Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

The challenge of integration in environmental information systems

G. Schmick,[a] G. Ernst,[a] H. Humer[a] & R. Denzer[b]
[a] *Austrian Research Center Seibersdorf, Department of Information Technology, Division of Instrumentation and Information Technology, A-2444 Seibersdorf, Austria*
[b] *University of Kaiserslautern, Am Forleinwald 3, D-69521 Gaiberg, Germany*

Abstract
The paper will describe in a more general way the purpose of environmental information systems and its practice. To show how difficult integration of different environmental data, for example of different monitoring systems is we will specify an integration approach for such systems and demonstrate a concept how integration can be done.

1. Introduction

Environmental protection, management and research today is only possible when based on valid information. This information is gathered, stored and used in many computer systems. Many people at many different sites need and use information about the current state of the environment. It is obvious that computers play an important role in every day's work of environmental protection. From the beginning of the 1980's, an increasing number of public authorities and research groups in Germany began to use software systems more extensively for this purpose. Within the last few years, it has become more and more clear that environmental software systems have to deal with a number of specific information problems, e.g. the variety, size and distribution of the data sets, the complexity of the models and the appropriate graphical presentation of environmental information.

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

2. Purpose and practice of environmental information systems

If we use the term Environmental Information System (EIS) we mainly think of systems used in public authorities. These systems serve many purposes and we want to mention only a few of them.

- Monitoring
- Coordination in case of catastrophes due to plant accidents
- Environmental impact assessment (EIA)
- Every day planning and decision making
- Gathering information to prepare political decisions

For an overview see [1,2]: Most of these systems have in common that due to the complexity of the decision processes involved, data from many different sources *should have to be* combined. The reason for this is the increasing environmental awareness which is reflected in many new laws and which forces the decision takers in every public authority to consider aspects they did not have to consider ten years ago. Many of these new aspects are related to data or information which comes from elsewhere and that is exactly where the integration problem begins. If we wrote above "should have to be" in italics it is for the reason that today's practice is far from what we would like it to be.

The reality of EIS today is best described by the image of many small islands in a vast ocean. Many of the public authorities noticed that they would not be able to carry out the increasing number of tasks related to environmental protection without computer support. Thus many systems grew on the whole range of the scale, from local to national projects. Most of these systems are incompatible to each other in every thinkable aspect. If you want to keep an overview about which data has been where produced and for which purpose it will be necessary to combine them and make their information widely available.
We do not believe that it will be possible in our case to integrate several hundreds of systems using a standard approach by central definition of data structures or by using the same programs at every site. This may work in business institutions but it will not work in our case for two reasons: first, many of the institutions are autonomous and second, the tasks which have to be carried out are in many cases much more complex than anything we dealt with before. Therefore an approach is needed which combines information islands leaving them as far as possible as they are but nevertheless making their data sources available for other purposes.

3. The technical challenge to integrate different environmental data

Our first ideas date back to 1989 when the discussion about integration in EIS was still a personal interest of few people. We were sure that this is a problem which cannot be solved by one institution or research group.

Thus we prepared several cooperations which included practical projects and research work. In the following sections, the overall problems and concepts are introduced. We think the task of an EIS is twofold

1. to make information available describing data gathered at many sites
2. to make the data available for use

So an EIS must include a metainformation system as well as an information system. Both of them must be strongly decentralized because of the fact that an EIS only for one federal state will include several hundred components which are at different levels of realization and which change dynamically over time. Therefore we believe that it will be impossible to make an overall system design. The only practical way will be to build an intermediate system in-between. The concept of the intermediate system will be discussed in the next section.

3.1 Concepts

The architecture is introduced in Figure 1. On the right side we have several local systems which shall be integrated into an environmental network (on the left). The local systems treat their data using their local conventions (which means storage technologies, software concepts, semantics, etc.). They have been free before and they keep being free in doing what they want. They are designed for the local purpose.

The intermediate system will consist of two interfaces and a number of services which bind the systems together. These are the so-called *public services*. They must reside on every node in the network connected to the EIS. The first interface (*S0*) is the interface which is used by the public services to communicate. This interface is called *environmental data protocol* (EDP) and must be publicly agreed upon. The second interface (*S1*) is the local interface connecting the local applications to the public services.

The most difficult task in the design of the EIS, we call *open,* is to find a useful concept to translate the local conventions into valid global conventions and vice versa. The particular difficulties reside in the semantics of the data. On the way to the open EIS there are three main integration problems:

- *Integration problem 1*: First we must design the network interface *S0* for the public services. *S0* must be very general and flexible and the services must provide high functionality.
- *Integration problem 2*: The second problem is to design the interface *S1* to the local applications such that we can embed whatever local information.
- *Integration problem 3*: Integration problem 3 deals with the public services themselves and the node specific services. These are applications which are only of local interest but which need to use remote data.

