SeaDataNet

A Pan-European Infrastructure for
Ocean and Marine Data Management

# IMDIS2008 Overview and Outputs

**Yannick Beaudouin, Stathis Balopoulos, Sissy Iona, Robert Keeley, Catherine Maillard, Giuseppe Manzella, Gilbert Maudire, Steve Miller, Peter Piessersens, Dick Schaap,**

## 1. Introduction

The International Conference on Marine Data Management and Information Systems IMDIS2008 was held in Zappeion Conference Centre of Athens (Greece) from March 31 to April 2, 2008, jointly organized by the SeaDataNet Consortium, the Intergovernmental Oceanographic Commission of Unesco (IOC/IODE) and the Hellenic Centre for Marine Research.

More than 200 participants met to overview the information systems on marine environmental data, and show the progresses on development of efficient infrastructures for managing large and diverse data sets.

The Conference has presented different systems for on-line access to data, meta-data and products, communication standards, adapted technologies, interoperable platforms.

Information systems have an increasing role in the society and a strong impact on science, technology and business. They are moving resources in software, hardware and telecommunication networks for managing information, allow the access to products through services. The fundamental principle defined during the Conference is that data should be a public good. The implication is that it is necessary to provide services to end users.

The Conference allowed to map out and analyze the major information systems so as to identify improvement and requirements for improved systems. Mission of the existing data network is to facilitate a long term and sustainable access to interoperable, high quality data necessary to understand the geological, biological, chemical and physical behavior of the seas and oceans.

Due to the diverse technologies and methodologies, interoperability has been defined as one of the priorities, in order to ease the user access to different kind of products in oceanography, fishery, biology, geology. Another priority discussed during the conference is the development of

common vocabularies/conventions. Also education has been considered an important issue to be considered in all programs.

IMDIS 2008 has represented the 'agora' for the different communities working on information systems. It has created collaborative links and understandings of common problems.

# 2. Session 1: MARINE ENVIRONMENTAL DATA BASES - INFRASTRUCTURES, METADATA AND DATA SYSTEMS

The session was introduced by two keynotes presenting some major on-going initiatives for the preservation and 'Access of marine data across Decades and Disciplines'.

After recalling the current interest in global processes and long-term changes, S. Miller underlined that:

- data in the marine sciences are being used by a far broader community than ever before (Table 1).
- Metadata are critical for recording the provenance of marine data, including processing steps and quality control certification (who, what, where, when, which, why, how).
- auto-harvesting the information, selecting standards, exchanging metadata, and synchronizing metadata with evolving standards and needs are difficult.
- a basic international goal is the definition of a standard set of metadata to define an oceanographic cruise, and a mechanism for exchanging the information.

Several US Universities and research institutions have recently convened a Data Management Best Practices Committee and digital libraries that support data sets from cruises, submersible dives, ROVs and IODP ocean drilling throughout their lifecycles, from initial ideas through proposal review and on to operations, as well as for educational use. Much greater use is made of controlled metadata vocabularies and XML-based procedures that deal with the diversity of raw metadata values and map the information to agreed-upon standard values, in collaboration with the MMI community (S. Miller).

The SeaDataNet project, presented by the coordinator G. Maudire, is a EC project to develop a sustainable Pan-European infrastructure for ocean and marine data management. It aims to secure and manage the large and diverse datasets collected by about 1100 laboratories from 35 countries by networking the National Oceanographic Data Centres and several satellite Data Centres. It cannot be emphasized enough that meeting the needs of 1100 laboratories is a daunting task. We face a wide range of institutional legacies, procedures, data types, and languages. Every component of the infrastructure must be sensitive to a wide range of user and contributor expertise. Success will not be possible without a coherent approach, with controlled vocabularies for data types and metadata names. Two interface strategies may emerge.

1) Mappings may be needed to migrate institutional content to SeaDataNet standards, especially legacy holdings.
2) A very popular and easy to use SeaDataNet set of interfaces will need to be available for laboratories to adopt.

These two major recent initiatives, as well as the other systems presented in the session, pointed out basic needs for:

- multi-disciplinarity
- completeness of data and meta-data
- bringing-in of the new/adapted information and communication technology
- sustainability of the data management infrastructures.

## *2.1 Multidisciplinarity*

Almost every following presentation emphasized also the requirements to meet the need of multi-disciplinary users for ecosystem approach.

ICES has developed the EcoSystemData to function as an integrated repository, where different types of data can coexist and interact for multiple purposes, e.g. marine spatial planning, ecosystem management decision support tools, climate change studies and fisheries management. For biological databases (C. Arvanitidis) and fisheries information systems (C. Papaconstantinou, J. Barde), multidisciplinarity and the use of common taxonomy are critical.

