WWW-UDK: A Web-based Environmental Meta-Information System

Ralf Kramer, Ralf Nikolai, Arne Koschel, Claudia Rolker, Peter Lockemann
Forschungszentrum Informatik (FZI), Karlsruhe, Germany
{kramer, nikolai, koschel, rolker, lockemann}@fzi.de

Andree Keitel
Landesanstalt für Umweltschutz Baden-Württemberg, Karlsruhe, Germany
Andree.Keitel@lfuka.um.bwl.de

Rudolf Legat, Konrad Zirm
Umweltbundesamt, Vienna, Austria
{Legat, Zirm}@uba.ubavie.gv.at

Abstract

The environmental data catalogue *Umweltdatenkatalog* UDK is a standard meta-information system for environmental data for use by state authorities and the public. Technically, the UDK consists of a database together with a front-end tailored to the needs of environmental specialists. FZI's contribution has been to develop a front-end that makes the UDK database available using the tools and techniques of the World-Wide Web. Among the features of WWW-UDK are several query modes for UDK objects and addresses, an environmental thesaurus, on-line access to some of the underlying data (e.g., databases and environmental reports), multilingual query and result forms, and an on-line help system. Currently, several installations of WWW-UDK are used in Austria and in Germany on the Internet and on Intranets. WWW-UDK can be easily integrated into a federation architecture which is based on CORBA, WWW, and Java.

1 Introduction

Information required to answer environmental questions (e.g., concerning environmental protection) is often found in different knowledge domains and on different systems. In order to determine whether and how a particular question can be answered a user has to find out *which* information is available and *where* and *how* to find it. That is, we need some metadata [8], which is essential for data retrieval and use of data by persons other than the data collectors. Thus, a meta-information system is required that consists of two main parts, the metadata itself and an application program that allows users to access the metadata.

In Germany, the environmental data catalogue *Umweltdatenkatalog* (UDK) has been developed as a meta-information system for environmental data [27, 9]. Technically, UDK consists of two parts, the metadata itself and a front-end tailored to the needs of environmental specialists. Currently, the UDK is used in most German states, Austria and Liechtenstein. Further countries, e.g., Switzerland, currently evaluate the UDK. The catalogue of data sources (CDS), which is being developed under contract to the European Environment Agency (EEA), is based on the UDK technology.

The World-Wide Web (WWW) is an attractive means to make data of any kind available for heterogeneous clients, because browsers are at hand for all major operating systems. Starting at the end of 1994, project teams at FZI took a leading role in integrating UDK data using the tools and techniques of the World-Wide Web [16, 14, 11, 18]. Currently, customized versions of WWW-UDK 3.0 are used in several Austrian and German Ministries of the Environment and state agencies for environmental protection both on the Internet for public access and on Intranets.

This paper is organized as follows. In Section 2, we provide an introduction to the UDK. The requirements, general approach, and the functionality of WWW-UDK are described in Section 3. Current limitations and future enhancements that

Figure 1: UDK object representing a Measurement Station

take into account most recent technological advances of Internet-related technologies, namely Java, Java Database Connectivity (JDBC), and the Object Management Group's Common Object Request Broker Architecture (CORBA) are described in Section 4. In Section 5, we outline how WWW-UDK is incorporated into a federation architecture that uses these technologies and that is based on the server site architecture of the environmental information system (UIS) of the German state Baden-Württemberg [23]. Section 6 concludes the paper.

2 The UDK

Data organization in the UDK is based on a three-way object model [9]: environmental objects (e.g., rivers, factories), data sets that describe environmental objects (e.g., measurement data), and a description of these data sets, i.e., metadata. In these metadata, data sets are decribed as so-called UDK objects. In order to give structure to the wide variety of metadata about environmental data, nine UDK-classes such as *Empirical Data, Data Collection/Expert Evaluations/Reports,* and *Maps* have been created [27, 9]. Additionally to UDK objects, the UDK comprises the addresses of the personnel in charge of these data.

UDK objects are organized in tree-shaped catalogues. In Germany, UDK objects are organized in separate so-called primary catalogues for each state. Each UDK object corresponds to exactly one node

of one primary catalogue. In the upper part of the tree, the edges between the nodes reflect the administrative structure of the state agencies and their responsibilities for the data sets. In the lower part, the edges reflect the part-of-relationship between aggregated data and the involved data sets. Additionally to these primary catalogues, secondary catalogues can be used to organize the UDK objects that are already part of primary catalogues according to specific subjects.

