CoastBase: an information system for easier access to dispersed marine databases

P. Picco[1], W. Kazakos[2] and A. Schmidt[2]

[1] ENEA Marine Environment Research Centre S. Teresa, La Spezia, Italy
[2] FZI Forschungszentrum Informatik, Karlsruhe, Germany

ABSTRACT   CoastBase was an EU-IST project aimed at developing and testing an information system to access dispersed and heterogeneous marine and coastal data and information through an internet-accessible, multilingual, interoperable architecture. In contrast to the centralised databases, CoastBase relies on a "virtual" or mediated approach, that allows one to leave both data and metadata at its original location, and overcomes the problem of updating the database and metadata catalogues. This paper describes the main functions, such as catalogue search, data access, download and manipulation as well as the available data sets. The advantage for this kind of approach for marine and coastal planners and researchers is shown in the test case scenarios given.

Introduction

Availability of data and information is the basis for research and a sound environmental assessment. Data collected in the framework of coastal and marine research and during monitoring programmes are scattered in different institutes, archived in different systems with different formats and in metadata models. This often hampers an efficient data flow and exchange among these organisations and a wider user community, with the result that relevant data and information for environmental assessment, other research programmes, and policy making are lost. Several programmes have been devoted to promoting standardisation and facilitating data flow; some of them aim at gathering dispersed marine data into a single centralised database [MODB, MEDATLAS/2002: MEDAR Group (2002)] or at compiling catalogues and directories (Sea-Search). Nevertheless, after the project has ended, additional efforts are required to maintain the system, with the risk that the results achieved, if not regularly updated, might lose their value.

The main objective of the EU-IST project CoastBase - The Virtual European Coastal and Marine data Warehouse - (http://www.coastbase.org) was to design and test a prototype system to facilitate the search for and the access to dispersed marine and coastal information (Niesing, 2001). The approach of CoastBase is to leave data and information where they are, archived in their information system and catalogued according their metadata model. The innovative aspect of CoastBase is that there is neither a centralized database for the data nor for the metadata: instead, the existing Internet access points of the sources to be connected, are used. The system is customised to meet the requirement of a broad range of users working in the field of the marine and coastal environment such as scientists, coastal planners, policy makers, providing them with a tool for their work. CoastBase functionalities are described here from a user's point of view, through three different examples of possible applications.

CoastBase general description

CoastBase is accessible through Internet at the address: http://www.coastbase.org. The system's architecture (Fig.1) can be described as composed by the CoastBase client, server and data extraction (Kazakos *et al.*, 2000a, 2000b). Interoperability between each part of the system is ensured by a common infrastructure, the Middleware, which is based on HTTP for external communication such as the user interface, CORBA for the internal and XML mainly for querying, transformation, and data exchange. From a functional point of view, main modules include data search through the virtual catalogue, data access and manipulation, and feedback.
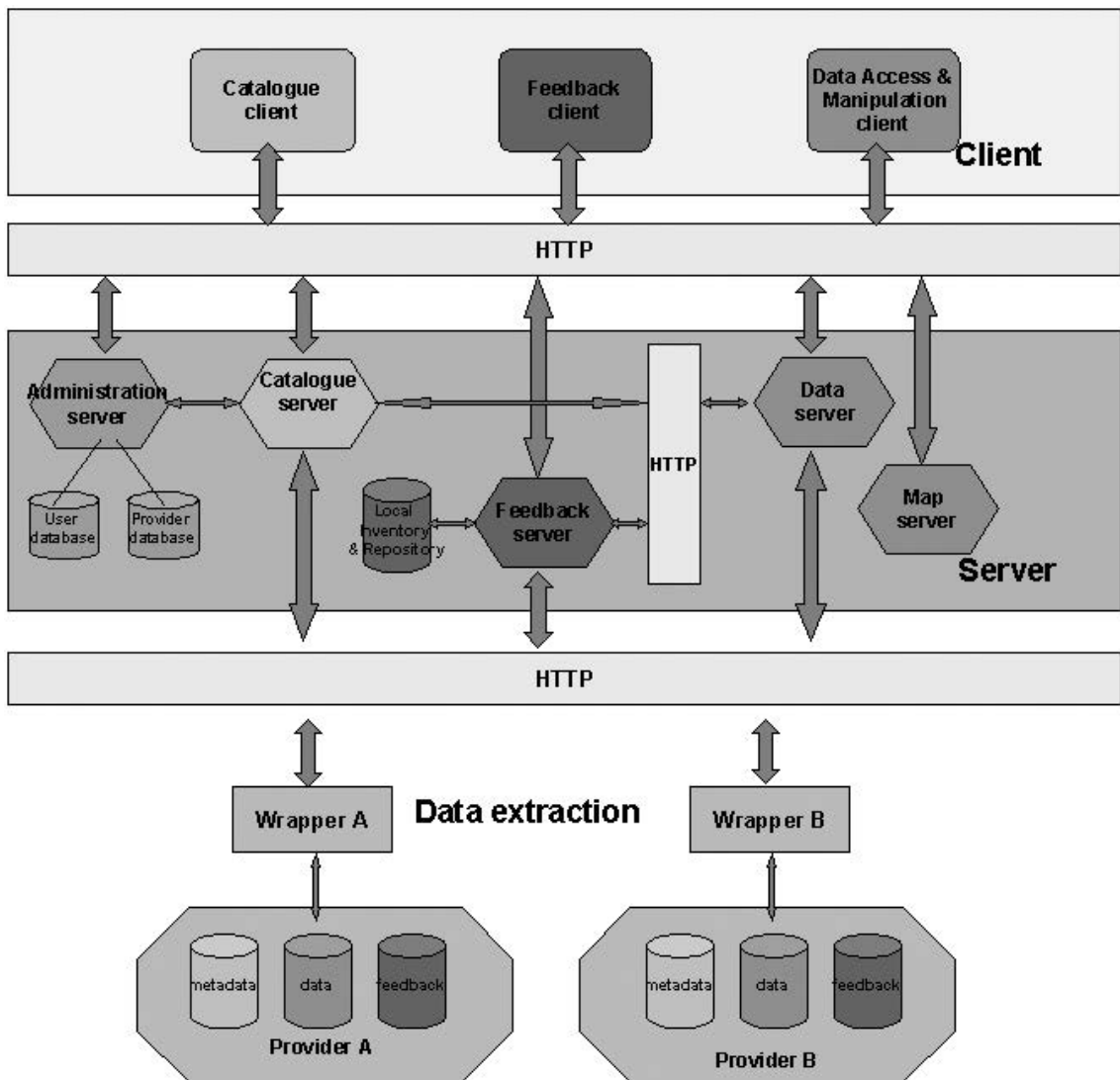

