

La comunidad geoespacial y el acceso a los datos oceanográficos.

Matias Bonet, Rosario Cañas, Marta Fuster, Bartolome Garau Pujol, Miguel González, Joaquín Tintoré, Guillermo Vizoso

Instituto Mediterráneo de Estudios Avanzados. IMEDEA (CSIC-UIB). sig.imedea@uib.es

RESUMEN

En la actualidad, las comunidades oceanográfica y geoespacial son mundos estrechamente relacionados pero que han seguido caminos paralelos a la hora de realizar un almacenaje, distribución, modelado y análisis de sus datos. Lo cual se traduce en la implementación de distintos modelos de datos para la representación del mismo fenómeno. Mientras que los datos geoespaciales tienen 2 ó 3 parámetros dimensionales los datos oceanográficos tienen múltiples parámetros como temperatura, salinidad, corrientes marinas, color del océano... que son multidimensionales. Como consecuencia, esta diferencia en los modelos de datos que tienen ambas comunidades, puede provocar dificultades a la hora de mostrar y analizar los conjuntos de datos procedentes de las dos disciplinas.

Esta dificultad afecta de manera directa al IMEDEA (CSIC-UIB) ya que en este organismo se tratan datos oceanográficos procedentes de CTDs, boyas, gliders..., y datos geoespaciales por lo que se refiere a la gestión integrada de la zona costera. En este artículo se presenta la aproximación a la resolución de este problema mediante la implementación de un catálogo THREDDS (Thematic Real-time Environmental Distributed Data Services) el cual permite el acceso a los datos a través del protocolo estándar para acceso a datos espaciales Web Coverage Service, dentro del marco del proyecto europeo ECOOP (European COastal sea Operational observing and Forecasting system) cuyo objetivo es consolidar, integrar y desarrollar un sistema pan-europeo de observación operacional y predicción encaminado a la detección de los cambios ambientales y climáticos.

ABSTRACT

Nowadays, Oceanographic and Geospatial communities are closely related worlds. The problem is that they follow parallel paths in data storage, distributions, modelling and data analyzing. This situation produces different data model implementations for the same features.

While Geospatial information systems have 2 or 3 dimensions, the Oceanographic models uses multidimensional parameters like temperature, salinity, streams, ocean colour... This implies significant differences between data models of both communities, and leads to difficulties in dataset analysis for both sciences.

These troubles affect directly to the Mediterranean Institute for Advanced Studies (IMEDEA (CSIC-UIB)). Researchers from this Institute perform intensive processing with data from oceanographic facilities like CTDs, moorings, gliders... and geospatial data collected related to the integrated management of coastal zones.

In this paper, we present an approach solution based on THREDDS (Thematic Real-time Environmental Distributed Data Services). THREDDS allows data access through the standard geospatial data protocol Web Coverage Service, inside the European project (European Coastal Sea Operational Observing and Forecasting system). The goal of ECOOP is to consolidate, integrate and further develop existing European coastal and regional seas operational observing and forecasting systems into an integrated pan- European system targeted at detecting environmental and climate changes.

Key words: netcdf, thredds, WCS, CSW, TMOOS, ECOOP Jornadas, SIG, software libre, Girona.

INTRODUCCIÓN

Tradicionalmente son conocidas las diferencias entre los modelos de datos utilizados por la comunidad geoespacial y por la comunidad oceanográfica [1]. Estas diferencias en los modelos de datos que tienen ambas comunidades, puede provocar dificultades a la hora de mostrar y analizar los conjuntos de datos procedentes de las dos disciplinas. Fundamentalmente se puede decir que las dos comunidades no hablan el mismo lenguaje, aunque distintas iniciativas internacionales como GALEON IE (Geo-interface for Atmosphere, Land, Earth, and Ocean netCDF Interoperability Experiment) [2] se han puesto en marcha para abordar y solventar estos problemas.

