The Design of a Multi-catalog System for a Public Environmental Information Network

Wolf-Fritz Riekert[1,2]

Abstract

A multi-catalog approach is presented for the German Environmental Information Network (GEIN) which is being constructed by the German Federal Environmental Agency UBA (Umweltbundesamt) and other environmental administrations in Germany. Towards this end, a Web-based facility, also referred to as Locator and Communication Service Environment (German acronym: VKSU), was designed that allows users to search for and directly access environmental information resources provided by various suppliers. The core of VKSU is a Broker that integrates a number of existing catalog and meta-information systems rather than build one single 'universal catalog' which records as many resources as possible. This allows the VKSU to make use of the existing German meta-information infrastructure that already comprises a variety of information catalogs and meta-information systems. The results of the project are currently being used by a commercial IT solution provider for the further development of the GEIN system.

1.   Starting Point

Since the mid-eighties, there has been an increasing demand for reliable and value-added environmental data and information. So a large number of German administrations, agencies, research organizations, and non-governmental organizations have built powerful information systems to collect, manage, analyze, and present information about the environment. Information servers have been developed for the World-Wide Web to make this information available to a wide community of users from politics, administration, industry, research and the public. To date, a variety of German Web sites provide a rich supply of multimedia documents, information services, and data on environmentally-related topics.

In addition, various governmental agencies and research organizations allow their information to be searched for by installing digital catalogs or meta-information systems on the Web. An example for this approach is a meta-information system referred to as *Information Locator* (figure 1) which was developed by the FAW Research Institute for Applied Knowledge Processing (FAW) for use in the Environmental Information System Baden-Württemberg and the German Environmental Information Network GEIN (Riekert et al. 1997, Tochtermann et al. 1997). Other examples comprise the Environmental Data Catalog UDK ('Umweltdatenkatalog') (Günther et al. 1996), the European Catalogue of Data Sources CDS (Kramer et al. 1997), and the Global Environmental Information Locator Service GELOS (GELOS 1999). In all these systems, catalog entries describing environmental information resources can be searched for, a result list can be displayed, and the original information, if available on the Web, can finally be accessed through hyperlinks.

The challenge on the road towards a fully operational German Environmental Information Network is to integrate all these catalogs and meta-information systems. Towards this end, the German Federal Environmental Agency UBA ('Umweltbundesamt') started the project Locator and Communication Service Environment VKSU ('Verweis- und Kommunikationsservice Umwelt'). The UBA commissioned a consortium leaded by FAW Ulm to elaborate a conception for this integration and initiate exploratory steps for its realization (Riekert et al. 1998).

1 University of Applied Sciences Stuttgart, School of Library and Information Science, Wolframstr. 34, 70191 Stuttgart, Germany, Email: riekert@hbi-stuttgart.de

[2] FAW Research Institute for Applied Knowledge Processing, PO Box 2060, 89010 Ulm, Germany

Figure 1: The *Information Locator* developed for the prototype of the German Environmental Information Network and the Environmental Information System Baden-Württemberg. Sources of information can be searched for by specifying various search criteria. Geographical references may be entered and processed by using a *Gazetteer* (Riekert et al. 1997).

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 at other sites. 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 by the metadata, whenever possible.

As a first approach, a demonstrator for the intended system was developed on the basis of the existing Information Locator developed for GEIN (figure 2) and put up for discussion. Then the user requirements were ascertained and documented thoroughly in a number of interviews and workshops with members of various departments in the UBA and other environmental administrations. In consequence, we decided to make use of concepts and functions from many existing developments including the digital catalogs and meta-information systems mentioned in the last section. Rather than operating these functions in parallel, VKSU should provide a common front-end to integrate these functions.

Therefore, we designed VKSU as a multi-catalog system. The idea is to integrate a number of specialized digital catalogs and meta-information systems rather than build one single 'universal catalog' which records as many resources as possible. This allows us to make use of the existing German environmental information infrastructure that already includes a variety of digital catalogs and meta-information systems for Internet resources as well as bibliographic catalogs for non-digital printed media. We felt that VKSU should handle both kinds of catalogs (for Internet resources and printed media) in a similar way although the latter kind of catalog does not provide direct access to the original resources. In our design, we also decided to consider stock index servers and Internet search engines (e.g., Altavista) as special cases of catalogs, since these systems potentially open up the complete spectrum of environmental information resources on the global Internet.

