Marine databases: a step towards integrated requests

Gilbert MAUDIRE [(1)], Marcel LE DUFF[(1)], Eric MOUSSAT[(1)],
Christian BONNET[(2)], Hélène BEUCHER[(2)]

[(1)]IFREMER, Centre de Brest, BP 70, 29280 Plouzané, France
☎33 (0)298224216, Fax 33 (0)298224644
Gilbert.Maudire@ifremer.fr
Marcel.Le.Duff@ifremer.fr
Eric.Moussat@ifremer.fr

[(2)]IFREMER, Centre de Nantes, rue de l'île d'Yeu, BP 21105, 44311 Nantes Cedex 03, France
☎33 (0)240374102, Fax 33 (0)240374133
Christian.Bonnet@ifremer.fr
Hélène.Beucher@ifremer.fr

Keywords
Marine data, distributed databases, web services, XML fluxes, normalisation.

Abstract
IFREMER, according to its missions of providing facilities to French oceanographers, manages continuously data bases to archive and disseminate data and information collected during sea operations from research and opportunity vessels, submarine vehicles, buoys, lagrangian drifters and coastal monitoring networks.
Dedicated data management systems have been developed for 25 years to handle these various data types in order to aggregate long term series, control the quality of the archived data sets and distribute them both in real time and delayed mode according to user access rights.
But, integrated projects, impact studies and models need to access comprehensive and integrated data sets including several parameters. In order to meet this requirement, a web portal have been set up which allows to access several databases from a single request.

Résumé
IFREMER, dans sa mission de service aux océanographes français, entretient plusieurs bases de données pour archiver et diffuser les données et informations acquises lors des opérations à la mer effectuées par les navires de recherches, des navires d'opportunité, des véhicules sous-marins, des bouées et flotteurs lagrangiens ainsi que les réseaux d'observations côtières.
Des systèmes de gestion de données dédiés ont été développés depuis 25 ans pour ces types de données variés afin d'agréger des séries à long terme, de contrôler la qualité des jeux de données archivés et de les diffuser en temps réel comme en temps différé dans le respect des droits d'usage.
Cependant, les projets intégrés, les études d'impacts et les modèles ont besoin de jeux de données complets et intégrés comportant plusieurs paramètres. Dans le but d'atteindre cet objectif, un portail web a été développé qui autorise l'accès à plusieurs bases en une seule requête.

1. Introduction

The oceanographic data archiving at IFREMER is a distributed system with a high number of data types and parameters to manage. All the main databases intend to centralise in a homogeneous standardised system the data of the same types to allow the compilation of long series of data of the same types.
However, the data types are heterogeneous in terms of acquisition (automatic process, scientific work), volume and characteristics, and, as a consequence, managed in different thematic data bases: geophysics, physical oceanography, coastal monitoring networks, fisheries monitoring, …
For practical as well as budget reasons, dedicated data management systems have appeared to be simpler to set up. But, with such an organisation, a user who want to access several parameters for impacts studies or coupled biological and physical models, needs to know all the data bases and to query them one by one. In order to simplify data access, a unique portal is necessary, to get data in coherent standardised format without querying all data repositories. The purpose of the new development made at IFREMER is to open a single direct on line access to the data.

2. Existing Data Bases

IFREMER manages several "Marine Data Bases", by which we indicate major series of basic data types archived for the long term. More specifically, it operates SISMER, the Designated National

Oceanographic Data Centre for France (French NODC) for the International Oceanographic Data Exchange programme (IODE) of UNESCO Intergovernmental Oceanographic Commission, and the data collected for the coastal and fishery monitoring. The data management includes compilation of data and meta-data, perennial safeguarding and dissemination. It includes also information on methodology and formats (metadata), and the implementation of quality checks to insure comparability of data from various sources.
The data archived have been collected during national or international scientific projects, especially in the frame of pan-European projects, which are in general multi-disciplinary. In this way, several hundreds of parameters are currently managed at SISMER. Specific short/middle term databases are designed for these projects such as the DORSALES/RIDGE geophysical database (Deplus et al, 2000), the Mediterranean databases MATER (Maillard et al., 2002) and MEDAR/MEDATLAS (Fichaut et al., 2002). However, for the long term, all the data of the same types are archived in the data banks of the respective same types. This requires that the different data banks are interrelated and can be accessed by a common portal.
Several key main databases are managed by IFREMER:

1. The French National Geophysical Data Bank,
2. The French National Physical and Bio-chemical Data Bank and the Coriolis Data Base,
3. The Coastal Monitoring Data Bank,
4. The Fishery Monitoring Data Bank.

