⁸ (Balzer/Nouhuys 1998). In addition to these systems, the designed solution was greatly influenced by developments such as the European Catalogue of Data Sources CDS (Kramer et al. 1997) of the European Environment Agency (EEA) and the Global Environmental Information Locator Service GELOS (Kleih 1998) from the G7 project Environment and Natural Resources Management (ENRM). Systems like these contribute essential concepts and functions to the proposed solution. Rather than operating these functions in parallel, VKSU will integrate these functions through a common front-end.

For this purpose, we designed VKSU as a multi-catalog system. The idea is to integrate a number of specialized digital catalogs rather than build one single ‘universal catalog’ which records as many resources as possible. This allows us to make use of the existing information infrastructure in the UBA that already includes a variety of information catalogs⁹ such as the bibliographic catalogs BIBLIODAT and ULIDAT, the research project database UFORDAT, and the FINDEX meta-information system for chemical databases, as well as the metadatabases of the systems CDS, UDK, and GISU mentioned above (Umweltbundesamt 1996).

This collection of catalogs will be connected to each other so that queries can be processed simultaneously in all the different catalog systems. Our research has shown that a *Broker* is to be put in between the user interface also referred to as front-end and the digital catalog systems forming the back-end of the projected system. The broker translates uniform search queries from the front-end into the query languages of each of the connected digital catalog systems at the back-end. In a similar way, the broker unifies search results contributed by the catalog systems at the back-end before they are forwarded to the user front-end.

On the front-end side, the Web technology has been chosen for the communication between the users and the broker. The broker will be installed on a Web server in the Intranet of the UBA and will be accessible through a Web interface. The users in the UBA can interact with the broker by using a standard Internet browser without any specific software installation on their computers. Therefore, the chosen approach could also be used to supply the general public with environmental information. This will serve to join the future development lines of the Internet-based GEIN system and the Intranet-based VKSU system. For this reason, the conception of VKSU greatly influenced the requirement specification for GEIN 2000, i.e., the next devel-

⁶ The German acronym UVIS stands for ‘Umwelt- und Verkehrs-Informationsservice’.

⁷ The German acronym UDK stands for ‘Umweltdatenkatalog’.

⁸ The German acronym GISU stands for ‘Geographisches Informationssystem Umwelt’.

⁹ In the sequel, the term catalog will be used in a sense that tacitly includes meta-information systems and metadatabases.

opment step of GEIN which is also to be presented at the World Exposition EXPO 2000 in Hanover.

A special solution has been chosen on the back-end side to connect the broker with the digital catalogs to be integrated. For this purpose, we compiled a synopsis of the metadata schemas of various existing catalog and meta-information systems in order to identify corresponding data fields in the respective systems. It turned out that VKSU should support at least the following search criteria: thematic keywords, temporal and spatial coverage, suppliers of the information, and a classification of the information resources to be searched for. We found out that the Web protocol HTTP is not well suited to transmit these search criteria since HTTP lacks most of the features which are necessary for the mediation between heterogeneous catalog systems. In particular, the broker needs to know what attributes can be searched for in the digital catalog systems on the back-end in order to translate the query syntax supported on the front-end side into the various formats which are understood by the respective systems on the back-end side.

As the most promising solution to transmit these search criteria we suggest the use of Z39.50 (Z39.50 1998) or one of its profiles, in particular GILS (GILS 1998) or GELOS (Bjarnason et al. 1997). The broker would serve as a Z39.50 client while each of the connected digital catalogs would serve as a Z39.50 server. Both sides communicate with each other to define what attributes the broker can search for in each of the digital catalogs. Z39.50 or a Z39.50 profile would avoid different names being used for the same attribute in different catalogs (e.g., 'time period' and 'temporal coverage' as attribute names for the temporal coverage of the information resource). The GILS and GELOS profiles also provide spatio-temporal extensions of the Z39.50 protocol which support the transmission of spatially and temporally-related search criteria which are particularly important for retrieving many types of environmental information resources.