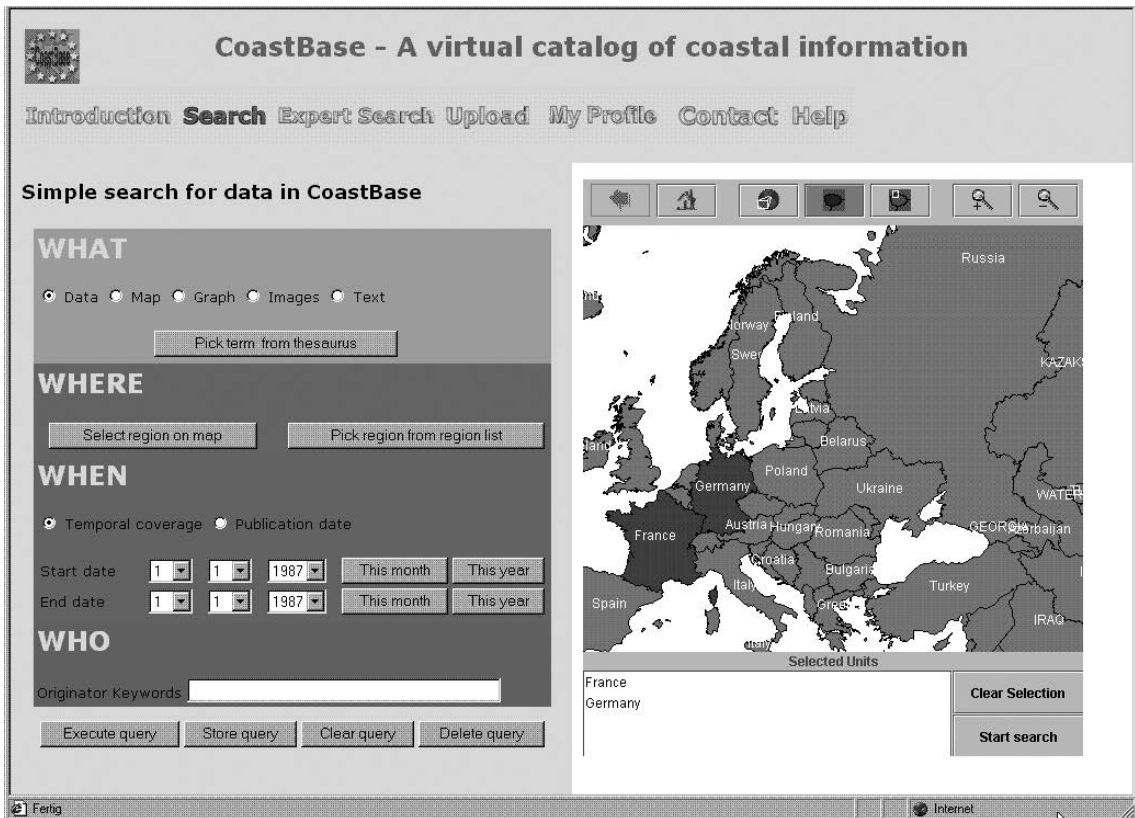Fig. 1 - CoastBase system architecture.

