Management of Distributed and Heterogeneous Information Resources for Environmental Administrations

W.-F. Riekert, G. Wiest, M. Gaul, and B. Münst
FAW (Research Institute for Applied Knowledge Processing)
PO Box 2060, D-89010 Ulm, Germany
Phone: (++49) 731 501 500, Fax: (++49) 731 501 999
Email: riekert | wiest | gaul | muenst@faw.uni-ulm.de

I. Henning
Ministry of Environment and Traffic Baden-Württemberg
PO Box 1034, D-70182 Stuttgart, Germany
Phone: (++49) 711 126 2502, Fax: (++49) 711 126 2881
Email: Inge.Henning@x400.uvm.um.bwl.de

Abstract

Since the mid-eighties, environmental administrations have accumulated an extensive collection of information and knowledge in the form of data, methods, and multimedia documents on a variety of computer systems. In the Environmental Information System (UIS) of Baden-Württemberg a service-oriented approach has been chosen to bring this treasure of information and knowledge to users in public administrations, industry, and to the general public. The guiding idea was to identify data, functionalities, and multimedia information that can be used in many different ways and to supply these information resources as self-contained network services. Implementation is based on WWW technology and the utilization of meta-information. To date, a comfortable solution has been found that can be used by decision-makers in government, administration, and industry as well as by scientists and the general public at regional, national, and international levels.

Keywords

Information resource management, environmental administration

1 INTRODUCTION

Since the mid-eighties, highly developed information systems in both their content and functionality have been installed at environmental administrative centers in many parts of the world. They have subsequently proved their suitability for practical. An extensive collection of information and knowledge has been accumulated at many sites, which is available in the form of data, methods, and multimedia documents on a variety of computer systems.

We must now pay increasing attention to the task of bringing this treasure of information and knowledge to users in public administrations, industry, and the general public. For more than three years, this task has been the subject of various research activities for the further development of the *Environmental Information System* (UIS) of the State of Baden-Württemberg (Riekert et al. 1997). Currently, we are developing a fundamentally revised version of the *Environmental Management Information System* (UFIS II) for the Ministry of Environment and Traffic of the State of Baden-Württemberg. The purpose of UFIS II is to provide decision-makers in the executive branch with comprehensive information on the environment.

In these research activities it has become clear that the use of TCP/IP-based Internet/Intranet technology allows us to integrate the individual UIS resources and to combine them in a distributed information network. This facilitates the economical usage of hardware and software resources throughout all UIS sites. A large supply of information can be made available to users without having it stored or maintained locally. Thus reasonably priced and easily maintainable PC systems can be used as UIS client platforms.

2 SOURCES OF ENVIRONMENTAL INFORMATION

Existing information systems used in environmental administrations offer basically three kinds of information source: data, methods, and multimedia documents.

- The primary source of information is *data* on the environment, e.g., measurement series, census data, or data from surveying activities. This data can also exist in edited form as corrected data, aggregated data, or complex data (such as geographical data).
- Many government agencies have developed computer-based *methods* for the interpretation of environmental data. These methods serve as information sources by generating information from existing data. Examples are environmental simulations, database application programs, statistical analysis functions, or presentation functions (e.g., to generate reports or maps).
- Rather than fall back on the original data, it often makes sense to consult the results of earlier interpretations of the data in question. These results are usually available in the form of reports and maps, i.e., *multimedia documents*. Reports and maps are very important sources of information because – implicitly – they
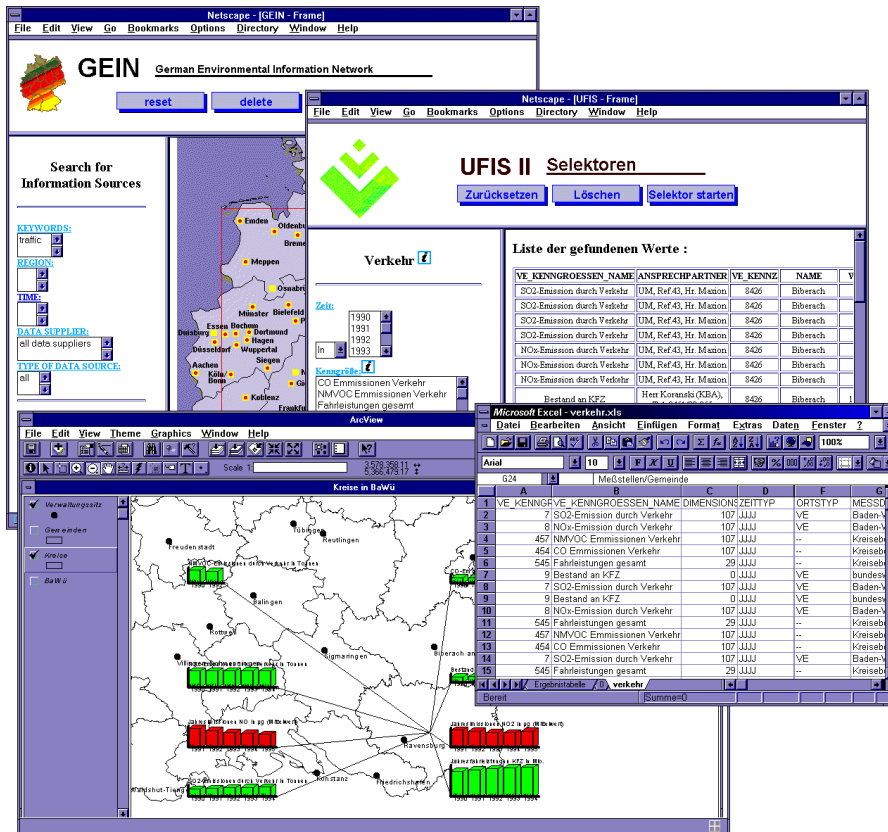
Figure 1 Accessing environmental databases with the UFIS II system through a WWW browser. The query result can be downloaded into local applications such as a spreadsheet or a desktop-mapping system. The upper left window shows the Information Locator of the German Environmental Information Network GEIN.