3. The Multi-catalog Architecture

Figure 1 shows the multi-catalog architecture which underlies the design of VKSU.

1. On top of the schema, the *VKSU Broker* integrates various catalog systems and meta-information systems such as the environmental data catalogs UDK and CDS, the documentation databases BIBLIODAT, ULIDAT, and UFORDAT, as well as the meta-component of the geographic information system GISU.
2. These *catalog systems* provide specific metadata, i.e., descriptions of environmental information resources such as printed publications, databases, multimedia documents and information services.
3. In many – but not all – cases, these information resources are made available in the World Wide Web through various *information servers*.

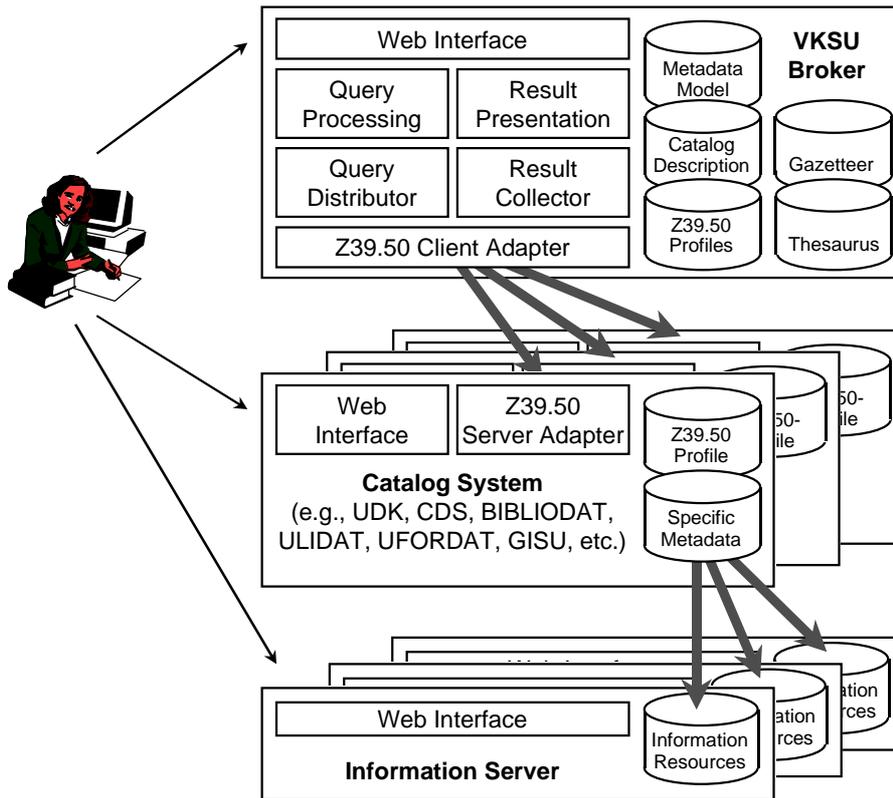


Figure 1

The VKSU broker integrates catalog systems and information servers.

The design of the VKSU broker comprises six functional components:

1. A *Web Interface* allows users to interact with the broker. The query formulation should be possible at various levels of complexity ranging from simple textual searches to logical combinations of specific criteria such as thematic keywords, temporal coverage, geographic regions, suppliers, and type information.
2. A *Query Processor* translates the user's inputs into an intermediate representation that is independent of the specific style of interaction with the Web interface.
3. The *Query Distributor* determines which catalog systems are concerned with the query. The query is then transformed according to the specific Z39.50 profiles of the respective catalog systems.
4. A *Z39.50 Client Adapter* transmits the query to the catalog systems to be addressed by using the Z39.50 transfer format. The client adapter also receives the

results of the query and translates it into the internal metadata format of the VKSU broker.

