In: F. Huber-Wäschle, H. Schauer, P. Widmayer (Hrsg.): *GISI 95 – Herausforderungen eines globalen Informationsverbundes für die Informatik*. pp. 618–625. Springer-Verlag. Berlin – Heidelberg – New York, 1995.

Cooperative Management of Data and Services for Environmental Applications

Wolf-Fritz Riekert, FAW Ulm[1]

Abstract: In recent years, systems for processing environmental information have been evolving from research and development systems to practical applications. Many of these systems already support environmental activities of the public sector. Most of these systems, however, were developed as island solutions. The integration of such heterogeneous systems and the support of cooperation on a regional, national and international scale is a big challenge of today. It requires new methods to access and make use of the services and data offered by the agencies of such a distributed system. Metadata (such as in data and service dictionaries) and codata (such as spatiotemporal references) must be provided in order support the proper usage of these data and services. Systems integration techniques are required in order to overcome the heterogeneity of the existing systems. An approach to fulfill these requirements is shown by examples from recent work at FAW that has been conducted in the context of the Environmental Information System Baden-Württemberg.

1. Introduction

Environmental protection and environmental monitoring require a great deal of information processing. Computer based support, known as environmental information management, can be given for conducting these information processing tasks. These tasks include environmental research activities, the observation of environmentally relevant features and actions to take care of, preserve, and restore the environment. Information technology can help both to conduct these activities an to confer and decide upon these activities [RADE94]. The relevant tasks can be displayed in a cyclic diagram, shown in figure 1. This cycle starts at the environment and produces data, information, and finally plans to restore and preserve the environment, thus returning to the beginning of the cycle.

- Environmental data, which are the origin of the information processing cycle described, differ from traditional data in many respects. This concerns spatiotemporal aspects as well as the concepts of codata and metadata. Interpretation processes are necessary to derive the information of interest from existing data. Data processing programs, also known as *services*, can support the necessary processing steps on the road from raw data to user-oriented information.
- Planning actions is a very important task in cooperative environments. It includes decision support activities such as communicating information, raising issues, stating problems, claiming objectives, simulating and forecasting future scenarios, and elaborating plans. The action that results from these plans may affect the environment directly or indirectly. The former case includes activities, such as protecting or restoring the environment. The latter case includes environmental monitoring work, such as collecting more data or developing new data interpretation methods.

2. Cooperativity

A special organizational and technical framework is necessary to conduct these activities by a network of cooperating agencies known as *collaboratory*. Cooperativity aspects are very important for the overall coordination of information processing activities for the environment. In the past, the cooperative aspect of environmental information processing has been neglected in

[1] FAW (Research Institute for Applied Knowledge Processing), P.O.B. 2060, D-89010 Ulm, Germany, Tel.: ++49-731-501500, Fax: ++49-731-501999, E-Mail: riekert@faw.uni-ulm.de

most cases. Many efforts resulted in island solutions which were locally optimized but failed when applied for real world problems. Environmental monitoring and environmental protection are not done by one single hierarchical organization under a centralized leadership. Many independent governmental and non-governmental organizations take part in these processes and need to cooperate with one another. These organizations include, in particular, governmental departments for ecology, economy, natural resources, and agriculture as well as universities, research institutes, associations of environmentalists, etc. and are not limited to one single nation. The number of these organizations is still increasing. Examples are nuclear or defense research agencies that are targeting at new tasks in the environmental domain. Therefore the employment of information technology in the environmental domain will be of limited use, unless it supports virtual organizational structures which go across the existing hierarchies. This leads to the concept of national and international collaboratories, consisting of independent sites. From an informational point of view, these sites are nodes in a communication network. Any node provides data and services which may be accessed from any other node in the network. For supporting this network approach, new software solutions are required that are modular, distributed, and interoperable. The solutions presented in the following sections are intended to meet this requirement.

3. Environmental Information and Environmental Data

In the regulation on freedom of access to information on the environment, the European Council lists the following types of environmental information:

"`Information relating to the environment´ shall mean any available information in written, visual, aural or database form on the state of water, air, soil, fauna, flora, land and natural sites, and on activities (including those which give rise to nuisances such as noise) or measures adversely affecting, or likely so to affect these, and on activities or measures designed to protect these, including administrative measures and environmental management programs."

