



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

HACIA UNA IDE MARINA: UNA METODOLOGÍA PARA LA CREACIÓN DE METADATOS OCEANOGRÁFICOS

Soto, Sara ⁽¹⁾; Hernández, Alberto ⁽²⁾; Olivé, Joan ⁽²⁾; Chic, Óscar ⁽²⁾; Sorribas, Jordi ⁽²⁾.

⁽¹⁾ Instituto de Ciencias del Mar ; ⁽²⁾ Unidad de Tecnología Marina - CSIC.

El concepto de la IDE se ha desarrollado mucho para el contexto terrestre y en estos momentos se sirven miles de mapas con capas de información muy diversa. Más difícil está siendo la creación de la IDE para el mundo marino por sus especiales características de dimensionalidad (ej. 4D), la necesidad de poder visualizar modelos (ej. evolución de volúmenes 3D con el tiempo) y por el poco interés mostrado hasta ahora por la comunidad oceanográfica en el uso de estas herramientas.

Dentro del proyecto de elaboración de una IDE Marina para el Centro de Mediterráneo de Investigaciones Marinas y Ambientales, se establece el objetivo de proporcionar los datos a través de sus descriptores o metadatos. Tras un estudio previo de las tecnologías existentes se ha visto la necesidad de mejorar la integración entre estas, para la creación y la explotación de metadatos.

El desarrollo de un Servicio de Catálogo Web (CSW) es uno de los primeros pasos dados hacia la creación de una IDE Marina. Servirá para buscar y consultar los metadatos asociados a los datos generados en los diferentes proyectos. En esta primera fase se decide integrar los metadatos que pueden ser generados en las campañas oceanográficas de los buques gestionados por la Unidad de Tecnología Marina (BIO Hespérides y B/O Sarmiento de Gamboa). Las herramientas elegidas son MIKADO y GeoNetwork, como editor de metadatos y aplicación para la elaboración del CSW respectivamente.

La no existencia de un estándar único para el mundo marino que siga una ontología determinada se convierte en una dificultad añadida. El que algunas instituciones utilicen sus propios estándares (ej. SeaDataNet a través de MIKADO) dificulta la integración de los metadatos. En general, es aconsejable que la creación de metadatos asociados a los datos oceanográficos se haga lo más cercana en el tiempo de la propia adquisición. A menudo los editores de metadatos son demasiado complejos para ser utilizados en entornos marinos. Es por ello que se han desarrollado unas Aplicaciones web que hacen más sencilla la creación de los metadatos durante en el transcurso de las campañas, bajo las plantillas XML de MIKADO. Dichos metadatos son integrados en el CSW usando tecnología XSL que permite cumplir con los estándares de la ISO19139 que sigue GeoNetwork.



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

La creación de un servicio de catálogo, como primera fase del proyecto de creación de una IDE Marina es una contribución al sistema I+D+i importante ya que permitirá, a través de los metadatos, descubrir y acceder a los datos.

HACIA UNA IDE MARINA: UNA METODOLOGÍA PARA LA CREACIÓN DE METADATOS OCEANOGRÁFICOS

Sara Soto ¹, Alberto Hernández ², Joan Olivé ², Óscar Chic ¹ y Jordi Sorribas ²

¹ Observatorio Oceanográfico Costero (COO), Instituto de Ciencias del Mar (ICM-CSIC)
coo@listas.csic.es

² Depto. de Tecnologías de la Información y Comunicación, Unidad de Tecnología Marina (UTM-CSIC)
utmtel@utm.csic.es

1 Introducción

Hasta hace poco, la comunidad oceanográfica y la geoespacial han modelado de manera distinta la realidad. La primera entendiendo la naturaleza como ecuaciones con parámetros y la segunda como polígonos, puntos... que pueden ser manipulados desde una base de datos. En la actualidad, las IDEs están permitiendo la convergencia hacia unos estándares aceptables para ambas comunidades que van a permitir la interoperabilidad de las aplicaciones para la gestión de los datos [1]. La visualización de los datos vía web permitirá a la comunidad científica su divulgación y el uso científico de los mismos y, por su parte, la comunidad geoespacial añadirá herramientas muy poderosas desarrolladas por la otra comunidad.