Los factores que dificultan la interoperabilidad entre la comunidad geoespacial y oceanográfica [1] [3]:

- Diferencias en cuanto a la terminología utilizada por ambas comunidades. Por ejemplo, el término *atributo* representa el valor que tiene un objeto geoespacial mientras que para los datos científicos multidimensionales define el metadato para una variable concreta o para todo el archivo.
- La representación de la evolución de los datos en el tiempo. Tanto en el campo de la oceanografía como en el de la meteorología, la evolución temporal de un determinado fenómeno es tan importante como su localización espacial, por el contrario la evolución temporal de los datos geoespaciales (gis) es mucho más lenta. Los sistemas de información geográfica pueden manejar la variable temporal como atributo de una capa pero, normalmente, las series temporales no son soportadas.
- El aspecto geográfico. La información geográfica distingue entre fenómenos observados y su ubicación en la tierra, el grado de precisión por lo que se refiere a la localización espacial es más elevado en los datos gis, tan importante es el fenómeno como su ubicación. Por el contrario los datos oceanográficos mantienen el aspecto espacial de los datos de la manera más simple posible y, en la mayoría de los casos no se incluye el datum en los metadatos.
- Diferencias estructurales entre los datos oceanográficos y los geoespaciales. Mientras que en los primeros se manejan arrays multidimensionales, los segundos optan por bases de datos relacionales. Ver figura anexa.

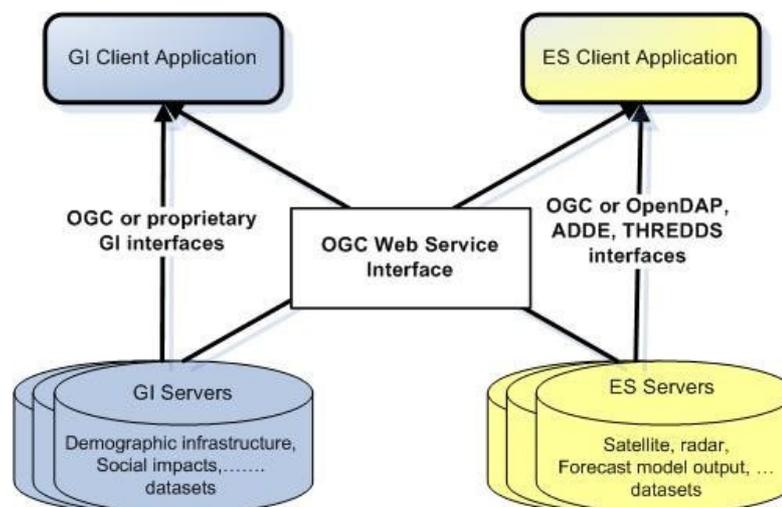


Figura 1. Estructuras de datos de "Earth Science" y "Geographic Information" e interoperabilidad mediante estándares OGC [3]

En este artículo y, concretamente, desde el Centro Operacional de Datos Oceanográficos (CODO), perteneciente al departamento de Tecnologías Marinas, Oceanografía Operacional y Sostenibilidad (TMOOS) [4] y del Servicio de Sistemas de Información Geográfica [5] del Instituto Mediterráneo de Estudios Avanzados (IMEDEA (CSIC-UIB)) [6], se pretende definir la línea de trabajo adoptada para que la información oceanográfica generada en el IMEDEA (CSIC-UIB) sea también accesible para la comunidad geoespacial.

DATOS OCEANOGRÁFICOS

Dentro del mundo de la oceanografía física, es muy común tener diversos modelos numéricos (HOPS, ROMS, Symphonie...) corriendo para obtener unos resultados sobre una misma área del océano. Normalmente, el resultado de estos procesos genera una elevada cantidad de información, del orden de GB, cuyo acceso debe ser fácil, concreto, rápido y eficaz.

Actualmente los datos oceanográficos generados o almacenados en el TMOOS superan los 2.5 TB. Esta cantidad de información tiene su origen en:

- Grids: entendidos como mallas de información referentes tanto a la salida de modelos oceánicos como a imágenes satélite.
- Trayectorias: representaciones de ubicaciones espacio-temporales de mediciones a lo largo de un recorrido procedentes por ejemplo de gliders (planeadores submarinos).
- Series temporales: procedentes de boyas y CTDs (perfiladores de conductividad, temperatura y profundidad)

Además, se debe tener en cuenta que la cantidad de información aumenta diariamente ya que por ejemplo, un modelo de predicción oceánica para el mar Balear puede llegar a originar entre 2 y 3 GB diarios.

