EnviroInfo 2000: Umweltinformatik ’00 Umweltinformation für Planung, Politik und Öffentlichkeit
Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

CoastBase – The Virtual European Coastal and Marine Data Warehouse

Wassili Kazakos, Ralf Kramer and Andreas Schmidt[1]

Abstract

CoastBase is a project within the IST programme of the European Commission, which started at the beginning of this year. CoastBase aims to improve European marine and coastal management with regard both to the search for and access to a broad range of information sources. In this paper we present the project’s main objective, which is the development of a user friendly, virtual, multilingual, multi-platform, Internet-accessible architecture for searching and querying distributed coastal and marine information sources.

1. Introduction

1.1 Motivation

As described in the recent reflection paper of the European Commission ”Towards a European ICZM Strategy” (1999), coastal areas and their natural resources (marine and terrestrial) have a strategic role to play in meeting the needs and aspirations of current and future European populations. Important examples of functional uses include: tourism, shipping, industry and energy, fishing and mariculture, coastal defence and natural development. It is essential that these functions can develop in a sustainable way to support conditions for human health, employment and regional development and environmental quality. The physical and administrative units related to coastal systems and its environmental problems may vary from local up to transboundary and supranational.

Authorities responsible for the management of these areas need to consider both socio-economic and environmental aspects of these functions. Human pressures pose the risk of destroying habitats and the resource base of the coastal zones, and with them the ability of the coastal zone to perform many of its essential functions.

[1] Forschungszentrum Informatik (FZI), Haid-und-Neu-Straße 10-14, D-76131 Karlsruhe, Germany; kazakos | aschmidt@fzi.de; http://www.fzi.de/dbs.html. Current address of 2nd author: G.Braun electronic media services GmbH, Karl-Friedrich-Str. 14-18, D-76133 Karlsruhe, RalfKramer@gbraun-ems.de)

The EU Demonstration Programme for ICZM and the abovementioned Reflection Paper suggest that integrated coastal management is needed, including communication between all players representing different sectors and between administrations at different levels.

Access to all relevant information is essential to this approach. This information is generally dispersed among coastal institutes and organisations and neither readily available nor accessible. This poor availability and accessibility of information hampers effective planning and decision making; the need to improve information provision is the main motivation of this CoastBase project

1.2 Outline

The paper is organised as follows. In Section 2, we introduce the basics. In Section 3, we place our work in the context of those recent projects that we feel are most closely related to CoastBase. In Section 4, the main part of this paper, we outline the general approach. Preliminary conclusions and an outlook on future work conclude the paper in Section 5.

2. Basics

2.1 Catalogues

In order to respond to complex environmental problems, it is necessary to have information at hand covering different fields of application, and in most cases this will be located at different sites. Only if the respective information is sufficiently available is it possible to find complex interrelations, to effectively survey environmental laws, and also share existing information. This avoids collecting data that already has been collected and keeping data without any further due to a lack of knowledge about the data.

Finding the answer to a certain question means discovering *what* information is available, *where* the information is managed, *how* this information can be obtained, and *how* to interpret the information correctly. The information that is necessary to obtain abovementioned information is called *meta-information*. This information can be compared to the information of classical index cards in library catalogues, which describe books but are not the books themselves. The present explosion in data volumes makes it even more important for the user to be able to rely on information about existing data in order to find what he or she needs. The equivalent of the library catalogues containing index cards, meta-information systems or *catalogue systems* are the more general electronic form. Typical examples of environmental catalogues are the German/Austrian Umweltdatenkatalog (UDK) and the Catalogue of Data Sources (CDS) of the European Environment Agency. Both

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

deal with descriptive information about environmental resources (Kazakos et al. 1998; Swoboda et al. 1999).

2.2 Information Integration

Integration of information from heterogeneous sources has been a major topic in database related research for several years. Roughly speaking, there are two possible approaches (Widom 1996):

- *The materialised (or Data Warehousing) approach.* Information that may be of interest is extracted/exported from each source and – after filtering, harmonisation and fusion with information from other sources – stored in a (logically) centralised repository (the data warehouse). User queries are evaluated at the central repository without connecting to the individual sources.
- *The virtual (or mediated) approach.* When a user poses a query, this query is sent directly (after necessary transformations) to the appropriate sources for evaluation. Their results are then filtered, harmonised and fused and presented to the user.

Whereas the materialised approach permits optimisation of response times for certain applications and offers reliability of access, it has serious drawbacks for updating the repository when the content of the information sources changes. The autonomy of the information sources has to be restricted to permit implementation of updating policy. The virtual approach on the contrary allows the content to be left in its original place. Changes to the information sources are immediately reflected in the query results.

