



Gliders Data Management Issues

B. Garau, M. Bonet, S. Ruiz, A. Pascual,, B. Casas, G. Vizoso, A. Alvarez, J. Tintoré

IMEDEA(CSIC-UIB), Institut Mediterrani d'Estudis Avançats, Mallorca, SPAIN

Outline



- Introduction
- European context of Ocean Data Management
- TMOOS Data Management Goal
- TMOOS Activities
 - Platform management and communications
 - Metadata
 - Data Archive
 - Discovery
 - Downloading, distribution and access
 - Viewing
- CODO development methodology

Introduction



Data Management definitions:

- ❑ Data Management is the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of an enterprise (The Data Management Association International, DAMA).
- ❑ Marine or oceanographic data management is the process of entry, quality control, archival and dissemination of research or monitoring data collected in the world's seas and oceans (Jan Seys et al. Ocean Challenge, 2004 Vol. 13, No. 2).

Introduction



Key issues:

- ❑ Diversity of variables and platforms gathering information.
- ❑ Huge dataset (storage and distribution).
- ❑ Many organizations and data providers (Different formats and protocols).
- ❑ Different users and access profiles.

Introduction



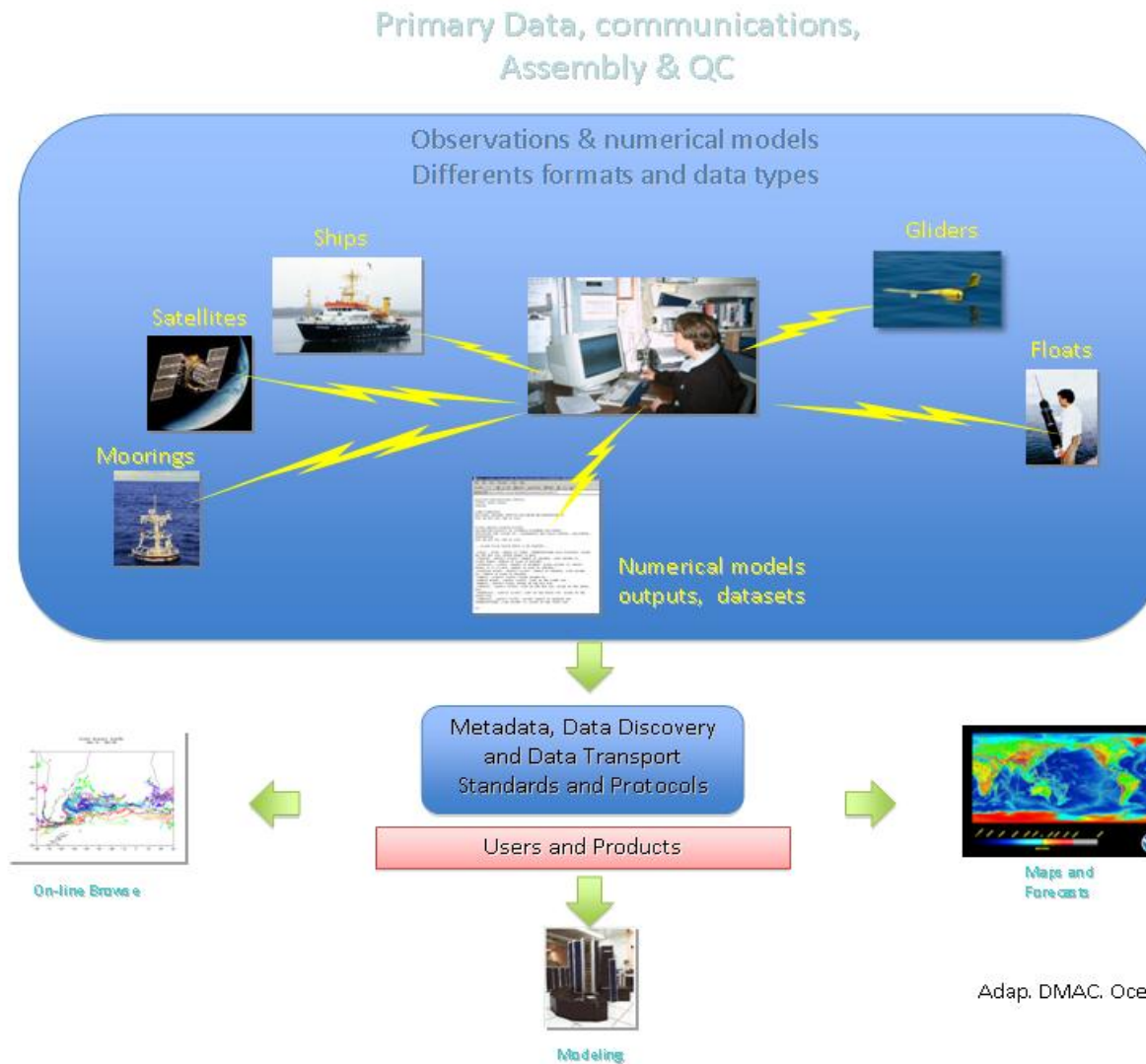
Key issues:

- ☐ Data sources and platforms
 - CTD
 - XBT
 - ADCP
 - Gliders
 - Moorings
 - Drifters
 - Satellites
- ☐ Variables
 - Hydrographic
 - Bio-geo-chemical
 - Atmospheric
- ☐ Quality controls
 - Spikes
 - Gradients
 - Ranges
- ☐ Data organization
 - Time series
 - Profiles
 - Trajectories
 - Gridded
- ☐ Data formats
 - Particular ASCII
 - Particular binary
 - Matlab
 - Database models
 - NetCDF
- ☐ End users
 - Authorities
 - Maritime security
 - Scientist
 - General public

Introduction



Data management & communication system



Adap. DMAC. Ocean.us

European Context of Ocean Data Management



Marine Environment and Security for the European Area: to provide an integrated service of global and regional ocean monitoring and forecasting to intermediate users and policy makers in support of safe and efficient offshore activities, environmental management, security, and sustainable use of marine resources.



SESAME - Southern European Seas: Assessing and Modeling Environmental Changes.



European Coastal sea Operational observing and Forecasting system: to build up a sustainable pan-European capacity in providing timely, quality assured marine service (including data, information products, knowledge and scientific advices) in European coastal-shelf seas.



MyOcean: an European Commission "GMES"(Global Monitoring for Environment and Security) project. To deliver regular and systematic reference information on the state of the oceans and to provide the common denominator data for all users in the marine sector, in other words the information for existing & new downstream services.

TMOOS Data Management Goal



TMOOS (Marine Technologies, Operational Oceanography and Sustainability) **goal** is to provide a **system** that allows the users to **locate and download the data** of interest in order to **visualize, analyze** and manage the information.

❑ Specific Tasks:

- Platform management and communications
- Metadata: wide descriptions of data sets (standards compliant)
- Data Archive: ability to securely archive data and metadata and retrieve them on demand

❑ Services:

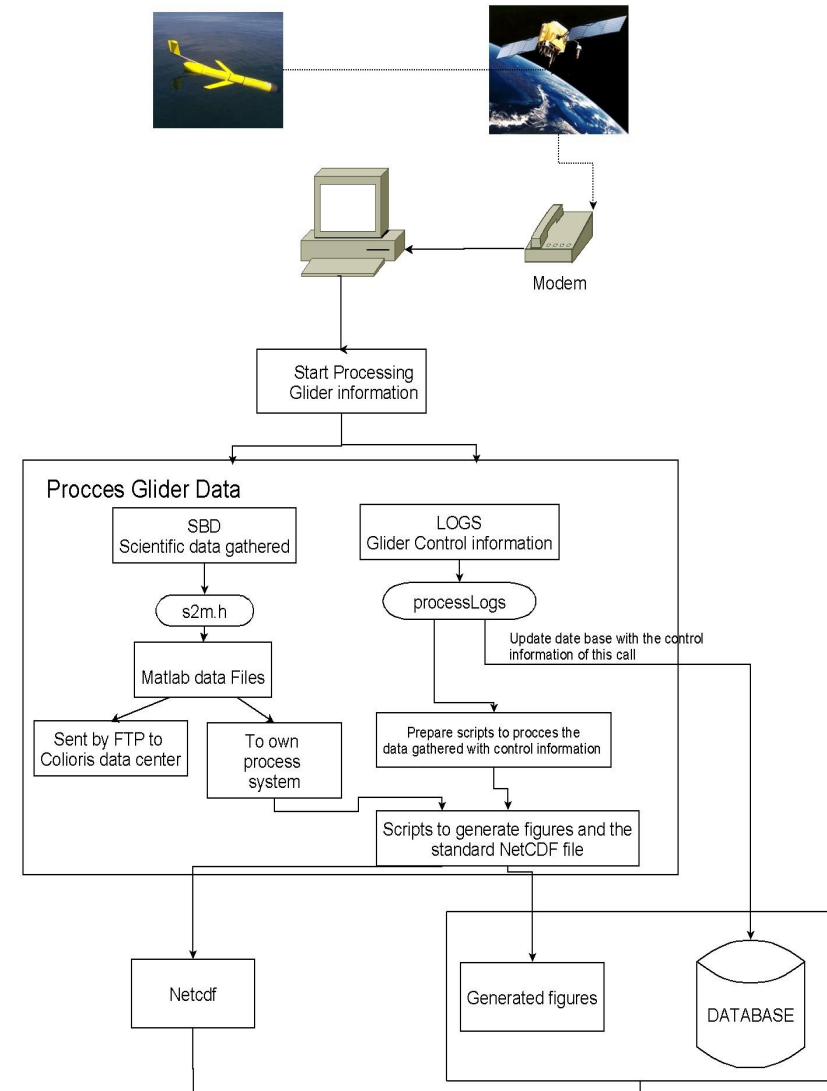
- Discovery: the ability to search for and find data sets of interest
- Downloading, distribution and access: the ability to (easy) access the data in an interoperable manner from client applications
- Viewing: capability to provide an integrated viewing service (Web Browsers)