¿Cómo se gestiona esta información?

Tal cantidad de información hace que su gestión no sea fácil, por ello el TMOOS ha adoptado el formato NetCDF [7], como estándar para todos los datos almacenados y creados en el departamento ya que es ampliamente utilizado por la comunidad científica a escala mundial, la recomendación para metadatos *Climate and Forecast (CF) Metadata Convention* [8], y la implantación de un catálogo THREDDS [9] para poder acceder de manera eficiente a la información ya sean usuarios tradicionales de datos oceanográficos como usuarios gis.

NetCDF

NetCDF (network Common Data Form), desarrollado por por Unidata [10], son un conjunto de interfaces y librerías distribuidas libremente para los lenguajes de programación C, Fortran, C++, y java entre otros, que permiten el acceso a datos almacenados en vectores y matrices. Conjuntamente, interfaz, librerías y formato, permiten la creación, acceso e intercambio de datos científicos.

Los archivos NetCDF son:

- *Autodescriptivos*, es decir, incluyen metadatos en el propio archivo.
- *Portables*, pueden ser accedidos desde todos los sistemas operativos incluso desde lenguajes de programación distintos.
- *Escalables*, permite el acceso por partes al fichero.

- *Agregables*, los datos pueden ser agregados correctamente a un archivo netCDF sin copiar el dataset o redefinir sus estructuras.
- *Intercambiables*, el autor y múltiples lectores pueden acceder simultáneamente al mismo archivo netCDF.
- *Archivables*, mantiene el acceso a versiones anteriores de datos netCDF.

CF conventions

Las Climate and Forecast Metadata Conventions (CF Conventions) son recomendaciones diseñadas para promover el procesamiento e intercambio de archivos. Estas recomendaciones definen los metadatos que proporcionan una descripción definitiva sobre que datos representan cada variable así como sus propiedades espaciales y temporales. Esto permite a los usuarios de datos procedentes de distintas fuentes decidir que variables son comparables y como se pueden representar en el espacio.

Dado que netCDF no define ninguna relación entre las variables y sus dimensiones si se estructura el fichero en base a las convenciones CF se consigue además de unir metadatos con los datos, su relación espacial.

En el TMOOS se ha adoptado pues el NetCDF/CF ya que además de las ventajas descritas son una extensión de otras convenciones como COARDS y GDT conventions [x].

THREDDS (TDS, Thematic Real-time Environmental Distributed Data Services)

Es un middleware, implementado en java, entre el servidor de datos y el cliente o usuario. TDS proporciona el acceso a los metadatos y a los datos científicos. A través de las siguientes tecnologías y protocolos:

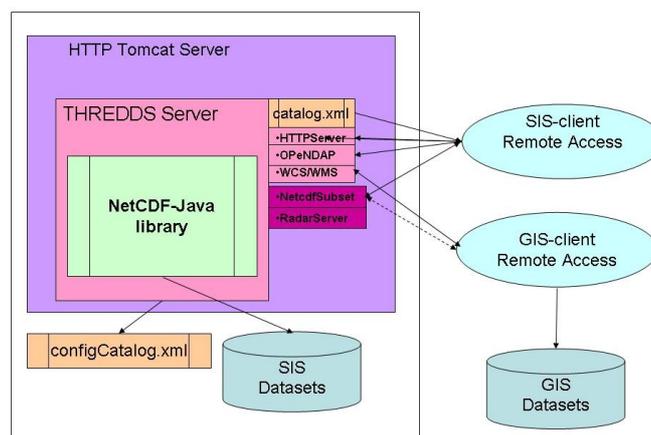


Figura 2. Repositorio de Datos Oceanográficos (Science Information Systems) y su conexión a clientes SIG (GIS).