2.1. French National Geophysical Data Bank

This data bank consists in data of the geophysical parameters indicated in Table I, since 1977. In addition to these data, the archiving of seismic data has been undertaken and will start operationally at the end of 2002. They are collected underway on board of research vessels and transferred systematically to the data centre on return of the ship. For the multibeam bathymetry, it represents roughly a distance of $1x10^6$ miles and a surface of $14x10^6$ square kilometres covered (Fig. 1) worldwide (Fig. 2). It represents also the largest volume of data increasing rapidly with the management of the seismic data. The data dissemination is limited due to different scientific, defence and economical rules. The large volume of the files limits also the direct access of data on ftp files to users equipped with good network facilities.
The SISMER data centre archives systematically all the pre-processed raw data with some quality control. The data management system is a combination of a RDBS system for the meta data (cruise information and files descriptions) and big scientific files at the IFREMER CARAIBE NetCdf format.
For some projects, it also archives higher processing levels of data like digitised terrain models (DTM).

Table I : Marine Geophysical data in the French national data bank (data collected underway)

CODE	DATABASE	NB OF CRUISES
NAVI11	PROCESSED NAVIGATION DATA	614
BATM11	BATHYMETRY: MULTIBEAM ECHO-SOUNDINGS	322
BATS01	BATHYMETRY: SINGLE-BEAM ECHOSOUNDING	428
GRAV01	MARINE GRAVITY DATA	276
MAGN01	MARINE MAGNETIC DATA	273
MGD77	DATA AT THE INTERNATIONALMGD77 FORMAT	217
IMAG01	IMAGERY DATA OF MULTIBEAM ECHOSOUNDERS	122
In development	SEISMIC DATA	