The access front-end of UDK uses a relational database system. At the user interface, users may first select a catalogue. They then can access the metadata directly or via additional search terms that are attached to the metadata via a thesaurus or as free keywords. In response of such queries, the metadata objects found are presented. Since not all of the underlying environmental data is accessible on-line today, addresses and telephone numbers are provided to obtain the desired information 'offline'.

3 WWW Access to UDK

3.1 Requirements

The requirements can be summarized as follows:

- Enhancements to the UDK data model have to be compatible with the existing data model and front-ends.

- Access across the private or public network is mainly read-only.

- Multilinguality must be taken into account especially on the Internet.

- Currently, adding new UDK metadata and updating existing metadata is primarily left to the UDK front-end that runs on Unix workstations and on PCs. This program can be used for local read-access as well.

- The functionality of the new front-end across the net should be similar to that of the existing UDK program. However, user friendliness has to be improved.

- The system architecture must be based on WWW communication, i.e., the hypertext transfer protocol http, and its tools.

3.2 Approach

There are several ways to make information which is managed by a DBMS available on the World-Wide Web. If open and portable systems are required, which have to run in a heterogenous system environment (e.g., different WWW servers, database

management systems, operating systems), proprietary solutions are not acceptable. Hence, we used classical WWW technology (WWW clients, HTTP, CGI scripts) and C programs with embedded SQL statements for access and integration of information sources. The approach is service-based, providing several more or less small (independent) services for information access, information processing, information presentation etc. To provide more complex functionalities, several services are combined. The effort for integrating the services into an n-tier architecture is comparative low.

3.3 Querying Metadata

Several search modes for UDK objects and addresses are available. Hyperlinks provide easy navigation among the objects and to the hypermedia help system. Furthermore, a poly-hierarchical thesaurus with environmental terms renders help in determining the search keywords appropriate for the problem at hand. This thesaurus had originally been developed as part of an information retrieval system for scientific research at the Umweltbundesamt (UBA), the German Federal Environment Agency [1]. Query and result forms can be presented in different languages including German and English. User search terms are mapped to (a sequence of) SQL queries, whose results enter into dynamically generated pages. Several search modes support both expert and novice users. In simple search modes, the user enters a single search term that is in turn mapped to several attributes of several relations. In complex search modes, the user makes use of attribute groups and specifies which ones are to be searched; searches in several attribute groups can be combined using boolean opertors. Case-sensitive and case-insensitive searches can be easily configured.

As a result, users first receive a list of those UDK objects that have been found. Then users can choose those objects they are interested in. Figure 1 shows an example of the complete metadata that is finally presented.

3.4 On-line Access to Environmental Data Services and Reports

The underlying UDK data model has been extended in order to provide WWW access primarily to environmental databases and reports within the UIS of Baden-Württemberg (henceforth referred to as WWW-UIS [11]). Nevertheless, the WWW access to UDK data is still fully compatible with the original UDK data model.

An UDK object can be linked to several base data sources via *typed links*. These are URLs extended by type specifications that indicate the type of information (e.g., database service, report, map,

image) the URL provides. The type specification gives useful hints about the information provided so that the user is protected from following worthless links. E.g., a user who is interested in measurement values may not be interested in photographs of the measurement station (see Figure 1).

Environmental reports are split into static HTML pages [5]. Currently, keywords are extracted automatically from these pages and inserted into the UDK database. One of the databases that is available on-line on the Intranet is the limits value database of WWW-UIS. This database features imprecise querying techniques based on fuzzy sets [15].

4 Current Limitations and Future Enhancements

WWW-UDK is based on the standard communication protocols of WWW. Subsequent HTML pages start programs at the WWW server site, which in turn access the database via the common gateway interface (CGI). Each of these database accesses has to perform its own connection to and disconnection from the database. Another severe limitation is that the hypertext transfer protocoll (HTTP) does not support more complex interactions with database systems such as database transactions. Thus, especially database transactions that access and modify more than one database relation are extremely difficult, if at all possible, to implement. Furthermore, the connection-less communication model of HTTP and HTML (re-)imposes restrictions on the user interface. Common techniques such as checks of user inputs, context-sensitive lists of values or pop-up menus are currently out of the reach of the WWW standards.

The problem of complex database interactions across the net can be solved by using Java applets, i.e., small programs that are executed within WWW browsers [26, 7, 4]. These applets can communicate with server processes using more powerful communication protocols than HTTP, which, among other possibilities, may support a session concept and encryption.