- THREDDS Dataset Inventory Catalogs: documentos XML que proporcionan directorios virtuales sobre los datos disponibles y sus metadatos asociados.

III Jornadas de SIG Libre

- Librería NetCDF-Java: permite la lectura de archivos NetCDF, OpenDAP, HDF5 así como otros formatos binarios como GRIB o NEXRAD en CDM (Common Data Model). Los archivos que pueden ser leídos a través de la librería Netcdf-Java son denominados CDM datasets.
- TDS puede utilizar NetCDF Markup Language (NcML) para modificar y crear agrupaciones virtuales de CDM datasets.
- TDS proporciona acceso OPeNDAP (Open-source Project for a Network Data Access Protocol) a cualquier CDM dataset.
- TDS proporciona acceso a través del protocolo http con cualquier tipo de archivo.
- TDS proporciona el **acceso a los datos a través del protocolo Web Coverage Service (WMS)** del OpenGis Consortium (OGC) [11] para archivos netCDF/CF.
- Servicio NetCDF Subset que permite acceder a subconjuntos de los datos a través de la selección del área geográfica (lat/lon) y de los rangos de variables definidas en el archivo NetCDF, para archivos netCDF/CF.

Los dos últimos servicios requieren ficheros NetCDF/CF puesto que necesitan la situación espacial de las variables.

Una vista general del THREDDs instalado en el TMOOS se puede ver en las siguientes figuras (Figura 3 y Figura 4)

