DISMAR: Data Integration System for Marine Pollution and Water Quality

T. Hamre, S. Sandven, É. Ó Tuama

Nansen Environmental and Remote Sensing Center, Thormøhlens gate 47, N-5006 Bergen, Norway
(Torill.Hamre@nersc.no, Stein.Sandven@nersc.no)

Coastal & Marine Resources Centre - ERI, University College Cork, Ireland
(e.otuama@ucc.ie)

ABSTRACT

A distributed system for monitoring and forecasting of the marine environment has been developed in line with INSPIRE and GMES recommendations for a Spatial Data Infrastructure (SDI) architecture. This system, DISPRO, enables integration and distribution of multi-source data from satellites, aircraft and in situ instruments, as well as results from numerical models. Geographic data and metadata are stored on a set of distributed computer nodes, and retrieved and integrated by a web GIS client by means of Web Map Server (WMS) technologies. DISPRO has been demonstrated in six European coastal zone and ocean regions in the spring and summer of 2005.

1 INTRODUCTION

The main goal of the DISMAR project was to develop a distributed system for monitoring and forecasting of the marine environment, integrating data from various observing platforms and modelling systems. A distributed GIS offering access to all available observations and forecasts in the area of interest provides valuable support for public administrations and emergency services responsible for prevention, mitigation and recovery of crises such as oil spill pollution and algal blooms. Establishment of such a system on an operational basis can make a significant contribution to improved management of natural or man-made pollution crises in the coastal and ocean regions of Europe.

During the DISMAR project, a prototype decision-support system, DISPRO, was developed, enabling integration and distribution of multi-source data and results from numerical models. The DISPRO architecture is consistent with INSPIRE's general model of an SDI (Spatial Data Infrastructure). DISPRO is a multi-tier system with four main groups of components: user applications, geo-processing and catalogue services, catalogues and content repositories. Implementation is guided by INSPIRE recommendations [1], conforming to Open Geospatial Consortium [2] and W3C [3] standards, and using Open Source software where available. Metadata plays a central role in DISPRO. All data products and services are described in an accompanying metadata file. The product metadata are defined in a profile of the ISO 19115 geographic metadata standard restricted mainly to the core 'discovery' metadata elements [4]. Metadata are provided in XML format, and validated against an XML Schema. This provides platform independent metadata that can easily be transformed to HTML for presentation in common web browsers using XSLT stylesheets.

DISPRO provides a single entry point, via a web portal, to several services delivering satellite data and other observations as well as model results, conforming to international standards for both metadata and data. The system has been demonstrated in six coastal zone and ocean areas in Europe where Web Map Servers are installed: (1) North Sea / Skagerrak area, (2) German coast, (3) coast of Italy, (4) coast of France, (5) coast of UK, and (6) South-West Ireland. Users in each of the demonstration areas have been involved in testing and evaluation of the distributed system, accessing a number of EO (Earth Observation), in situ and model data products via the Internet. Target user groups include, among others, decision-makers in the public sector and industry, environmental agencies, scientists and the general public.

Satellite ocean colour data and SAR images have been used in combination with ferrybox data, oil drift model and ecosystem models to show how algae blooms and oil spills can be monitored and forecast. These data and predictions typically cover entire demonstration areas, facilitating early warning and estimates of spread and evolution of oil spills and potential algae blooms. Aircraft observations using infrared and ultraviolet sensors, as well as coastal radar, are used to observe water quality parameters on a local scale, providing higher resolution data for smaller areas.

1 http://www.nerc.no/Projects/dismar/
2 DISPRO ARCHITECTURE

2.1 The SDI model for DISPRO

The DISPRO architecture follows the INSPIRE design (Fig.1) constituting a multi-tier system (client, middleware, data layer) with four main groups of components: user applications, geo-processing and catalogue services, catalogues and data/content repositories. User applications in the client tier provide the interface to the system. These will usually be web-based and provide a range of functions such as querying the catalogues and viewing results (e.g., metadata and maps), performing spatial analyses, or database administration (Fig.1:1). Direct access to data and associated metadata can also be provided (Fig.1:6).

The middleware tier provides the various computational services. These services include metadata search and retrieval of both data and services, and serving up transformed data, maps and other content through linking to the data repositories (Fig.1:5). The services can be chained (Fig.1:2). The catalogues supply the data (catalogue entries) for the catalogue services (Fig.1:3), and in turn build their content from metadata provided by the content repositories (Fig.1:4). The data tier provides all the data both geo-spatial (vector, raster, tabular) and other (e.g. documents, multimedia files).

Catalogue services are singled out as a special type of geo-processing service: they provide the functionality to allow the organisation, discovery and access to geo-spatial data and services, and as such, form the core of any SDI. A catalogue consists of a collection of indexed, searchable catalogue entries each providing a description of some resource. Entries usually take the form of a subset (a discovery set) of the complete metadata element set of the resource they describe.

