DISMAR: Data Integration System for Marine Pollution and Water Quality

T. Hamre [a,], S. Sandven [a], É. Ó Tuama [b]

[a] Nansen Environmental and Remote Sensing Center, Thormøhlensgate 47, N-5006 Bergen, Norway – (Torill.Hamre@nersc.no, Stein.Sandven@nersc.no)

[b] 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 is being 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, aircrafts 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 technologies. DISPRO will be demonstrated in six European coastal zone and ocean regions in the spring and summer of 2005.

Keywords: Marine pollution monitoring, oil spills, algae blooms, web portal, web map server, INSPIRE, GMES, OGC.

1. INTRODUCTION

The main goal of the DISMAR[1] project is to develop a distributed system for monitoring and forecasting of the marine environment, integrating data from various observing platforms and modelling systems. This distributed system will be used to improve the management of natural or man-made pollution crises in the coastal and ocean regions of Europe, supporting public administrations and emergency services responsible for prevention, mitigation and recovery of crisis such as oil spill pollution and harmful algal blooms. DISMAR will provide 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.

A prototype decision-support system, DISPRO, is being developed, for 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 (Smits et al., 2002), conforming to Open Geospatial Consortium (OGC, 2005) and W3C (W3C, 2005) 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 (Cox, 2001). 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.

[1] http://www.nersc.no/Projects/dismar/

The DISPRO system will be 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 will get online access to a number of EO (Earth Observation), in situ and model data products.

Satellite ocean colour data and SAR images will be used in combination with ferrybox data, oil drift model and ecosystem models to show how algae blooms and oil spills can be monitored and forecasted. Aircraft observations using infrared and ultraviolet sensors and coastal radar will be used to observe water quality parameters on local scale. Users in each of the demonstration areas will be involved in testing and evaluation of the distributed system. Target user groups include, among others, decision-makers in public sector and industry, environmental agencies, scientists and the general public.

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).

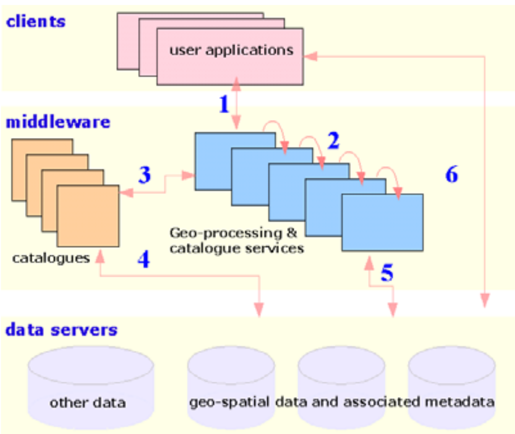

Figure 1. A generic model for a spatial data infrastructure (based on INSPIRE).

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 of the complete metadata element set of the resource they describe.

Established standards, like the Dublin Core metadata element set (DCMI, 2005), developed by the Dublin Core Metadata Initiative (DCMI) provide a set of standard core elements specifically for discovery metadata, i.e., for cross-domain information resource description, and includes a 'coverage' element for defining spatial location and temporal period. Likewise, the ISO 19115 (ISO, 2003) standard provides an extensive and generic definition of metadata for geo-spatial data sets while the forthcoming metadata implementation specification (ISO 19139) 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 for open use yet, the DISMAR project has developed its own XML 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.

2.1 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 catalogue browser*. This provides an interface to the catalogue, which allows the user to browse and search for available geo-spatial data. The catalogue entries are generated from the capabilities files returned by each participating node and include a limited set of fields (e.g. title, abstract, bounding box, spatial reference system), with a link to the full metadata file residing on the parent node.
- *The map viewer client*. This provides the user interface, through which the user interacts with the GIS.
- *The news viewer*. This displays information on new and updated data, pollution events and service notices (e.g. a notification on completion of an oil drift model run). Live news, syndicated from the Europe Media Monitor[3] is also displayed here.

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 a 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). All nodes are polled regularly for updates.

In addition, a node must store full metadata files on their server where they can be linked to from the metadata catalogue.

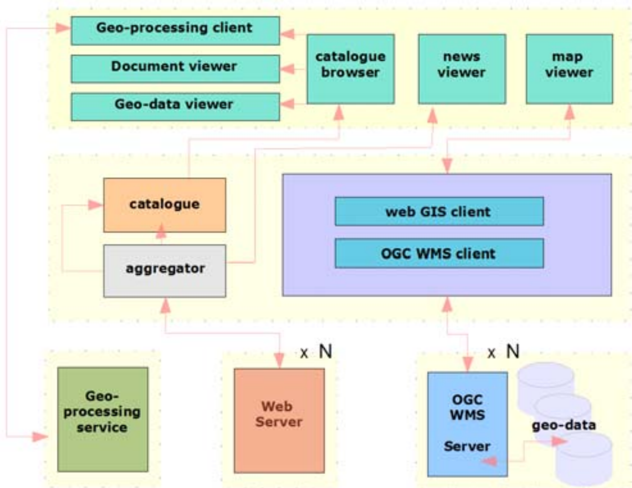

Figure 2. DISPRO architecture.