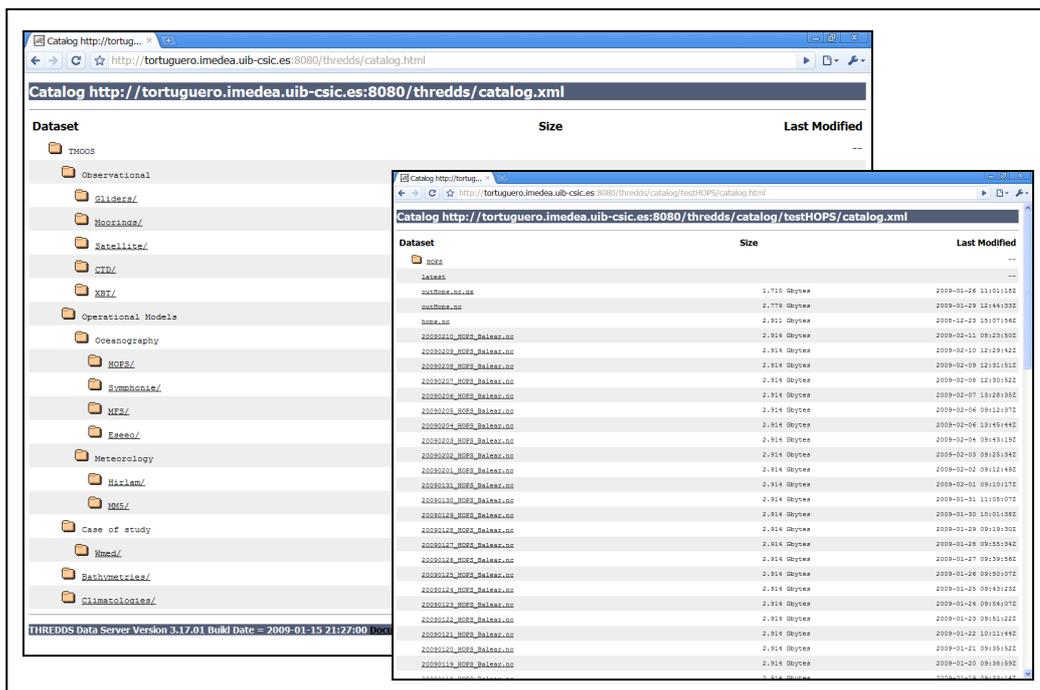


Figura 3. THREDDs

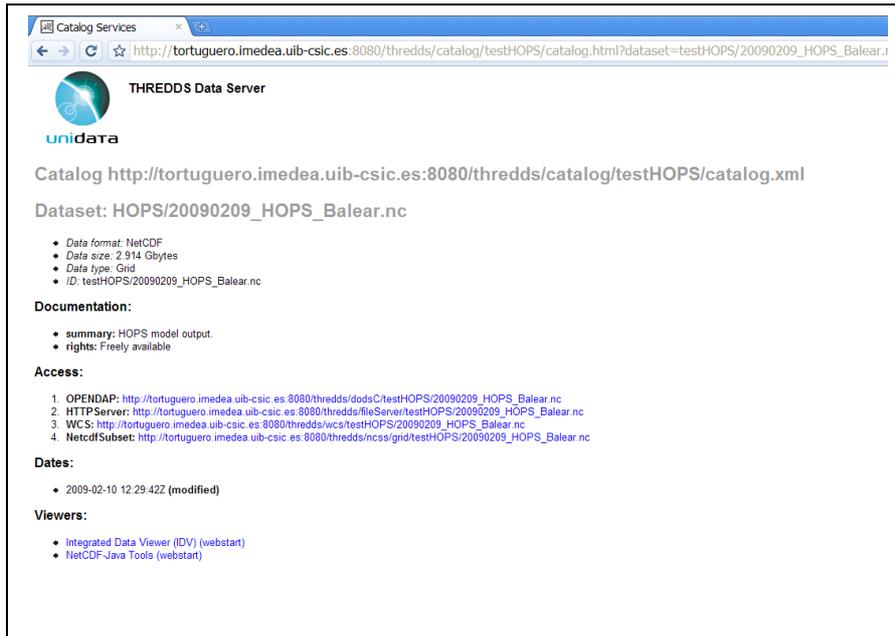


Figura 4. Acceso al archivo seleccionado.

El acceso a la información

a) THREDDS WCS

Como se ha visto, TDS permite el acceso a los datos de diversas maneras, bien sea mediante la descarga directa de todo el archivo (HTTP Server), bien sea mediante la descarga de parte del mismo (OPeNDAP, Netcdf Subset). La comunidad de usuarios gis puede acceder también a través de la descarga de los archivos en formato NetCDF, ya que hoy en día existen programas gis que permiten la lectura de estos formatos (como es el caso de ArcGis 9.2), o bien, a través del protocolo Web Coverage Service, no obstante THREDDS WCS Server todavía se considera experimental [12].

La versión actual TDS 3.17 implementa WCS Server 1.0 Esta versión posee ciertas limitaciones que dificultan el acceso: interpolación no disponible, todos los sistemas de coordenadas están catalogados como WGS84, el sistema de coordenadas sólo es horizontal, sólo se puede hacer una selección temporal de los datos, cada cobertura sólo puede tener un campo rango, y soporta la respuesta GetCoverage con los formatos GeoTIFF, GeoTIFFfloat y NetCDF3 (archivo netCDF que cumplimenta la convención CF-1.0).

El acceso URL a la información a través del protocolo WCS:

Punto de acceso:

`http://servername:8080/thredds/wcs/`

Definición del archivo:

`http://servername:8080/thredds/wcs/test/sst.nc`

URL necesaria para el cliente WCS

`http://servername:8080/thredds/wcs/test/sst.nc?
service=WCS&version=1.0.0&request=GetCapabilities`

b) Godiva2 ncWMS

Godiva2 [15] es una página web dinámica que proporciona acceso visual a varios terabytes de información científica (archivos oceanográficos 4D). Godiva2 se apoya en un servicio OGC Web Map Service (WMS) adaptado para archivos NetCDF/CF-compliant, ncWMS [14], el cual fue diseñado para generar imágenes de archivos NetCDF/CF de manera muy rápida. ncWMS es compatible con WMS 1.1.1 y 1.3.0.