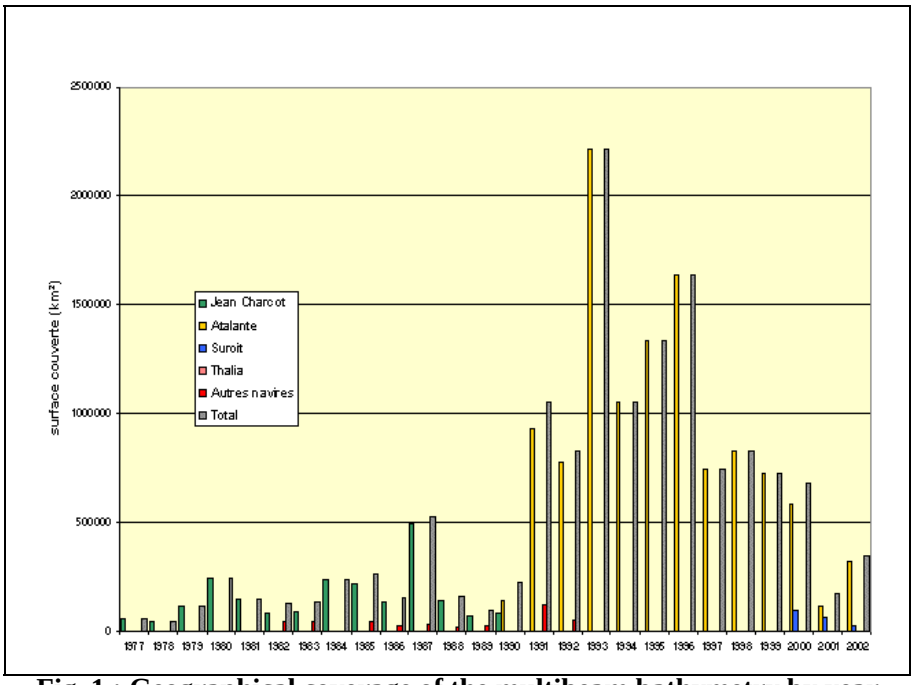

Fig. 1 : Geographical coverage of the multibeam bathymetry by year

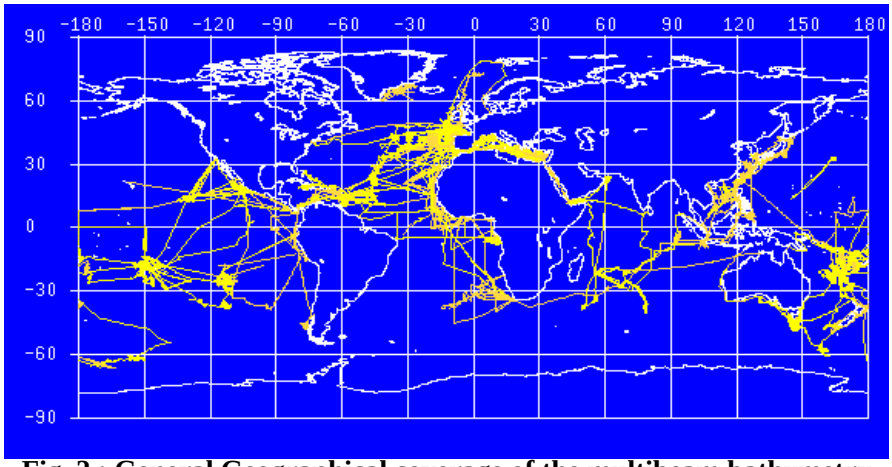

Fig. 2 : General Geographical coverage of the multibeam bathymetry

2.2. The French National Physical and Bio-chemical Data and Coriolis data base

The physical oceanographic data archived at SISMER concern two different types of measurements: hydrography and currentmetry. The data are collected either underway as the above mentioned temperature and salinity data (Table II), or in vertical profiles (table III and IV) by sending a CTD probe and/or sampling bottles on a cable (fig. 3), or in time series (table V)°. The time series themselves can be at fixed mooring or from drifting equipment. The data are collected during French or co-operation scientific sea cruises. They are validated by the scientific teams responsible for collecting them and afterwards formatted, crosschecked and archived at the data centre (in delayed mode archiving and dissemination). The different data types are the following:

- CTD (fig. 3),
- Hydrological cruises (bottle casts) (fig. 3),
- Bathythermograph (XBT, low resolution temperature vertical profiles),
- Current meters time series (in fixed moorings or drifting),
- Thermistor chains time series ,
- Sea level time series ,
- Sediment traps time series (in fixed mooring or drifting).

The physical and bio-chemical data bank contains the longest time series of data as it includes data from 1909 (Fig. 4), and the coverage is also worldwide, but in majority in the East Atlantic, Mediterranean and Tropical world Ocean. Each data type corresponds to several scalar parameters measured like temperature, salinity, current components, pressure, oxygen content, all the chemicals concentrations, biomass etc. The vertical profiles are normally referred to pressure in dbar, which is close to the depth in meter. The time series are referred to time. The data sets are archived from a few month after the end of the in situ data collection to several years, after data requests to the source laboratories. Accordingly the number of observed parameters is high, currently several hundreds, mainly associated with the bottle vertical casts and the fixed and drifting sediment traps time series. The data management system is similar to the geophysical database, with meta-data in RDBS which is common with the geophysical data bank, and the data in scientific files, organised as provided by the source scientist, but in a unique format (auto descriptive ASCII MEDATLAS).

Fig. 3 : CTD probe equipped with a rosette of water bottle to sample bio-chemical parameters.

During the last past year, a new type of CTD measurements has been made by using drifting profilers in the frame of global and regional operational oceanography programmes like Coriolis/Argo (Global scale) and Mediterranean Forecasting System. The data are automatically transferred in real time to the data centre for real-time or quasi-real time (no more than 48 hours delay) formatting, checking and dissemination. For these data, IFREMER/SISMER is a world data centre and manages the international data set. At present, this data management concerned only temperature and to some extend salinity. The low-resolution XBT profiles are for practical reasons managed in the same system, which is presented in more details in another (Carval and Petit de la Villeon, 2002). These data are entirely managed in another RDBS system: Coriolis.