Fig. 2 - CoastBase users graphical interface.

CoastBase client provides uniform, multilingual and interactive access to all CoastBase services. It uses a combination of HTML for the basic functions and (optionally) Java applets e.g. for more convenient map selection. Three languages are available at the moment: English, German and French. A graphical user-friendly interface (Fig. 2) helps the user to formulate the query and to use the available functions to visualise and manipulate the selected items.

User search is made specifying *WHAT, WHERE, WHEN, WHO* (optional). After choosing the data aggregation level of the desired information (data set, map, graph, image, text), the user is helped in his/her search by the on-line catalogue GEMET.2 (GEneral Multilingual Environmental Thesaurus). GEMET was developed to create a multilingual directory of environmental information for the European Environmental Agency (EEA, 1999; ETC/CDS, 2001). It was conceived as a general thesaurus merging the terms of the best existing multilingual thesauri both on a conceptual and a formal basis. GEMET includes more than 6000 descriptors which are arranged in a classification scheme made of 3 super-groups (natural and anthropic environment, human activities and products; effects on the environment, social aspects; environmental policy measure) each of them containing discipline-oriented groups adopted to approach an environmental management perspective and five accessory groups (general terms, functional

terms, personnel, acts, programmes: Table 1). Each descriptor is organised in a hierarchical structure headed by a Top Term. Further, there is a thematic order with 40 themes that are thematically related but scattered in different groups reflecting the EEA (European Environmental Agency) DPSIR (Driving force, Pressure, State, Impact, Response) dataflow scheme. It also takes into account main topics of the Dobris Assessment (EEA, 1995) and other sources such as ETCs (European Topic Centres) and EIONET (European Environment Information and Observation Network). GEMET can be thus accessed through the group-hierarchical list, the thematic list or through the alphabetic list. The CoastBase user accesses the thesaurus through the menu (Fig. 3) and is guided in the search by a tree-structure with the list of 35 groups at the top and several levels to refine the search. To speed up the process, a search engine for a free-term search in the thesaurus is also available. A short definition appears on the screen for each selected descriptor.

To demonstrate the capability of customised applications in the field of the marine and coastal environment, two specific thesauri -one for the assessment of eutrophication in the sea and another for Integrated Coastal Zone Management- are also implemented. More recently, in the framework of the EUROSION project, a catalogue addressing coastal erosion issues was added.

A geographical selection (*WHERE*) is possible by picking the name of the country, or of the coastal region, from a tree-structured list complying with NUTS (Nomenclature des Unités Territoriales Statistiques) codes up to level 2 (EUROSTAT, 1995), or by making a graphical map-based selection (Fig. 4), which is more convenient in the case of coasts or open sea. The available map represents European Seas and countries from the Arctic to the Mediterranean and the Black Sea; panning and zooming functions can make it easy to focus on the area of interest, while the coordinates of the bounding box selected by clicking on the map are visualised on the left side of the screen.

Table 1 - List of Super-groups and Groups from GEMET Thesaurus.

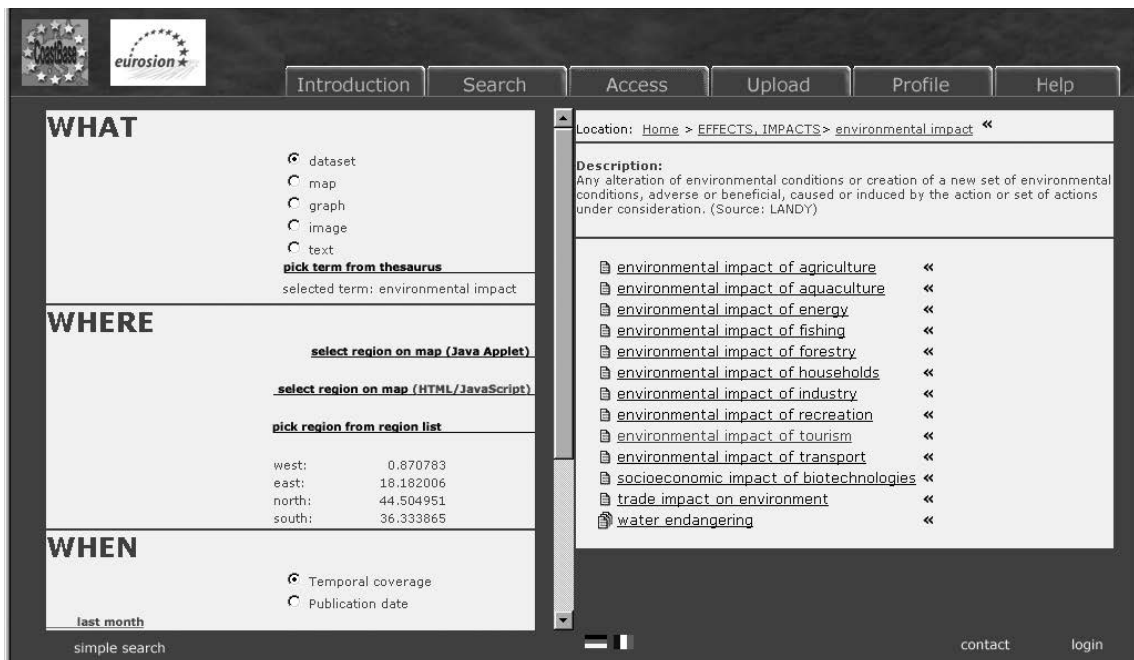| Natural Environment, Anthropic Environment	Human Activities and Products, Effects on the Environment	Social Aspects, Environmental Policy Measures	Accessory Groups
Environment Time Space Atmosphere Hydrosphere Lithosphere Land Biosphere Anthroposphere	Chemistry, Substances, Processes Physical Aspects, Noise,Vibration, Radiations Energy Resources Products, Materials Agriculture, Forestry; Animal husbandry; Fishery Industry, Crafts; Technology; Equipments Trade, Services Traffic, Transportation Wastes, Pollutants, Pollution Effects, Impacts	Economics, Finance Legislation, Norms,Conventions Administration, Management, Policy, Politics, Institutions, Plannings Environmental Policy Information, Education, Culture, Environmental Awareness Research, Sciences Health, Nutrition Risks, Safety Society	General Terms Functional Terms Personnel Acts Programmes