We try to differentiate the terms "data" and "information", although the two terms are often used as synonyms. While data can be anything that can be archived on the storage media of a computer, information, as a result of an interpretation process, has the special property that human users can make use of it and apply it for their decisions. Although information can be stored as data on a computer, it requires special presentations in order to be easily understood by humans. Typical presentations of information are text, graphics, sound, voice, and video. These presentations can be made available via modern communication media.

Environmental data differ from traditional data in many respects. Often the data are related to a certain location and time interval. Measurement data that are acquired for discrete locations and points in time belong to a special category; these data need to be interpolated if continuous profiles are required (e.g., air pollution data or elevation models). A special case is image data (e.g., satellite sensor data), which typically consist of very large raster data sets. NASA experts often handle terabytes of satellite data; this is more than two orders of magnitude greater than the amount of data managed today in large international financial transaction systems. This requires new storage models for environmental databases, e.g., the use of automatic tape archives and CD ROM jukeboxes as tertiary storage medium.

The most demanding problem concerns the problem of deriving information of interest from existing data. In the ideal case, an environmental database contains all necessary data that are needed to derive the information requested by the user. Apart from the raw environmental data, however, this requires the availability of additional data. These additional data are often referred to as *metainformation* or *metadata* [RADE91]. *Codata* is another term which is also used in this context. In this paper, we try to differentiate between metadata and codata. Metadata provide abstract descriptions of the data structures and data formats used in the underlying system, whereas codata include additional instance-specific data about location, time, precision, and revision date of the data under consideration. Location and time are very important kinds of codata which are discussed in more detail in the next section.

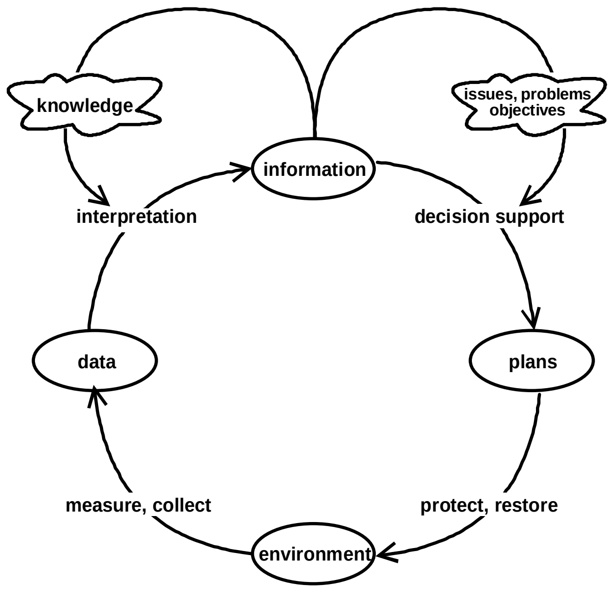

Fig. 1: Information processing tasks in environmental protection

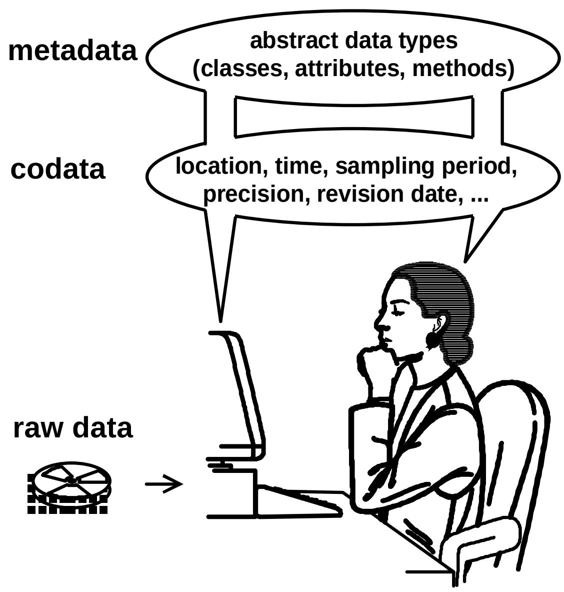

Fig. 2: The importance of metadata and codata for environmental information processing

Metadata and codata are missing today in most existing environmental information processing systems. Often the only data source consists of a magnetic tape containing the raw data (e.g., image data or measurement data) to be analyzed. All of the additional information that is necessary for the correct interpretation of the raw data is implicitly coded in the application programs or must be contributed by the user (figure 2). Such a tape may become completely useless after a change of the personnel or of the data processing software. Therefore an effective management of metadata and codata is of crucial importance for environmental information processing systems.