Table III : Physical and bio-chemical data in the French national data bank collected in station (Delayed mode vertical Profiles)

CODE	DATABASE	NB OF CRUISE/FILES	NB OF PROFILES
HYDR01	DISSOLVED CHEMICALS IN THE WATER COLUMN	989	37314
CTDF02	CTD DATA - VERTICAL PROFILES	371	19616
CTDF03	COASTAL OR LOW RESOLUTION CTD	71	4658

Table IV : Temperature and salinity vertical profiles of the Operational Oceanography - Real Time and Delayed Mode Data (Coriolis Data Base)

DATABASE	NB OF PROFILES
PROVOR PROFILES	2510
OTHER PROFILES RECEIVED FROM THE GTS	267159
REAL TIME PROFILES SENT BY RESEARCH VESSELS	4682
TOGA/WOCE/CLIVAR Global Subsurface Data Centre	703227
MFSPP XBT	4263

Table V : Time series in the French national data bank

CODE	DATABASE	NB OF MOORINGS /LAUNCHINGS	NB OF TIME SERIES
STBD01	LAGRANGIAN FLOATS TRAJECTORY DATA	6	62
STCH01	CHEMICALS TIME SERIES	3	13
STCM01	CURRENT METER TIME SERIES	386	1916
STCT01	CTD TIME SERIES	7	14
STMA01	TIDE/SEA LEVEL TIME SERIES	10	24
STME01	METEOROLOGICAL TIME SERIES	22	133
STPD01	DRIFTING SEDIMENT TRAPS DATA	2	24
STPG01	SEDIMENT TRAPS TIME SERIES AT FIXED MOORING	19	39
STTH01	THERMISTOR CHAIN TIME SERIES	37	117

Table VI : Physical data in the French national data bank collected underway

CODE	DATABASE	NB OF CRUISES
THTQ01	SEA SURFACE TEMPERATURE DATA FROM THERMOMETER	149
THSA01	SEA SURFACE SALINITY DATA FROM THERMOSALINOGRAPH	174
In development	ADCP CURRENT DATA	

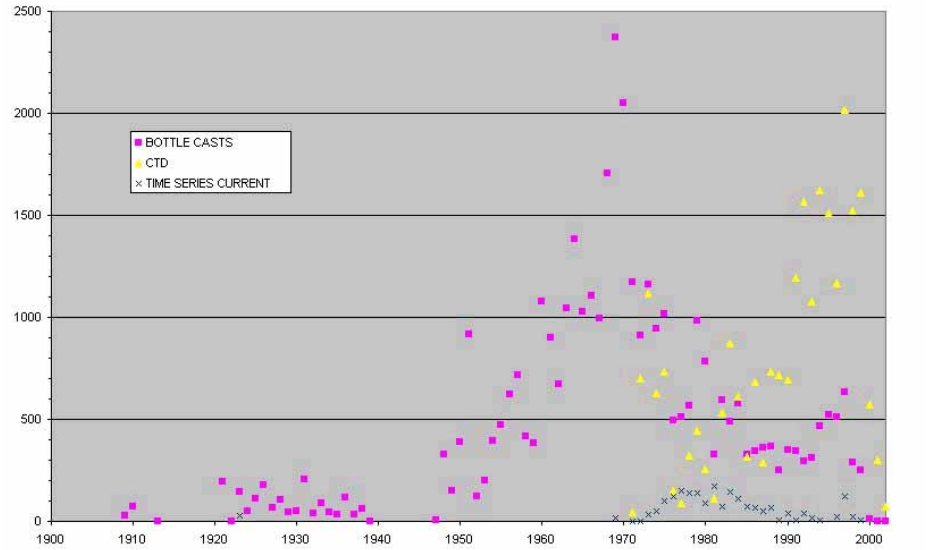

Fig. 4 : Annual distribution of the main data types in the French Data Bank

2.3. Coastal Monitoring Databases

Since 1974, data have been collected regularly (typically each 15 days) over 42 sites around the metropolitan territory (Fig.5) for four networks (for a general description see IFREMER, 1997 report):