Fig. 3 - Selection of Term from GEMET.2 Thesaurus.

The time period covered by the data sets or the year of publication of reports and documentation (*WHEN*) can be defined from a sliding menu indicating start and end date (day, month, year).

Even though, with the virtual catalogue approach, the user does not need to know which are the available connected sources, the option *WHO* can be used when one wants to limit the search to a particular data provider or to the CoastBase local repository. At the end, the user can confirm the selection and send the query to the server or clear and start again.

The result of the query appears as a list of the available items also including a short description of its contents (metadata) and can be saved for further use.

The server includes the main functional blocks, namely the virtual catalogue, the data access and manipulation, and the feedback. The catalogue server implements the main functionality of the virtual catalogue (Kazakos *et al.,* 2001; Valikov *et al.,* 2001) that is able to access distributed but homogeneous sources with the same domain model. To overcome the heterogeneity due to the different data models of the connected sources, wrappers are employed.

The data access and manipulation services are responsible for transforming and aggregating the data. They also have to address security problems: the administration server is invoked directly by the catalogue server for the authorised data access requests. The data server permits one to access and retrieve the data, to download the product in the desired format and unit by converting the requested data, and provides aggregation and manipulation functions such as the production of maps. The Map Server handles all the requests related to a geo-location. It uses *LuciadMap* (www.luciad.com) as a cartographic library, allows the visualization of geo-
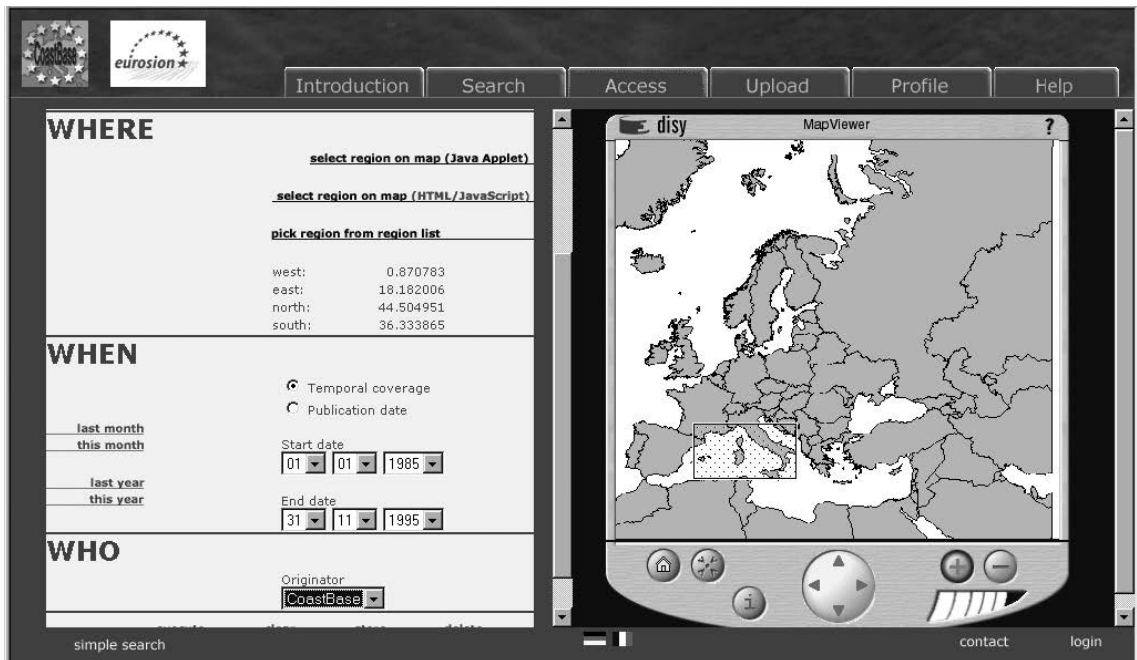
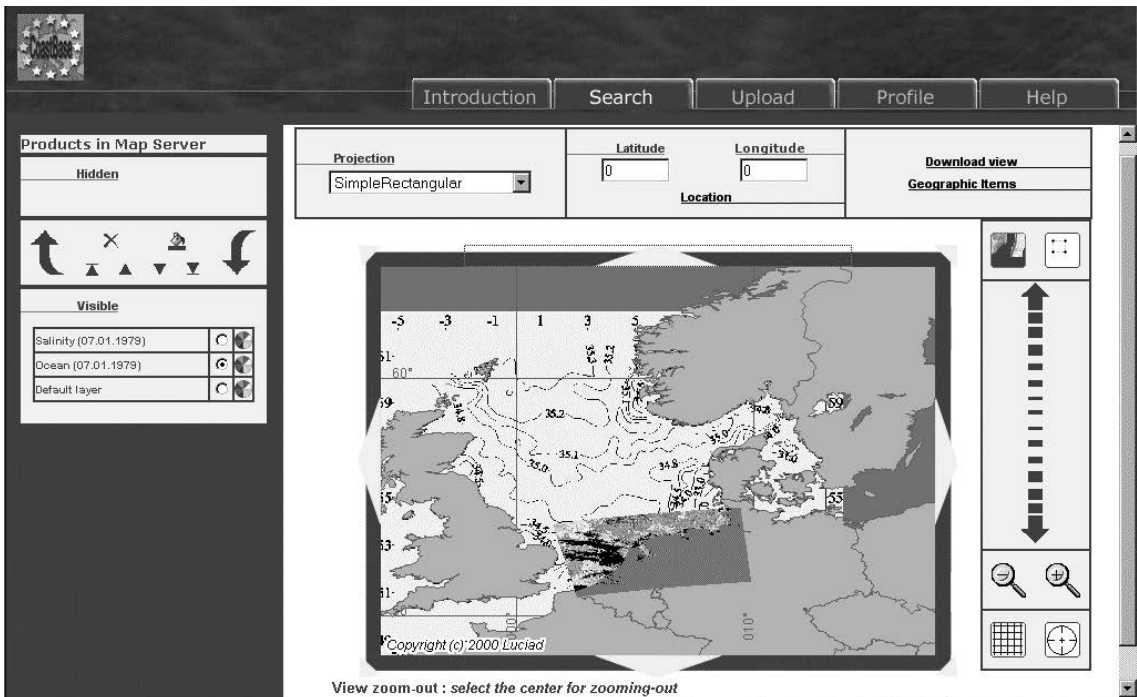Fig. 4 - Map-based *WHERE, WHEN* and *WHO* selection.

Fig. 5 - Map Server application: overlay of Chlorophyll Satellite images with Salinity distribution in the North Sea.

