

Un data server per l'accesso unificato ai dati scientifici, nell'ottica di un Sistema Integrato per le Osservazioni dell'Ambiente Marino

T. Minuzzo, A. Bergamasco, S.Carniel, M. Sclavo
CNR ISMAR, Castello 1364 31022 – Venezia Italy
e-mail: tiziano.minuzzo@ismar.cnr.it

1- Introduzione

L'ambiente marino costiero offre una gamma di processi di grande impatto sulla nostra società, dal “sea level rise” al “coastal flooding”, dai bloom algali agli “oil spills”. Inoltre, allo stato attuale, non si comprendono completamente le dinamiche dei cambiamenti climatici che stanno avvenendo a scala globale, le loro cause e ed i loro molteplici potenziali impatti sulla zona costiera.

Molti enti ed organizzazioni nazionali ed internazionali operano sul fronte delle acquisizioni di dati - sia con sensori in situ, come correntometri, mareografi e boe ondametrichi, che con sensori remoti spesso a bordo di satelliti - allo scopo di monitorare e descrivere lo stato di salute dei nostri mari. Tuttavia, gli sforzi attuali sono solo graffiano appena la superficie di un mondo di conoscenze ancora da esplorare per arrivare ad una adeguata valutazione dell'impatto antropico sul tempo meteorologico, sulle variazioni climatiche e sulle sue correlazioni con i processi tipici degli ecosistemi marini e lagunari.

Inoltre, molti di questi sistemi di osservazione sono pensati per collezionare dati e informazioni in modo specifico, o per analizzare alcune problematiche solo da certi punti di vista, e frequentemente non sono di facile accesso. Spesso il ricercatore quindi impiega molto tempo a localizzare e recuperare i dati disponibili ed una volta acquisiti i dati una loro visione generale risulta difficile e spesso una valutazione di tutti i dati nel loro insieme incompleta. Un Sistema Integrato di Osservazione dell'Ambiente Marino, IOOS (*Integrated Ocean Observing System*), rappresenta un tentativo - coordinato dal NOAA in U.S.A. (vedasi <http://www.usnra.org>) ed ancora agli inizi in Italia - nella direzione di rendere i dati oceanografici utilizzabili con maggiore facilità ed efficienza.

2 - Cos'è, chi usa e perché è importante un IOOS

Un IOOS è una rete coordinata di tecnici e tecnologie che lavora insieme per acquisire, organizzare, registrare e disseminare dati osservazionali e modellistici sulle acque marine costiere e oceaniche, assicurando che tutte le informazioni possano essere usate in modo corretto e permettendo conseguentemente ai ricercatori del settore di sviluppare una più completa caratterizzazione e comprensione degli ambienti marini e costieri.

Una volta completata la rete nazionale di un IOOS, per la realizzazione della quale il CNR- ISMAR sta già contattando potenziali partner scientifici, questa diventerà una infrastruttura al servizio di tutti i soggetti interessati a vario titolo a monitorare e predire cambiamenti negli ambienti e negli ecosistemi marino costieri. E' anche possibile sviluppare sistemi integrati a scala sopranazionale, collegando l'IOOS italiano a quello di altri paesi. Ovviamente questo richiede l'uso di un approccio comune e standardizzato ai

protocolli di accesso standard, OPeNDAP, quali Open Geospatial Consortium (OGC) Web Coverage Service e OPeNDAP service in linea con la Climate and Forecast (CF) convention. Una siffatta infrastruttura risulta fondamentale per agevolare lo scambio di risultati ed output di modelli, e quindi per aiutare a comprendere e porre rimedio a problemi derivanti da eventi meteorologici eccezionali, da variabilità climatiche globali-regionali, a rischi di carattere naturale e antropico. Un esempio di impatto a scala globale può essere quello della predizione degli eventi come “El Nino”, mentre esempi a scala regionale-locale possono essere la ciclogenesi (formazione e sviluppo di minimi di pressione in zone preferenziali) mediterranea o l'acqua alta a Venezia.

Le informazioni raccolte attraverso un IOOS possono aiutare la comprensione, la caratterizzazione, il monitoraggio e la gestione degli ambienti costieri e marini. Alcuni potenziali utilizzatori potrebbero essere ad esempio: a) gestori delle risorse costiere e marine, b) gestori delle emergenze, incluse le ricerche a mare e le pianificazioni per burrasche e tempeste, c) Studiosi degli oceani ed ecosistemi costieri, d) “Policymakers” a livello internazionale, nazionale, regionale e locale, e) Educatori e insegnanti consci dell'importanza delle risorse oceaniche, f) Chiunque “usa” il mare sia per lavoro che per divertimento.