4. Spatiotemporal Aspects

Environmental data often describe environmental objects with a spatial and temporal extent. These objects, also known as geographic objects, possess a geometry consisting of point-form (0D), linear (1D), flat (2D), or solid (3D) features. Often geographic objects also have a lifetime which is given by their dates of construction and destruction. During the lifetime of an environmental object, its attributes may change. That is, the values of the attributes are only valid during a certain time interval, and a given attribute (e.g., the population of a city or the land use of a parcel) may possess many possible values during the lifetime of the object. A special case is the change of the geometry of a geographic object such as the growth of a city or the shrinking of a lake or a forest.

Spatiotemporal data are important examples of codata in the environmental domain. The management of spatiotemporal data is a special challenge to applied computer scientists. Abstract data types are required for representing concepts such as the partonomy, the topology, and the spatiotemporal extension of geographic objects as well as the thematic information associated with these objects, such as alpha-numeric attributes and relationships to other geographic objects. An important task is also the (carto-)graphic presentation of these objects. In most cases, non-standard data types are required for representing these kinds of information and, therefore, a trend towards systems which allow the definition of such data types, such as extended relational or object-oriented databases [CACM91], can be recognized in current developments [GÜNT93].

The special nature of environmental data also requires a query language with special characteristics. Apart from typical SQL-like questions, the forthcoming object-oriented database management systems [CATT94] also allow navigational queries and queries that include user-defined predicates. In particular the latter allows the usage of spatiotemporal predicates in queries. The optimization of spatiotemporal query-processing is a scientifically demanding problem which concerns both storage models and indexing techniques for multi-dimensional data. The integration of these techniques with existing databases is a hard problem which will still require major research activities in the future.

5. Environmental Information Management in the Public Sector

In recent years, environmental protection as an objective of public activities has reached quite a high standard in Europe, particularly in Germany. Communal tasks, for example, include land use planning, approving compliance with environmental standards, management of hazardous waste, and water and energy management for public facilities. Several states and regions of the European Union have already installed effective environmental information systems and powerful sensor networks.

Official systems such as the Environmental Information System (German acronym: UIS), Baden-Württemberg [MAYE93] support environmental tasks at various levels: Decision support systems are provided for the high-level environmental management, reporting and planning systems are available for middle management, while basic components support the acquisition and management of specific environmental data at the operational level. In addition, interdisciplinary information systems are being used in public environmental management; these systems – such as the topographic or cadastral information systems of the surveying offices – are not restricted to environmental tasks.

UIS is the organizational, informational, and task-oriented framework for the supply of environmental data and for the processing of both department-specific and interdisciplinary tasks in the environmental domain of the State administration. UIS consists of a large number of components that are implemented on various hardware and software platforms and are operated by various departments at distributed locations. In this context, the INTEGRAL project, which is presented in the following section, aims at a user-friendly and economical way of accessing the functionality of these distributed system components.

6. INTEGRAL

The INTEGRAL (Integration of Heterogeneous Components of the Environmental Information System of Baden-Württemberg) project has been conducted since 1993 at FAW in cooperation with the Institute for Nuclear Energetics and Energy Systems (IKE) of the University of Stuttgart under commission of the Ministry of the Environment, Baden-Württemberg. The central goal of this project is to continuously increase the integration of UIS components and to improve the ease of use of these components from remote sites [RIEK94]. This is done by making available the functionalities of these components through an electronic network in the form of data and services.

The networking concepts available in open systems such as the Transmission Control Protocol / Internet Protocol (TCP/IP), Remote Procedure Calls (rpc), Remote Shells (rsh), Common Object Request Broker Architecture (CORBA), World Wide Web (WWW) and the X Window system are promising options for the necessary integration, because they provide highly effective mechanisms for sharing data and functionality within a computer network. Open networking concepts were chosen for the INTEGRAL project since they are supported by all commercial operating systems at least to a large extent.