referenced data (raster, vector) on a map, performs the conversion between different coordinate systems, supports standard map projections, implements basic GIS functionalities such as overlay (Fig. 5), panning, zooming and provides the possibility of downloading data in GIS format for external visualisation.

The feedback module permits the data providers to track their information as each data access request is logged automatically at the CoastBase server and is then sent to the provider. Local Inventory and Repository is considered as part of the feedback module as it can be used for storing uploaded products together with their metadata as well as new data or compiled products generated by the users within or outside the CoastBase system. Also in this case, a user interface, structured as a simple metadata form to be filled in, helps the user to better describe the items to upload; in addition, it permits one to send comments that can be shared by CoastBase users.

The data extraction implements the wrappers for data and information sources. The wrappers can be considered an interface between the CoastBase and the data providers as they transform data from the format of the provider to the CoastBase one. They adapt the access protocol, translate the results into XML and transform the structure to fit the CoastBase domain model. Wrappers should be specifically developed for any connected provider, which uses a different data model. There are two types of wrappers, the metadata wrapper, as needed for the catalogue module, and the data access wrapper; both types are needed to connect to a new data source. Existing metadata catalogues of the data providers are wrapped and integrated via mediators into the CoastBase metadata schema. This was designed as an extension of GELOS (Global Environmental Locator Services), a metainformation scheme internationally used for the description and location of information resources and is compliant with EDMED (European Directory of Marine Environmental Data). Data providers, even though they represent a crucial part of CoastBase, cannot be considered exactly a component of the system as they make their data and products available without having standalone components for that.