5. The *Result Collector* collects and merges the lists of results from the various catalog systems thus producing a unified list of information resources matching the user's query.
6. The *Result Presentation* component generates appropriate views of the list of results and displays them by using the Web Interface. The results are shown as hyperlinks if the underlying information resources are available in the Web. The presentation component also includes an export interface to generate external representations of the retrieved metadata, e.g., for the purpose of reporting to the European Environment Agency.

Being a broker rather than a metadatabase, VKSU does not persistently store metadata on environmental information resources. Nevertheless VKSU makes use of five major databases:

1. The *Metadata Model* describes the meta-object classes understood by VKSU including attribute descriptions and inheritance hierarchy relationships. The basic idea is to keep the metadata model small, thus reducing the metadata schemas of the integrated catalog systems to a common denominator.
2. VKSU possesses *Catalog Descriptions* for each integrated catalog. In particular, this includes the information about which classes of metadata are provided by which catalog.
3. A database of *Z39.50 Profiles* is used by the Query Distributor and the Z39.50 Client Adapter for the translation and transmission of the user's queries.
4. A *Thesaurus* is used by the query processor and the Web interface to support the formulation and processing of thematic search queries.
5. Similarly, a *Gazetteer* containing names and geometries of geographic regions supports the textual and graphic formulation and processing of geographic search queries.

The operation of the existing catalog systems to be integrated by the VKSU broker may remain unchanged. However the catalog systems need to be enhanced by an additional Z39.50 server adapter that understands the transfer format used by the Z39.50 client adapter of the VKSU broker. If necessary, the Z39.50 server adapter makes use of a *Z39.50 Profile* which is appropriate for querying the metadata managed by the catalog system. In addition, a *Web interface* is very useful, if the users want to directly access the catalog system rather than get filtered information via the VKSU broker. Most of the existing catalog systems already possess such Web interfaces, which will be accessible from the VKSU Web interface through hyperlinks.

Many of the information resources managed by the catalog systems integrated by VKSU are supplied by information servers in the World Wide Web. The Internet

addresses of these information resources are known to the respective catalog systems from where they can also be retrieved via the VKSU broker. Therefore it is possible both for the VKSU broker and the catalog systems to put the user through to the original information resources by using the hyperlink mechanism.

4. Results and Outlook

Based on the conception described, the 'treasure' of information in the UBA can be utilized more intensively and in an easier way than in the traditional fashion. Through its Locator facility, VKSU will provide a comprehensive survey of all released information resources in the Intranet of the UBA and on selected information resources in the global Internet. The users can search for the desired information through a comfortable and easy-to-use Web interface that runs on all major computer platforms. The search aids provided by the VKSU Locator include the Environmental Thesaurus of the UBA, a Gazetteer supporting a geographic search vocabulary, and a classification of the available environmental information resources, as well as general descriptions of the integrated catalogs and the information resource types available in their holdings.

After being found with the help of the Locator, released and validated information resources can be immediately accessed through the communication component. The use of the Multimedia Internet Mail Exchange (MIME) facility allows the accessed information to be further processed by using standard office software components such as word processors, graphics programs, and spreadsheets.

VKSU is not a central 'super-system' but an information service which is easy to use and which is only activated on demand. VKSU allows a powerful integration of the documents, data collections, databases, and information systems in the UBA without infringing on the independence of their owners. The responsibility remains at the respective departments of the UBA who own and maintain the information resources. They still decide which of their information resources will be released for use in the Intranet of the UBA.

The conception of the VKSU will be stepwise implemented in the future. For this purpose, the implementation of VKSU is to be continued jointly with the further development of the German Environmental Information Network GEIN. The next step will be a detailed functional specification for the broker solution and the development of a pilot version for the general public, also referred to as GEIN 2000, which will be installed in the Internet and presented at the world exposition EXPO 2000 in Hanover. Finally, the intranet-based VKSU system will be derived from this solution and will be installed at the UBA where it will be successively extended by an increasing number of integrated meta-information systems, catalogs, databases, and information systems.