The idea behind INTEGRAL is to make the functionality from existing programs available as a collection of self-contained services. These services include services for navigating through the available functionality as well as services for data retrieval, data selection, data processing, data presentation, and simulation. Examples are the services to retrieve the content of the database that underlies the Environmental Management Information System UFIS and to invoke the CRAYSIM pollution transport simulation service (figure 3). Another example is the prototype of a geodata server that has been developed in cooperation with the GODOT project (figure 4) [RIEK95a].

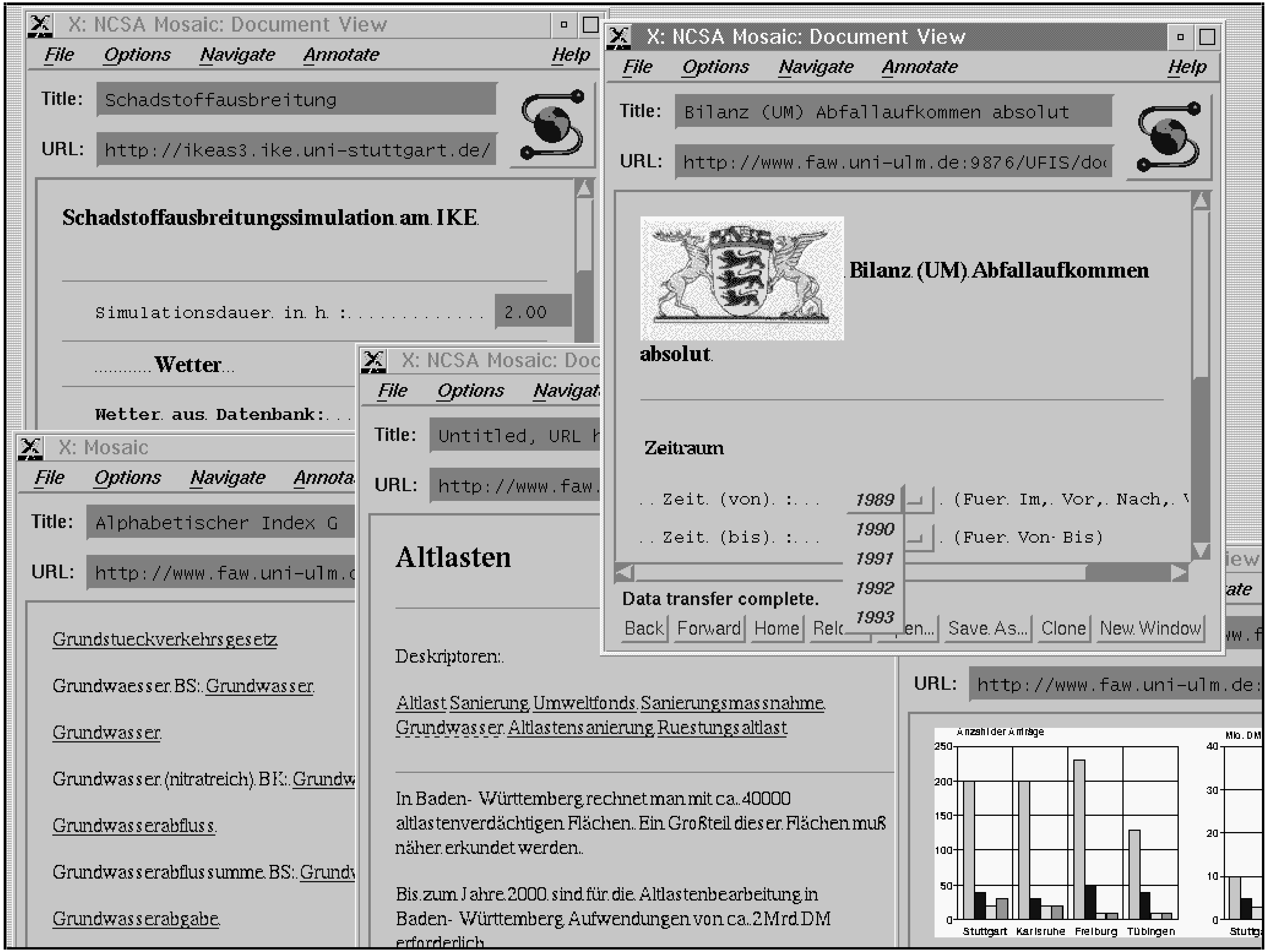

Fig. 3: INTEGRAL: Integration of Heterogeneous Components of the Environmental Information System of Baden-Württemberg