The available data

The resources available for this project limited the number of data providers to four; these four were chosen from the participating partners. Some of the determining factors for the choice was the sufficient variety of data sources in terms of database structure and content, data aggregation levels and cataloguing systems: according to the data selection criteria (*WHAT*), data sets, maps, images and texts are included. In addition, the data available from the selected data-providers are able to assure a good spatial and temporal coverage of the North Sea, which was preferred as test area to demonstrate the functionalities of the system to combine different data/parameters and to permit temporal-trend analysis. The institutions whose database is connected to CoastBase are:
- ICES (International Council for the Exploration of the Sea, Denmark) the inter-governmental body coordinating researches in the marine environment in the North Atlantic and acting as data handling for OSPARCOM and HELCOM;
- IMR (Institute of Marine Research), the Norwegian National Oceanographic Data Centre and national thematic centre on oceanography, marine biology and ecology;
- RIKZ (National Institute for Coastal and Marine Environment, The Netherlands) that maintains the national information infrastructure for the data related to water management;
- JRC-SAI-ME (Joint Research Centre- Space Application Institute- Marine Environment Unit,

Italy) the EU research institute developing and promoting the use of satellite data for their application to the ocean and coastal areas.

CoastBase also allows one to locally store other relevant data and information; this is used to store documents and texts, as such kind of information is not present in the connected sources.

The database that can be accessed through CoastBase consists of:

- raw data of ocean temperature and salinity profiles for the North Sea for the period 1979-2001;
- about 300 GIF-format compiled maps with isolines obtained from spatial and temporal aggregated data of salinity, temperature, phosphate, nitrite and nitrate, silicate on the North Sea ranging from the 1980's until today;
- DONAR (Data Opslag Natte Rijkswaterstaat) database that contains data used to assess the eutrophication for the Dutch coasts: water level, ocean temperature, salinity, phosphate, nitrate, nitrite, oxygen, and chlorophyll;
- more than 7000 satellite images of ocean colour for Northern Europe, North East Atlantic, the Mediterranean, Black Sea, and North West Africa, from the CSCZ (Coastal Zone Colour Scanner) instrument on board NIMBUS-7 for the period 1979-1985;
- nearly one hundred metadata records of text documents including environmental reports, standard documents of European or regional relevance related to integrated coastal management processes and EU policy, hyperlinks to information holding centres, signposts to institutions that can provide information or manage data sets off-line.

About data quality, it was decided not to perform any kind of check, leaving this responsibility to the data provider: the information on the performed controls is reported in the metadata relating to each database.

The user-driven approach

The CoastBase prototype was developed keeping in mind the needs of possible users operating in the field of coastal and marine management and research. To this end, an investigation aimed at determining users' profiles and their requirement regarding the kind of data and information, aggregation level, as well as their interests about some system functionalities was performed in the first phase of the project (Eleveld *et al.*, 2003). Raw data and indicators are mainly used by scientists working on data analysis and interpretation, while complied products from higher aggregation level data such as maps and tables summarising results from analysis are of interest for those involved in the preparation of a general assessment. The availability of reports and grey literature regarding environmental assessment, policy and management at regional, national or EU level is intended for a wider category of users, including policy makers and the general public.

To demonstrate CoastBase functionalities to a broad range of users and to test the system, three case studies, describing different environmental scenarios and methodological approaches, have been prepared. The three examples of applications were chosen in order to match user requirements with data availability from the connected data sources and trying to enlighten the main functionalities and the potentiality of the system. Different people representing categories of the identified users were also directly involved in the feedback process, as they were asked to test the CoastBase prototype and to give suggestions for further improvements.

The first "case study" is meant to fulfil the needs of a process study for the assessment of the eutrophication in the North Sea. Through this application, it is possible to test the access to distributed information sources with disparate data models and access methods, to demonstrate some functions of the Map Server and to focus on the use of a set of indicators to describe the state of the eutrophication. Once the period and the area of interest is defined, the user can take advantage of a "customised" thesaurus dedicated to eutrophication assessment, which also includes a list of the indicators adopted by the European Environment Agency as eutrophication state indicators, namely: the total concentration of nitrogen in mg/l, orthophosphates in mg/l, and dissolved oxygen in mg/l. In the example, the user selects and downloads a map with the spatial distribution of phosphate in the North Sea in winter 1985 and a satellite image reporting chlorophyll concentration in the Northern Europe an seas in September 1985, which are then saved on the local disk. Using the Map Server functions the scientist is able to overlay different cartographic items within the same geographic reference, to zoom in to enlarge the area of main interest, and to save the current view as displayed in the map server on a local disk.

The second illustrates how the information that can be accessed through CoastBase can support the evaluation of environmental and socio-economical impact due to the presence of an airport on an artificial island. In particular, it focuses on the effects of changes in sediment and nutrient flow due to the artificial island on the coastal erosion and on the impact of reduced food web and import of toxics on economically important fish stocks species and on ecologically valuable areas. Also in this case, the area of interest is in the North Sea, while no temporal interval is specified. The search for information is performed from the list of terms in the GEMET 2.0 thesaurus. As an example, selecting "data set", that includes data at the lower level of aggregation (raw data), and choosing the term "fish", CoastBase allows one to access several databases from different data sources. The result of the search is the list of the available data such as the ICES database of catches in the North Atlantic Area or the Yearbook of fish population compiled by RIKZ, along with the description (metadata) of the content of each selected data set. A new search changing the term "fish" with "sediment" gives as a result the ICES inventory of contaminant in sediment data set. The data sets of interest can be then downloaded and stored for further analysis.