Prominent virtual integration projects like TSIMMIS (Hammer et al. 1997; Garcia-Molina et al. 1997), Information Manifold (Levy 1998), or MIX (Baru et al. 1999) can be more or less accurately subdivided into three main logical components according to the I³ reference architecture's terminology (Arens et al. 1995):

- *Wrappers* overcome the technical and syntactical heterogeneity of the individual sources.
- *Mediators* overcome the semantic differences (schema/information model) between the sources and fuse equivalent information artefacts.
- *Facilitators* select the sources needed to satisfy a given user need and combine them appropriately.

3. Related Work

In this section, we place our work in the context of those projects that we feel are most closely related to it.

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

3.1 AirBase

AirBase (Sluyter et al. 1997) is a European Environment Agency (EEA) project which seeks to establish a multi-level information system for storing data about air quality collected in a pan-European air quality monitoring network (EUROAIRNET). AirBase has a layered architecture: The *basic layer* consists of a single database into which the data collected from the member countries are imported after they have been converted to the appropriate format. The *common layer* is an extract from the database of the basic layer and can be distributed in different versions, e.g. CD-ROM. For web-based access to the database, a Java applet was implemented at the *strategic layer* that allows users to search for available air quality measurements and to view visualised and aggregated versions of these measurements.

Unlike CoastBase, AirBase uses a materialised approach. The collected data is uploaded by member countries into a centralised database. The information sources (member countries) must ensure they keep this centralised database up to date.

3.2 Remssbot

Like CoastBase, Remssbot focused on the integration of different environmental sources without incorporating their content in a centralised data warehouse. Remssbot’s front-end is HTML based. Communication between the different system components is achieved by a CORBA-based middleware layer. In order to locate relevant information for the user, Remssbot uses a centralised meta-data repository which is an extension of an early version of the CDS (Kazakos et al. 1998) data model.

Whereas Remssbot only keeps the data at its original location, CoastBase goes one step further and tries to keep even the meta-data at its original location, i.e., Remssbot employs the virtual integration approach only for the data; CoastBase employs this approach for the catalogue as well. Furthermore, CoastBase permits not only data retrieval but also data processing by aggregation.

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

4. General Approach

4.1 Overview

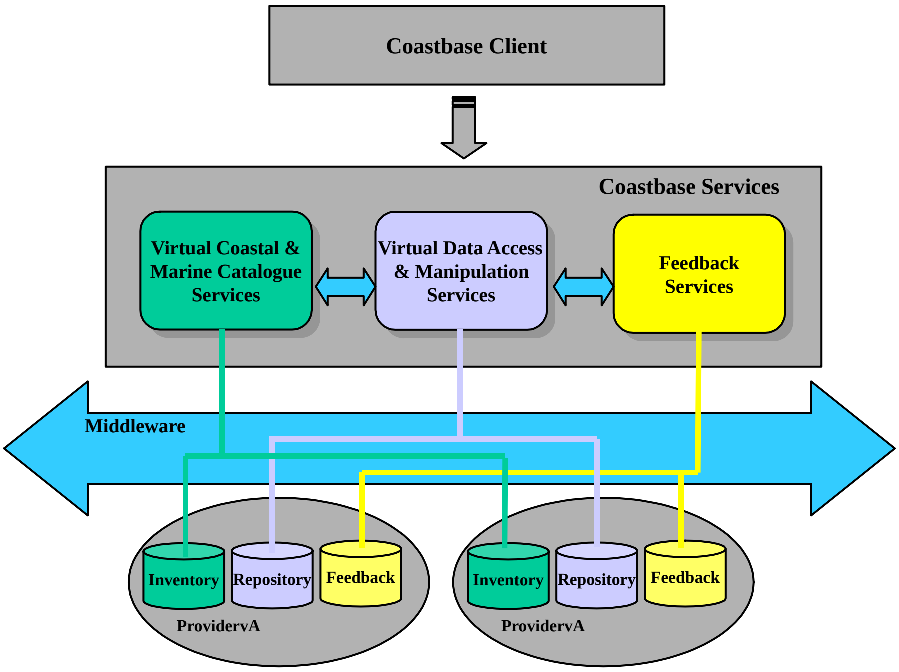

Figure 1
CoastBase Architecture -- Overview

The general Architecture of CoastBase provides for three building blocks corresponding to its three functional units, search, access and manipulation, and feedback, as shown in Figure 1. The CoastBase services will be accessible for the users through one common client.

4.2 CoastBase Client

The CoastBase client will be the user front-end and offers the functionality of CoastBase in an easy to use way. It will be implemented using a combination of HTML for the basic functionality and Java for the representation of specific data, like vector graphics. The multilingual HTML based user interface will be implemented using Java Server Pages. For this purpose some parts of the WebCDS

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

can be reused, but the main parts have to be re-implemented to achieve a common CoastBase look and feel, to allow access to real data and to include the virtual catalogue system.

