



THREDDDS



**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, Venezia

MOTIVAZIONE

- Potenzialità dell'ambiente marino-costiero
- Mancanza di sistematicità nella catalogazione
- Necessità di una piattaforma comune per data management
- Esempio di un *Integrated Ocean Observing System*

IOOS - *Integrated Ocean Observing System*

- rete 'mondiale' coordinata di tecnici e tecnologie per garantire un approccio comune e standardizzato ai protocolli di accesso
- Scala italiana (CNR)
- Scala soprannazionale (paesi del Mediterraneo)
- Scala Globale



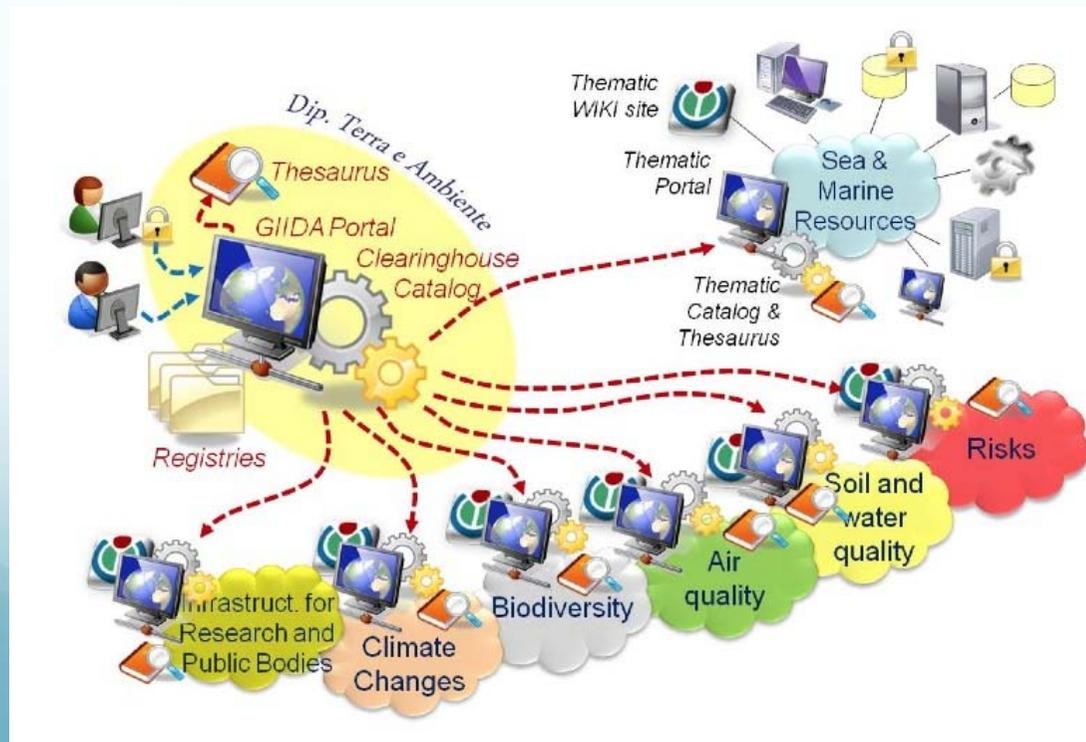
Le informazioni raccolte attraverso un IOOS possono aiutare la comprensione, la caratterizzazione, il monitoraggio e la gestione degli ambienti costieri e marini

Potenziali utilizzatori di un IOOS:

- gestori delle risorse costiere e marine
- gestori delle emergenze
- Studiosi degli oceani e degli ecosistemi marini
- “Policymakers” di livello internazionale, nazionale, regionale, comunali
- Chiunque usi il mare per lavoro o per divertimento

Un esempio italiano

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.



GIIDA 
Gestione
Integrata e
Interoperativa dei
Dati
Ambientali

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”