Recientemente, se ha producido una revolución de la geociencia con la aparición de *Google Maps* y *Google Earth*. Se ha demostrado que todo lo georreferenciable puede estar en la web. El concepto de la IDE se ha desarrollado mucho para el contexto terrestre y en estos momentos se sirven miles de mapas con capas de información muy diversa. Más difícil está siendo la creación de la IDE para el mundo marino por sus especiales características de dimensión (ej. 4D), la necesidad de poder visualizar modelos (ej. evolución de volúmenes 3D con el tiempo) y por el poco interés mostrado hasta ahora por la comunidad oceanográfica en el uso de estas herramientas.

Los metadatos son críticos en la interoperabilidad y la búsqueda de los datos. De estos depende la localización, el buen uso y la organización de los datos. Este proyecto está orientado a la mejora de la integración de las tecnologías existentes para la creación y explotación de los metadatos.

El desarrollo de un CSW es uno de los primeros pasos dados hacia la creación de una IDE Marina. Servirá para buscar y consultar los metadatos asociados a los datos generados en los diferentes proyectos. En esta primera fase se decide integrar los metadatos que pueden ser generados en las campañas oceanográficas de los buques gestionados por la UTM-CSIC (BIO Hespérides y B/O Sarmiento de Gamboa). Así como



la conexión a otros servicios, para cargar los metadatos contenidos en ellos, como son el caso de *THREDDS* y *GeoServer* (figura 1).

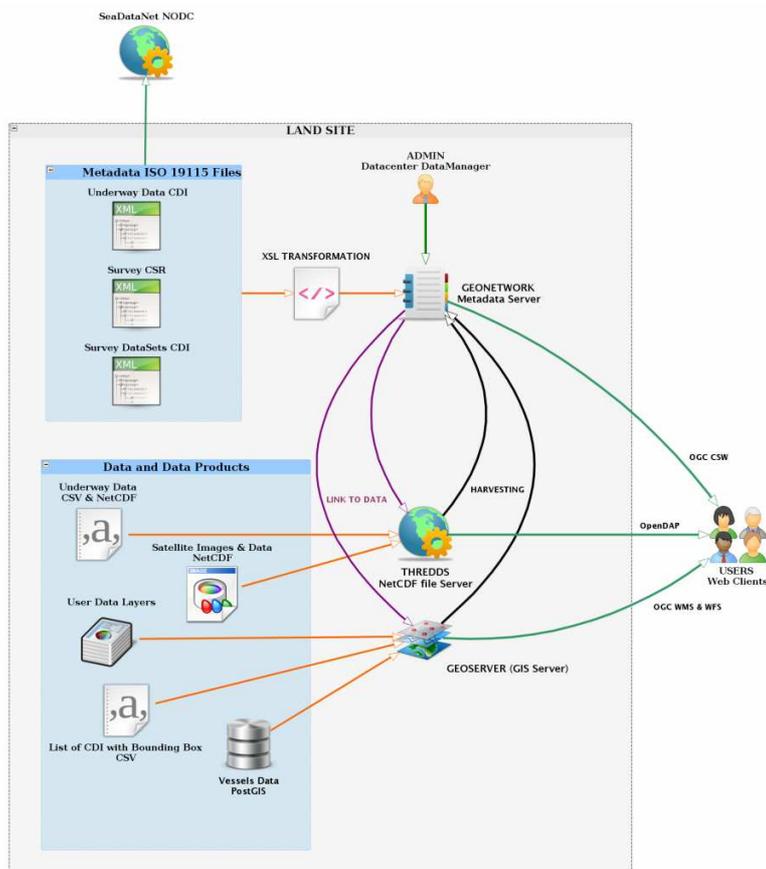


Figura 1. Sistema.