[2] http://dispro.ucc.ie/apps/dismar/index.xml

[3] http://press.jrc.it

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 the 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 Earth observation 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 Internet[4].

In order to early detect and forecast the development of possible harmful algae bloom a monitoring scheme including several sources of information - traditional field observations, satellite Earth Observation technologies and numerical ocean and phytoplankton modelling tools.

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.

The satellite Earth Observation data covers large areas, but information about the surface chlorophyll-a concentration is only possible to derive during "cloud free" conditions. The principles used for remote sensing of chlorophyll from space is 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 (Skogen and Søiland, 1998) are able to predict the development of a phytoplankton bloom

3.2 Product examples

Massive coastal and offshore blooms of the specie 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 qualitatively 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 Earth Observation data products, as well as to evaluate new sensors such as MERIS.

Figure 3 shows an example of the algae bloom conditions in 2005. This type of product will be made available to end-users in the North Sea/Skagerrak region, and will be delivered by the NERSC DISPRO node. In addition, other remote sensing, in situ and model data products will be 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 will ensure seamless data integration and provide a uniform GIS interface to all users.

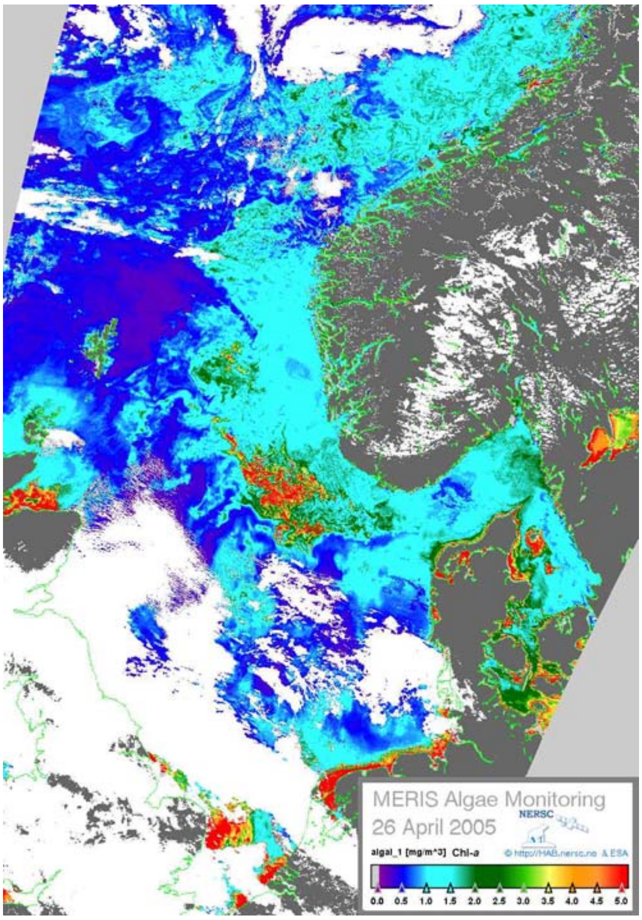

Figure 3. Chlorophyll-a derived from MERIS, 26 April 2005.

---

[4] E.g. http://HAB.nersc.no

4. CONCLUSION

A prototype decision-support system, DISPRO, for integration and distribution of multi-source data and results from numerical models is under development. 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 by pre-selecting area of interest, map projection for display, and which layer types to include. These preferences can be bookmarked so that when the user revisits DISPRO, the same settings are turned on by default.

Demonstrations of DISPRO are ongoing in the six coastal zone and ocean regions selected by the DISMAR project, serving combinations of measured, derived and predicted data to target end-users. This will include remote sensing data from satellites and aircrafts, 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, providing 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. ACKNOWLEDGEMENTS

The DISMAR project has been co-funded by the IST Programme under contract no. IST-2001-37657.

6. REFERENCES

Cox, S., 2001. Summary of some Geospatial Metadata Standards. http://www.ned.dem.csiro.au/research/visualisation/metadata/geospatial/

DCMI, 2005. Dublin Core Metadata Element Set, Version 1.1: Reference Description. http://dublincore.org/documents/dces/

ISO, 2003, ISO 19115:2003, Geographic information – Metadata.

OGC, 2005. Open Geospatial Consortium. http://www.opengeospatial.org/

Pilgrim, M., 2002. What is RSS? http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html

Skogen, M. D. and H. Søiland. 1998. A user's guide to NORWECOM V2.0. The Norwegian Ecological Model System. Fisken og Havet, 18:42 pp.

Smits, P. C., et al., 2002, INSPIRE Architecture and Standards Position Paper. http://inspire.jrc.it/reports/position_papers/inspire_ast_pp_v4_3_en.pdf

W3C, 2005. World Wide Web Consortium. http://www.w3.org/