TMOOS Activities: Platform management and communications



Glider data processing:

- ❑ Data conversion chain: .sbd (binary) to .m&.dat (ASCII) to work in Matlab
- ❑ Trajectory estimation
- ❑ Distance computation
- ❑ Inflections removal
- ❑ Profiles identification
- ❑ Thermal Lag correction
- ❑ Quality control (under development)
- ❑ Regular depths interpolation
- ❑ Mean profiles generation
- ❑ Drift currents estimation
- ❑ Output to:
 - ASCII files in Saclant format
 - NetCDF
- ❑ Several plots (for web):
 - Vertical sections of hydrographic variables
 - TS Diagrams
 - Mean profiles
 - Trajectory + currents



TMOOS Activities: Platform management and communications



Glider data processing: Thermal Lag issue on CTD data

- Situation:
 - Temperature sensor measures sea water temperature **outside** of the conductivity cell.
 - Conductivity sensor measures sea water conductivity **inside** of the conductivity cell.
 - Conductivity cell **walls store heat**, warming up in-cell water when moving to cooler waters, and cooling it down when moving to warmer waters.
 - Conductivity is affected by temperature.
- Result:
 - Salinity is computed from out-cell temperature and in-cell conductivity.
 - Salinity present important errors on strong temperature gradients (thermocline).

TMOOS Activities: Platform management and communications



Glider data processing: Thermal Lag issue on CTD data

- Previous works have studied this problem and proposed techniques to mitigate it:
 - *Thermal Inertia of Conductivity Cells: Observations with a Sea-Bird Cell* (Lueck & Picklo, 1990).
 - *The Correction for Thermal-Lag Effects in Sea-Bird CTD Data* (Morison et al., 1994).
 - And many more references...
- (Slocum) gliders arise new problems that make difficult to apply traditional techniques without modifications:
 - CTD is not pumped. Therefore, flow speed depends on glider surge speed.
 - CTD sampling has low temporal resolution (~ 0.5 Hertz).
 - CTD sampling interval is not perfectly regular.