2 Metodología

En este proyecto se apuesta por herramientas con las características siguientes: interoperabilidad, software libre, multiplataforma y multilenguaje, bajo las especificaciones del *Open Geospatial Consortium* (OGC) y los estándares establecidos por la *International Organization for Standardization* (ISO). En torno a las IDEs y a los Sistemas de Información Geográfica (SIG) son muchas las tecnologías existentes y desarrolladas que se pueden encontrar. Tras hacer un estudio comparativo, se optó por *GeoNetwork*, como herramienta para el desarrollo del CSW.



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

GeoNetwork es una aplicación de catálogo de software libre y código abierto para administrar recursos georreferenciados. Permite la edición de estándares de metadatos, aplicar funciones de búsqueda, así como la visualización web de mapas interactivos. Se apoya en una serie de estándares y normas: de metadatos (*ISO19100*, *FGDC* y *Dublin Core*), de interfaces de catálogo (perfiles de cliente y *OGC-CSW Server 2.0.2*, *OAI-PMH Server/Client*, *GeoRSS Server*, *GEO Server OpenSearch*, *WebDAV Harvesting*, *GeoNetwork* a apoyo *GeoNetwork Harvesting*); e interfaces de servicios como el servidor de mapas incrustado *GeoServer*. Se utiliza actualmente en numerosas iniciativas de IDEs en todo el mundo, y es parte de la *Open Source Geospatial Foundation* (OSGeo) [2].

El catálogo está conectado a dos fuentes de datos que proporcionan el servicio de catálogo y el acceso a otros servicios complementarios como *Web Map Service* (WMS), *Web Feature Service* (WFS),... Como se ha indicado anteriormente, estas fuentes son el servidor *THREDDS* y el *GeoServer* (figura 1).

2.1 Servidor *THREDDS*

THREDDS (*Thematic Real-time Environmental Distributed Data Services*) es un *middleware* que une proveedores de datos con usuarios finales. Facilita la búsqueda de datos y su uso científico, permitiendo referenciar estos datos en publicaciones y materiales educativos. Inicialmente se pensó como una tecnología para encontrar los conjuntos de datos de la propia institución, acceder a los datos y poder bajarse solo una parte de éstos. Para que ello sea posible es necesario que los suministradores de datos publiquen la lista de datos, la descripción de estos y cómo utilizarlos. El concepto clave que está en el corazón de *THREDDS* es el catálogo [3].

Los catálogos son documentos XML que describen conjuntos de datos y que pueden contener los metadatos que se quieren. El generador de catálogos de *THREDDS* puede escanear una o varias colecciones de datos locales o remotos. Además, utiliza el *Unidata's Common Data Model* (CDM) que une los modelos de *OPeNDAP*, *netCDF*, y *HDF5* para crear una API común para todos los tipos de datos.

En el Observatorio Oceanográfico Costero (COO) [4] existe un servicio de catálogo *THREDDS* con imágenes desde 2001 del sensor AVHRR de los satélites NOAA (N16 hasta N-19) y *MetOp2* (ESA-EUMETSAT). Estos datos fueron recogidos hasta el 2008 por una antena de recepción de imágenes satélite sita en el tejado del CMIMA y a partir del 2008 por un sistema *EUMETCast*. Los datos son procesados con un software *Terascan* (SeaSpace) y convertidos por un lado a formato *netCDF* y por otro a un mapa de temperaturas en formato imagen *PNG*.



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

Los datos *netCDF* son postprocesados para permitir que cumplan las convenciones de metadatos CF 1.0 y después cargados en *THREDDS* [5]. Desde *GeoNetwork* se logra la visualización de dichos metadatos y acceso a los datos.