contain also those metadata which are necessary for interpreting the original data that underlie the document in question. Especially for the public and for high-level decision-makers, this kind of information source is indispensable.

In order to supply these information sources – data, methods, and multimedia documents – in an efficient and decentralized manner, a service concept was designed and implemented. The available system components are broken down into individual functions which can be addressed as self-contained network services in a wide-area network. The World-Wide Web (WWW) technology was chosen for implementing these services. WWW provides transfer services for typed data through the Multimedia Internet Mail Extensions (MIME) facility, allows for the activation of methods through the Common Gateway Interface (CGI) and Java applets, and is especially designed for the presentation of multimedia documents.

3 SUPPLY OF INFORMATION FROM ENVIRONMENTAL DATABASES

An important objective of the research activities described here is to support the supply and usage of information from environmental databases. Towards this end, we designed a software solution called *WebQuery*. It is used in the UFIS II project for accessing data from existing relational databases in the UIS. WebQuery makes possible the installation of network services, also referred to as *Selectors*, which allow for the retrieval of information on specific topics from a database. These selectors are presented to the user as data entry forms through a WWW browser. WebQuery transforms the content entered on the form into a database query. The result of such a query is a text file representing a table. This table is shown by the WWW browser as a dynamically generated hypertext document and, optionally, can be transferred (as data) to the client system. There it can be used in a spreadsheet system, a desktop-mapping system or a word-processing system, thus forming a building block for generating studies and reports (see figure 1).

One does not have to know a procedural programming language to define selectors for an environmental database. To link a database to the WWW, it is sufficient to create a set of description files. This task only requires knowledge of the WWW's hypertext definition language HTML and the database query language SQL. The WebQuery interpreter uses these descriptor files to process the user's queries. WebQuery comprises – independently of the database connected – general functions concerning session management, multi-user mode, user management, statistics, management of intermediate results, download, etc. without the necessity of additional programming. The existence of these features is particularly important for the provider of the database since their implementation is not easy owing to the stateless Hypertext Transfer Protocol (HTTP) that underlies the WWW.

The selectors generated by WebQuery are network services that are self-contained, individually addressable, and independently executable. Different access privileges can be given to individual user groups for the activation of these services. Thus the database can present itself to each specific user group (individuals, departments, institutions, the public, etc.) in a customized appearance in order to fit their needs concerning both content and desired query complexity.

