A Web Accessible Environmental Model Base: a Tool for Natural Resources Management

Giorgio Guariso*, Emma Tracanella*, Luigi Piroddi*, Andrea E. Rizzoli‡
*Dipartimento di Elettronica e Informazione, Politecnico di Milano
Milano, Italy
{guariso, tracanel, piroddi}@elet.polimi.it
‡IDSIA, Istituto Delle Molle di Studi sull'Intelligenza Artificiale
Lugano, Switzerland
andrea@idsia.ch

Abstract An environmental model base, accessible via the World-Wide Web, is presented. The information stored in this model base can be retrieved using various methods such as structured queries, full text search, exact and approximate keyword search. These search tools help the environmental modeller to find a suitable model to solve their modelling problem. The modeller is then presented with a set of information, regarding the models matching the query, including source and executable code, user manuals, and references to supplemental documentation. The client-server architecture of the model base enables the maintainers to provide frequent and regular upgrades, thus giving the users access to a single data source which contains up to date and constantly increasing information. This model base has been developed as part of the GAIA project, sponsored by the European Community, which is aimed at expanding the role of Information Technology in developing countries.

1. INTRODUCTION

We describe the design and implementation of a model base covering a number of environmental models which can be queried on the World-Wide Web (WWW) via an easy and flexible user interface. Powerful search tools such as structured queries, full text search, exact and approximate keyword search (the latter is performed by successive relaxation of keywords) are provided to access the environmental model base.

This work is part of the GAIA project, which is supported by the European Community. The aim of the GAIA project is to produce a multi-media tool for natural resource management and environmental education. This tool shall meet the needs of developing countries; the expected outcome is to increase the role of Information Technology (IT) in the management of natural resources, thus increasing the ability of decision makers in using IT tools to evaluate and assess the rapid changes affecting the environment of these areas.

The importance of this project is particularly relevant if we consider that, while developed countries can only think about preservation of limited areas of their pristine environment, in many cases developing countries are still in time to decide how they want to change their natural environment. Our hope is that correct and up-to-date information can lead to better decisions, avoiding many of the errors which have impaired the preservation of the natural environment in developed countries.

In this framework, an environmental model base may give the decision makers the possibility to access a wide base of catalogued environmental models, documented by their technical and user's manuals, and often in executable form. The users, who must solve a given problem (e.g. predicting the dispersion of sulphur dioxide in a urban environment), are "driven" towards the model which can solve their case by the various search tools.

In the remainder of the paper we briefly review the use of environmental model bases. Then, we describe the design of the GAIA model base and detail the model base implementation. This is based on a client-server architecture which employs, on the server side, the MiniSQL database and various Perl CGI (Common Gateway Interface) scripts and, on the client side, the Netscape internet browser with Javascript extensions. Finally, we describe a sample session on GAIA.

2. ENVIRONMENTAL MODEL BASE STRUCTURES

Nowadays, several hundred environmental models are available, both as scientific papers and as computer executables, and, therefore, model selection is one of the key activities in environmental analysis and management (Guariso and Rizzoli, 1996).

Usually, each model has been developed to represent a limited set of environmental conditions and thus its reliability is limited to such circumstances. However, understanding whether a given situation can be handled within the model capability is a difficult task, normally involving a large body of expert knowledge, typically expressed in qualitative terms (Guariso and Werthner, 1986, Del Furia and Petrucci, 1994). This remark suggests that expert systems may be a suitable tool to store the modellers' knowledge and to bring it to the users' fingertips. Indeed, such an approach has been widely used and the experience of the model developers has been translated into large knowledge bases. Most of the rules in these bases were devoted to match a set of problem descriptors to the applicability conditions of the model. Given the variety of possible environmental conditions, these rule bases can grow to a very large extent with a minimum of inference activity and may cause obvious maintenance problems.