2.2 Servidor GeoServer

Abordo de los buques de investigación oceanográfica se adquieren datos de forma continua. Sería el caso de datos meteorológicos (velocidad del viento, humedad relativa del aire, temperatura), de temperatura, de salinidad y fluorimetría superficial del agua y de gravimetría entre otros valores. Toda esta información, que evidentemente tiene un componente espacial, se almacena de forma automática en una base de datos *PostgreSQL-PostGIS* [6] [7].

Es *GeoServer* el servidor encargado de ofrecer esta información geográfica [8]. Este hace accesible mediante un WMS las capas adquiridas durante la navegación y otras, como son las del plan de campaña, con los correspondientes puntos de trabajo y las líneas de navegación que ha de realizar el buque. Estas capas son accesibles a través de un WFS servido por *GeoServer*.

Es importante señalar que los datos almacenados en *PostgreSQL-PostGIS* son utilizados para calcular la delimitación geográfica o *bounding box* del metadato de cada evento de adquisición de una campaña oceanográfica. A partir de la fecha de inicio y de final de cada uno (ver *Web Application* en la figura 2), se buscan las posiciones en ese intervalo obteniendo las longitudes y latitudes máximas y mínimas. A su vez, el listado de dichos metadatos de eventos generados por la aplicación *WebForestAdmin*, sobre la que se profundiza en el siguiente apartado, se utiliza para ser servido por *GeoServer*.

2.3 Buques Oceanográficos

La integración de los metadatos durante el transcurso de campañas oceanográficas se consigue gracias al desarrollo de un sistema que permite su generación de manera semiautomática.

Esto implicó el estudio de las distintas tecnologías para la edición de los metadatos; la recopilación de la información histórica de las campañas oceanográficas gestionadas por la UTM-CSIC así como la edición de sus metadatos; la implementación de aplicaciones Java para la automatización de la edición de metadatos en los buques; y por último la integración de dichos metadatos en la aplicación de catálogo.



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

2.3.1 Elección de la Tecnología

Para cubrir las necesidades de edición de metadatos se elige *MIKADO* como herramienta clave. *MIKADO* es una herramienta desarrollada por el *IFREMER* para la generación de metadatos en el marco de los programas europeos *SEA-SEARCH* y *SeaDataNet*. Este último ha desarrollado una Infraestructura para la Administración de Datos Marinos formada por centros de datos que constituyen una red Pan-Europea que proporcionan bases de datos centralizadas *on-line* de calidad estandarizada [9].

Con esta herramienta se generan archivos XML que siguen Directivas Europeas: *Marine Environmental Datasets* (EDMED), *Marine Environmental Research Projects* (EDMERP), *Cruise Summary Report* (CSR), *Common Data Index* (CDI) and *Ocean Observing System* (EDIOS). La interoperabilidad es la clave del éxito del sistema de gestión de datos distribuida y *SeaDataNet* lo ha conseguido gracias a usar vocabularios comunes, adoptar la norma ISO19115, proporcionar herramientas de software estándar y protocolos de calidad.

2.3.2 Histórico de campañas

Se ha desarrollado un duro trabajo de rescate de metadatos de las campañas oceanográficas de los últimos 20 años. La recopilación de información para las campañas más antiguas ha sido muy complicada, debido a que los soportes, los dispositivos y el software eran obsoletos.

Se recopila la información necesaria para la creación de los metadatos del histórico de las campañas oceanográficas de BIO Hespérides y B/O Sarmiento de Gamboa Infraestructuras Científicas y Tecnológicas Singulares (ICTS) Marinas existentes y gestionadas por la UTM-CSIC.