1. National Observation Network of Sea Water quality (RNO), including the following up of the general seawater masses characteristics (temperature, la salinity, nutrients, chlorophyll-a, phaeopigments and occasionally supplementary parameters like suspended matter and turbidity) and contaminants of the biota and sediment (metals and toxic organic constituents).
2. Phytoplankton surveillance network (REPHY), following up of all the microscopic algae, especially the toxic ones, and the phycotoxins in the shellfishes.
3. Micro-biology surveillance network (REMI) created in 1989 to control the sanitary conditions over the areas of shellfish production according to the regulations. It includes a system of surveillance and a system of alarm.
4. Impact of the large energy equipment on the seawater (IGA).

Not all of these networks have the nominal space and time sampling. All the data are managed in the same RDBS system Quadrige and the number of data points is given in Table VI.

Table VI : Number of Data Points in the coastal monitoring database Quadrige (mars 2002)

Network	**Number of data points (observations & counting)**
RNO (contaminants)	838 720
REPHY (phytoplankton & phytoplanctonic toxins)	444 598
REMI (microbiology)	108 701
IGA (all kind of observations in the vicinity of the power plants)	708 843
Other regional networks	84 571
Total	**2 185 433**

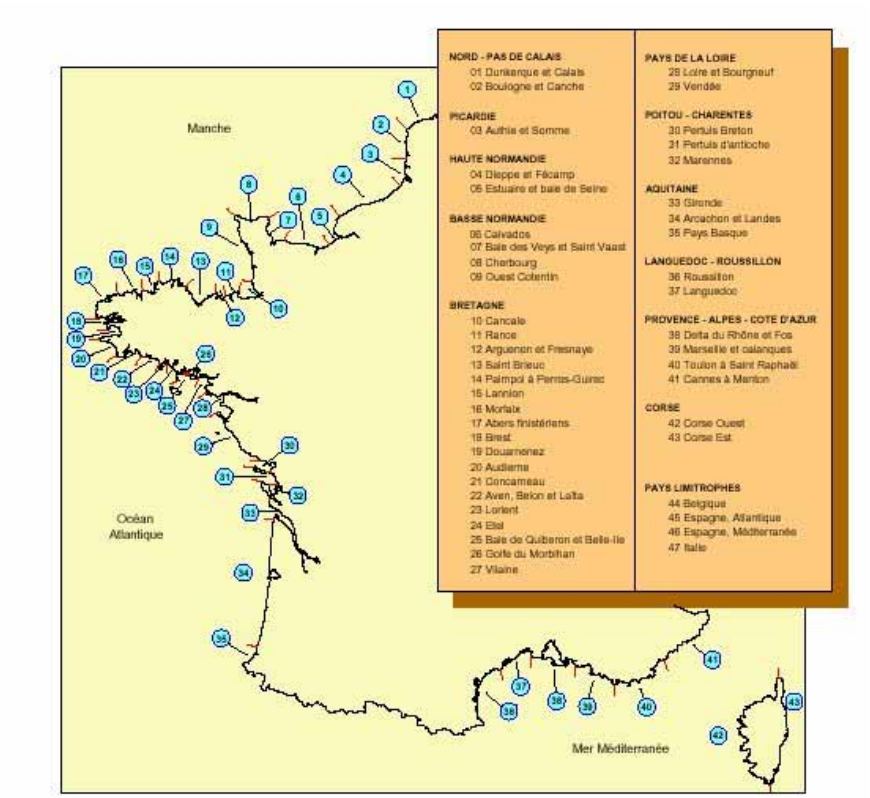

Fig. 5 : Positions of the surveillance sites along the French shoreline

An automated coastal environment monitoring network is under development to collect observations from coastal automatic stations on instrumented buoys or pylons and disseminate the data in real time on internet. Two sites, Mer d'Iroise (Fig.6) and Baie de Seine are already operated, with the possibility of about 20 basic parameters (surface meteorology, physics and bio-chemistry). The data are archived in a dedicated database system.