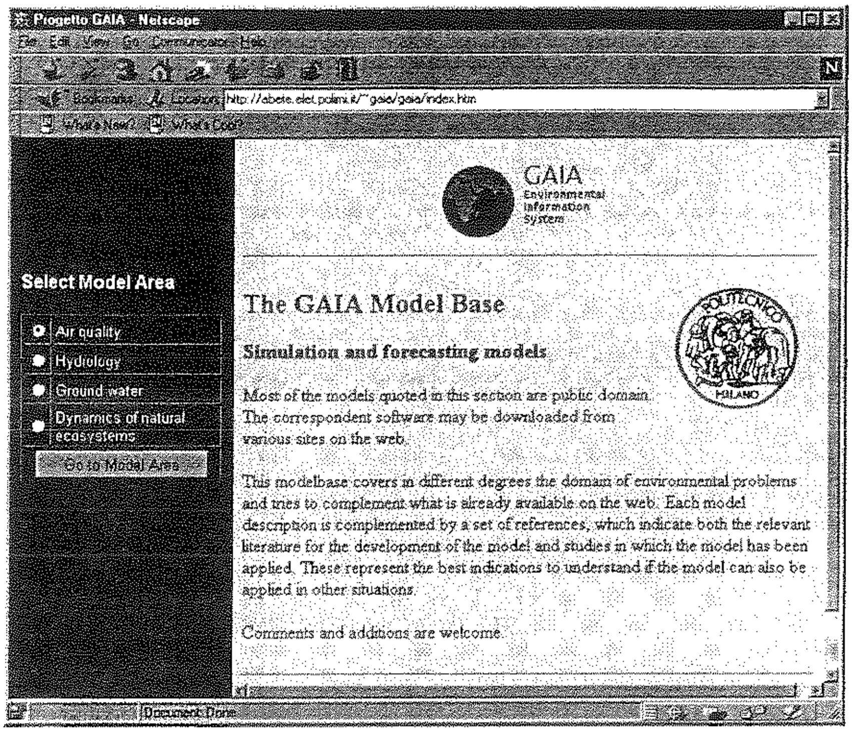

Figure 1. The entry of the GAIA model base. The user can select the model area of interest.

A different approach can be used to solve the problem of model applicability. It is based on a more structured classification of the problem in terms of a number of descriptive attributes. Each problem is identified by a set of values assigned to these attributes and can be stored in a classical database table together with a reference to the suitable models/codes.

Our model base uses this last approach, based upon the experience we acquired in the creation of model selection environments (Calori *et al.*, 1994a and b). Efficiency in model search was improved by the adoption of a relational database with enhanced query capabilities

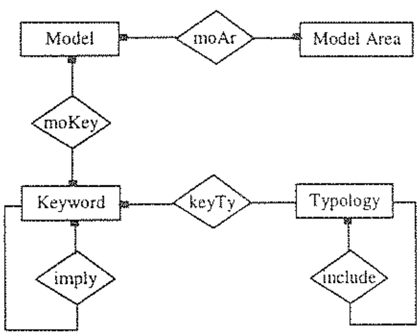

Figure 2: the Entity-Relationship diagram for Models, Keywords and Keyword Typologies (the black box indicates a one-to-many relationship).

3. THE DESIGN OF THE GAIA MODEL BASE

The GAIA model base was designed with the idea of hypertextual navigation in mind and for this purpose its design was tackled from two sides: its static structure, and its dynamic, "navigational" structure.

The static structure defines the data base implementing the model base contents and the semantic relationships among these contents. This structure was designed using the Entity-Relationships (ER) schema of analysis.

The dynamic structure is a hypermedia model of data: it displays not only the contents and the semantic relationships, but also the navigational relationships and the access methods to the contents of the model base. This analysis was performed using the Hypermedia Definition Model (HDM) schema (Garzotto and Paolini, 1993).

The dynamic structure is accessed using WWW tools, such as Internet browsers. An introduction to these concepts and to the terminology used in this paper can be found in Berners-Lee *et al.* [1994].