The development of the World register for marine species WorMS (F. Hernandez) represents a powerful advance in this regard.

Multidisciplinarity and common vocabularies are also the foundation of other databases created for other specific economical use, such as the following up of coastal substrate extractions (G. Sutton), with similar multidisciplinary requirements. To reduce the barriers for participants from other fields, jargon and acronyms that are commonly used in one discipline must be accompanied by easy access to explanatory definitions.

Multidisciplinarity should be extended beyond the marine domain. Marine processes are influenced by terrestrial and atmospheric processes, so our databases must be aware of vocabularies and data sets in other fields. We need to be prepared to insure interoperability with other communities as requested by GEO. Such efforts have been undertaken for GENESI (L.Fusco) and the International Polar Year (IPY) (A. Kuznetzov).

## *2.2 Metadata and Data Volume and Completeness*

The new release of the World Ocean Atlas 2006 (O. Mishonov) shows that the amount of available data is increasing fast, and in a near future it is expected to produce a ¼ degree resolution in the gridded fields instead of 1 degree. However many data are still missing, for the classical parameters (temperature, salinity, nutrients, bottom depth) but mainly for all the other types of observations.

Contribution of data for public access needs to be encouraged. Peer pressure may be effective, when others see contributed content presented attractively, heavily used, and cited. Credit for contribution needs to be made available for career advancement.

We need to be aware also of non-traditional sources of data, e.g. industry (Schaap). The navies and private firms (telecommunication, oil, environmental companies etc.) should be encouraged to contribute with declassified data. Joint projects and meetings could provide better perspectives and would accelerate the data circulation.

The geophysical data management from specific sensors like seismic (M. Schaming) and satellite (G. Guevel) have specific requirements and systems due to the size of the files and the conditions applied to the access to data. Their data products, their validation, their access and availability statistics, and their usage should be facilitated, in particular by fostering new collaborations and consolidating existing research groups (P. Diviacco).

We need to be aware of innovations (and limitations) of sensor capabilities (F. De Strobel, G. Manzella). The comparability of the data collected with different sensors should be assessed via appropriate documentation of the experimental conditions and quality control applied.

A basic tool for data inventory, management and retrieval is the Cruise Summary Report (CSR) database (F. NAST). With the growing number of observatories and autonomous systems (V. Solovyev) with real time and delayed mode data management systems, there will be need for maintaining a similar international inventory.

## *2.3 Technology*

The transformative innovations in information technology (Ioannidis) and grid resources (Fusco) should be investigated to facilitate the data management taks. A key point is the possibility of using a dynamic archiving system in order to be independent of the archiving media like the magnetic tapes. Other key tools of the information technology (RDBS and GIS, WWW etc.) are now currently used and should be taken into account to develop integrated *in situ* and remote sensing databases and data sets and facilitate their access to users.

On the other hand, it is critical to make users aware of tools that are available. Examples include seismic conversion (M. Schaming), collaboraties (P. Diviacco), and many others.

Commercial and open source software is available to equip new data centers. However their set up and maintenance is not simple and expert engineers are needed in the data centers (A. Stefanov).

## *2.4 Sustainability*

Underlying all of our activities is the goal of sustainability. National repositories represent a key resource to insure the security, long time stewardship, subject to the national mandates, policies and support beyond the time duration of usual projects. However they need to receive funding proportionally to the data they have to manage. An argument can be made that any proposal for

developing a new sensor, building a new ship or a new observatory, carrying out a field observation program should include a data management plan with a budget of a few percentage of the project cost, as successfully implemented during the former EU MAST framework program. The most effective strategy is still "demand." If the databases are used heavily, by critical audiences, then it will be much easier to appeal for future funding.

**Table 1. Components of an information system must meet the needs of a range of audiences.**

| | Scientific Researcher | Data Manager | Policy Decision Maker | Industry | Education | Media | Public | Funding Decision Maker |
|---|---|---|---|---|---|---|---|---|
| Introductory Information | X | | X | X | X | X | X | X |
| Case studies | X | X | X | | X | X | X | X |
| Human Access User Interface | X | X | X | X | X | | X | X |
| Web Services | | X | | | | | | |
| Data upload interface | X | X | | X | | | | |
| Common Vocabulary | | X | | | | | | |
| User tools | X | X | | X | | | | |
| Standards | | X | | | | | | |
| Best practices | | X | | | | | | |
| Information architecture specification | | X | | | | | | |
| Help Desk | X | X | | X | X | | | |
| Related projects | X | X | | X | X | | X | |

# 3. Session 2: Standards and Interoperability

The session began with an invited presentation on the results of the IODE / JCOMM Standards Forum held in late January in Oostende (R. Keeley). The most important outcome was the agreement of a formal standards recommendation process for both IODE and JCOMM with a plan for wide community involvement.