Established standards, like the Dublin Core metadata element set [5], provide a set of standard core elements specifically for discovery metadata, i.e., for cross-domain information resource description, and include a ‘coverage’ element for defining spatial location and temporal period. Likewise, the ISO 19115 [6] standard, adopted for DISPRO, provides an extensive and generic definition of metadata for geo-spatial data sets while the forthcoming metadata implementation specification, ISO 19139 [7], will define the definitive UML interpretation and XML schema (XSD - XML Schema Language) of ISO 19115. However, as the ISO 19139 specification is not available yet, the DISMAR project has developed its own metadata schema in XSD. The XML data format provides platform independence and is easily transformed, e.g. to HTML, by means of standard W3C technologies (XSLT). All data sets in DISPRO must provide a discovery metadata companion file according to the defined ISO 19115 profile.

The content repositories provide the data including geo-spatial (raster and vector) and other (documents, images, multi-media files, etc.). Each data set / item must have an associated metadata file which contains amongst other fields, a geo-reference, to enable discovery through the catalogue service. The experience of the end user (as described in the general model) is that a single database is being accessed. However, the databases are actually distributed and database interoperability issues form an essential part of any implementation of the model.
2.2 Implementing DISPRO

DISPRO uses an architecture consistent with the INSPIRE general model of an SDI. The architecture constitutes a multi-tier system (Fig.2) featuring diverse end user applications communicating with various application servers, which are in turn linked to the data repositories. There are several components in each tier of the system.

The portal\(^2\) is the front-end to the prototype providing access to all features of the system through a web browser. The key components of the portal are:

- The **map viewer** provides the interface for interacting with the web GIS client. It is restricted to the basic functionality of the browser (i.e. a thin client relying on HTML and JavaScript).
- The **notes viewer** displays interpretation notes and abstracts (including legend graphics) of the map layers.
- The **updates viewer** displays abstracts of the most recent data sets available for a geographic area of interest (definable by the user).
- The **profile client** allows the user to set preferences for geographic bounding box, map projection, and layer groups, and to bookmark their selection for future, easy access.
- The **metadata browser** provides an interface to the metadata catalogue which also provides links for downloading data (when made available for download).

![DISPRO Architecture Diagram](http://dispro.ucc.ie/apps/dismar/)

Fig. 2. DISPRO architecture.

The WMS (Web Map Server) client provides the basic functionality of the GIS interface, communicating with distributed WMS servers to retrieve map layers, performing queries, and actions such as panning and zooming. In response to requests from WMS clients, the Web Map Servers generate and serve up maps in image (raster) format. There are several map servers in operation across the demonstration areas, all communicating via OpenGIS protocols.

Each data provider contributing to DISPRO must set up an OGC compliant WMS server, and ensure that the capabilities file (a metadata file returned in response to a GetCapabilities request describing the WMS server's capabilities, e.g. URL, map layers, formats, SRS, bounding box co-ordinates, etc.) returned by their server includes the metadata fields essential for integration in DISPRO. Returned capabilities files are stored in the portal catalogue’s database and provide all the information to populate a GIS client based on an end user's preferences (e.g. bounding box, SRS, layer groups). In addition, a node must store full metadata files on their server where they can be linked to from the metadata catalogue.

The portal catalogue is thus responsible for integrating the DISPRO WMS system (Fig. 3). An aggregator continually retrieves the most current capabilities files from the participating nodes and stores them in the catalogue.

\(^2\) [http://dispro.ucc.ie/apps/dismar/](http://dispro.ucc.ie/apps/dismar/)
database. All the web-based applications (e.g. metadata query, profiler, bookmarker, web browser GIS client) available to the end user are based on queries to the catalogue. The browser GIS interface, in turn, interacts with the web GIS and WMS client components on the server to retrieve maps from various WMS nodes using the GetMap request and present them in the browser GIS interface.

Fig. 3. DISPRO implements an OGC compliant WMS system.

3 NORTH SEA/SKAGERRAK DEMONSTRATION

3.1 Introduction

Phytoplankton is an important and necessary part of the annual biological cycle of coastal and marine waters of the global oceans. However, some algae blooms may have harmful or even toxic effects on the natural life and impact human activities in marine waters. The need for monitoring and prediction of so-called Harmful Algae Blooms (HAB) is therefore evident.

Since 1998 the North Sea and Skagerrak have been regularly monitored using various satellite EO technologies, including the Orbview-2 SeaWiFS, NOAA AVHRR and more recently also the European Envisat MERIS (Medium Resolution Imaging Spectrometer) sensors. Information about the marine chlorophyll-a (phytoplankton) distribution have been processed, analysed and published for research purposes in several national and international projects and distributed via Internet3.