Fig. 6: Coastal Monitoring Station MAREL in the Iroise Sea

At middle term, the Quadrige database of water samples and the data of the automated stations should be merged in a common system. However, like for the previous data bank, the management of heterogeneous data, and of several versions of the data sets, with delayed mode data replacing the real time data is not simple. Theses data are not managed by the data centre, but by the scientific division in charge of the data collection. However, SISMER should insure a link to it and facilitate the access to them in an integrated way, which means unique data request and unique output format.

2.4. Fishery monitoring data

IFREMER has been involved for more than twenty years in data collection from fishery monitoring. This includes:

- fishing effort,
- fish production,
- economical studies,
- environmental and biological studies on fishing areas. This includes data sets produced by cruises series done aboard french research vessels driven by ICES working groups (IBTS, …).

These data sets are aggregated in a framework including several French governmental bodies and managed by a dedicated data management system. However, not all of these data are publicly available and several data sets are still under confidentiality.

3. SISMER Web Portal

3.1. Requirement description

Since data are handled in thematic databases as close as possible of their acquisition and primary usage, the same type of information can be archived in several databases. By example, a very basic parameter such as “sea water temperature” is measured and used by all oceanographers to qualify life environment, compute circulation, calibrate echo sounders, … and, as a consequence, recorded in all thematic databases. However, even if the same type of information can occurs in several databases, measurement data are unique and not duplicated.

In this scheme, a user of a parameter (e.g. sea water temperature) must be aware of all the IFREMER distributed archive in order to retrieve all the interesting information by querying several web sites and data repository. To avoid this complexity, the need of a unique **“SISMER Web Portal”** was clearly expressed by the French oceanographic major actors: modellers, fisheries and environmental monitoring experts, universities ...

The web portal must be able to accept a request from the user and to forward it to all thematic databases archiving data with the desired type. The main needed selection criteria are: data type, time and space coverage, data sources (scientist, laboratory), measurement equipment, data accuracy, state and quality.