The second presentation, also invited, stressed the value of encoding semantic information with data and the great importance of a formal governance process for maintaining the semantic frameworks developed (R. Lowry, W. Geoff).

The next presentation described the delayed mode quality control standard used within Argo, and reported that all Argo data had been reprocessed from 2002-2006 and are now available from the Global Data Assembly Centres of this project (C. Coatoan, E. Brion).

The meeting was presented with a description of the S-100 standard that has developed through the IHO. This is an improvement on S-57 in that it has greater flexibility and extensibility and allows for inclusion of registries of information from sources beyond IHO. In fact, the sea ice community is building such a registry within the S-100 framework (R. Ward).

The following presentation noted that most data sources are not semantically enabled and this makes it difficult for machine to machine processing (M. Erdmann). IMDIS participants were shown a test procedure that examines data sources and builds semantic information to describe them. This was followed by an explanation of a system being under development where xml structured data files are the core data structure used to feed all applications needed.

The next presentation discussed the tiers of data management from raw data, through processing to products. The meeting then heard of the links made between THREDDS and underlying netCDF files and the CDI of SeaDataNet (S. Nativi).

The last presentation explained the updates underway and completed by ICES in managing the ship and country code lists (N. Holdswarth, H. Parner)

## *3.1 Standardization*

All of the presentations revolved around the theme of standardization of data and information. It is necessary to take these ideas and to get the discussion and buy in from the wider community. This is happening to a degree from groups such as MMI and the CF group. The Standards Process noted in the first presentation is another avenue. Journal editors should be approached to get their help in insisting that standards be used in the underlying data of journal articles. Funding agencies could also get involved by insisting that funding be tied to data that complies to international standards.

### *3.2 How to deal with duplication*

Many data are exchanged among partners and these data are stored away for future use by many. This creates myriad duplications and near duplications that must be dealt with. It was suggested that one strategy is that evolving through SeaDataNet with each national centre being responsible for serving their own data. Another method would be to implement versioning on data collections.

## 4. Session 3: User oriented services and products

Three main types of data users reflected in the session:

1 - Mandate driven user: working within the context of international treaties; focused approach, ”inside a box”

2 - Innovation driven (i.e. research, science): broader approach, ”bigger box”

3 - Curious/novice user: (i.e. education, general public): broadest approach, no ”box”

Each user type represents a different set of needs, but they:

1 - all depend on ease of access to as large a dataset as possible in their area of interest (topical and/or geographical)

2 - increasingly rely on multidisciplinarity.

3 - require, in many instances, access to real- or near real-time, quality data

Increasing complexity of databases and integrated data networks can sometimes marginalise the user. The establishment of goals and product targets seems to be the main driver of consolidation and database integration (partnerships, consortium etc…).

### *4.1 Technology*

Technology: we can conclude that most new system developments are moving towards and are adopting the ISO standards for metadata and related services as well as the OGC standards for GIS mapping services and also introducing SOA based architectures with Web services. Especially if we compare with the previous IMDIS conference, there is a clear land slide going on.

## *4.2 Interoperability*

Interoperability: There is a clear drive towards achieving interoperability and realising cross-sectorial access to data and information. Most projects and system developers underpin, that they need multidisciplinary data from various sources and disciplines, such as fisheries, hydrography, geology, ocean environment, oceanography, etc. to undertake their activities. Likewise they understand and are motivated to make their data accessible to other user communities.

## *4.3 Vocabularies*

The importance of common vocabularies is recognised as a key element for interoperability within sectors and using ontologies between sectors. These vocabularies must be alive with proper governance.

## *4.4 GIS*

Various new GIS Open Source software programmes are being used by several projects, such as Minnesota Map Server, Geoserver, SeaGIS, ... Most projects include a mapping functionality in their systems, thereby adding very useful presentation and retrieval functionalities.

## *4.4 EU Directives*

The EU Directives, such as INSPIRE and the new Marine Strategy, are considered as clear driving forces behind the urge to interoperability and data harmonization. Moreover, these Directives pose technical challenges, which are very inviting to the IT communities. We are in an exciting technological time period.

Technology is not the hampering factor for achieving interoperability. Most important is the organisation or human factor. The present climate is positive towards interoperability and most institutes and governments are keen to adopt the principles. However maybe we must be careful not to go too fast, because standards need their time to mature and also we should prevent that too many projects are trying to develop the same wheel, while some reviewing and analysis and closer cooperation could be very useful and much more effective.