It should be possible to search all these different catalogs[3] for information simultaneously. 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 resides on a Web server and is accessible through a Web interface. The users can interact with the broker by using a standard Internet browser without any specific software installation on their computers. Therefore, the chosen approach can be used both to support users in environmental administrations in their work and to supply the general public with environmental information.

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 coverage,
- spatial coverage,
- suppliers of the information,
- and a classification of the information resources to be searched for.

As the most promising solution to transmit these search criteria we suggest the use of Z39.50 (Z39.50 1999) or one of its profiles, in particular GILS (GILS 1999) 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).

---

[3] In the sequel, the term '(digital) catalog' will be used in a sense that tacitly includes meta-information systems, bibliographic catalogs, index servers, and search engines.
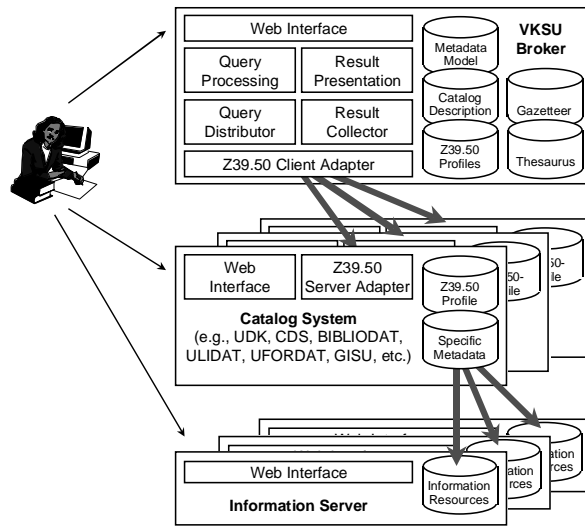
Figure 2: The VKSU broker integrates catalog systems and information servers.

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.

Although the use of Z39.50 appears to be the best choice for the communication with the digital catalogs, special solutions could be implemented for special purposes. In particular cases, it might be easier to develop an HTTP-based gateway to existing catalogs on the Web rather than developing a Z39.50 server interface. This is particularly true for systems such as Internet search engines which are not under the responsibility of the UBA or any other environmental administration in Germany.

3. The Multi-catalog Architecture

Figure 2 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 including the environmental data catalogs UDK and CDS, the bibliographic databases BIBLIODAT, ULIDAT, and UFORDAT, as well as stock index severs and search engines.

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*.

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. Alternatively, an HTTP-based gateway can be provided in order to adapt foreign meta-information systems or the big Internet search engines which cannot be expected to provide a Z39.50 server adapter.

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 under-

lying 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 self-contained meta-information system, 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 multilingual environmental *Thesaurus* is used by the query processor and the Web interface to support the formulation and processing of thematic search queries. This also includes semantic transformations and translations into foreign languages (figure 3).

5. Similarly, a *Gazetteer* containing names and geometries of geographic regions supports the textual and graphic formulation and processing of geographic search queries (see also figure 1).