The problem of using context-sensitive presentation techniques within the user interface can be approached using Java or JavaScript. JavaScript [20] is a scripting extension to HTML. Unfortunately, in sharp contrast to Java, JavaScript still lacks important security features and strict definition as a standard.

Within catalogues, navigation along the tree-shaped hierarchy of UDK objects can be made more intuitively appealing with a graphical user interface. Further functional enhancement is possible by the use of the environmental thesaurus. First, the rela-
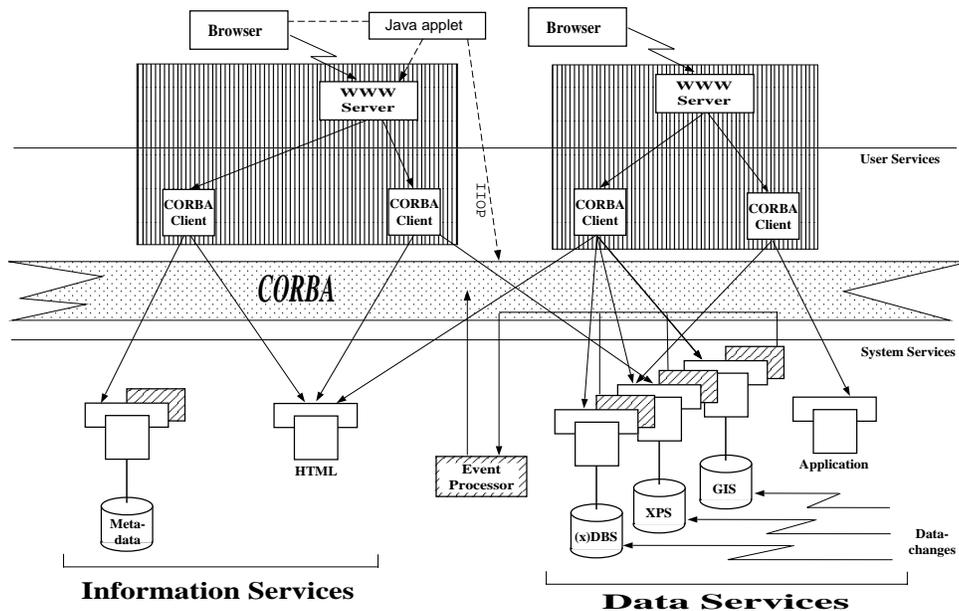
Figure 2: Federation Architecture incorporating the World-Wide Web, CORBA, and Java

tionships within a thesaurus can be used to broaden the search space of a query. Second, multilingual thesauri facilitate multilingual retrieval [17]. This requires a slight change of the current coupling of UDK objects and thesaurus descriptors. Thus, the combination of a standardized environmental data catalogue and the World-Wide Web is a workable solution which addresses metadata management and on-line retrieval of both metadata and the underlying environmental data, even when language barriers have to be surmounted.

5 WWW-UDK in a Federation Architecture

Within the UIS Baden-Württemberg, WWW-UDK is embedded in the federation architecture of WWW-UIS, an architecture of loosely coupled components that retain their autonomy. With respect to databases, federated DBMS [24, 2] represent the idea of loosely coupled, potentially heterogeneous, and autonomous databases.

However, driven by technological advances – not the least in middleware – and organizatorial changes, information systems nowadays are not just collections of (heterogeneous) databases [3, 22]. They are rather composed of almost arbitrary information sources, which all have to be used together. Now, to allow for a higher degree of heterogeneity and thus flexibility, we developed a service-based federation architecture for heterogeneous (environmental) information systems [11]. As a middleware layer we use OMG's Common Object Request Broker Architecure

(CORBA) [21, 25], a common industry-standard for object-oriented integration of distributed, heterogeneous information systems [12]. This allows for technical interoperability between (almost) arbitrary information sources, covering heterogeneity and distribution of sources at the server site using wrappers. These wrappers provide interfaces for information exchange using CORBA's interface description language (IDL).

WWW-UDK is part of this federation architecture [18], as are databases like the limits value database (Section 3.4) and geographical information systems (GIS). All these can be accessed via WWW-UDK. Equally important, the CORBA-based middleware layer of WWW-UIS allows to overcome some of the deficiencies of the World-Wide Web, e.g., the lack of combining the results of several DBMS queries within a single result HTML page. For this environment, we also developed an distributed event processing facility that uses event-condition-action (ECA) rules similar to active DBMS [29] for complex situation monitoring [28].