To support early detection of potentially harmful algae blooms, a monitoring scheme should include several sources of information - traditional field observations, satellite EO technologies, as well as numerical ocean and phytoplankton modelling tools for predicting the development and spread over time.

In situ sampling is discrete and coarse in sampling, but has a high level of accuracy and also provides species identification information. However, this sampling is typically done at or near shore locations and may be hampered by local ocean circulations effects. The sampling is often conducted at aquaculture sites and samples are preserved for analysis in laboratories. Additional in situ data may be available from ships of opportunity, which also allow for offshore sampling along transects.

3 E.g. http://HAB.nersc.no
The satellite EO data covers large areas, but it is only possible to derive information about the surface chlorophyll-a concentration during "cloud free" conditions. The principles used for remote sensing of chlorophyll from space are based on the fact that the coloured pigments of the phytoplankton in the waters are detectable from a satellite sensor 800 kilometres above the ocean surface. Pure water is "blue" and with increased chlorophyll concentrations the water becomes more "greenish". This fact is used to estimate the chlorophyll-a concentrations from the satellite images. However, in coastal regions both sediments and dissolved organic matters contribute to the ocean colour and the accuracy of the retrieved chlorophyll concentrations becomes degraded.

Numerical ocean circulation models have improved significantly over the last decade and coupled to ecosystem or plankton models such as the NORWECOM model [8] are able to predict the development of a phytoplankton bloom.

3.2 Product examples

Massive coastal and offshore blooms of the species *Chattonella* have been detected early in SeaWiFS satellite images in 1998, 2000 and 2001 in the coastal Danish and Norwegian waters. During all these HAB event years the development peak and decay of the blooms were monitored by integrated use of satellite and in situ data. A qualitative consistency was observed between the in situ observations and the satellite data of the bloom extent and development cycles. Regional algorithms for retrieval of phytoplankton chlorophyll, suspended sediments and dissolved organic matter are applied in order to improve the use of ocean colour EO data products, as well as to evaluate new sensors such as MERIS.

Fig. 4 shows examples of the algae bloom conditions in spring and summer of 2005. This type of product has been provided to end-users in the North Sea/Skagerrak region by the NERSC DISPRO node. In addition, other remote sensing, in situ and model data products have been available from other providers serving data to this demo region. The user, however, will not notice any difference between products delivered by different providers. All products will be accessible through the web GIS client, and be displayed in the same, user chosen, map projection. Furthermore, all data layers will be accompanied by commonly structured metadata, which is searchable through the catalogue browser, independent of the physical location and configuration of the provider’s node.

Adhering to OpenGIS protocols allows multiple providers to deliver products to the same user without specialised software for handling delivery from individual providers. DISPRO ensures seamless data integration and provides a uniform GIS interface to all users. Fig. 5 shows an example of how data from multiple providers can be combined in DISPRO. In this case, a wind forecast provided by the Norwegian Meteorological Institute is overlaid on top of chlorophyll-a data from MERIS provided by NERSC. Having access to ‘simultaneous’ chlorophyll-a observations and predicted ocean conditions (esp. wind, waves, currents) facilitates an assessment of a potential HAB reaching vulnerable areas, such as fish farms and other aquaculture sites. DISPRO also provides access to ferrybox data provided by the Norwegian Institute for Water Research. Ferrybox data constitute another valuable input for HAB monitoring, as the species identification derived from these sensors gives information about the toxicity along the vessel’s route, thereby complementing the wide-area satellite data.

Fig. 4. Chlorophyll-a derived from MERIS, 26 April, 12 May and 7 June 2005.
4 CONCLUSION

A prototype decision-support system, DISPRO, for integration and distribution of multi-source data and results from numerical models has been developed. The first version of DISPRO was demonstrated to end-users in summer-fall 2004. Based on user feedback, DISPRO has since been enhanced and extended to provide, among others, capabilities for customising the user interface to a set of user preferences, inclusion of raster/vector legends, automatic grid lines with labels, information on newly added products, and improved display capabilities. The user preferences can be bookmarked so that when the user revisits DISPRO, the same settings are turned on by default.

The second version of DISPRO has been demonstrated in several European coastal zone and ocean regions in 2005, serving combinations of measured, derived and predicted data to target end-users. Products have included remote sensing data from satellites and aircraft, in situ data from various types of instruments, among others, the ferrybox sensor, parameters derived from remote sensing and situ data, as well as model results such as prediction of oil spill spread and algae bloom development.

Our experiences with the Web Map Server and other Open Source technologies have shown that this is a viable option for building future GMES services. Different data and service providers can ‘hook’ into the DISPRO network, and provide pollution monitoring and prediction products to any user through the DISPRO portal and web GIS client. Providers can set up these nodes using public domain tools, which makes the entry cost for new providers low. Users need no expensive tools to access DISPRO either; only a common web browser is needed.

5 REFERENCES