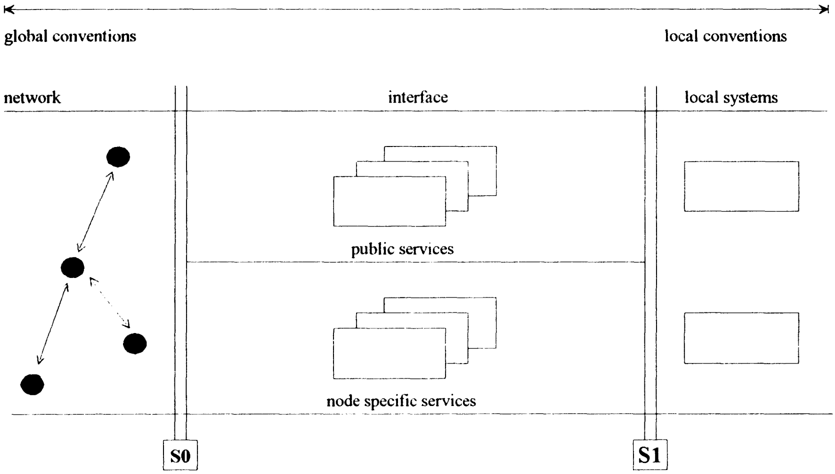

Figure 1: General approach of an open EIS architecture

4. The Austrian Ozone Network

Forced by a new law (Ozone Law §5, , BGBl. No. 1992/210, i.e."Bundesgesetz über Maßnahmen zur Abwehr der Ozonbelastung und die Information der Bevölkerung über hohe Ozonbelastungen"), the existing air pollution measurement networks operated by the Austrian federal states had to be combined to a system observing the overall air quality.

The system designed by the Austrian Research Center Seibersdorf was implemented in October 1993 and is called Austrian Ozone Network. The system really provide the whole range of measurements available and not only ozone [3]. Before talking about the system itself let´s have a view on some integration problems we were faced last year [4] before designing an integration concept for the overall network:

- Hardware, operating systems and local network technology are different in nearly every monitoring system.
- Storage technology differs on a wide range (file systems, conventional databases, sophisticated non-standard databases for time series storage and retrieval).
- Data models are different.
- Names and naming conventions are handled differently.
- Data quality and control values are treated differently.

- Measurements of equal components are made with different methods and are built in different units.
- There are national thresholds and local thresholds. Alarm handling is different.
- The systems change over time.

It should be mentioned that for integration the most important problems are that the systems are operated autonomously and that they change dynamically as new components are measured, new stations are built or measurement methods of components change. To treat all the above mentioned problems and requirements you are forced to develop your system as an open one.

4.1 Open approach

Figure 2 gives an overview on the network itself. There are two main interfaces: first, the interface in-between the 10 Ozone Network Servers (ONS) and second the interface between each ONS and its corresponding Local Computer System (LCS).

The Ozone Network Server as front end machine (communication computer system) does the work of integration. It makes all the transformations form local to global conventions, services the network, communicates with the local hosts (via LAN) and provides all the necessary interfaces for metainformation updates, data transport and visualization applications.
The system is totally decentralized. There is no central database storing data and the services are decentralized, too. It is possible that every monitoring system uses directly the information gathered in every other monitoring system.

The system consists of a number of communication computers which act as the services for the existing monitoring systems and for each other. Data, messages and changes in the structure of the monitoring systems are transferred to the communication computers and from there to the other communication computers which make the information available for other monitoring systems.

The system is totally decentralized. There is no central database storing data and the services are decentralized, too. It is possible that every monitoring system uses directly the information gathered in every other monitoring system. The local monitoring systems are connected to the network through a local application programming interface, which is implemented using remote procedure calls (RPC). The monitoring system acts as server for the network as well as the communication computer acts as server for the monitoring system and other local hosts.

For a user either on the local host or on the communication computer, it will always look as if there is one global data catalogue although the catalogue is distributed. The local part of the catalogue is updated from the metainformation of the monitoring system if the monitoring system stores such metainformation,

if not, the local metainformation is updated by the system administrator on the communication computer. For the catalogue exists an interactive program to query the catalogue contents. Every request goes through the catalogue.
To provide the needed functionality, we need several databases: an administration database and a cache memory. These databases are operated upon by the services. A central service is the catalogue service, which is responsible for keeping the data catalogues up-to-date.