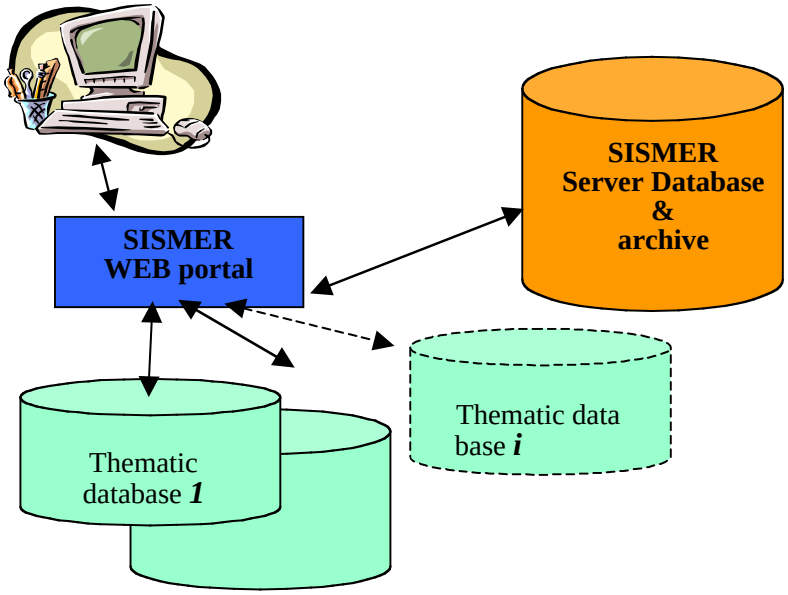

Fig. 7: Access to several databases in a single request

3.2. Technical implementation

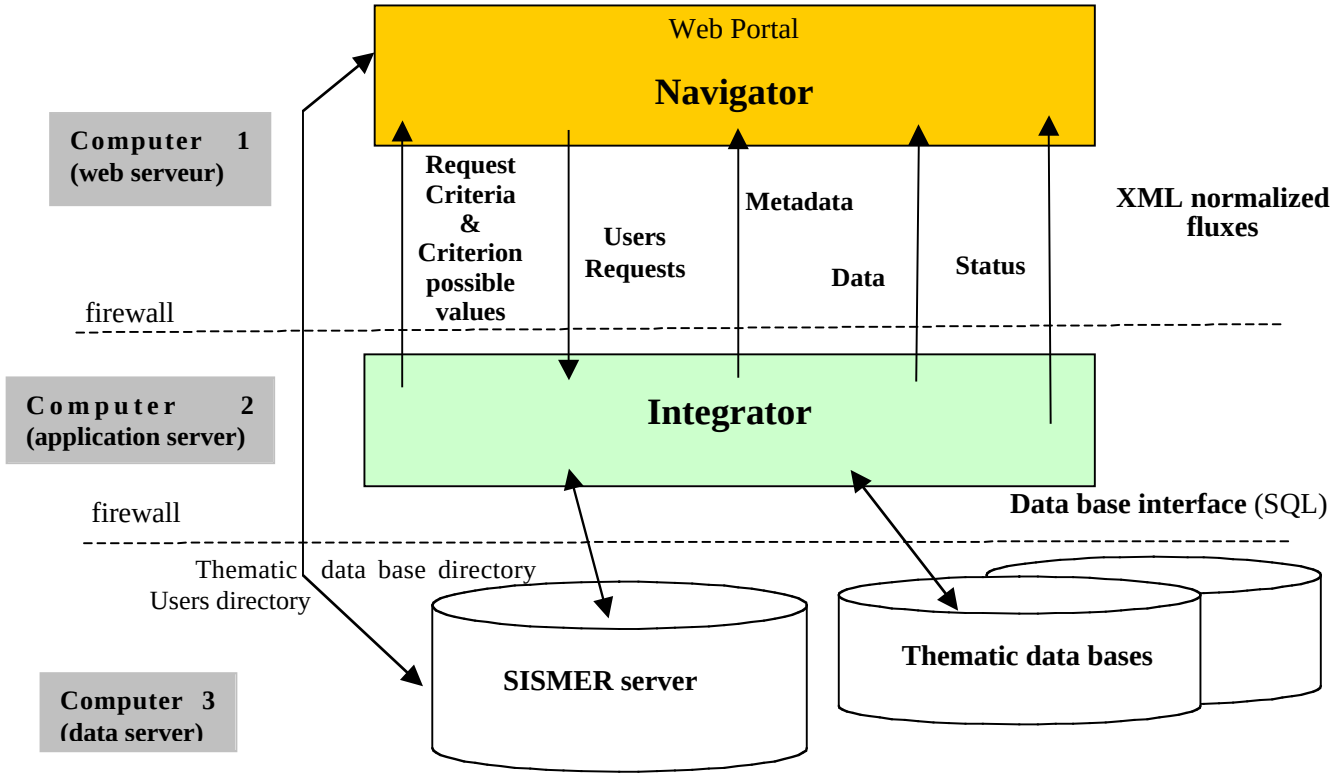

Fig. 8 : The two layers of the web portal

On the technical point of view, the "Sismer web portal" is divided in two layers (fig 8):

- the "Navigator" layer. The Navigator has to:
  - identify the user and his access rights. A user directory is managed by SISMER but a new user is granted to subscribe online (with restricted access rights),
  - display the user interface. This interface is dynamically generated since selections criteria are proposed by the thematic databases themselves. The only mandatory criteria are the data type and the spatial coverage of the request,
  - receive the user query,

- identify all the thematic databases archiving the requested data type (and/or data from a specified equipment). The entry point in the data base directory managed by SISMER is the data type.
- Forward the user's request to all identified thematic databases.
- wait for the responses from the thematic data bases
- display meta data transmitted by the data bases, including drawing the location map of the requested data sets,
- propose data to be downloaded.

- the "integrator" layer.
  Integrators implement dialogue with the navigator in a uniform way. A dedicated integrator is written to interface each thematic database.