4 SUPPLY OF MULTIMEDIA DOCUMENTS ABOUT THE ENVIRONMENT

In many cases it is useful to consult the results of earlier interpretations rather than retrieve non-processed data from environmental databases. These results exist, as a general rule, in the form of reports and maps, in other words, in the form of multimedia documents. With the help of adequate tools, such multimedia documents can be supplied in the WWW as hypertext. By using a commercial WWW browser it is possible to navigate through such a document along its structure.
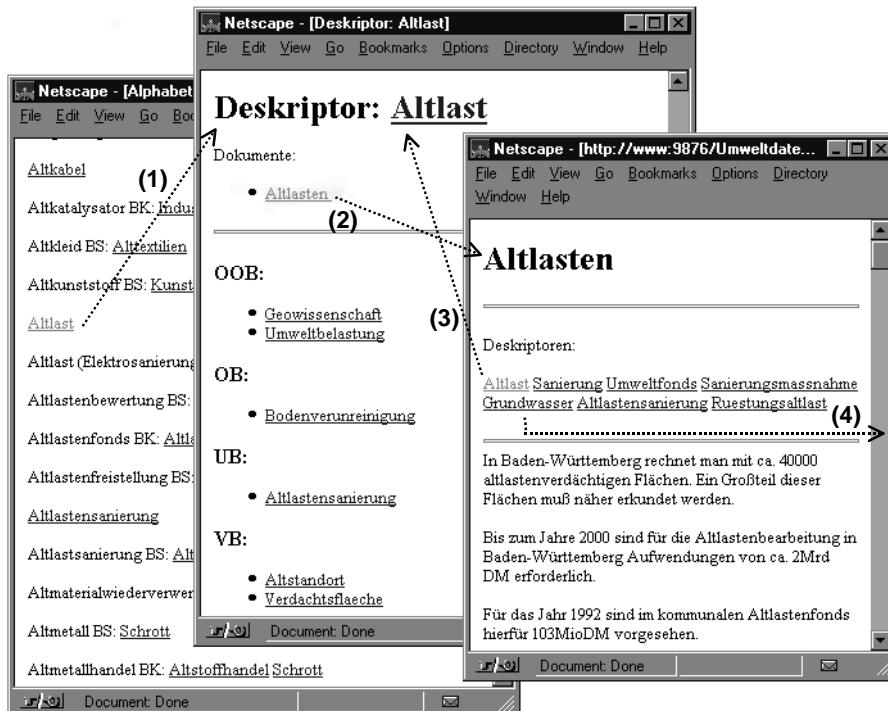
Figure 2 Thesaurus-based access to an environmental report: The alphabetical index (left window) contains the keywords that can be used for accessing specific sections of the report. Some of these keywords are synonyms which stand for other keywords (indicated by the BS tag) or keyword combinations (BK). By clicking on one of these keywords (in this case: "Altlast" = residual pollution) one gets to a hypertext document (1) which shows the selected keyword as a thesaurus term (middle window). From there it is possible to either browse within the thesaurus to broader (OB, OOB), narrower (UB), and related (VB) terms or navigate to related sections of the report ("Dokumente") (2). The terms associated with such a section (here: the section on "Altlasten" in the window on the right) are listed as hyperlinks under the section heading. Thus it is possible to return to the term *Altlast* (3) or to navigate to a different term, for example *Grundwasser* (ground water) (4).

Beyond navigating within the document structure, users often desire keyword-based access facilities to particular pages. We have developed a prototype system with such a functionality, using environmental reports as examples. Towards this end, we made use of a structured keyword catalog, in this case the polyhierarchical Environmental Thesaurus of the German Federal Environmental Agency (Batschi 1994). The thesaurus was transformed into a hypertext representation and its terms were used by an efficient string-matching program to generate an index for the pages of the report (Riekert 1996). Figure 2 gives an example of how the hypertext

representation of the thesaurus is used for retrieving information in the report "Umweltdaten" (environmental data) of the Landesanstalt für Umweltschutz (State Agency for Environmental Protection) Baden Württemberg.

5 RETRIEVAL OF ENVIRONMENTAL INFORMATION

It is to be expected that the supply of environmentally relevant information in the WWW will steadily increase. This concerns information from environmental databases and multimedia documents (as discussed in the previous two sections), as well as general information services (e.g., ozone measurements) or computing services (e.g., pollutant transport simulations). All in all, the future will bring a complex supply of environmental information resources on the Internet. Some parts of this complex supply will be offered by public administrations, others by external institutions.

In the public sector alone an increasing number of servers can be counted at regional, national and international levels. In Germany, this fact served as impetus to launch the *German Environmental Information Network* GEIN, which integrates the servers operated by German environmental administrations at state and federal levels (Seggelke 1996), as well as a number of other relevant servers. In 1996, the FAW was commissioned by the German Federal Environmental Agency UBA to develop the GEIN prototype, which now contributes to the pilot project *Environment and Natural Resources Management* (ENRM) of the G-7 initiative for a *Global Information Infrastructure* (GII).

The increasing supply of environmentally relevant services in the Internet leads to the problem that individual information resources are often difficult to find. What is needed, therefore, is an *Information Locator* facility on the basis of meta-information, which enables users in the administration and among the general public to search for relevant information. We have developed a prototype of such a locator as part of the UFIS II project. A multilingual version of this locator is also in use for the German Environmental Information Network GEIN (figure 1).

The locator contains an index of all information resources available (data, service programs, documents, etc.). These resources can be searched for by specifying keywords, geographical references, and temporal references, as well as the suppliers or the type of the desired information resources. The data descriptions available in the database of the *Environmental Data Catalogue* UDK (Günther et al. 1996) are included in the locator. The result of a search is a list of information resources. By simply using the mouse it is possible to activate detailed descriptions for every resource in the list. For this purpose, the locator offers a simplified view of the UDK data stock. It is also possible to access the original information, as long as it is offered by a server in the WWW.