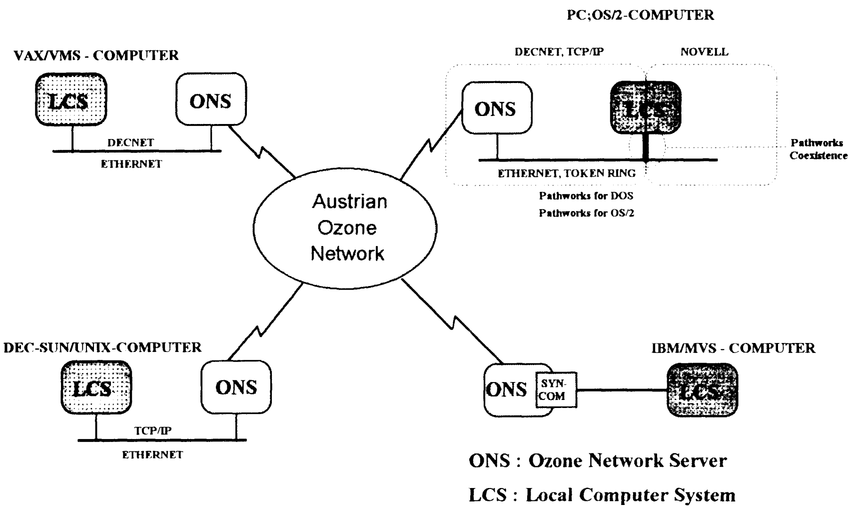

Figure 2: Austrian Ozone Network an open approach for EIS

The second central service is the data service for queries and data transport. It is completed with a visualization service, which serves for quick overviews, and a data processing service , which is responsible for data aggregation and calculus.
In the next section we want to talk about some of the difficulties in building the system.
The general task of the system is to make remote data available *and* to assure that remote data is handled correctly. This is a difficult task, as different measurement methods, different units, different representation of the data are used. Therefore one of the main functions the system must provide is a valid transformation of local data into a global data and vice versa. This transformation includes: name transformation, unit transformation and transformation of validity stamps on the data. We must also assure that local thresholds are processed in another way than global thresholds.

The next main difficulty lay in the design of the local interface. Thereby you have to standardize in a way that you can transport everything you need (data and metainformation) without interfering with the local systems in a way that they are affected too much.

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

To handle the above mentioned difficulties we installed a set of servers and describe them now in a very brief overview (for detail see [4]).

4.2 Overview of the ozone network servers

The ONS is implemented as a set of RPC-servers. Each main application task is coupled with a server implementation. Therefore we have servers for the transport of metainformation, values (data), communication, database updates and alarming. Each server is strictly based on RPCs. The network interfaces on both ends (LAN,WAN) handle the RPCs sent by the different servers. In the following we will discuss some of the server processes we implemented for the Austrian Ozone Network:

- **Metainformation Server**: The metainformation server handles the insertion into, update of and retrieval of metainformation from the metainformation database. If changes in the database occur at one node, the metainformation server broadcasts automatically all new metainformation, in order to care, that the metainformation databases of all ONS's in the network are always up to date.
- **Communication Computer System (ONS):** The Communication Computer System is the central communication and control application of the ozone network. All other servers, services and functions depend on the existence and functionality of the ONS. It controls the exchange of values via the Formula Server as well as it enables the export and import of metadata via the Metainformation Server and communicates with the local computer system (LCS).
- **Formula Server:** For data query a high speed retrieval mechanism for relational databases is implemented into the query server called Formula.
- **Alarm Server:** If in two measuring stations of an ozone monitoring area the ozone concentration arise about 100 ppb an alarm schema is automatically activated which is handled by a separate alarm server. It manages also the broadcasting of all alarm messages.
- **Message Server:** It handles the automatically broadcasting of different messages within the Austrian Ozone Network and is used to store a message history.
- **Update Server (Low Level Files):** This server allows only the transport of data (values) by file transfer from LCS to ONS.
- **Data Server :**The data server communicates with the formula server and enables formula to get data from remote nodes.
- **LCS-Data Server (local computer system interface):** The LCS-Data Server is a service program installed as an interface for the local computer system and handles requests from the communication computer system (ONS).

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

4.3. Visualization applications

In order to deal with the complex system, it was necessary to design an acceptable man-machine interface. Therefore the user interface is based on a set of visualization applications. These applications are highly interactive and react flexible and dynamic on changes of the system over time (i.e. new stations are built in). It consists currently of five main components:

- user profile management
- map with integrated metainformation display
- time series display
- daily report generation tool
- alarm display.

Database administration and maintenance is also done by a separate visualization application under SMARTStar Vision. For further details see [5].

5. Conclusions

In this paper we talked about general integration problems in Environmental Information Systems and showed a real design of a system which is currently working under pressure of an ozone period. The challenge to solve much of the problems discussed above and to do a good work was really enormous. Our main conclusion is that if your task is to implement a system integrating a number, perhaps a large number, of heterogeneous, dynamic and autonomous systems, you are still in a difficult position. For EIS there is still a lack of concepts, methods and tools for such a purpose in general.

References

[1] M. Hälker, A. Jaeschke (eds.), 6. Symposium Computer Science for Environmental Protection, Munich, December 1991, Informatik-Fachbericht 296, Springer, 1991

[2] R. Denzer, Graphics for Environmental Decision Making, IEEE Computer Graphics & Applications (IEEE CG&A), March 1993

[3] G. Schimak, H. Humer, R. Denzer, Ozone Monitoring in Austria, in: P. Zannetti (ed.), Air Pollution 94, Barcelona, Spain, September 1994, Elsevier, 1994

[4] G. Schimak R. Denzer, , H. Humer, Integration in the Austrian Ozone Network in R. Güttler, Ralf Denzer (ed.) Integration von Umweltdaten 1994 (Integration of Environmental Data 1994), 2. Workshop, Dagstuhl Castle, February 1994, to be published

[5] G. Schimak, R. Denzer, Visualization in the Austrian Ozone Network in Ralf Denzer (ed.) Visualisierung von Umweltdaten 1994 (Visualization of Environmental Data 1994), 4. Workshop, Dagstuhl Castle, January 1994, to be published