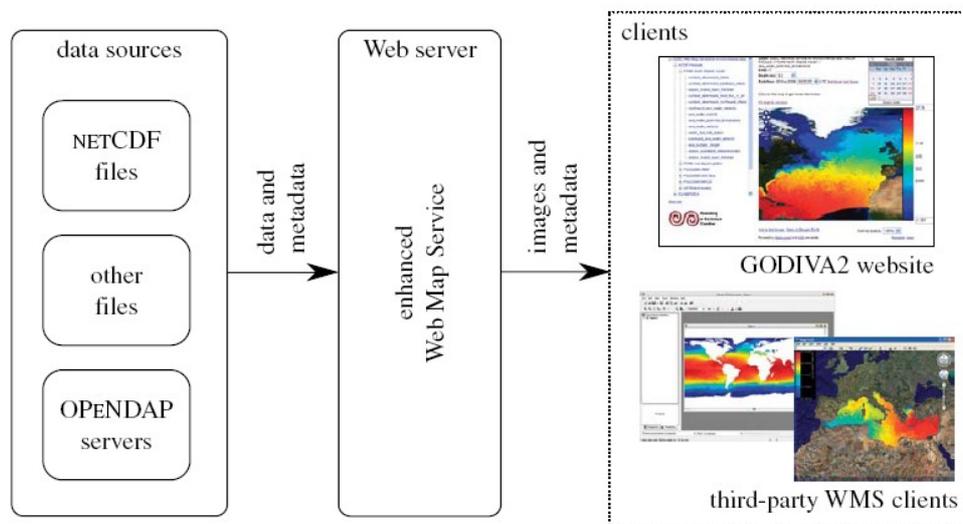


Figura 5. Arquitectura del sistema Godiva2 cliente-servidor [15].

Las fuentes de datos para Godiva2 pueden ser:

- Ficheros locales.
- Directorio que contenga ficheros NetCDF/CF.
- Ficheros servidos a través de OPeNDAP.
- Experimentalmente, catálogo THREDDS con archivos NetCDF/CF.

El sistema Godiva2/ncWMS se utiliza actualmente en dos grandes proyectos europeos sobre datos operacionales y de predicción oceánica en conformidad a la directiva europea INSPIRE: MERSEA [16], para océanos abiertos, y ECOOP [17] para zonas costeras. Actualmente el Instituto Mediterráneo de Estudios Avanzados se encuentra implicado activamente en el proyecto ECOOP (European COastal sea Operational observing and Forecasting system). A día de hoy el IMEDEA (CSIC-UIB), dentro del marco de ECOOP, ofrece los datos del modelo oceanográfico HOPS (predicción de corrientes, temperatura, salinidad) a través de este sistema [18] sirviendo sus datos a través de THREDDS y OPeNDAP. (Figura 6).