For entering keywords, we use the environmental thesaurus of the German Federal Environmental Agency that has been already mentioned above. Semantic relations between keywords (such as associations with synonyms, broader, narrower or

related terms) are evaluated and used by the system during the search for information resources. In addition, the GEIN locator is equipped with a multilingual thesaurus extension for the retrieval of information in foreign languages.

Entering geographical references is supported by a structured geographical index known as a *Gazetteer*. With the help of this gazetteer a geographical search for environmental information is possible either by giving a rectangle defining the area of the search or a geographical name. Topological relations between geographical references are represented in the gazetteer, thus allowing the search for information using an enclosing, enclosed, or, in the most general case, overlapping geographical reference.

The entry of meta-information on information resources can be done best by the suppliers themselves. The locator offers services concerning the entry and maintenance of this meta-information. These services can be used via the WWW. The extraction of the meta-information from the information resources is to be increasingly automated. The current solution already associates keywords with textual documents automatically by comparing the texts in question with the environmental thesaurus. The indexing of geographical references with the help of the gazetteer is planned to work analogously. Further requirements arise when, apart from the textual information, data and service programs are to be subjected to the indexing process. Here, too, we are planning to introduce automatic methods, e.g., via the evaluation of database contents and schemas, or the analysis of description files used to define selectors through the WebQuery tool.

6   SUMMARY AND FURTHER OUTLOOK

The service-oriented approach chosen has resulted in promising steps towards the management of heterogeneous and distributed information resources. Distributed resources from various supplier sites have been combined in an integrated network. The guiding idea was to identify data, functionalities, and multimedia information that can be used in multiple ways and to supply them as self-contained network services. Implementation is based on WWW technology and the utilization of meta-information. Recent software developments in the Internet could easily be adapted, often without additional cost. This will help to make the use of the available hardware and software resources more economical and, in addition, allow for a relatively unrestricted and easy access to the existing environmental information resources.

To date, a comfortable solution has been implemented that supports the supply, distribution, and usage of environmental information. It can be used by decision-makers in government, administration, and industry as well as by scientists and the general public. It is on the way to surpassing the boundaries of Baden-Württemberg, thus forming a supranational solution which can be used in Germany, Europe, or even world-wide within the framework of the G-7 countries for a Global Information Infrastructure.

REFERENCES

Batschi, W.-D. (1994) Environmental Thesaurus and Classification of the Umwelt-bundesamt (Federal Environmental Agency), Berlin, in *Environmental Knowledge Organization and Information Management* (eds. P. Stancikova and I. Dahlberg), INDEKS Verlag, Frankfurt/Main.

Günther, O., Lessing, H., and Swoboda, W. (1996) UDK: A European Environmental Data Catalogue, in: *Proc. Third International Conference in Integrating GIS and Environmental Modeling*. National Center for Geographic Information and Analysis, Santa Barbara.

Riekert, W.-F. (1995) Cooperative Management of Data and Services for Environmental Applications, in *GISI 95 – Herausforderungen eines globalen Informationsverbundes für die Informatik* (eds. F. Huber-Wäschle, H. Schauer, and P. Widmayer), Informatik aktuell, Springer-Verlag, Berlin – Heidelberg – New York.

Riekert, W.-F. (1996) Thesaurusgestützter Zugriff zu Umweltinformationen in einem netzübergreifenden Hypermediasystem, in *Integration von Umweltdaten* (eds. R. Güttler and W. Geiger), 3. Workshop, Dagstuhl, Metropolis-Verlag, Marburg.

Riekert, W.-F., Mayer-Föll, R., and 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).

Seggelke, J. and Lessing, H. (1996) Globales Umweltinformationsnetz: Eckpunkte, Chancen und Gefahren, in *Informatik für dem Umweltschutz* (eds. H. Lessing and U.W. Lipeck), 10. Symposium, Hannover, Metropolis-Verlag, Marburg.

BIOGRAPHY

Wolf-Fritz Riekert is a Senior Scientist at FAW (Research Institute for Applied Knowledge Processing) in Ulm, Germany, where he heads the Environmental Information Systems Division. He holds a degree in mathematics and a doctorate in computer science from the University of Stuttgart. *Gerlinde Wiest* (diploma in computer science technology from the Technical College Ulm) is project manager for the UFIS II and GEIN projects at FAW Ulm. *Margit Gaul* (diploma in computer science from the Technical College Gießen) and *Bernhard Münst* (diploma in mathematics/economics from the University of Ulm) are research scientists at FAW Ulm. *Inge Henning* is a project manager at the Ministry of Environment and Traffic Baden-Württemberg; she is responsible for the steering of various research and development projects for the Environmental Information System (UIS) Baden-Württemberg.