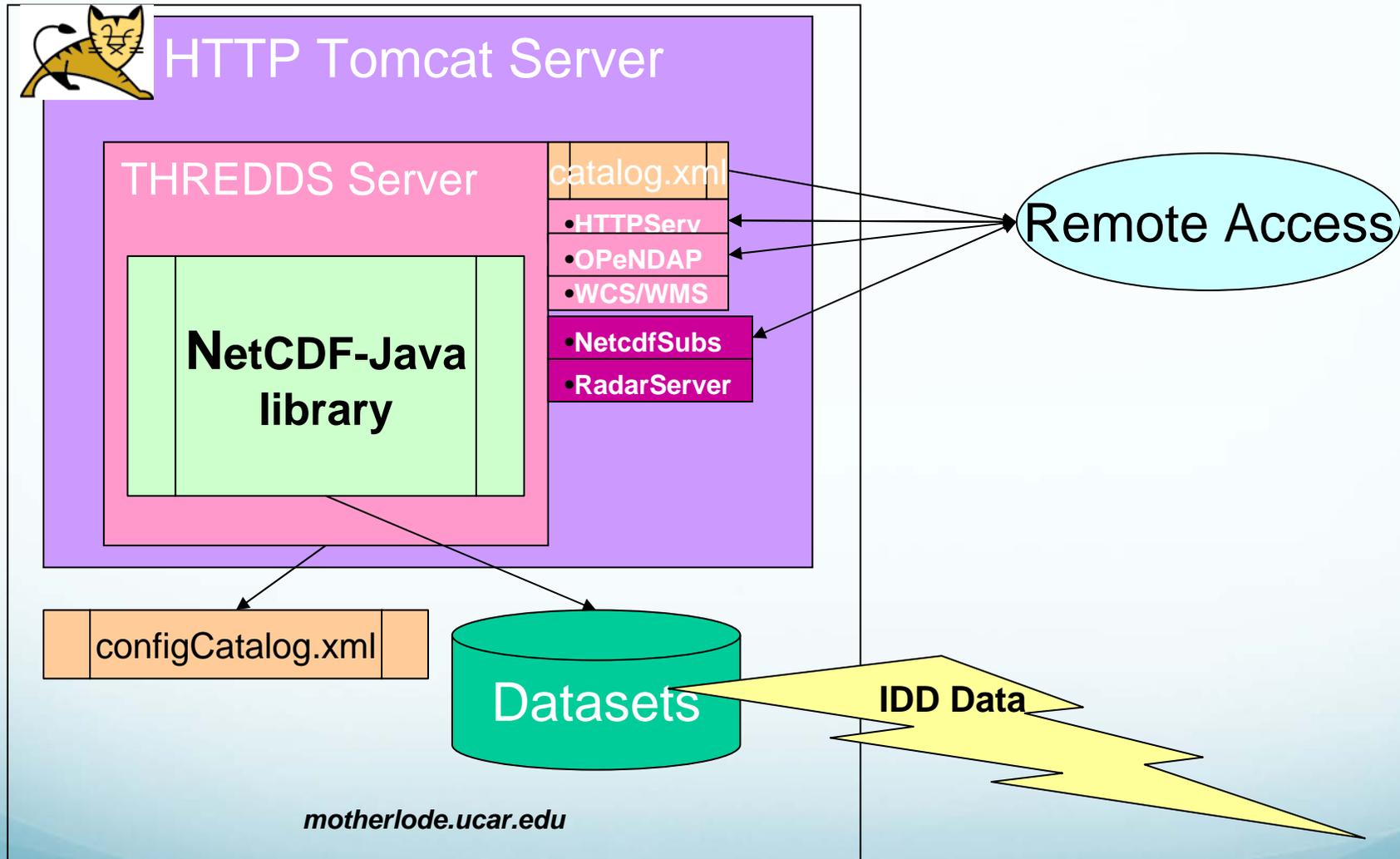
Obiettivo: semplificare l'uso dei dati scientifici