The last test case again concerns coastal zone planning and it is mainly devoted to support a policy maker in the task of planning the future development of a coastal area with high conflicts of use. It aims at demonstrating the role of the local inventory as an additional source of information and providing an example of search on texts and documents, that can be considered the information at the higher level of aggregation. This Mediterranean example deals with integrated planning of Strymonikos Bay in eastern Greece, an area rich in natural resources, landscapes and cultural features. The resident population is 16,860, but rises to over 150,000 in the summer. In addition to tourism, the area's economic activities include mining, fisheries, aquaculture, and forestry. Pollution and environmental degradation have increased in recent years, and tourism is expected to grow: authorities need to plan in an integrated fashion. To this end, the user is helped in the search by the thesaurus devoted to Integrated Coastal Zone Management, which is developed considering the DPSIR approach to coastal zone planning. The example includes the map-base selection, the choice of text as data type and the selection of several terms from the ICZM thesaurus. It starts with: "driving forces and pressures" / "uses of

the coast"/ "recreation & tourism"/ "tourism" to obtain a list reports on sustainable tourism planning included in the CoastBase repository such as European Commission publications, MAP (Mediterranean Action Plan) guidelines. Then, under "state of environment & environmental impacts"/ "pollution" it is possible to have access to reports on the state of the environment, while selecting "Response" / Policy and Management" / "environmental policy concerning"/ "bathing water", CoastBase conducts the metadata search and produces a result list also including hyperlinks to the EU Bathing Water Directive and to the Ministry of Environment of Greece. The user may view the metadata about each document by clicking on the titles under "description" and downloading the reports. The search process to obtain reports, texts, regulations and general information relating to ICZM and environmental policy permits to enlighten the structure of the thesaurus and the advantage of the additional information stored in the server local repository.

Discussion

The CoastBase prototype has demonstrated that data dispersed in different data sources can be efficiently reached and exchanged among external users without changing the information system of the data provider. With respect to existing systems developed in the framework of past Information Technology projects that aim at facilitating environmental data exchange such as REMSSBOT (Regional Environmental Management Support System Based on Telematics), CoastBase represents a step ahead since neither the databases nor the metadata catalogues are centralised. The main advantage of such an approach is that, after the initial effort of developing the wrapper to connect the data provider system to the CoastBase server, additional work is not required to make its data available nor is it necessary to update the catalogue, since any changes to the information sources are immediately reflected in the query results. The use of XML has tuned on to be well suited for information exchanges between heterogeneous systems. In fact, it is useful to describe standard document formats for information interchange between all open systems and especially for web-based ones. Nevertheless, relevant heterogeneity in the metadata domain-model of the connected data sources may increase the complexity of the development of the wrappers, raising the problem of the need for some level of harmonisation. In this regard, a positive result of the CoastBase project is that it has permitted us to demonstrate and promote the use of GEMET as an environmental metadata catalogue and to foster metadata standardisation in marine environmental networks. These issues, among others have been recently addressed by ICES/IOC-IODE (Intergovernmental Oceanographic Commission- International Oceanographic Data Exchange) who has set up a working group with the aim of developing a framework and methodology for marine data exchange (ICES/IOC, 2002) throughout the network of the National Oceanographic Data Centres, adopting a common standard using the XML technology.

The CoastBase prototype was devised to meet the requirements of a broad range of users involved in coastal and marine management and research: scientists, consultants, administrators and policy makers, as well as the general public. It facilitates the search and access to data and information related to several aspects of the marine and coastal environment – from environmental state indicators to socio-economic and legislation – at different aggregation levels such as raw data, satellite maps, compiled products, reports which can be integrated and

harmonised, with the final aim of providing an easy and efficient tool for an integrated approach to coastal zone management.

The prototype is customised for managing marine and coastal environmental data, but as it is modular and open to connections to other data sources and to new elaboration functions, it is flexible enough to be used for any kind of thematic network and other applications. Organisations with highly distributed information sources but centralised decision-making mechanisms that need to integrate information from their own organisation or network, data providers interested in improving access to their information products for either the public or co-operation partners can exploit systems like CoastBase.

Even though the system prototype has successfully fulfilled the objectives and requirements of the project and is keeping up well beyond the lifetime of the project, it is still far from being a fully operating system. The bottleneck is the limited information now available, that cannot meet the needs of any of the identified users: much effort is then required to enlarge the CoastBase network, linking new data-sources. A step in this direction is represented by the application of CoastBase in the framework of the EUROSION project for the dissemination of EUROSION data sets towards end-users; this will permit us to update the system and to further promote its use.

Appendix

COASTBASE - The European Virtual Coastal and Marine Data Warehouse. An open-system architecture for integrated distributed coastal and marine information search and access.