El BIO Hespérides perteneciente a la Armada Española tiene su base en Cartagena. Donde fue construido y botado el 12 de marzo de 1990; desde entonces ha realizado más de 150 campañas oceanográficas. El B/O Sarmiento de Gamboa perteneciente al Ministerio de Ciencia e Innovación (MICINN) [10] tiene su base en Vigo. Fue botado el 30 de enero de 2006, realizando más de 20 campañas. En ambos la UTM-CSIC es la responsable del mantenimiento del equipamiento científico del buque y aporta el personal técnico de apoyo para la realización de dichas campañas. La gestión científica del buque recae en la Comisión de Coordinación y Seguimiento de las Actividades de Buques Oceanográficos (CCSABO) del MICINN.

Se parte de la creación de los metadatos de todas las campañas oceanográficas hasta llegar a la actualidad. Y a partir de entonces, gracias al sistema establecido los metadatos son creados de forma sencilla y semiautomática en los buques durante las campañas oceanográficas.



2.3.3 Sistema

Como se ha comentado anteriormente tras estudiar cada una de las posibilidades que la tecnología ofrecía y determinar con precisión las distintas necesidades de nuestro sistema se optó por utilizar *MIKADO* para la generación de los metadatos.

La herramienta *MIKADO* es visual con ventanas y pestañas donde se van rellenando campos, muchos de los cuales son seleccionables de vocabularios comunes y sincronizados con *SeaDataNet*. Para todas las bases de metadatos este es un requisito previo importante para la coherencia y la interoperabilidad. Son los campos obligatorios que contienen la información esencial y que junto con la instrumentación y los parámetros forman nuestro núcleo de información.

El producto de la herramienta es un metadato en formato XML que respeta la norma ISO19115 y que van acompañados de campos propios de *MIKADO*. Los formatos de los archivos metadatos generados por nosotros son *Cruise Summary Report* (CSR) y *Common Data Index* (CDI).

CSR describe temporal, geográfica y temáticamente la campaña, especificando la fecha, el lugar, el objetivo, el barco, los parámetros, los instrumentos y los organismos implicados. Así mismo, CDI describen los conjuntos de datos que los investigadores embarcados consideren (corresponden a estaciones, líneas, muestras o cualquier evento de adquisición con diferente grado de granularidad) especificando dónde, cuándo, qué parámetros, cómo (qué instrumentos) y quién (qué organismos).

Aunque *MIKADO* es una herramienta muy sencilla de usar, no es óptima para crear metadatos con ella durante las campañas, por la inversión de tiempo que supone. Dado que es aconsejable que la creación de metadatos sea lo más cercana en el tiempo a la propia adquisición de los datos, se apuesta por una creación semiautomática de su descripción.

Por estas razones se desarrollaron dos aplicaciones web Java (*WebForestUser* y *WebForestadmin*) que permiten generar fácilmente metadatos durante las campañas usando plantillas hechas con *MIKADO* (figura 2).

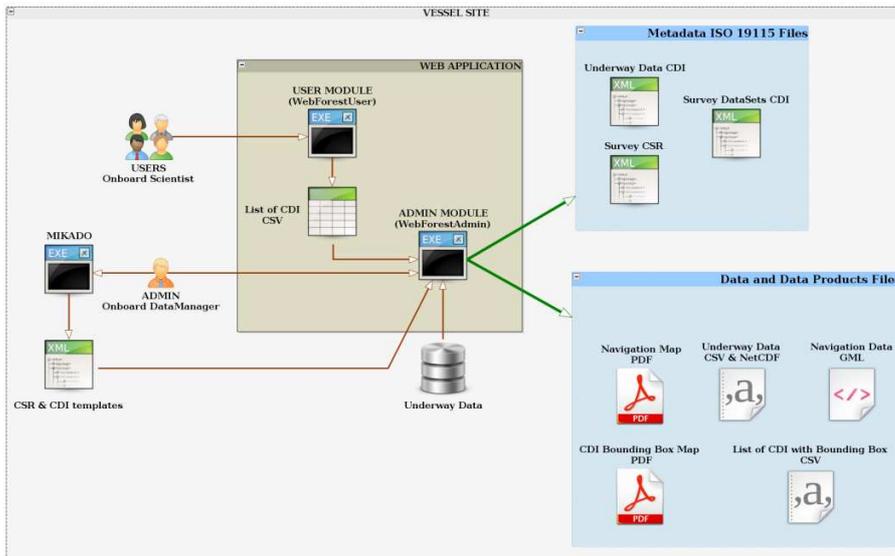


Figura 2. Sistema en Barco.