4.3 Virtual Catalogue Services

Separate descriptions of the data are needed to permit users to search efficiently. These descriptions are often referred to as meta-data or catalogue data since they are not operational data, but data about data and are used by the data provider for cataloguing purposes. In heterogeneous and distributed environments, there are several ways to manage the meta-data depending on the needs of the system as to scalability, distribution, actuality, and on the policies of the participants. To elaborate on that, we must clearly define the point of departure. Within CoastBase the meta-data from a diversity of distributed and heterogeneous data sources have to be managed and must remain up to date. Classical approaches involving centralised repositories usually perform poorly on scalability and updates, and they are difficult to enforce while incorporating different policies at the local to European level.

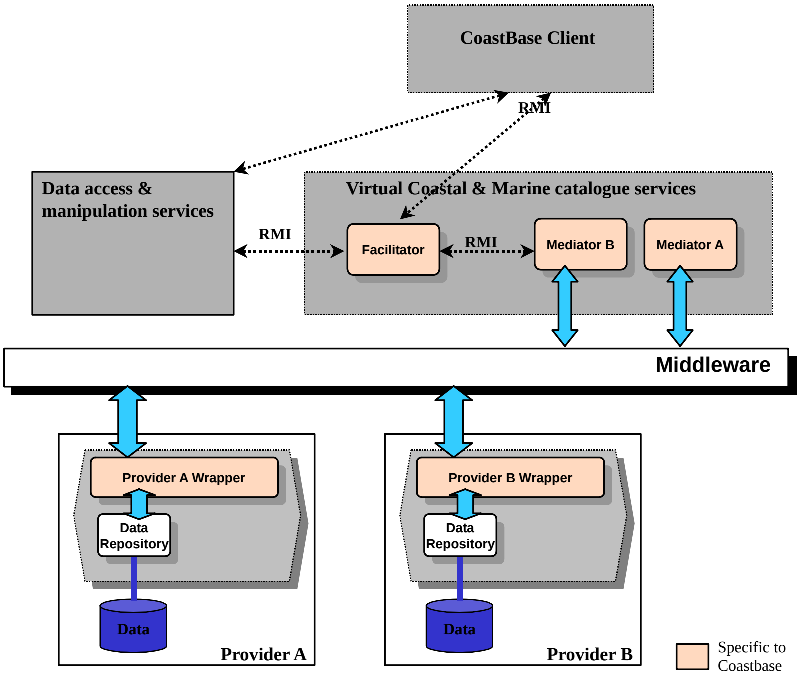

Figure 3
Virtual Coastal & Marine Catalogue Services

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

On the other hand users are not interested in details about distribution and heterogeneity of the data and meta-data, but want just one simple user interface to search for the data they need for their daily work. To allow distribution of catalogues on the one hand and to give the user the impression of accessing just one single source CoastBase will introduce the concept a Virtual Catalogue, as shown in figure 3.

4.4 Virtual Data Access and Manipulation Services

The Virtual Data Access & Manipulation Services will be built on top of the Coastal Information Server which was used also in several projects such as DESIMA and AVID.

The architecture for the data access and manipulation services is derived from the Coastal Information Server concept. CIS is composed of a *Core facility* which implements the basic services for the data access and manipulation in the CoastBase system and several *Extension facilities* which are installed locally at each provider's site. Interoperability between these facilities is achieved by a common infrastructure, the *Middleware*.