The project was supported by the European Commission within the V Framework Programme- Research and Technology Development DG Information Society IST-1999-11406. It was launched on January 2000 by a consortium of eleven organisations and ended on December 2001. Coordinator was:

RIKZ - National Institute for Coastal and Marine Management-The Hague, (NL);
Partners:
ENEA - National Agency for New Technologies, Energy and the Environment - Roma (I)
EUCC - The European Union for Coastal Conservation, Leiden (NL)
FZI - Forschungszentrum Informatik an der Universität - Karlsruhe (D)
ICES - International Council for the Exploration of the Sea - Copenhagen (DK)
IMR - Institute of Marine Research - Bergen - (N)
JRC - Joint Research Centre of EU - Space Application Institute - Ispra (I)
MATRA - Matra Systèmes & Information - Toulouse (F)
MELS - Ministry of Environment of Lower Saxony - Hannover (D)
MIG - Marine Institute - Gdansk (PL)
PETA - Information Training Local Development - Athens (GR)

Additional information on the project can be find at the project web site http://www.coastbase.org or requested directly from the project coordinator: Hugo Niesing coastbase@rikz.rws.minvenw.nl

The prototype is still operating thanks to FZI, which is maintaining the server, and to the data providers (ICES, IMR, JRS-SAI, RIKZ), which make their database available. The prototype is available on line and can be tested at the project web site.

REFERENCES

DONAR; *Data Opslag Natte Rijkswaterstaat*, http://www.donarweb.nl

EDMED; *European Directory of Marine Environmental Data*, www.sea-search.net/edmed.

EEA; 1995: *Europe's Environment: The Dobris Assessment*. Stanners D., Bourdeau P. (eds), Office for Official Publication of EC, Luxembourg, 712 pp.

EEA; 1999: *GEMET, GEneral Multilingual Environmental Thesaurus, V.2.0.*, EEA, Copenhagen, pp. I-XXV.

EIONET; http://eionet.eu.int/

Eleveld M.A., Siegert A.G. and Schrimpf W.B.H.; 2003: *User requirements and information definitions for a virtual coastal and marine data warehouse.* Ocean and Coastal Management, **46**, 487-505.

ETC/CDS; 2001: *Preface to the 2001 version of GEMET.* 10 pp., http://eionet.eu.int/GEMET .

EUROSION; *EUROSION: a European Initiative for Sustainable Coastal Erosion Management*, http://www.eurosion.org.

EUROSTAT; 1995: *REGIO Banque de données régionales. Description du contenu.* Office for Official Publication of EC, Luxembourg, 97 pp.

GELOS; http://registry.gsdi.org/

ICES/IOC; 2002: *Report of the ICES/IOC Study Group on the Development of Marine Data Exchange System Using XML*. ICES CM 2002/C:12, Ref. ACE, ICES Copenhagen, 63 pp.

Kazakos W., Kramer R., Schmidt A.; 2000a: *CoastBase- The Virtual Coastal and Marine Data Warehouse-*. In: Cremers A. and Greve K. (eds), Computer Science for Environmental Protection 2000, Vol. 2, pp. 646-654.

Kazakos W., Schmidt A., Denis G. and Ma Paw Youn K.; 2000b: *System specification document*. CoastBase Project Report, 40 pp.

Kazakos W., Schmidt A., and Paoli H.; 2001: *XML based Virtual Catalogue Module in CoasBbase*. In: Hilty L. and Gilgen P. (eds), Sustainability in the Information Society. 15th International Symposium Informatics for Environmental Protection, Marburg, Oktober 2001, volume II, 513 - 520. Metropolis-Verlag.

MAP; Mediterranean Action Plan, www.unepmap.gr

MEDAR Group; 2002: MEDATLAS/2002 database. *Mediterranean and Black Sea database of temperature salinity and bio-chemical parameters.* Climatological Atlas. IFREMER Edition (4 Cd roms).

MODB; *Mediterranean Oceanographic Database* http://modb.oce.ulg.ac.be

Niesing H.; 2001: *From scattered data towards European level assessment regarding Marine and Coastal Environments* Proceeding of EurOCEAN 2000 EUR 19408, 215-219.

REMSSBOT *Regional Environmental Management Support System Based on Telematics* http://www1.hellasnet.gr/remssbot/

Sea-Search; http://www.sea-search.net/

Valikov A., Kazakos W., and Schmidt A.; 2001: *Building Updateable XML Views On Top Of Relational Databases*. In: Lasker G.E. and Dahanayake A. (eds), Proceedings of the International Symposium on Systems Integration. The International Institute for Advanced Studies in Systems Research and Cybernetics, Aug 2001, pp. VII - 1, VII - 8.

Corresponding author: Paola Picco
ENEA Marine Environment Research Centre S. Teresa
CP 224, 19100 La Spezia
phone +39 0187 978278; fax +39 0187 978213; paola.picco@santateresa.enea.it