TMOOS Activities: Platform management and communications



Glider data processing: Thermal Lag issue on CTD data

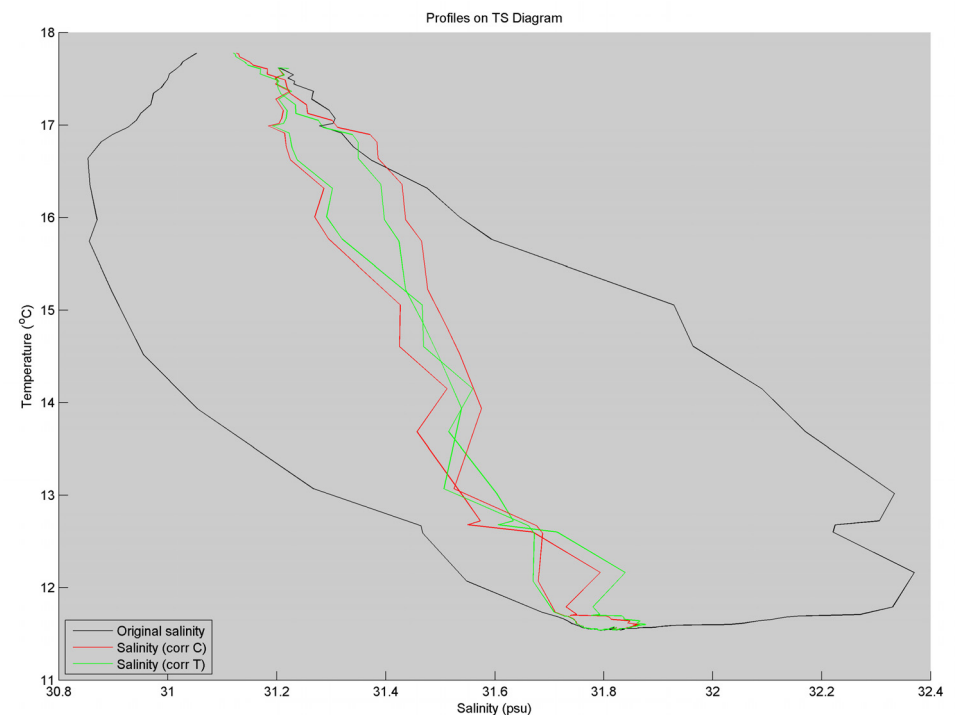
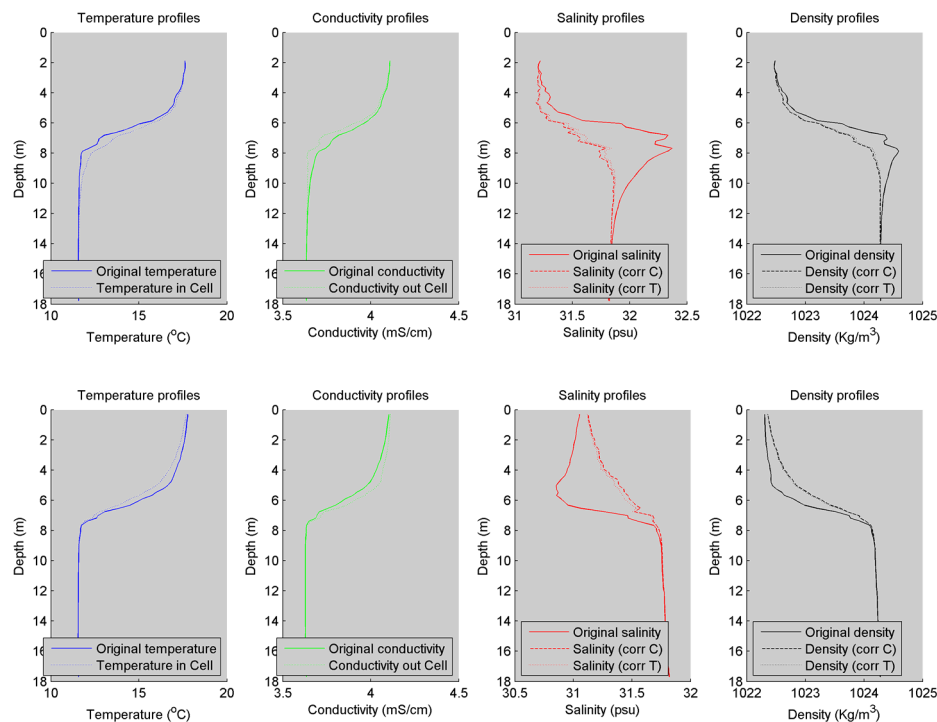
- ❑ Some generalizations to correction techniques and their parameters estimation methods are under development to improve glider CTD data quality.
- ❑ Time “constant” and magnitude are no longer constants dependent on CTD sensor, but also related to glider speed and attitude.
- ❑ Estimation of their relation is performed through a minimization process:
 - Minimum area between two consecutive glider profiles in the TS diagram (assuming same water).
 - Minimum area between glider profile and CTD cast (ground truth).