3.1 The static structure

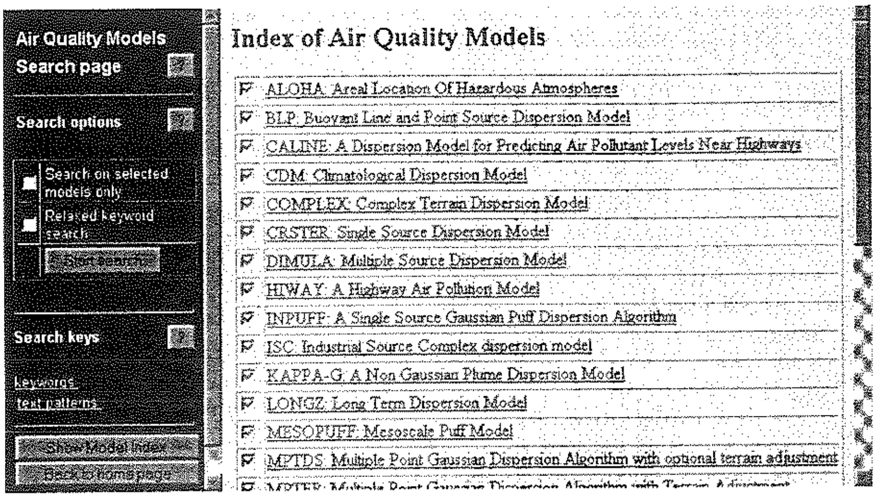

Figure 3. The area for Air quality models. Note the navigational frame at the left.

The model base is constituted by a set of database tables. The following ones identify the contents and the semantic relationships used to classify the models:

- the Model table, characterised by its unique identifier (UID), model area identifier, name, authors, acronym and a URL (uniform resource locator) reference to the HTML pages with all the informational details;

- the Typology table, which has a UID and a name, and contains both the model areas and the keyword typologies;

- the Keyword table, identified by a UID and a name, which contains the whole set of searchable keywords;

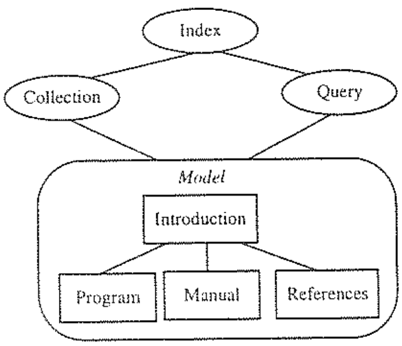

Figure 4. The simplified HDM schema for the model base.

The semantic relationships among the entities as represented in Figure 2, are as follows:

- a model belongs to one or more model areas, and a model area contains one or more models;

- a keyword describes one ore more models (e.g. the keyword "saturated soil" describes a wide set of groundwater models) and a model can be described by one or, often, more keywords (e.g. a river eutrophication model can be described by the following keywords: "ecological", "river", "eutrophication", "algae", and many more); each keyword is associated with one or more keyword typologies, for instance the keyword "river" is associated with the "hydrological" keyword typology; a complete classification of the model specifies at least one characteristic per keyword typology.

Moreover the data model can describe "mutual exclusion" between keywords belonging to the same typology and "implication" relationship between keywords and typologies (e.g. the keyword "eutrophication" implies the "ecological" keyword).

3.2 The dynamic structure

The HDM model defines a hierarchical structure of web documents which mediate the access to the entities in the model base tables. The main access is through the main index which allows either for the selection of one of the collections of models (model areas) or the set up of a model base query. In the GAIA model base the model collections are Air quality, Ground water, Hydrology, and Dynamics of natural ecosystems. The available queries are based on the specification of keywords freely entered by the user or by the definition of a hierarchical structure of constraints which can be later relaxed if the query was too narrow (e.g., no model matched the query). The keyword search can be enriched by a full text search on the root HTML pages of the models. A distinctive feature of our model base is the possibility of searching a model by relaxation of keywords and, at the same time, looking for a given text pattern.

After the user has selected an entry point, the GAIA model base displays one or more links to the HTML hypertext part, which contain all the information and available documentation concerning the models. Selecting one of these links, makes it possible to access the model's home page, which contains the following information (see Figure 4):

- the model description (the `Introduction` box in Figure 4), with keywords, authors, references (the `Reference` box) and a list of validation studies;

- the associated software description and availability (the `Program` and `Manual` boxes), with the URL

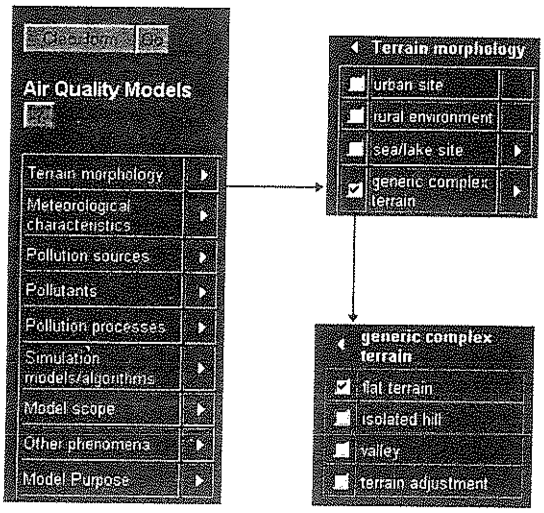

Figure 5. The refinement of the query by keyword.

referring to the location of the software on the Internet;

- a link to the software manuals and validation studies.