Figure 2 summarizes this federation architecture which combines emerging technologies like Java, CORBA, and WWW. Horizontally, we distinguish between user-level and system-level services. System-level services provide basic functionality, such as accessing databases or geographic informations systems and preparing HTML pages. Also non-data sources, like computing services, are integrated. User-level services basically combine several system-level services into higher-level services that are eventually made available to the user. Java ap-

plets communicate with their server site based counterparts using CORBA's stateful IIOP communication protocol.

Vertically, we distinguish between information and data services. Information services are based on UDK metadata, and help the user in locating relevant environmental data. Both information services and data services deliver results that in turn are either used as input for the generation of HTML pages or are directly presented by Java applets.

6 Conclusions

In this paper, we presented WWW-UDK, a Web-based environmental meta-information system that is currently used in Austria and Germany. Three customized versions of WWW-UDK 3.0 (State Agency for Environmental Protection Baden-Württemberg (LfU), Germany; Federal Ministry of Environment, Youth and Family, Vienna, Austria; Ministry of Environment of Niedersachsen, Germany) are available on the Internet (http://www.fzi.de/dbs/applAreas/ eis.html). Further installations are used on Intranets.

In the near future, we plan to use new technologies like Java to further improve query functionality as outlined in Section 4 and to provide basic data collection functionality via WWW. Access to the European Environment Agency's Catalogue of Data Sources (CDS) is currently under development at FZI as well [13]. On an international level, interoperability with other catalogue standards like the Global Information Locator Service (GILS) [6] will become increasingly important.

References

[1] W.-D. Batschi. Environmental Thesaurus and Classification of the Umweltbundesamt (Federal Environmental Agency), Berlin. In P. Stancikova and I. Dahlberg, editors, *Environmental Knowledge Organization and Information Management*, pages 57–62. INDEKS Verlag, Frankfurt/Main, 1994.

[2] M. W. Bright, A. R. Hurson, and S. H. Pakzad. A taxonomy and current issues in multidatabase systems. *IEEE Computer*, pages 50–60, March 1992.

[3] M. L. Brodie. Interoperable Information Systems: Motivation, Challenges, Approaches, and Status. In *19th International Conference on Very Large Data Bases (VLDB'93)*, Dublin, Ireland, August 1993. Tutorial.

[4] David Flanagan. *Java in a Nutshell: A Desktop Quick Reference for Java Programmers*. O'Reilly, 1996.

[5] W. Geiger, M. Reissfelder, and R. Weidemann. Das WWW-basierte Altlasten-Fachinformationssystem AlfaWeb. In Lessing and Lipeck [19], pages 211–220.

[6] The Government Information Locator Service. http://www.llnl.gov/gils/gilsexec.html, December 1995.

[7] J. Gosling and H. McGilton. The Java Language Environment: A White Paper. Technical report, Sun Microsystems, 1995. http://java.sun.com/ whitePaper/ java-whitepaper-1.html.

[8] K. Greve and A. Häuslein. Metainformationen in Umweltinformationssystemen. In L.M. Hilty, A. Jäschke, B. Page, and A. Schwabl, editors, *Informatik für den Umweltschutz; 8. Symposium, Hamburg 1994; Band I*, Umwelt-Informatik Aktuell, pages 169–178, Marburg, 1994. Metropolis.

[9] O. Günther, H. Lessing, and W. Swoboda. UDK: A European Environmental Data Catalogue. In *Third International Conference/Workshop Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, January 1996. http://ncgia. ucsb. edu/ conf/ SANTA_FE_CD-ROM/ sf_papers/ guenther_oliver/ my_paper.html.

[10] F. Huber-Wäschle, H. Schauer, and P. Widmayer, editors. *Herausforderungen eines globalen Informationsverbundes für die Informatik; 25. GI-Jahrestagung und 13. Schweizer Informatikertag, Zürich, 18.-20. September 1995/GISI 95*, Informatik Aktuell. Springer, 1995.

[11] A. Koschel, R. Kramer, R. Nikolai, W. Hagg, and J. Wiesel. A Federation Architecture for an Environmental Information System incorporating GIS, the World-Wide Web, and CORBA. In *Third International Conference/Workshop Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, USA, January 1996. http://ncgia.ucsb.edu/ conf/ SANTA_FE_CD-ROM/ sf_papers/ nikolai_ralf/ fedarch.html.