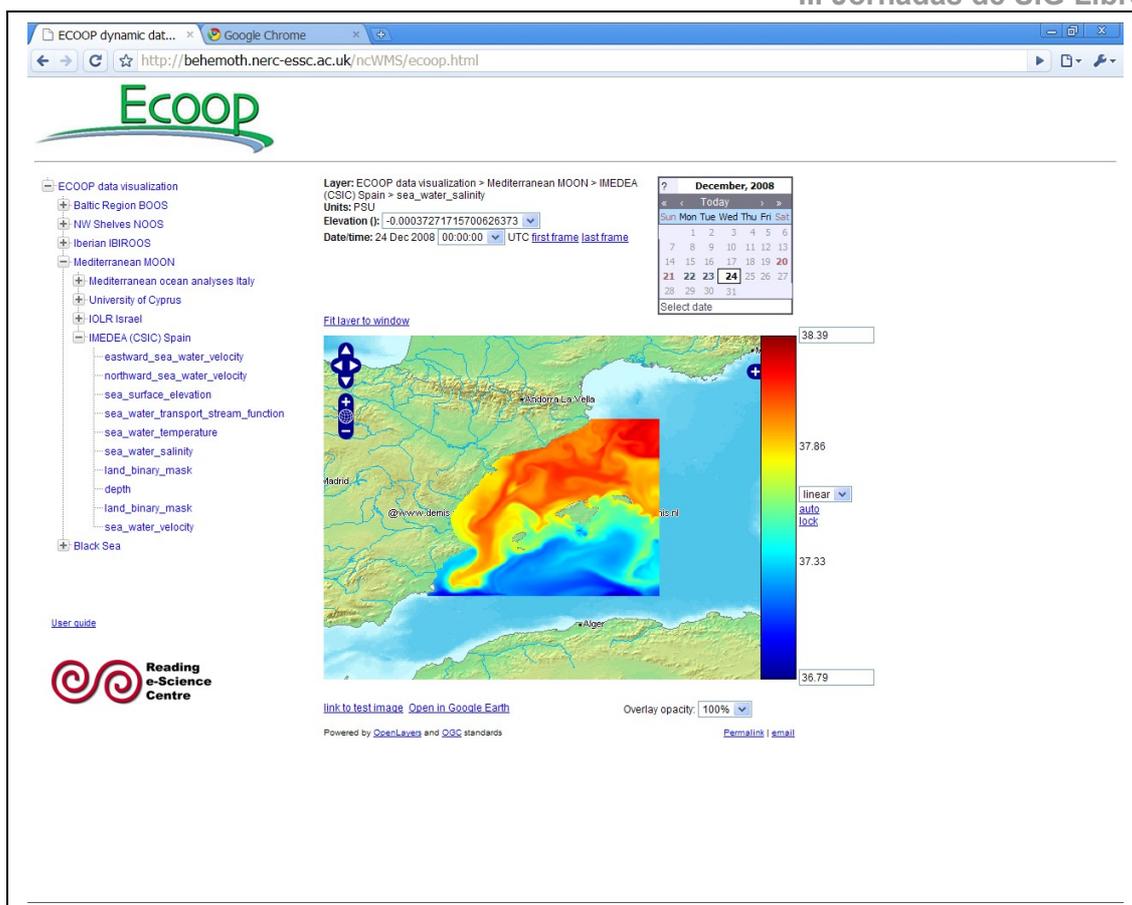


Figura 6. ncWMS ECOOP dynamic data visualization site

La interfaz de Godiva2 (Figura 5) permite seleccionar el archivo y la variable deseada a través del menú desplegable de capas disponibles. A su vez el calendario permite ver el mismo archivo y variable para otro periodo de tiempo, seleccionando un rango temporal se puede generar animaciones.

CONCLUSIONES Y TRABAJOS FUTUROS

El desafío está en lograr la interoperabilidad entre los dos modelos de datos, pero la pregunta podría ser: ¿es necesaria esta interoperabilidad? Los argumentos a favor son:

- No se trata de dar toda la operativa SIS sobre un sistema GIS, sino usar en cada momento la herramienta adecuada para el problema concreto, así como encontrar un campo común cuando fuentes de datos de ambos entornos tienen que convivir en un proyecto concreto.
- Mejora del acceso por parte de otros grupos de investigación y la sociedad en su conjunto.
- Presentación de resultados concisos frente a Teras de información oceanográfica. Paso de sistemas n-dimensionales a bidimensionales o 2 y medio dimensionales. Secuencias temporales no continuas.
- Creación y explotación de servicios OGC (WCS/CSW)
- Resolución de problemas geográficos. Geoproceso (perfiles de inundación, etc)
- Creación de soluciones mixtas (SIS/GIS) con la explotación de servicios OGC para completar datos SIS. I.e. modelos de clorofila sobre WMS de ortofotografía del IGN.

- Vinculación de catálogos vía Threeds y CSW, lo que permite “descubrir” información de ambos “universos” y explotarla conjuntamente.

En este sentido, el trabajo a afrontar está en la explotación de estas ventajas mediante la explotación de los estándares en desarrollo. Sirviendo datos a clientes SIG mediante servicios NetCDF/WCS (THREDDS WCS Server) y “publicando” esta información mediante servicios CSW vinculados a los datos oceanográficos generados por el departamento TMOOS del IMEDEA (CSIC-UIB).