3.3 Implementation details

The outlined architecture of the model base is supported in its static structure by a relational database manager, while its dynamic structure has been implemented with a set of HTML documents, enhanced with Javascript extensions (Netscape, 1997).

The relational database engine we chose is the MiniSQL database engine, which comes with a ready to use CGI programming interface, for easy access of the database via the Internet. The database engine resides on the server (in our case an HP workstation, with the HP-UX UNIX flavour and an Apache 1.0 web server) and provides access, via SQL queries, to the database tables described in section 3.1.

The server also houses a set of HTML documents which can be interactively downloaded using an internet browser by the clients, which are the remote users of the environmental model base. Using Javascript extensions to standard HTML, we were able to let the client application (the Internet browser) keep track of the user's choices while forming the query. The HTTP (Hyper Text Transfer Protocol) (W3C, 1997) is an atomic protocol, which means that is "memoryless" and each transaction is independent on the previous one. On the other hand, the query strategy we devised is based on refining the query through a series of user choices which are dependent on each other, using the keyword and keyword typology data structures. The Javascript program, embedded in the HTML documents, performs the task of remembering the user selections, thus avoiding to create "ad-hoc" CGI programs on the server to keep track of the user's history and reducing the quantity of information to be transferred at each HTTP transaction.

The application works with Netscape 3.01 and above and with Internet Explorer 3.02 and above.

4. NAVIGATING IN THE MODEL BASE

The home page of the GAIA model base is composed by an interaction area (navigational bar) on the left side of the window and of a wider area for displaying results (display frame) which occupies the remaining part of the screen (see Figure 1).

The first step is to select the area of interest of the model (Air Quality, Hydrology, Ground water, Dynamics of Natural ecosystems). In this example, we enter the Air quality area (see Figure 3). In the display frame a list of all the available air-quality models is shown. This list is composed of active hyperlinks which lead directly to the models' home pages. In the navigational bar, search options are offered to the user. The search options define the extent of the query: either on a subset of the available models listed in the display area and/or using relaxed keyword search. The search keys define a search based either on keywords or on full-text. The full-text search retrieves any model where an occurrence of the specified word is found. In the display area, the list of matching models is presented. The keyword search is more articulated and we explain it using the example. In Figure 5, we report the successive navigational bar which appear in sequence when specifying a keyword typology. Clicking on the small arrows at the left or right of the keyword typologies the user can jump back and forth from different keyword specification levels. In our example, the keyword typology "Terrain morphology" is selected and a new navigational bar appears: the user can now either select on of two values ("urban site" and " rural environment") or further specify the query entering one of the additional keyword typologies ("sea/lake site" or "generic complex terrain"). When the user is satisfied with the set of keywords and keyword typologies she has entered, she can go back to the "Air quality model search page" and select a search option (see Figure 6). We select a "relaxed keyword search" and the results appear in the display area: three models match exactly the keywords "point sources", "flat terrain" and "simulation", but relaxing the "point sources" constraint and extra model is found.