WebForestUser es una aplicación que genera el listado de eventos de adquisición de datos que dan lugar posteriormente a un metadato CDI. Emplea una base de datos MySQL con una tabla donde guarda las entradas de *start date*, *end date*, *station name* y *the name of the template*; que será usada posteriormente para generar los CDI definitivos.

Estas plantillas están creadas con *MIKADO* y se denominan con el nombre del instrumento o el sistema de adquisición; e internamente describe detalladamente los sensores y los instrumentos empleados, y los parámetros estudiados. Posteriormente dicho listado es cargado en la aplicación *WebForestAdmin* para generar de manera automática los siguientes productos finales, que detallan las actividades realizadas durante la campaña:

- Los ficheros de metadatos CDI detallados en el listado.
- El listado de entradas CDI y otro con la posición integrada.
- El mapa de localización de los CDI cajas geográficas o *bounding box* que engloban los datos).
- Los ficheros de metadatos CDI correspondientes a los datos de propósito general (*underway*).
- El fichero de metadato CSR, también se genera a partir de una plantilla previamente generada con *MIKADO*.
- El fichero de navegación de la campaña en *GML*.
- El mapa de la navegación.



- Los ficheros de los datos de propósito general integrados (*underway*) con la posición

Todos los ficheros generados durante la campaña se guardan en la base de datos de los barcos que se sincroniza con la de la estación en tierra que está en el Centro Mediterráneo de Investigaciones Marinas y Ambientales (CMIMA-CSIC) de Barcelona a través de enlace vía satélite.

2.3.4 Integración de XML de *MIKADO* en *GeoNetwork* usando plantillas XSL.

Como hemos señalado anteriormente, *MIKADO* genera ficheros XML, siguiendo como estándares la ISO19115 y unos metadatos propios de la *SeaDataNet*. Sin embargo, el objetivo es conseguir unos metadatos fijados por la familia ISO19100, que es lo que INSPIRE establece. En concreto el aplicar la ISO19139, que determina lo referente a los metadatos en relación a su aplicación en esquemas XML y así poder representarlos a través de *GeoNetwork*.

La integración entre ambas aplicaciones se consigue a partir de la aplicación de plantillas XSL. A través de la modificación de la hoja de estilo 19115to19139 de *GeoNetwork* fue posible la integración de los ficheros XML procedentes de *MIKADO*.

Su adaptación pasa por la eliminación de algunos metadatos opcionales como el caso del grupo de metadatos de información de extensiones de metadatos, así como la integración de otros metadatos de la ISO no considerados por *GeoNetwork* como la información agregada. En otros casos, la plantilla se ajusta a las necesidades de cambio de los metadatos *SeaDataNet* a la ISO, como el identificador del fichero de metadatos, palabras clave y algunos de los metadatos que componen las restricciones de acceso. Una última intervención, es la integración de los metadatos que determinan el rectángulo envolvente geográfico, es decir los metadatos que definen las coordenadas de latitud y longitud, que son representadas por *GeoNetwork* en la plantilla de visualización de los metadatos, gracias al servicio de mapa incrustado.

De esta forma los metadatos generados en los buques pasan a formar parte del CSW del CMIMA-CSIC.

2.4 Resultados

Con este sistema se consigue un Servicio de Catálogo Web, que contiene los metadatos de las campañas (CSR) realizadas por el BIO Hespérides y el B/O Sarmiento de Gamboa hasta la fecha actual, así como los metadatos de los registros de datos (CDI), tomados en las últimas campañas. Su número se estima en unos 3500 metadatos.