La tecnología opensource en proceso de evaluación para estos objetivos son:

- GI-GO Geobrowser [19]. Que actualmente soporta WCS 1.0, WCS 1.1, WMS 1.3.0, CSW/ISO 2.0.2 , THREDDS 1.0.1, GI-cat 5.x.
- Geonetwork [20] como cliente CSW
- Geoserver [21]

EL IMEDEA (CSIC-UIB)

El Instituto Mediterráneo de Estudios Avanzados es un (IMEDEA (CSIC-UIB)) es un instituto mixto entre la Universidad de las Islas Baleares (UIB) y el Consejo Superior de Investigaciones Científicas (CSIC), cuya principal seña de identidad es el énfasis que pone en las aproximaciones interdisciplinares a la investigación científica. Dentro del IMEDEA (CSIC-UIB), el Departamento de Tecnologías Marinas, Oceanografía Operacional y Sostenibilidad (TMOOS) [4] centra su investigación en la Física y Tecnología del Sistema Costero Oceánico: Observación, Predicción e Interacciones. La variabilidad de las escalas de los fenómenos que se dan en este sistema, desde metros a miles de kilómetros y de minutos a años, juntamente con sus interacciones no lineales, hacen que entender estos mecanismos sea un reto internacionalmente reconocido. Este reto es afrontado combinando modelos teóricos, observacionales (in situ y remotamente) y numéricos en particular (pero no sólo) para el mar Mediterráneo, un laboratorio ideal, un océano a pequeña escala para entender los procesos físicos y probar nuevas ideas.

Algunos Datos sobre IMEDEA (CSIC-UIB): 8,4 millones de euros de presupuesto anual (en el 2006). Una plantilla de 150 empleados, 31 investigadores permanente. 88 proyectos de investigación en marcha.

AGRADECIMIENTOS

Marta Fuster agradece la ayuda del programa de subvención para la contratación de Técnicos de Apoyo PTA2008-1435-I del Ministerio de Ciencia e Innovación del Gobierno Español.

REFERENCIAS

- ◆ [1] S. Nativi, M. B. Blumenthal, J. Caron, B. Domenico, T. Habermann, D. Hertzmann, Y. Ho, R. Raskin, and J. Weber, "Differences among the datamodels used by the Geographic Information Systems and Atmospheric Science communities," in *20th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography*.
- ◆ [2] <http://www.opengeospatial.org/projects/initiatives/galeonie>
- ◆ [3] Chen Hsiang-Yun, 2007, "Interoperability between Web Services for Geoinformation and Earth Sciences". (Thesis)
- ◆ [4] <http://imedea.uib-csic.es/tmoos/>
- ◆ [5] <http://www.imedeas.uib.es/gis/geoportal/>
- ◆ [6] <https://imedea.uib-csic.es/>

- ◆ [7] <http://www.unidata.ucar.edu/software/netcdf/>
- ◆ [8] <http://cf-pcmdi.llnl.gov/>
- ◆ [9] <http://www.unidata.ucar.edu/projects/THREDDS/>
- ◆ [10] <http://www.unidata.ucar.edu/>
- ◆ [11] <http://www.opengeospatial.org/standards/wcs>
- ◆ [12] <http://www.unidata.ucar.edu/projects/THREDDS/tech/reference/WCS.html>
- ◆ [13] <http://www.resc.rdg.ac.uk/trac/ncWMS/>
- ◆ [14] <http://www.resc.rdg.ac.uk/trac/ncWMS/>
- ◆ [x] <http://www.unidata.ucar.edu/software/netcdf/conventions.html>
- ◆ [15] JD Blower, K Haines, A Santokhee, and Liu CL. GODIVA2: Interactive visualization of environmental data on the web. *Philosophical Transactions of the Royal Society A*, 2008.
- ◆ [16] <http://www.mersea.eu.org/>
- ◆ [17] <http://www.ecoop.eu/>
- ◆ [18] <http://behemoth.nerc-essc.ac.uk/ncWMS/ecoop.html>
- ◆ [19] <http://zeus.pin.unifi.it/cgi-bin/twiki/view/Glgo/WebHome>
- ◆ [20] <http://geonetwork-opensource.org/>
- ◆ [21] Mickael Treguer (IFREMER), Martin Desruisseaux (Geomatys) 2007 "Managing WMS and WCS multidimensional NetCDF Datasets with Geoserver".