TMOOS Activities: Platform management and communications



Glider data processing: Thermal Lag issue on CTD data

- ❑ Glider downcast vs. glider upcast, promising results:
 - Salinity spikes reduced.
 - TS hysteresis reduced.

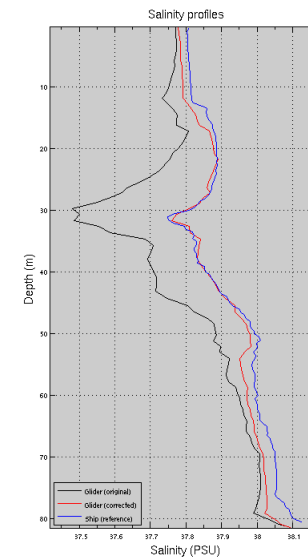
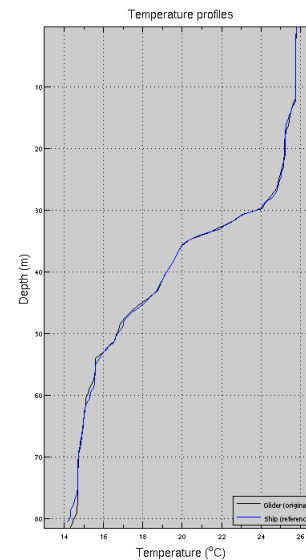
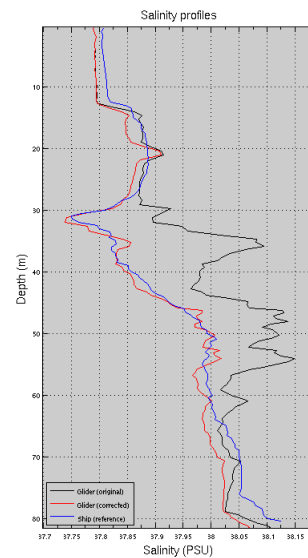
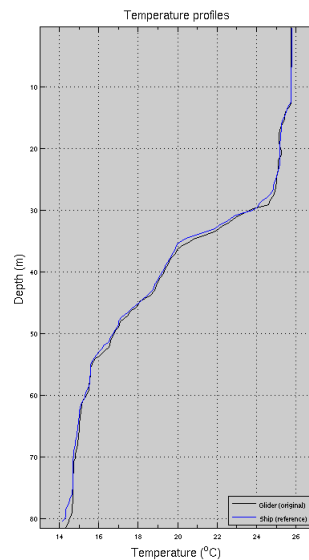


TMOOS Activities: Platform management and communications



Glider data processing: Thermal Lag issue on CTD data

- ❑ Glider vs. CTD cast, promising results:
 - Salinity spikes reduced.
 - Better profile fitting.





Metadata:

- ☐ Additional information about the data that makes it more meaningful.
- ☐ Answers questions like What, Who, Where, When, Why, How.
- ☐ The better the metadata is, the more benefits we can get from data.
- ☐ Some initiatives:
 - INSPIRE Directive (2007/2/EC)
 - ISO 19115
 - CDI Common Data Index
 - And many more...

TMOOS Activities: Metadata



Glider metadata handling:

- During the mission preparation, the PI fills in a form describing the mission name, goal, deployment day, waypoints, glider behavior, etc. This information is used to automatically generate the glider file metadata for this deployment.

Detalles de la campaña: SESAME 2008

1 - Parámetros de campaña

Parámetros de campaña	
Título de la misión que se utilizará en la parte visual de la web:	
SESAME 2008	
Descripción de la campaña:	
In the framework of the EU funded SESAME project, the specific objectives of this mission is: • To test the feasibility of the gliders technology usage in an area with intense current and high mesoscale variability • To study different adaptive sampling strategies for gliders in areas of high ocean variability based on optimization of new algorithms.	
Fecha de inicio de la misión:	2008-07-02 (aaaa-mm-dd hh:mm:ss)
Nombre y apellidos del investigador responsable	Simón Ruiz
Proyecto Asociado	SESAME
Reservas:	Hidrografía: densidad mínima de la zona de misión.
☒ Embarcación y Patrón	24.5 Temperatura
☒ Vehículo de transporte	36.5 Salinidad
Actualizar información	