Permettere pubblicazioni scientifiche e materiale educativo

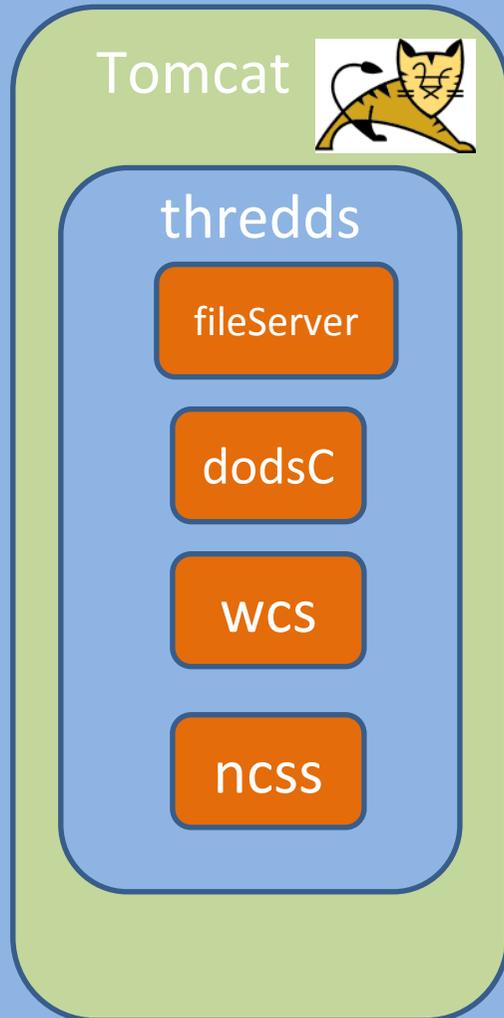


Prerequisito: creare e pubblicare i “catalogs”

TDS



TDS Data Services



Bulk File Transfer

HTTP Server (any file)

Remote access, subsetting CDM files

OPeNDAP (any CDM file)

Web Coverage Server (grids)

NetCDF Subset Service (grids)

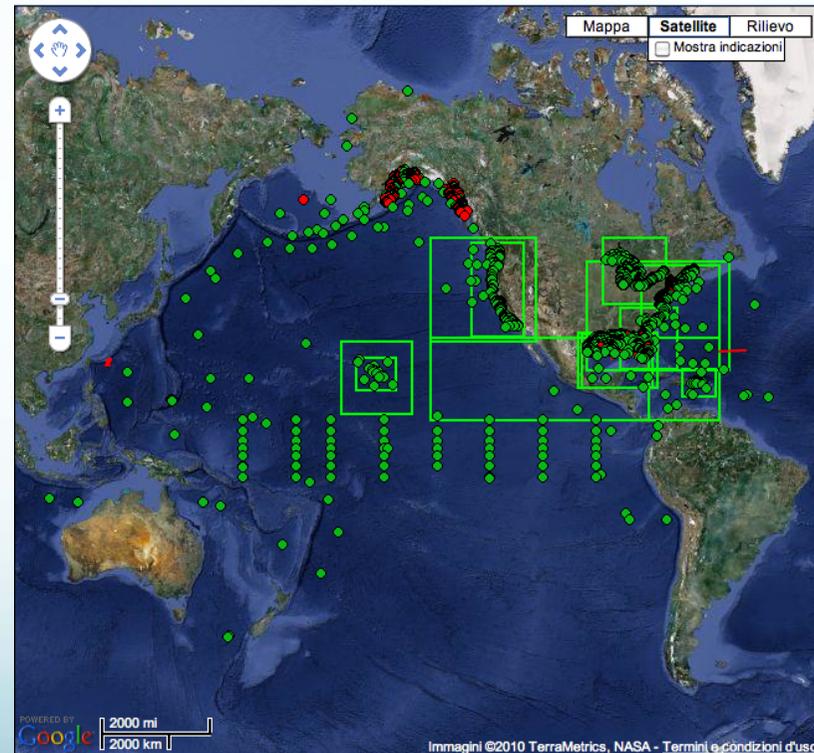
Web Map Server (grids) (soon)

`http://{server:port}/{contextPath}/{service}/...`

`http://motherlode.ucar.edu:8080/thredds/wcs/...`

Il cuore di THREDDS

I **catalogs** sono il cuore di THREDDS, sono files XML che descrivono datasets on line e che contengono metadati arbitrari.



Come funziona un generatore di cataloghi

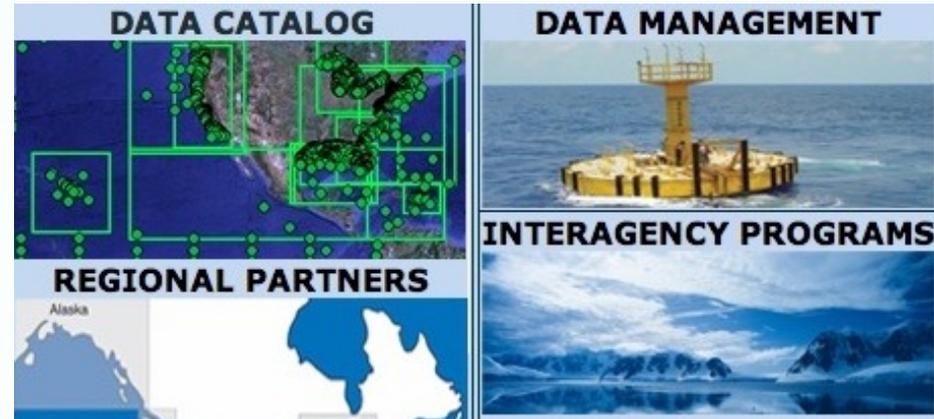
Un generatore di cataloghi produce i THREDDS Catalogs attraverso lo scanning di uno o più collezioni di datasets locali o remoti; tale produzione può essere periodica o “on demand”, usando opportuni files di configurazione che controllano quali directories possono essere analizzate e come i cataloghi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