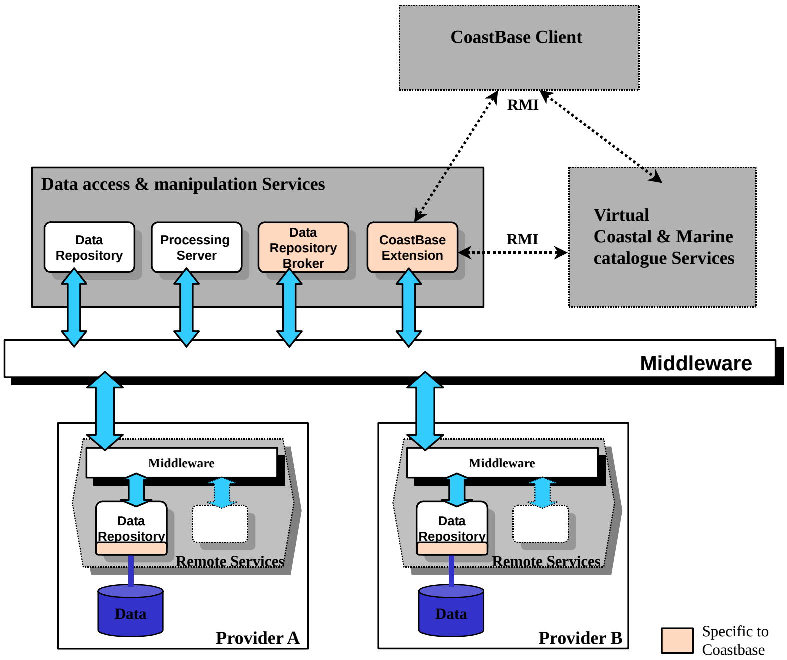

Figure 4
Data Access and Manipulation Services

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

The Core facility is part of the CoastBase system and provides the main entry point for data access and manipulation requests, like the *Data Repository* which manages the local repository, the *Processing Server* which is in charge of managing processing functions and the *Data Repository Broker* which is the key component that co-ordinates distributed and interoperable data access requests. It has the role of the main retrieval manager, which is responsible for handling all requests and performs the tasks of routing the requests to the appropriate provider.

Additionally, the *CoastBase Extension* provides the specific services for CoastBase functionality. In particular it implements the interface between the Virtual Data Access & Manipulation Services and the Virtual Coastal & Marine Catalogue Services. It uses the services offered by the other objects of the Core Facility.

The Extension facility is supported by every remote provider connected to the CoastBase system. In fact, it is a subset of the Core facility – installed at the provider's site – responsible for accessing the provider's data repository.

The Middleware is the communication node of the Core and Extension facilities. It enables distribution of the services. It is based on the CORBA ORB (Object Request Broker).

5. Conclusions and Outlook

In this paper, we have outlined the basic software architecture of CoastBase, the virtual European coastal and marine data warehouse. CoastBase is funded under the EU's IST programme. Similarly to classical business data warehouses, the CoastBase architecture comprises two main building blocks, namely the data access and manipulation part on the one hand side and the metadata or catalogue part on the other. Current work on the CoastBase project aims to flesh out this architecture taking into account user requirements.

6. Acknowledgements

The authors greatly appreciate the cooperation with all CoastBase partners (see http://www.coastbase.org section partners). Financial support of the European Commission under contract IST-1999-11406 is gratefully acknowledged.

7. References

AirBase Project Homepage, http://www.etcaq.rivm.nl/airbase/index.html

AVID Project Homepage, http://www.hydrostore.org

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

Arens, Y., Hull, R., King, R. (eds.) (1995): Reference Architecture for the Intelligent Integration of Information, Program on Intelligent Integration of Information, ARPA, Version 2.0

Baru, C., Gupta, A., Ludäscher, B., Marciano, R., Papakonstantinou, Y., Velikhov, P. (1999): XML-Based Information Mediation with MIX, Exhibitions Program of ACM SIGMOD 1999

DESIMA Project Homepage, http://desima.jrc.it/

Garcia-Molina, H., Papakonstantinou, Y., Quass, D., Rajaraman, A., Sagiv, Y., Ullman, J., Vassalos, V., Widom, J. (1997): The TSIMMIS approach to mediation: Data models and Languages, Journal of Intelligent Information Systems

Hammer, J., Garcia-Molina, H., Cho, J., Aranha, R., Crespo, A. (1997): Extracting Semistructured Information from the Web, Proceedings of the Workshop on Management of Semistructured Data. Tucson, Arizona, May 1997

Kazakos, W., Kramer, R., Nikolai, R., Rolker, C., Bjarnason, S., Jensen, S. (1999): WebCDS - A Java-based Catalogue System for European Environment Data. In Dogac, A., Özsu, M.T., Uluzoy, O (eds.): Current Trends in Data Management Technology, pp. 234 - 249. Idea Group Publisher, 1999.

Levy, A. (1998): The Information Manifold approach to data integration, IEEE Intelligent Systems, September/October 1998, pp. 11–16

Remssbot Project Homepage, http://www.netor.gr/remssbot/

Sluyter, R., Potma, C., Krognes, T., Petrakis, M., van Hooydonk, P. (1997): AirBase: 1997 Development Status and Extensions Foreseen, Second European Workshop on Air Quality Monitoring and Assessment, Brussels, 22-23 September 1997

Swoboda, W., Kruse, F., Nikolai, R., Kazakos, W., Nyhuis, D., Rousselle, H. (1999):The UDK Approach: the 4th Generation of the Environmental Data Catalogue for Austrian and German Public Authorities, Proc. IEEE Meta-Data'99, Bethesda, Maryland, USA, April 1999, http://computer.org/proceedings/meta/1999/papers/45/wswoboda.html

Widom, J. (1996): Integrating Heterogeneous Databases – Lazy or Eager?, ACM Computing Surveys 28(4), December 1996

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5