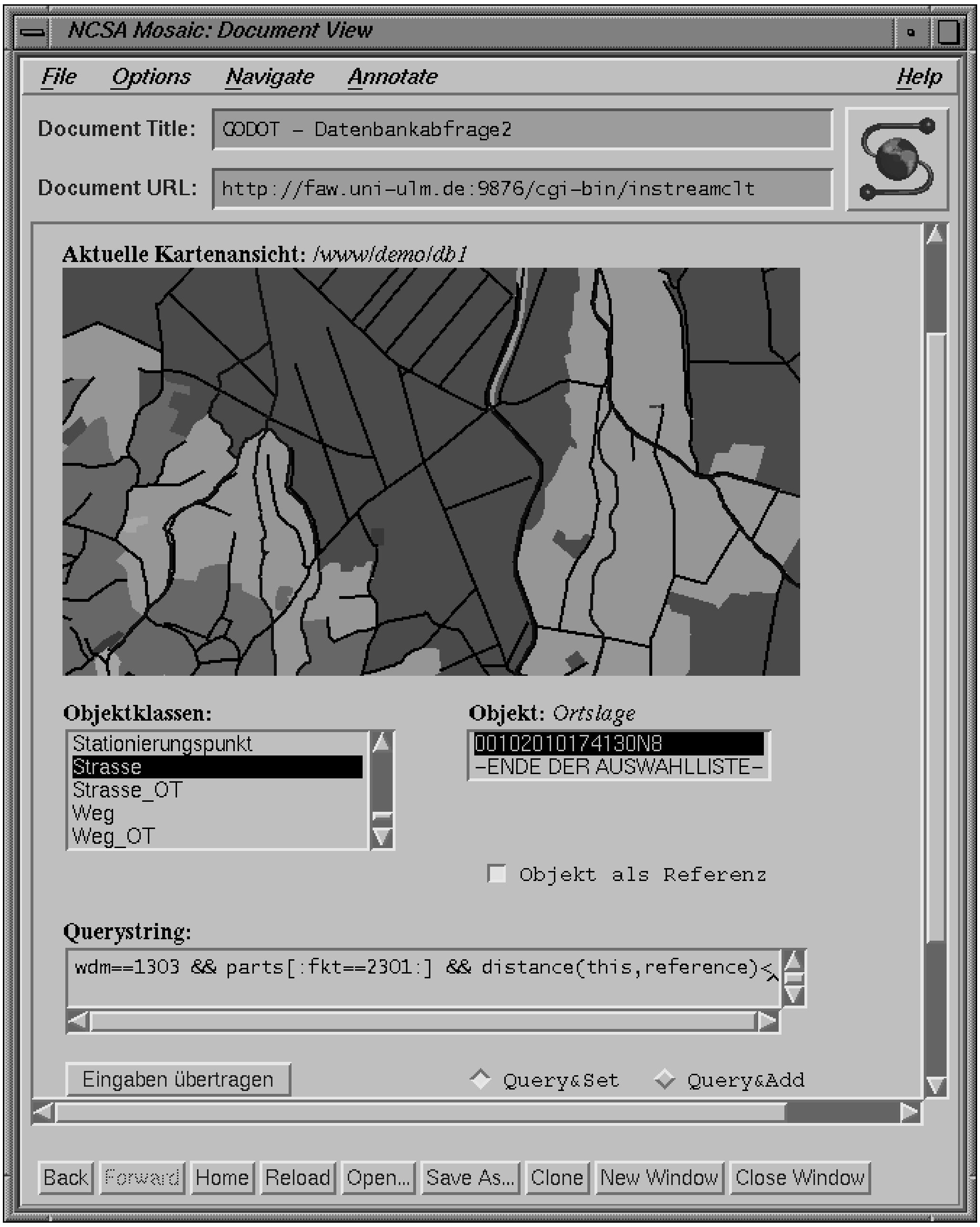

Fig. 4: The Geodata Server Prototype developed in project INTEGRAL

The access to these services and other information resorces such as multimedia documents and data is based on a hypertext system, namely the wide-area hypertext system World Wide Web (WWW). With this system, a network of hypertext documents may be spread out over the computer systems that are connected via an electronic network. WWW browser software such as *Mosaic* is available for all major computer systems including PCs and allows the users to navigate through the network of distributed hypertext documents. The navigation typically ends at some multimedia document, some data file or some service program offered by the participating agencies (figure 5).

Mosaic and other WWW browsers are able to present multimedia documents directly to the users. Data files may be retrieved using a file transfer facility and may be either stored in the file system or processed by a client-side application program (such as a spreadsheet, a desk-top publishing system, or a geographic information system). Server-side service programs (including interfaces to existing information systems) may be invoked through a form-based interface. The result of calling a service program may again be either a multimedia document or a data file, which can be further handled in the ways described above.

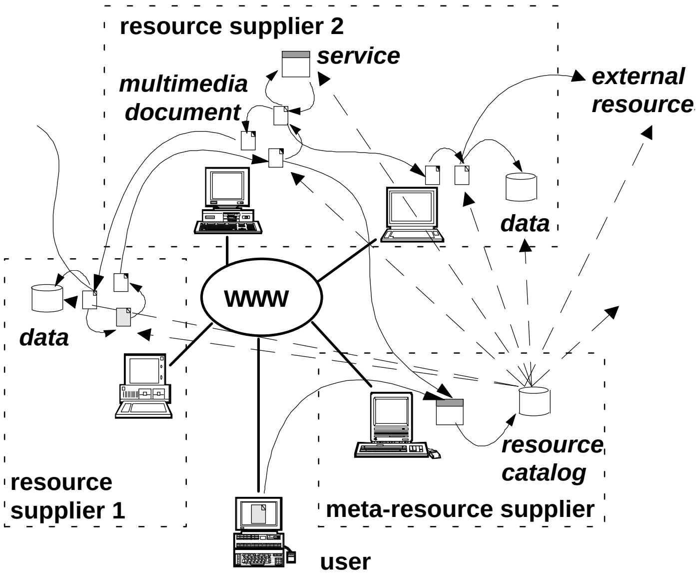

Fig. 5: Access to information resources of the Environmental Information System Baden-Württemberg (data, services and multimedia documents) through the distributed hypermedia system World Wide Web. Direct access to the resources is supported by an online Resource Catalog

The multimedia documents, services, and data of the Environmental Information System Baden-Württemberg may remain distributed among the various collaborating sites and are maintained locally by the responsible departments of the Ministry of the Environment and the State Office for Environmental Protection rather than by a centralized agency. Additional resources that are offered on the WWW by