# 5. Session 4: Databases and Tools for Education

This Session included papers that covered education of the ocean data management community and other communities in both developed and developing regions.

The Session was informed about the OceanTeacher project which has built an “encyclopaedia of knowledge” that covers all aspects of ocean data and information management, as well as training courses that are used during IODE training courses (W. Rommens,P. Piessersens). The system is now being reviewed and will be migrated to a new technology environment that will enable a broader range of content providers. In this regard participants in the Session were invited to contribute their knowledge to OceanTeacher. It was further suggested that OceanTeacher should cover not only oceanographic data and information management but should also welcome other related domains like marine geology and geophysics.

In the presentation on the African Marine Atlas the Session was informed that this product covers four domains (geosphere and atmosphere, hydrosphere, biosphere, human environment) with the aim of delivering a multi-disciplinary product that is of use to government decision makers. The project had clearly shown the complexity of sourcing data from such a broad range of domains (L. Scott). The success of the project had already led to the development of the Caribbean Marine Atlas. The Session had noted that the substantial expertise related to GIS-related services and products demonstrated at IMDIS 2008 offered many opportunities to further improve the African and Caribbean marine atlases.

The presentation on the Coastal and Marine Wiki (C. Simon) demonstrated a comprehensive information system aimed at communicating knowledge that addresses not less than 6 user audiences (policy makers, practitioners, scientists, students, public stakeholders and the wider public). As it uses the “wiki” technology it is easily assimilated by the current generation of computer users who are already familiar with Wikipedia. The system currently includes nearly 900 articles contributed by 257 editors. Additional editors are attracted during conferences, workshops and other events that bring together relevant experts. Content contributions are restricted to registered experts (registered in IMDIS).

The talk on Tools for education and capacity building in SeaDataNet presented the tools developed in the project to transform data file sin XML files and transform general ASCII files in MedAtlas or ODV formats. These tools are useful for the production of different catalogues that can be used for discovery projects and data. They are used to construct also the SeaDataNet CDI capable to link directly to data.

The presentation demonstrate that some tools must be available to trainers: internet access, services (e.g. access to data, images, media, data bases, …). There is the need to transfer experencies among the regions and promote education in data management in the academia (M. Fichaut).

# Conclusion

(G. Maudire, SeaDataNet Coordinator)

This Conference enlightens how fast the marine data management community is dynamic and the real progress that have been achieved since the former IMDIS Conference in 2005: new databases set up, new user interfaces and services to discover and access data, improved data management procedures have been applied and new ways for international collaborations have been explored. The overall impression of the participants has been very positive, as declared by many colleagues during and after the Conference.

Real advances have been made in the past three years, in many domains, namely:

1) the development of standards that ate used in distributed data management system, for search, discovery, visualization and dissemination of delayed mode and near-real time data. International standards, like ISO or OGC, are commonly used now, and also community standards like common vocabularies and data quality control procedures. The use of common vocabularies, which is starting now, represents a prerequisite for integrating of data sets data sets of different sources and allows interoperability among different information systems. The procedures for quality control that assess the data quality and comparability have improved a lot by using in routine, data visualization and geospatial analyses. These controls make us more confident on quality of data set data centres distributes.
2) useful new information technology tools and data processing software have been made available in the commercial domain and in freeware in the academic domain.
3) education in data management, has been enhanced by dedicated training courses both for the data managers and the data users

It is noticeable that the data management systems presented during the Conference are taking benefit of these advances to provide better security and stewardship of the marine data and services to scientists and other users. They have facilitated the building of integrated data systems, from observation to distribution in real time, that meet new user requirements like risk assessment, and continuous monitoring of the ocean for the evaluation of regional and global changes.

Beyond these advances, some concern has been raised for the long term data security and the sustainability of the data management systems. Whereas the national organizations supports the main infrastructures like national oceanographic data centres, the dependence of success in proposals for the maintenance of several systems represents a weak point. The main issue is to have to datasets used, and therefore to provide more data of assessed quality, to make them easier to discover and use and provide better associated services.

Finally, and maybe the most important, as the international and interdisciplinary cooperation appears as the basis of the enhancement of the data availability and standardization, new ways of collaborations have been explored among the different communities (physical oceanography, fishery, biology, geology) at the international level. The IMDIS Conferences belongs to this effort and the next Conference is foreseen in Spring 2010.

## Closure

The conference was closed with warmful thanks to the Hellenic Centre for Marine Research and the Zappeion Conference Centre for their outstanding organization that has facilitated the exchanges and new prospects of cooperation among the participants, and the other supporting agencies, the European Commission and IOC/IODE.