The dialogue between the Navigator and each Integrator is based on XML fluxes. Several levels of dialogue are identified for selection criteria, data, meta data and status. This XML fluxes have been defined by IFREMER according to its needs. However, even if no inclusive international normalisation have been set up yet for this kind of dialogue, some basic bricks or ideas have been found in:

- DODS (Distributed Oceanographic Data Systems), *http://www.unidata.ucar.edu/packages/dods/*
- FGDC (Federal Geographic Data Committee) *http://www.fgdc.gov/clearinghouse/clearinghouse.html*
- ISO 19115 (not adopted yet),
- OpenGIS consortium *http://www.opengis.org/*

On a practical point of view, the following tools and languages have been used:

- Apache Web Server as a standard web server and as object request broker (ORB) since Apache implements the XML/SOAP protocol.
- Java programming language. The standard object library (JDBC : Java Data Base Connectivity) is also used to write integrators in association with SQL queries.
- ORACLE ™ relational database management system (RDBMS). Since several Ifremer data bases are based upon this commercial RDBMS, the ORACLE XSU package have been use to write the ORACLE → XML transform within several integrators,
- Some XML fluxes are "zipped" in order to improve transmission efficiency between the Navigator and the integrators. In particular, geographical information may be heavy to transfer in real time through the network (even if the network is an efficient 100Mb local network).
-

Fig. 8 : Web portal user interface

4. Conclusion

In next November, the "SISMER web portal" first version will be opened for IFREMER users. At the end of the year, the web portal will be available on the internet (http://www.ifremer.fr/sismer). Before opening widely the web site, user identification will be improved using a Lightweight Directory Access Protocol (LDAP) user directory.

The design of the SISMER web portal could be a first step to implement large scale distributed data centres since it uses well known web tools and protocols like XML, SOAP, … In such a perspective, work is needed to define and standardise XML fluxes which can accurately describe marine data sets and requests. Better advances have been done in that way for other types of data like spatial data

(FGDC, ISO), geographical information system (OpenGIS), … Marine data community must take in account these advances to define its own XML protocol.
However, thinking about using these techniques for data exchange and marine data centre interoperability is on the way in the framework of:

- ICES (International Council for the Exploration of the Sea): Study Group on the Development of Marine Data Exchange Systems using XML (chair: B.Gelfeld, web: http://www.ices.dk/iceswork/workinggroups.asp),
- IOC (Intergovernmental Oceanographic Commission – UNESCO) and WMO (World Meteorological Organisation) joint commission (N.Smith: 2002).

5. References

CARVAL T., L. PETIT DE LA VILLÉON, 2002. *Coriolis: a data management service for operational oceanography.* Symposium Ensuring Long-Term Preservation and Adding Value to scientific and technical data. Toulouse 5-7 Novembre 2002, France

DEPLUS C., E. MOUSSAT, M.O. BOULANGER, A. BRIAIS, M. CANNAT, J.DYMENT, P. GENTE, R.GIORDMAINA, Y.LAGABRIELLE, C. ROMMEVAUX-JESTIN, E. RUELLAN et D. SAUTER, 2000 - *Banque de données "Dorsales" de bathymétrie et de Géophysique* - Journées Dorsales 2000 20-22 Septembre Roscoff.

FICHAUT M., M-J. GARCIA, G. MAUDIRE, M. RIXEN, E. VYAZILOV and MEDAR Group, 2002. *Added values to oceanographic multi-parameters data sets, Example of European concerted action : MEDAR/MEDATLAS 2* Symposium Ensuring Long-Term Preservation and Adding Value to scientific and technical data. Toulouse 5-7 November 2002, France

IFREMER/Direction de l'Environnement et de l'Aménagement du Littoral, 1997. *La surveillance de la qualité des eaux côtières à l'IFREMER*. Rap. Int. DEL/97-10. Issy-Les-Moulineaux, Novembre 1997.

MAILLARD C., E. BALOPOULOS, A. GIORGETTI, M. FICHAUT, A. IONA, M. LAROUR, A. LATROUITE, B. MANCA, G. MAUDIRE, P. NICOLAS AND J. -A. SANCHEZ-CABEZA, 2002, *An integrated system for managing multidisciplinary oceanographic data collected in the Mediterranean Sea during the basin-scale research project EU/MAST-MATER (1996¯2000),* Journal of Marine Systems Volumes 33-34, 1 June 2002, Pages 523-538

SMITH. et al., 2002, *An Ocean Information Technology Project*, N.Smith@bom.gov.au