Catalog http://150.178.42.70:8080/thredds/catalog.html

Dataset	Size	Last Modified
 Local Catalogs		--
 Bevano/		--
 Creus/		--
 Vector/		--
 Arpav/		--
 Remote Catalogs		--
 Bathymetry Catalog from USGS/		--

[back to top level at ISMAR](#)
THREDDS Data Server [Version 4.1.20100428.1437 - 20100428.1437] [Documentation](#)

TDS_1



Per unificare l'accesso ai dati scientifici l'idea di fondo è quella di costruire un “Common Data Model” (CDM).

UNIDATA, www.unidata.ucar.edu, un consorzio di Università americane per il calcolo scientifico nell'ottica Earth System Science propone di creare una API comune ai diversi tipi di dati usati dalla comunità scientifica.

Unidata

*Providing data services, tools, & cyberinfrastructure leadership
that advance Earth system science, enhance educational opportunities, & broaden participation*

GARR 2010

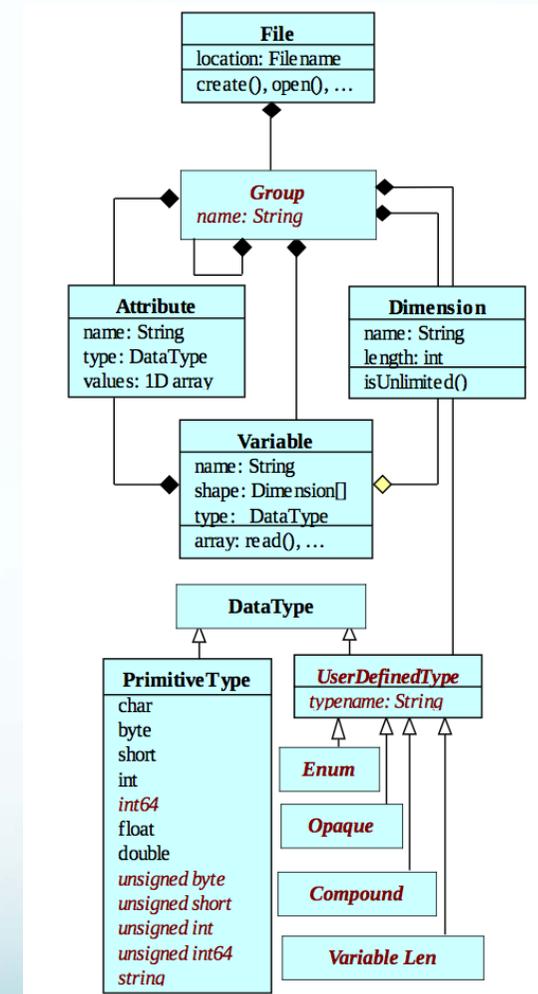
12

TDS_2

- Il THREDDS Data Server (TDS) è un web server che fornisce metadati e accesso al contenuto di datasets scientifici usando OPeNDAP, OGC WMS e WCS, HTTP.
- **TDS** fornisce i contenuti dei datasets oltre ai cataloghi e ai relativi metadati.
- **TDS** è in grado di aggregare più files in datasets virtuali svincolando l'utente dai dettagli di storage e di naming e semplificando in tal modo l'accesso a grandi collezioni di files.

