

DRAFT Theme session M – Environmental and fisheries data management, access, and integration

Conveners: Christopher Zimmermann (Germany), Helge Sagen (Norway), and Peter H. Wiebe (USA).

Enhanced integration of diverse fishery, oceanographic, and other marine environmental data, and tools to enable fishery and environmental assessments are needed to respond to the requirements for ecosystem-based management initiatives. The growing number of databases and data sources requires new approaches to be developed for efficient access to data – and especially to data in distributed databases. Management of the data also requires new approaches to meet enhanced requirements and take advantage of new technologies, for example, GIS-based systems and the internet. This Theme Session invited the participation of database specialists, distributed data specialists, visualization specialists, end users, and others.

The theme session initially attracted more than 50 contributions and thus was the most “popular” session during this year’s science conference. There were ultimately **22 posters and 21 oral talks** in the session. Participants in the session ranged from about 80 in the beginning to about 30 at the end; most sub-sessions had around 50 participants. The stage was set by an **invited keynote lecture by Lesley Rickards**, entitled “Marine Data - A big issue”, which was attended by approximately 200 – many of whom are not normally involved in data handling.

Theme session M was structured with 3 major sections. There were two papers section “**(1) Policies and general approaches to the problem of data access and integration**”. Julie Gillin (ICES) provided an introduction to the ICES data centre and the new ICES data policy, and how the data centre intends to serve the marine community in the future. In a second presentation, Fergal Nolan (EC) described current enforcement data collection within the EU (namely VMS-data) and ways in which they might be made available to the research community. The EU commission intends to amend their regulations to make the data available to science with unrestricted access in the near future.

There were a number of papers presented in section “**(2) Data integration - worked examples**” which was structured into three subsections with a total of sixteen papers. Four papers were addressing “**(a) general**” issues, with presentations on the Netherlands’ NODC-I infrastructure, a Canadian approach to fisheries data management, the UK marine data and information partnership, and an example from the ICES data centre’s DOME database. Six papers demonstrated different approaches to the integration of **(b) environmental** data. Several of these papers showed innovative GIS based visualization systems and one showed how this could be used in spatial planning of marine area use. One paper specifically addressed the issue of using near real-time data to create an environmental status report that was intended for use by assessment groups. Another six papers provided examples of **(c) fish and fisheries** data. Two methods for aggregating fisheries data were described and are expected to increase the quality of assessment input data. The issue of quality indicators was discussed in one talk, and the use of a web site with a WIKI designed to reach a consensus (about a definition or the meaning of a flag) was promoted. The last talk in this sub-section provided conceptual thoughts on visualisation of complex marine data sets.

The last section consisted of three papers dealing with “**(3) Technical approaches to data integration**”. Most of them stressed the need for interoperability (by the use of common vocabularies) for the integration of data. The fourth scheduled paper was not presented.

In addition to the presentations, there were **three animated discussion periods**, two for twenty minutes each and the final one for an hour. A number of issues raised in the presentations focused the discussions. **Highlights** of the discussion are:

People agreed that there is **an increasing need for data integration**. It was also emphasized that there was a need to **visualize data** to make them accessible and useful to the community, and that the visualization should be as **intuitive** as possible.

When **fishers** have direct access to the data they submitted, they become more cooperative, and there are examples where **they start to acknowledge benefits** in having submitted the data.

The issue of **open access to data** was discussed extensively and there were some opinions that restricted access needed to be maintained. However, a representative of the European Commission reminded the group that the EC policy was for open, free access to all data that was developed using public funds, and that this policy is implemented in the EU by most member states.

Even if policies are in place to ensure open access to data in national databases, the means to enable open access are often still rudimentary at best and not operational because of a **lack of interoperability**.

There is a perception that the marine community is lacking adequate **tools to describe biological data**. This may extend to more **elementary definitions** like what is a station?, What is a ship? In the case of biological data, the JGOFS and GLOBEC programs have been serving these kinds of data since their inception. There are biological and chemical data types described and protocols and dictionaries describing them. A major initiative within the SeaDataNet Program and MMI is bringing all of the **vocabularies** together so that they can be made interoperable. For the future, the vocabularies should be online and dynamic (not fixed), and existing vocabularies should be used whenever possible.

Data management should have a higher value in the **education of scientists**. Also, when scientists receive funding, the funding agency should provide templates that can be used to guide the **creation of metadata** as well as the **publication of data** that are ultimately produced.

The issue of **open source versus commercial software** was discussed, and it was generally agreed that both have positive and negative aspects. Proprietary standards or formats are often impediments to data sharing. It was concluded that commercial software should employ open standards, and when the product is very costly, “student” and/or “reader” versions should be made available so there is not an additional financial constraint for the distribution of data or evaluation results. Open source software often suffers from discontinued funding for further developments. One solution is to have a community based effort on such developments.

The EU framework program FP7, by virtue of an explicit emphasis on data management and integration, represents an opportunity for **funding** to significantly enhance and strengthen current initiatives to make data more freely accessible and visualisable and the distributed databases and data centre more **interoperable**. It was recognized that once the systems are interoperable, the collective data sets will provide a base from which **new kinds of products** can be developed that will serve science, management, and the public.

A recurrent theme was captured by the phrase “don’t re-invent the wheel” with regard to the development of software and standards. This was most apparent for map server applications during this theme session. The need for a public place to announce and demonstrate software like that for vocabularies would be desirable to **avoid duplication of effort**. **Communication** (as achieved in this theme session) is essential to this process.

Although no papers addressed it, the issue of **data recovery** of non-digitized data was recognized as an important problem. While there are major initiatives on-going, there are still technical impediments that might disappear as a result of technological advances.

At the conclusion of the theme session, the question of holding another theme session was raised. The audience was strongly in favour of holding another session within a two year time frame because substantial advances in the area of management of marine data were anticipated.