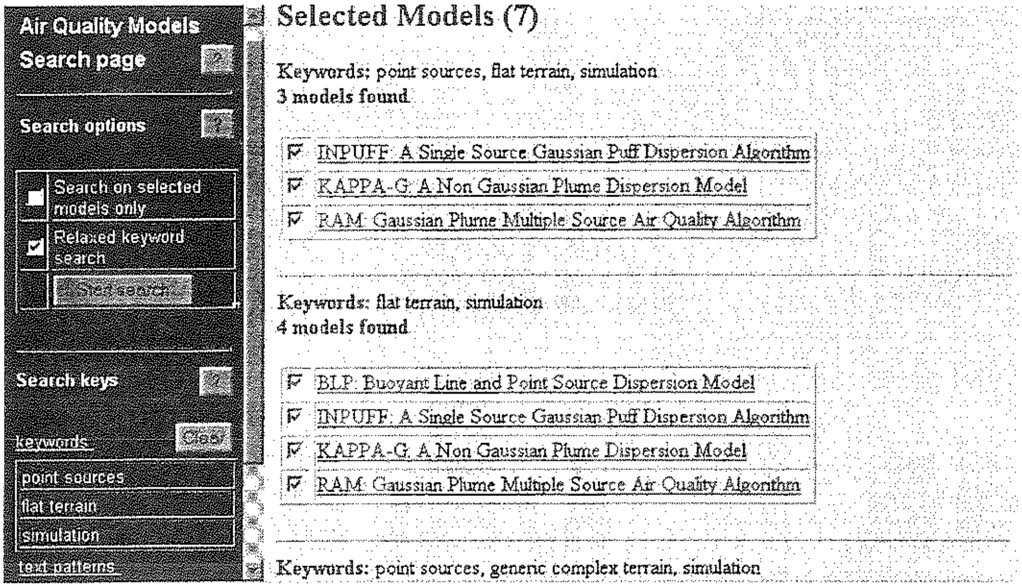

Figure 6: the results of the query.

5. AND THE FUTURE?

Our environmental model base is a first tentative, among others (see, for instance, Knorrenschild, 1996, IGWMC, 1997), to distribute information, providing easy access and centralised management of the data sources. The Internet, as it is today, allows for these objectives, but the rate of innovation is high and we can, helped by current research, try to envision what will be next. In the near future, there are suggestions (Abel *et al.* 1997a and 1997b, Bhargava *et al.* 1997 and Ba *et al.* 1997) that model bases will evolve in order to widen the spectrum of provided services.

Up to now, model bases have been implemented as static repositories the main purpose of which is to act as software libraries. The result of a query is usually a reference to model documentation, sometimes to its software implementation. To be used, the model has to be downloaded, sometimes implemented, and then executed on the user's computer.

Model bases are now evolving towards a more active role, allowing the user not only to find a model which suits his needs, but also to execute it and return the results, thus making modelling resources readily available for use. The Internet has been providing the tools to implement such a solution for a long time, using a traditional *client-server* architecture. The user gains access, via a client software, to the model base which is implemented on a remote server and stores model descriptions and their implementations, ready to be executed. This architecture has some weak points: each client has to be customised to access a specific server, thus making the interconnection of model bases quite hard, and moreover the model base has to bear the overhead imposed by model executions: it must provide not only a fast repository to access information, but a huge computing power to execute models.

Recent developments in distributed software architectures have proposed the *client-broker-server* approach. This architecture has already found widespread acceptance in the database field and, in particular, the CORBA approach has started to fulfill its promises (see, for instance, OMG, 1997, and TU Wien DSG, 1997). The main advantage of this architecture is that the model base acts as a *broker* between the user and the model. The modeller interacts with a local Model Management System (Blanning *et al.* 1993), running on a client workstation, connecting to a remote model base. The model base finds an appropriate model and contacts the server where its implementation is stored. The model executables, which are available on servers, different from the one which hosts the model base, will need to be encapsulated by an homogeneous model interface in order to be shared in the client-broker-server architecture. Research on the practical development of a common interface for model encapsulation is ongoing (see, for instance, Abel *et al.* 1997a, and Reed *et al.* in this volume).

The problems we have addressed in this paper, such as finding the model which suits some specific need, will play an important role in the "brokerage" services provided by future environmental model bases (Abel *et al.* 1997b).

6. CONCLUSIONS

The environmental model base which can be accessed via the Internet described in this paper aims at providing up-to-date and easy to access information to decision makers and modellers in the environmental field. The use of an environmental model base favours reuse of existing knowledge versus "reinventing the wheel", thus reducing development costs and time. Our model base has been developed as part of the GAIA project (ESS, 1997), aimed at providing a set of multi-media tools for natural resources management and environmental education for developing countries. It complements a set of case studies in using environmental models for actual developing countries problems, several data browsing facilities and on-line Java model implementations.

7. ACKNOWLEDGMENTS

The GAIA project is sponsored by the European Commission/Directorate General XIII within the programme "Cooperation with third countries and international organisations/Scientific and technological cooperation with developing countries". We thank Dr Kurt Fedra, ESS Austria, for his dedicated involvement as leader of the GAIA project and Dr Luca Mainetti, Politecnico di Milano, for his contribution to the design of the data structures in the model base.