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' provided by various German environmental information servers 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 information resources released by German environmental administrations as well as 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 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.
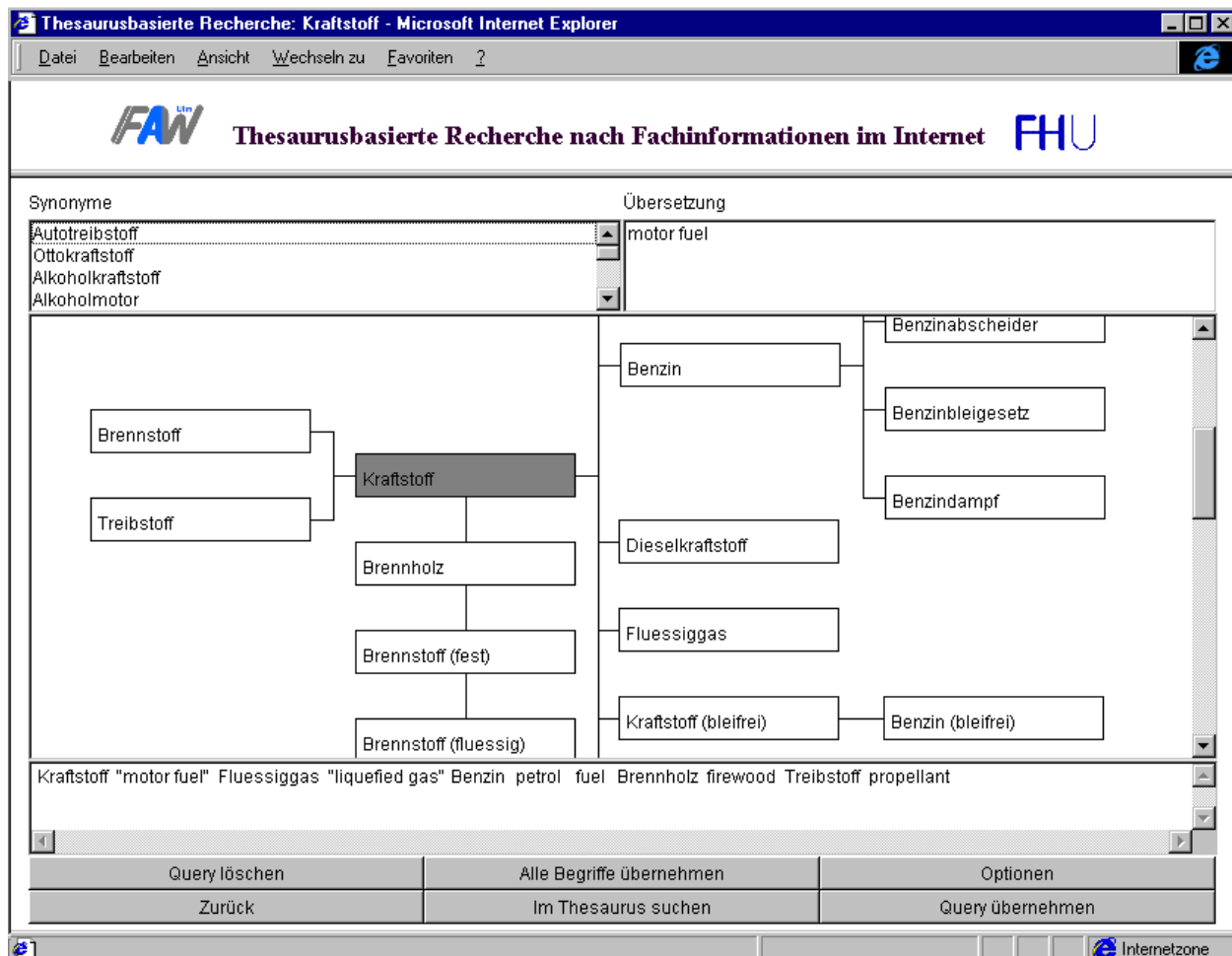
Figure 3: Thesaurus-based search: The user explores the thesaurus in the environment of the term Kraftstoff (i.e., motor fuel). Broader and narrower terms as well as synonyms, sibling terms and translations in foreign languages (here: English) are presented by a Java applet and can be included into the search query which is displayed in the lower part of the window Riekert et al. 1999).

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 Multipurpose 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.

The VKSU conception does not strive for 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 documents, data collections, databases, and information systems in environmental admini-

strations without infringing on the independence of their owners. The responsibility remains at the respective organizational units that own and maintain the information resources. They still decide which of their information resources will be released for use outside the department.

The chosen approach will be stepwise implemented in the future as part of the further development of the German Environmental Information Network GEIN. The next step which is currently being conducted by a commercial IT solution provider is 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. The final solution 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

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 .

GELOS (1999): http://ceo.gelos.org/.

GILS (1999): 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.

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.

Riekert, W.-F., Mayer-Föll, R., Wiest, G. (1997): Management of Data and Services in the Environmental Information System (UIS) of Baden-Württemberg. ACM SIGMOD Record, 26 (2), Special Issue on Environmental Information Systems (ed. O. Günther).

Riekert, W-F., Wiest, G., Günther, O., Röttgers, J., van Nouhuys, J., Mohaupt-Jahr, B., Seggelke, J. (1998): The Design of a Multi-catalog System for Distributed Environmental Information Resources. In: Haasis, H.-D., Ranze, K. C. (eds.): Computer Science for Environmental Protection'98. Metropolis-Verlag, Marburg, Germany.

Riekert, W.-F., Fuchs, C., Klingler, G. (1999): Erschliessung von Fachinformationen im Internet mit Hilfe von Thesauri und Gazetteers. In: Dade, C., Schulz, B. (eds.): Management von Umweltinformationen in vernetzten Umgebungen. 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.

Z39.50 (1999): http://lcweb.loc.gov/z3950/agency/.