References

- Balzer, H., v. Nouhuys, J. (1998): GISU – Geographisches Informationssystem Umwelt im Umweltbundesamt. In: (Riekert/Tochtermann 1998).
- Bjarnason, S., Cole, P., Aschbacher, J. (1997): Standard Element Set for GELOS Records. Prepared for the G-7 ENRM project by the Metainformation Topic Working Group. <http://www.g7.nasda.go.jp/G7/GelosEleSet6.html>.
- GILS (1997): <http://info.er.usgs.gov/gils/gils1p.html>.
- Günther, O., Lessing, H., Swoboda, W. (1996): UDK: A European Environmental Data Catalog. Proc. Third International Conference in Integrating GIS and Environmental Modeling. NCGIA, Santa Barbara (USA). http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/guenther_oliver/my_paper.html.
- Henning, I., Wiest, G., Gaul, M., Rainbold, E., Schultze, A., Riekert, W.-F. (1998): UVIS: Umwelt- und Verkehrs-Informationsservice Baden-Württemberg. In: (Riekert/Tochtermann 1998).
- Kleih, M. (1998): Die Informationsdienste des CEO: EWSE und GELOS-G7 – Zwei selbst-populierende, internetbasierte Systeme zur Förderung des Austausches auf den Gebieten der Fernerkundung und des Umweltschutzes. In: (Riekert/Tochtermann 1998).
- Kramer, R., Nikolai, R., Rolker, C., Bjarnason, S., Jensen, S. (1997): Interoperability Issues of the European Catalogue of Data Sources (CDS). Proc. of the 2nd IEEE Metadata Conference, Silver Spring, USA. <http://computer.org/conferen/proceed/meta97/papers/rkramer/rkramer.html>.
- Mohaupt-Jahr, B., Seggelke, J., Riekert, W.-F., Wiest, G., Günther, O., Röttgers, J., v. Nouhuys, J. (1998): Die Konzeption des Verweis- und Kommunikationsservice Umwelt im Umweltbundesamt. In: (Riekert/Tochtermann 1998).
- Riekert, W.-F., Tochtermann, K. (Hrsg.) (1998): Hypermedia im Umweltschutz. 1. Workshop. Metropolis-Verlag, Marburg.
- Riekert, W.-F., Tochtermann, K., Wiest, G., Gaul, M., Seggelke, J., Mohaupt-Jahr, B. (1997): Fach-, raum- und zeitbezogene Katalogisierung und Recherche von Umweltinformationen auf dem Internet. In: Fuhr, N., Dittrich, G., Tochtermann, K. (Hrsg.): Hypertext – Information Retrieval – Multimedia '97. Proceedings HIM'97. Universitätsverlag Konstanz, 1997.
- Röttgers, J., Günther, O. (1998): Nutzeranforderungen an Umweltinformationssysteme: Die Fallstudie VKS-Umwelt. In the same proceedings as this paper.
- Seggelke, J., Mohaupt-Jahr, B. (1997): Der Verweis- und Kommunikationsservice des Umweltbundesamts – Ein Modellfall für das Umwelt-Intranet? In: Geiger, W. et al. (Hrsg.): Umweltinformatik '97. 11. Int. Symposium, Metropolis-Verlag, Marburg.
- Tochtermann, K., Riekert, W.-F., Wiest, G., Seggelke, J., Mohaupt-Jahr, B. (1997): Using Semantic, Geographical, and Temporal Relationships to Enhance Search and Retrieval in Digital Catalogs. In: Peters, C., Thanos, C. (eds.): Research and Advanced Technology for Digital Libraries. Proceedings ECDL'97. Springer-Verlag, Berlin.
- Umweltbundesamt (1996): IT-Rahmenkonzept 1997, Berlin.
- Z39.50 (1998): <http://lcweb.loc.gov/z3950/agency/>.