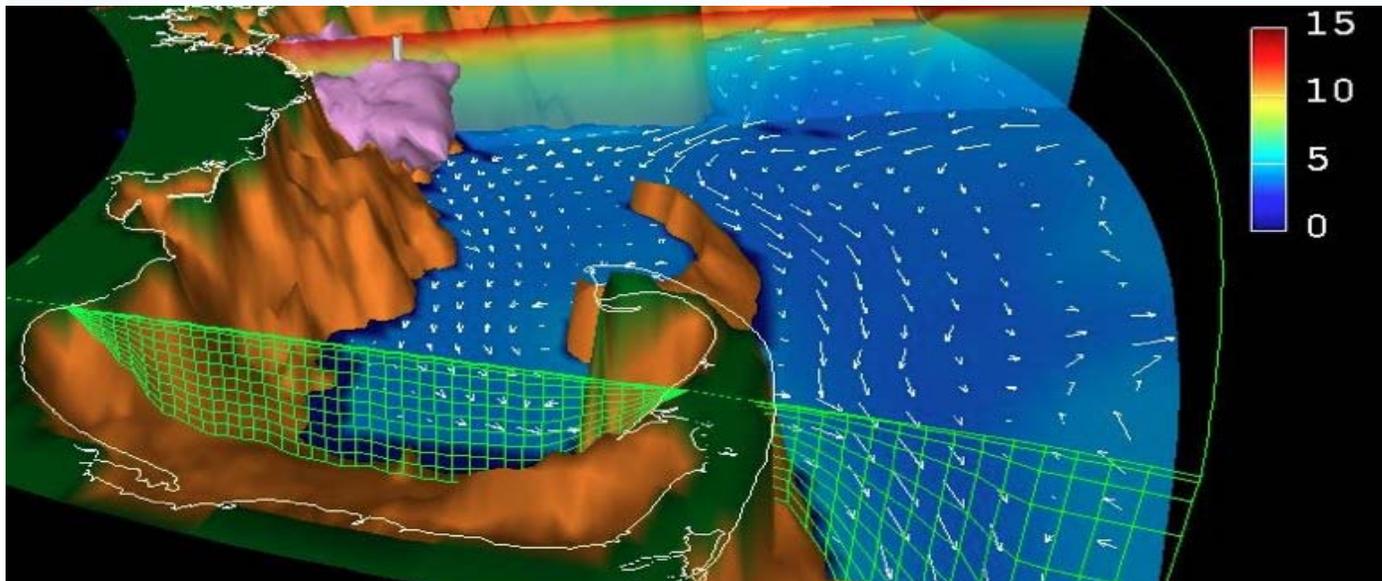
Con che tipo di files opera

- Con i **CDM Datasets**
- La libreria NetCDF-Java legge datasets NetCDF, OpenDAP e HDF5 oltre a qualsiasi formato binario come GRIB, o NEXRAD all'interno di CDM (Common Data Model).



Con chi si integra?

- Thredds integra un server OPENDAP per fornire accesso ai dati via HTTP
- Un ulteriore Server integrato fornisce accesso ai dati attraverso il protocollo [OpenGIS Consortium \(OGC\) Web Coverage Service \(WCS\)](#) per ogni "gridded" dataset fornito di un sistema di informazioni completo per le sue coordinate.



Il motore di THREDDS

- Il THREDDS Data Server è stato totalmente implementato con JAVA ed è contenuto in singolo war file.
- Si installa facilmente in un servlet open source come TOMCAT
- La configurazione avviene attraverso files xml
- La libreria è liberamente scaricabile e il codice sorgente è rilasciato sotto netCDF library



Esempio di implementazione

Una implementazione del sistema proposto è descritta in:

<http://150.178.42.4:8080/thredds/creus.html?dataset=creus/run5/his> ,
e rappresenta uno dei primi tentativi della comunità italiana;

All'indirizzo <http://www.ve.ismar.cnr.it/thredds> si trova 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.



Problematiche aperte

- **Performance** della procedura **non** completamente esplorato quando si manipolano ‘Grandi’ moli di dati
- Sintassi di estrazione non semplice
 - **Subsetting nello spazio delle coordinate**
 - **Subsetting nello spazio dei valori**

Conclusioni



THREDDS:

- consente di gestire un **accesso unificato ai dati** (“aggregation”)
- “spinge” alla produzione di **output numerici standardizzati** (p.e. NetCDF, CF compliant)
- consente la “**customizzazione**” anche in file di grosse dimensioni in ambiente 4D e il loro agile **trasferimento via rete** (p.e. MATLAB)
- favorisce **l’interoperabilità, l’utilizzo e il “quality check”** di risultati numerici spesso differenti per formato, struttura (p.e. long, lat, tempo, coordinate verticali, etc.)
- richiede una adeguata “**Data Policy**” per la gestione, manipolazione e pubblicazione dei dati

Requisiti minimi di sistema:

- [Java](#) 1.6 o superiore.
- [Tomcat](#) 5.5 o superiore.

Riferimenti:

<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>



Grazie per l'attenzione
!

tiziano.minuzzo@ismar.cnr.it