Un esempio italiano di gestione congiunta di dati ambientali è GIIDA (<http://www.dta.cnr.it/content/view/2735/244/lang.en/>), una iniziativa del CNR per una gestione interdipartimentale dei dati ambientali. In questa prima fase si sta cercando di creare un network tra i diversi Istituti CNR che operano in campo ambientale, implementando uno spazio comune di informazione e re-ingegnerizzando il sistema osservazionale e modellistico del CNR.

Il contributo che si vuole dare con il presente contributo va proprio in questa linea di sviluppo, mostrando alcune idee operative per la creazione di un **Data Server** adeguato, qui di seguito descritto come una possibile soluzione, parzialmente già adottata nel campo oceanografico-meteorologico.

3 - Come implementare un Data Server: THREDDS

THREDDS (THematic Real-time Environmental Distributed Data Services) è un middleware ideato per colmare il gap tra i “data providers” e i “data users”, vedi Fig.1. L'obiettivo è quello di semplificare l'uso dei dati scientifici, permettendo una più efficiente produzione scientifica e di materiale educativo. THREDDS permette ai “data users” di trovare i data sets pertinenti alle loro specifiche necessità scientifiche, accedere a tali dati e utilizzarli in modo semplice senza necessariamente eseguire un download del file intero nel loro sistema locale. Per ottenere questo, è necessario che i “data providers” pubblichino una lista (catalog) di quali dati sono disponibili per abilitare la loro individuazione ed il loro uso. I “catalogs” sono il cuore concettuale del sistema, documenti in formato XML che descrivono i data sets on-line; essi possono contenere metadati arbitrari che, una volta uniformati, andranno a popolare i discovery centers come GCMD, DLESE e NSDL.

Un generatore di cataloghi produce i THREDDS Catalogs attraverso lo scanning di uno o più collezioni di datasets locali o remoti; tale produzione può essere periodica o “on demand”, usando opportuni files di configurazione che controllano quali directories possono essere analizzate e come i cataloghi possono essere creati.

In sostanza, per unificare l'accesso ai dati scientifici l'idea di fondo è quella di costruire un "Common Data Model" (CDM). UNIDATA, un consorzio di Università americane per il calcolo scientifico nell'ottica "Earth System Science" (www.unidata.ucar.edu) propone di creare una API comune ai diversi tipi di dati usati dalla comunità scientifica come OPeNDAP, netCDF, HDF5. Il CDM è quindi un modello concettuale, mentre un valido esempio è fornito dalla libreria scritta da UNIDATA in NetCDF-Java (<http://www.unidata.ucar.edu/software/netcdf-java/>). Attraverso questa si può quindi popolare il CDM con informazioni risultanti da files NetCDF3, NetCDF4, HDF4, GRIB, GRIB2, OpenDAP, NcML, scritti secondo la convenzione CF, o modificati adeguatamente utilizzando NetCD Markup Language (NcML) che consente di modificarli attraverso XML.

Il CDM provvede inoltre uno standard API per la georeferenziazione dei sistemi di coordinate e queries specializzate per tipi di dato scientifico come datasets di tipo "Grid, Point e Radial". Il middleware THREEEDS è il componente principale del "THREEDS Data Server" (TDS), capace di produrre il contenuto dei datasets e provvede inoltre a produrre i cataloghi e i metadati per essi. Il TDS usa il CDM per leggere i datasets nei differenti formati d'origine e servirli attraverso i subset OPeNDAP, OGC Web Coverage Service, NetCDF e un bulk HTTP file transfer. I primi 3 permettono all'utente di ottenere agevolmente dei subsets dei dati, un aspetto cruciale nella gestione dei grandi datasets. Il TDS ha anche l'abilità di aggregare molti files in un unico file virtuale, semplificando di molto all'utente finale l'accesso a grandi collezioni di files. IL TDS è un codice open source 100% Java e gira dentro il popolare Tomcat Servlet.

Una implementazione del sistema proposto è descritta alla seguente url <http://150.178.42.4:8080/thredds/creus.html?dataset=creus/run5/his>, e rappresenta uno dei primi tentativi della comunità italiana; si rimanda all'indirizzo <http://www.ve.ismar.cnr.it/creus> per una descrizione dettagliata dell'implementazione con gli allegati multimediali dell'esempio effettuato. Nel sito si potrà trovare la simulazione di evento di sprofondamento di acqua densa neoformata nel Golfo del Leone e che sta sprofondando al largo di Cape de Creus.