8. REFERENCES

Abel, D.J., K. L. Taylor, D. Kuo, Integrating modelling systems for environmental management information systems, *ACM Sigmod, 26(1)*, 1997a.

Abel, D.J., V. Gaede, K. L. Taylor, X. Zhou, SMART: towards internet marketplaces, Technical Report, CSIRO Mathematical and Information Sciences, Canberra, Australia, 1997b.

Ba, S., R. Kalakota, A.B. Whinston. Using client-broker-server architecture for Intranet decision support. . *Decision Support Systems, 19(3)*, 1997.

Berners-Lee, T., R. Caillau, A. Lutonen, H.F. Neilsen and A. Secret, The World-Wide Web, *Communications of the ACM, 37(8)*, 76-82, 1994.

Bhargava, H.K., R. Krishnan, and R. Müller. Decision support on demand: emerging electronic markets for decision technologies. *Decision Support Systems*, *19(3)*, 1997.

Blanning, R.W., C.W. Holsapple and A.B. Whinston (eds.). Decision Support Systems, Special Issue on Model Management Systems, 9(1), 1993.

Calori, G., F. Colombo, and G. Finzi, FRAME: a Knowledge-Based Tool to Support the Choice of the Right Air Pollution Model, in *Computer Support for Environmental Impact Assessment*, edited by G. Guariso and B. Page, pp. 211-222, North Holland, Amsterdam, 1994a.

Calori, G., G. Guariso, and A. Rizzoli. From expert systems to expert databases: tools for model selection, paper presented at Resource Technology '94, Melbourne, 1994b.

Del Furia L. and M. Petrucci, LIVIA Project: a Computer Laboratory for Environmental Impact Assessment", in *Computer Support for Enviromnental Impact Assessment* edited by G. Guariso and B. Page, pp. 311-320, North Holland, Amsterdam, 1994

Garzotto F., Paolini P. and D. Schwabe, HDM - A model based approach to hypermedia application design, *ACM Trans. Off. Inf. Sys.*, 11 (1), 1993.

Guariso G. and H. Werthner, A Computerized inventory for water resource models, *Env. Software*, *1(1)*, 40–46, 1986.

Guariso, G., and A. Rizzoli. Developing environmental model bases, paper presented at Envirosoft 96, Como, 1996.

Guariso, G., and E. Tracanella. HYPERMODEL: a hypermedia guide to environmental models, paper presented at En+In., International Conference on Environmental Informatics, Budapest, July 1995.

Reed, M. S.M. Cuddy, A.E. Rizzoli, A framework for modelling multiple resource management issues - an open modelling approach, paper presented at Modsim 97, December 1997, Hobart, Tasmania.

9. ONLINE REFERENCES

CIRITA, Politecnico di Milano, 1997. The GAIA Model Base, Simulation and forecasting models (http://abete.elet.polimi.it/~gaia/gaia/index.htm).

Environmental Software Cooperative, 1997. The Environmental Software Resource Guide (http://www.envirosw.com/software.html).

ESS Austria, 1997. GAIA: multi-media for natural resources management, (http://www.ess.co.at/GAIA).

Hughes Technologies Pty. Ltd., 1996, Mini SQL Version 1 Library, (http://www.Hughes.com.au/library/msql1/).

IGWMC - International Ground Water Modeling Center, 1997. Ground-Water Modeling Software (http://magma.Mines.EDU/igwmc/all/software/).

Knorrenschild, M., R. Lenz, E. Foster, C. Herderich. UFIS: A database of ecological models. Ecological Modelling 86, n. 2-3, pp. 141-144, 1996 (http://www.gsf.de/UFIS/ufis/modelhom.html).

Netscape Communications Corporation, 1997. Javascript Authoring Guide, (http://home.netscape.com/eng/mozilla/Gold/handbook/javascript/index.html).

OMG, 1997, Object Management Group Home Page, (http://www.omg.org/).

TU Wien DSG, 1997, Distributed Object Computing with CORBA, (http://www.infosys.tuwien.ac.at/Research/Corba/).

W3C - World Wide Web Consortium, 1997, HTTP - Hypertext Transfer Protocol (http://www.w3.org/Protocols/).