Junto a estos, se encuentran disponibles los metadatos adquiridos gracias a la conexión a los servicios *THREDDS* y *GeoServer*. Estos incrementan el número total de metadatos, ya que hay que sumarle los 18000 metadatos aportados por el servicio de Imágenes de Satélite que tiene el COO-ICM-CSIC.

La implementación de las aplicaciones web Java (*WebForestUser* y *WebForestAdmin*) permite obtener metadatos de CSR y de CDI de una manera rápida, sencilla y en el momento de adquisición de datos. Poder tener estos metadatos disponibles significa el acceder a ellos de una manera multitemporal como de un registro continuo y tener una base de datos de cruceros históricos.

El desarrollo de una plantilla XSL que integra *GeoNetwork* y *MIKADO*, permite la catalogación de todos los metadatos elaborados en cada uno de las campañas que se llevarán a cabo en los barcos.

2.5 Conclusiones

La creación de un servicio de catálogo, como primera fase del proyecto de creación de una IDE Marina es una contribución al sistema I+D+i considerable ya que proporcionaría la creación de los metadatos de la información contenida en el centro, que son la forma de descubrir, acceder a los datos y a su descripción.

Con este servicio, un usuario tiene la capacidad de buscar, consultar y acceder a la información en relación con las diferentes campañas, gracias a las diversas herramientas que la aplicación de *GeoNetwork* ofrece a la hora de realizar las consultas sobre todos los metadatos contenidos.

El acceso a los datos y metadatos a través de internet brinda una nueva posibilidad al mundo científico gracias a la creación de publicaciones a través del registro de los datos bajo un DOI.

La implantación de este tipo de sistemas es de gran importancia dentro de la comunidad científico marina. Son numerosas las ventajas que una IDE Marina puede proporcionar, desde la fácil búsqueda y acceso a los datos, la actualización y no duplicidad de estos así como, la localización y disponibilidad de los mismos.

Agradecimientos. A Natalia Rodríguez por sus correcciones y comentarios.



II Jornadas Ibéricas de Infraestructuras de Datos Espaciales

Referencias

- [1] Hankin, S. & Co-Authors (2010). "Data Management for the Ocean Sciences - Perspectives for the Next Decade" in *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 1)*, Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E. & Stammer, D., Eds., ESA Publication WPP-306, doi:10.5270/OceanObs09.pp.21
- [2] *GeoNetwork Open Source* acceso en octubre de 2011. Disponible en <http://GeoNetwork-opensource.org>
- [3] *THREDDS data server (Unidata)* acceso en octubre de 2011. Disponible en <http://www.unidata.ucar.edu/projects/THREDDS/tech/TDS.html>
- [4] Coastal Ocean Observatory, Instituto de Ciencias del Mar-CSIC, acceso en octubre de 2011. Disponible en <http://coo.icm.csic.es/>
- [5] *CF Metadata* acceso en octubre de 2011. Disponible en <http://cf-pcmdi.llnl.gov/>
- [6] *PostgresSQL (OsGeo Project)* acceso en octubre de 2011. Disponible en <http://www.postgresql.org/>
- [7] *PostGIS* acceso en octubre de 2011. Disponible en <http://postgis.refractor.net/>
- [8] *GeoServer* acceso en octubre de 2011. Disponible en <http://geoserver.org/display/GEOS/Welcome>.
- [9] *MIKADO V2.1: user manual* [en línea]. [Issy-les-Moulineaux, France]: Iframer, 2011. Acceso en octubre de 2011. Disponible en http://www.seadatanet.org/content/download/8299/56192/version/1/file/sdn_Mikado_UserManual_V2.1.pdf
- [10] Ministerio de Ciencia e Innovación de España acceso en octubre de 2011. Disponible en <http://www.micinn.es/>