4 - Commenti

Prendendo spunto dal Progetto IOOS della NOAA, la comunità italiana ed in un certo senso tutta la comunità europea potrebbe fare sua questa sfida proponendo un IOOS per il Mediterraneo. Unificare l'accesso ai dati scientifici può sembrare solo un esercizio di tecnica informatica, ma, come converrà chiunque si sia confrontato con la ricerca e l'accesso a dati, è ormai indispensabile per eliminare quello che appare con sempre maggiore chiarezza un "collo di bottiglia" nella disseminazione e nel razionale utilizzo dei prodotti osservazionali e modellistici. Tuttavia, da non sottovalutare sono le problematiche "politiche" della gestione, manipolazione e pubblicazione dei dati che necessitano di una adeguata "data policy" e di un altrettanto adeguato standard di sicurezza per la rete ed il suo uso.

5 - Ringraziamenti

Ringraziamo tutti i colleghi di UNIDATA per gli spunti ed il supporto e il

Dr. R.P. Signell (USGS) per i consigli e il supporto. L'attività è stata in parte supportata dal progetto PRIN 2008YNPNT9_005 e il progetto FIRB cod. RBF08D825 "DECALOGO - Cascade di acqua densa: da osservazioni locali ad una parametrizzazione dell'oceano globale".

6 - Bibliografia

Signell R. P., Carniel S., Chiggiato J., Janekovic I., Pullen J., Sherwood C.R., 2008. *Collaboration tools and techniques for large model datasets*. Journal of Marine Systems, vol. 69 pp. 154 - 161.

<http://ioos.gov/>

<http://www.unidata.ucar.edu/software/netcdf-java/CDM/>

<http://www.unidata.ucar.edu/>

<http://www.unidata.ucar.edu/projects/THREDDS/>

<http://en.wikipedia.org/wiki/OPeNDAP>

THREDDS Data Server Architecture

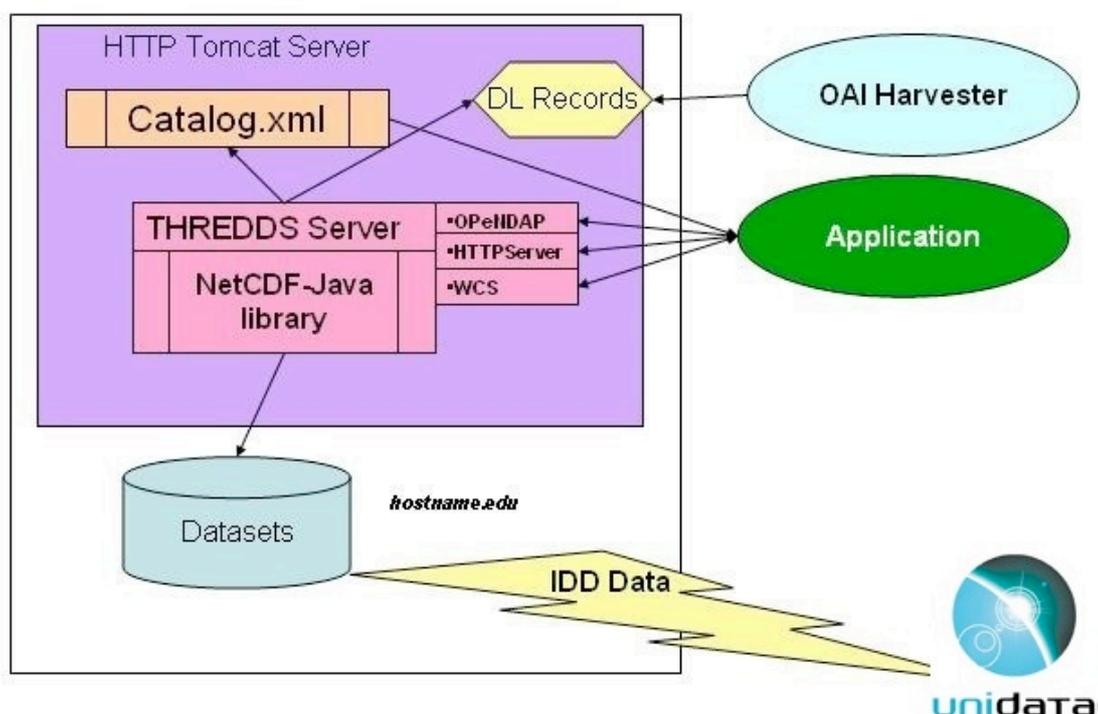


Fig.1 - Struttura di un tipico TDS (Thredds Data Server)