external suppliers at the regional, state, or European levels may be easily included into the system. All resources can be accessed from any workstation or PC in the network without the necessity of installing any software except a WWW browser. This server-side location of resources reduces hardware, software, and maintenance requirements for the client systems.

7. Overcoming the Heterogeneity

The harmonization of environmental information at national, European and worldwide levels is of central importance for gaining a reliable description of the environmental situation and, at the same time, is a basic requirement for the data and service management in this context. These requirements, however, are confronted with the existing heterogeneity of hardware and software environments, of database systems, of method bases, of network technology, and of various computer languages.

The task of overcoming heterogeneity particularly requires the availability of metainformation. Doubtless the development and promotion of standards are of particular importance in this respect. Experience shows, however, that we will still have to cope with competing standards in the future. In addition, technological advances will always produce new heterogeneity problems and will require strategies for migrating the software towards new solutions. Metainformation about data and methods can be used in order to do the necessary translations between different systems.

In the public sector, environmental metainformation systems, also known – with a reduced scope – as *environmental data directories* [SCHÜ93], are currently being developed to overcome the prevailing lack of metainformation in existing environmental information systems. In INTEGRAL, the environmental data directory concept is going to be extended towards a general *environmental resource catalog* that, apart from environmental data, also describes general information resources such as services and multimedial documents (e.g., reports or maps). This resource catalog is itself a service that can be accessed through the WWW (cf. figure 5).