[12] A. Koschel, R. Kramer, D. Theobald, and G. v. Bültzingsloewen. Evaluation and Application of CORBA implementations. In *ECOOP'96 Workshop: Putting Distributed Objects to Work, 10th European Conference on Object-Oriented Programming*, Linz, Austria, July 1996.

[13] R. Kramer. CDS, Thesauri, and the World–Wide Web: Experiences, Requirements, and Propositions. In *First International Workshop Catalogue of Data Sources (CDS) and Thesaurus – User Experiences with Environment Thesauri in CDS*, volume 8 of *Publication Series Umweltdatenkatalog*, pages 28–29, Vienna, Austria, April 1996. Federal Ministry for the Environment.

[14] R. Kramer and T. Quellenberg. Global Access to Environmental Information. In R. Denzer, D. Russel, and G. Schimak, editors, *Environmental Software Systems; Proceedings of the International Symposium on Environmental Software Systems, 1995*, International Federation for Information Processing (IFIP), pages 209 – 218, London, 1996. Chapman and Hall.

[15] R. Kramer and H. Spandl. Limits Database for an Environmental Information System – A Fuzzy Set based Querying Approach. In W. Litwin and T. Risch, editors, *Applications of Databases; First International Conference, ADB-94; Vadstena, Sweden; Proceedings*, number 819 in LNCS, pages 158–171. Springer, 1994.

[16] R. Kramer and H. Spandl. Metadatenzugriff in Weitverkehrsnetzen: Eine Realisierung am Beispiel des Umweltdatenkatalogs UDK. In Huber-Wäschle et al. [10], pages 610–617.

[17] R. Kramer and R. Nikolai. Accessing Multilingual, Heterogeneous Data Sources in Wide Area Networks - Requirements and Approach. In Henning Christiansen, Henrik Legind Larsen, and Troels Andreasen, editors, *Flexible Query-Answering Systems, Proceedings of the 1996 workshop (FQAS'96)*, pages 255 – 264. Roskilde Universitetscenter, Roskilde, Denmark, May 1996.

[18] R. Kramer, R. Nikolai, A. Keitel, R. Legat, and K. Zirm. Enhancing the Environmental Data Catalogue UDK for the World Wide Web. In Lessing and Lipeck [19], pages 59–68.

[19] H. Lessing and U.W. Lipeck, editors. *Informatik für den Umweltschutz; 10. Symposium, Hannover 1996*, number 10 in Umwelt-Informatik Aktuell, Marburg, 1996. Metropolis.

[20] Netscape Communications Corporation. *Netscape JavaScript*, 1996. http://www.netscape.com/comprod/prod-cts/navigator/version_2.0/script/.

[21] Object Management Group. The Common Object Request Broker: Architecture and Specification, Version 2.0. OMG Document, Object Management Group, Inc. (OMG), July 1995.

[22] R. Orfali, D. Harkey, and J. Edwards. Intergalactic Client/Server Computing. *BYTE*, pages 108–122, April 1995.

[23] W.F. Riekert, R. Mayer-Föll, and G Wiest. Management of Data and Services in the Bade-Württemberg Environmental Information System UIS. *this issue*, 1997.

[24] A. P. Sheth and J. A. Larson. Federated Database Systems for Managing Distributed, Heterogenous, and Autonomous Databases. *ACM Computing Surveys*, 22(3):183–236, September 1990.

[25] J. Sigel. *CORBA Fundamentals and Programming*. John Wiley & Sons, Inc., New York, USA, 1996.

[26] Sun Microsystems. *The Java Language Specification. Technical report*, July 1996. ftp://ftp.javasoft.com/docs/langspec-draft5.1.zip.

[27] W. Swoboda, H. Lessing, P. Grolimund, O. Günther, U. Haas, R. Legat, M. Vogler, and K. Zirm. Metadatenklassen im Umweltdatenkatalog (UDK). In Huber-Wäschle et al. [10], pages 601–609.

[28] G. v. Bültzingsloewen, A. Koschel, and R. Kramer. Active Information Delivery in a CORBA-based Distributed Information System. In Karl Aberer and Abdelsalam Helal, editors, *Proc. First IFCIS International Conference on Cooperative Information Systems (CoopIS'96)*, pages 218–227, Brussels, Belgium, June 1996. IFCIS, IEEE Computer Society Press, Los Alamitos, California.

[29] J. Widom and S. Ceri, editors. *Active Database Systems: Triggers and Rules for Advanced Database Processing*. The Morgan Kaufman Series in Data Management Systems. Morgan Kaufmann Publishers, Inc., San Francisco, California, U.S.A, 1996.