Beyond the standard navigation facility of WWW through explicit hyperlinks, the resource catalog supports direct retrieval strategies that allow a fast access to the desired resources. Towards this end, users can specify terms (including synonyms) from the thesaurus of the German Environmental Agency (Umweltbundesamt), apply full text retrieval techniques or indicate geographic references. In any of these ways, the resource catalog server can be used to identify relevant information content in the network and to put the user through to the appropriate information resource that is provided by a particular WWW server [RIEK95b].

The maintenance of the resource catalog is "outsourced" to the suppliers of data and services. To serve this purpose, the resource catalog possesses a WWW-based metainformation editing facility that can be used as a "resource advertising tool". For any information item to be offered in INTEGRAL, suppliers may provide metainformation about data ownership, pricing, and methods to access the information, as well as a related geographic reference and a collection of descriptors (terms) from an environmental thesaurus. For textual information sources, appropriate terms can be generated by a semi-automatic indexing tool.

8. Conclusion

The management of data and services for environmental applications using codata and metadata is of crucial importance for environmental information systems. The network and hypermedia-based approach chosen in INTEGRAL shows that the access to data and services in environmental information systems will be enhanced and standardized by the techniques proposed. The advantage arises from the simplified, uniform, and more user-friendly availability of remote data and services, which favors the cooperative work within national and international collaboratories. Beyond integrating computer systems, INTEGRAL thus connects people by empowering them to work cooperatively and to communicate efficiently by means of electronic media.

References

[CACM91] Special Section: Next Generation Database Systems. *Communications of the ACM*. 34(10):30-120, October 1991.

[CATT94] Cattel, R.G.G.: *The Object Database Standard: ODMG-93*. Morgan Kaufmann, 1994.

[GÜNT93] Günther, O.; Riekert, W.-F.: The Design of GODOT: An Object-Oriented Geographic Information System. *IEEE Data Engineering Bulletin* 16(3), September, 1993.

[JAES93] Jaeschke, A.; Kämpke, T.; Page, B.; Radermacher, F.J.: *Informatik für den Umweltschutz*. Springer-Verlag, Berlin – Heidelberg – New York, 1993.

[MAYE93] Mayer-Föll, R.: Das Umweltinformationssystem Baden-Württemberg; Zielsetzung und Stand der Realisierung. In: [JAES93], pp. 313-337.

[RADE91] Radermacher, F.J.: The Importance of Metaknowledge for Environmental Information Systems. In: Günther, O.; Schek, H.-J. (eds.): *Large Spatial Databases*. Proceedings. Lecture Notes in Computer Science 525, pp. 35-44. Springer, Berlin – Heidelberg – New York, 1991.

[RADE94] F.J. Radermacher, W.-F. Riekert, B. Page, L.M. Hilty. Trends in Environmental Information Processing In: R. Raubold, K. Brunnstein (eds.) Proceedings of the IFIP Congress 94 – Volume 2. Elsevier Science Publishers B.V. (North Holland), 1994.

[RIEK94] Riekert, W.-F.; Henning, I.; Schmidt, F.: Integration von heterogenen Komponenten des Umweltinformationssystems (UIS) Baden-Württemberg. In: *2. Workshop „Integration von Umweltdaten"*, Metropolis-Verlag, Marburg, 1994.

[RIEK95a] Riekert, W.-F., Ebbinghaus J.: Geodatenserver in Wide Area Networks. In: *Geographische Informationssysteme '95*. Tagungsband. Wiesbaden, 1995.

[RIEK95b] Riekert, W.-F.: Thesaurusgestützter Zugriff zu Umweltinformationen in einem netzübergreifenden Hypertextsystem. In: *3. Workshop „Integration von Umweltdaten"*, Metropolis-Verlag, Marburg, to appear.

[SCHÜ93] Schütz, T.; Lessing, H.: Metainformationen von Umwelt-Datenobjekten – Zum Datenmodell des Umwelt-Datenkataloges Niedersachsen. In: [JAES93], pp. 19-28.