2 - Definición de trayectoria

Definición de trayectoria					
latitud	longitud		corriente		
grados	minutos	grados	minutos	grados	metros/seg
0	37	28.68	-0	0	0 m/s
1	36	35.1	0	19.98	0 m/s
2	37	28.68	-0	19.98	0 m/s
3	37	28.463	-0	19.98	0 m/s
4	37	24.581	-1	12.124	0 m/s
5	37	28.463	-0	12.124	0 m/s

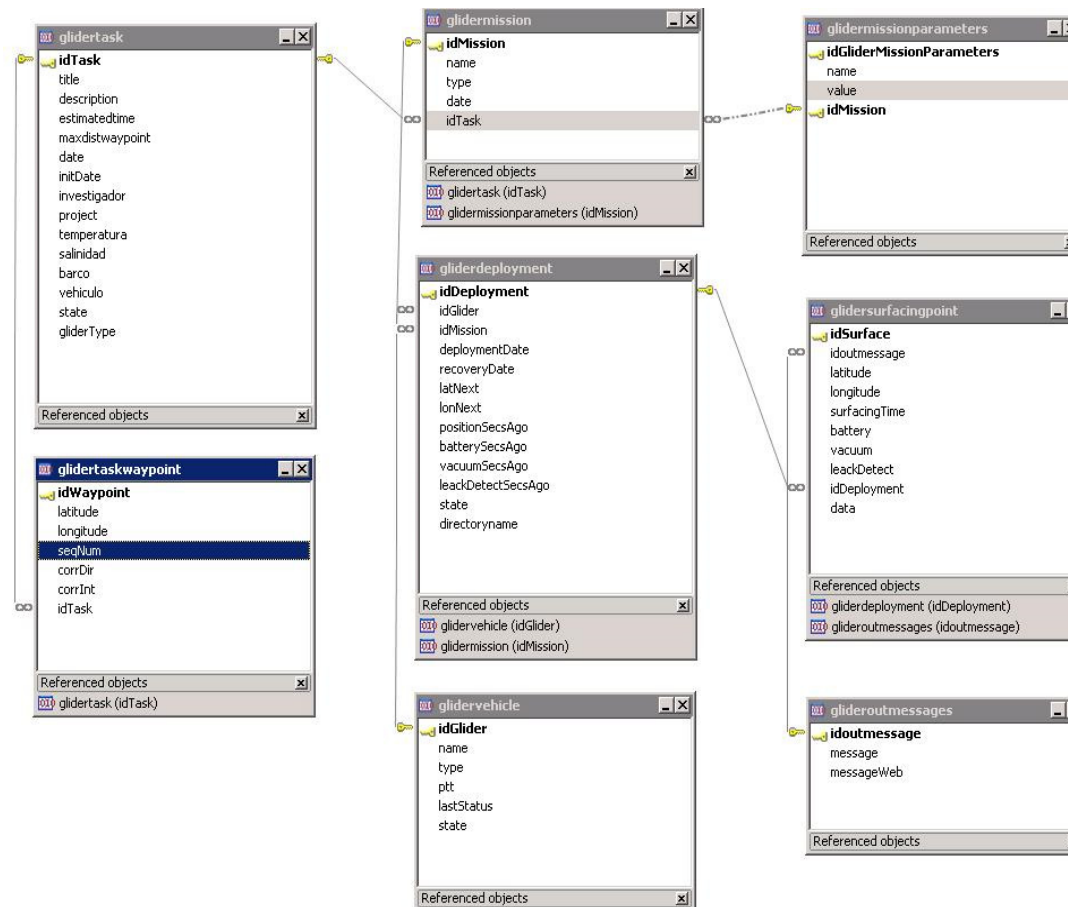
Calcular tiempo de misión
Ver sobre mapa
 Borrar mapa
 + otro Waypoint
 Quitar último
Tiempo estimado de navegación: 10d 20h 48m 23s
Distancia máxima de segmento: 115485.859331

TMOOS Activities: Metadata



Glider metadata handling:

- A relational database is used to query glider deployments information on the glider website



TMOOS Activities: Data Archive



Data Archive:

- ☐ Data organization:
 - Directory organization.
 - Naming and conventions.
 - Relational Data Base
- ☐ Data format:
 - in TMOOS, NetCDF (following ECOOP directives).
- ☐ Data security, backup systems,...



NetCDF format:

NetCDF (network Common Data Form): Set of software libraries and machine-independent data formats that support creation, access, and sharing of array-oriented scientific data.

NetCDF content (ver. 3):

- ☐ Set of dimensions (time, latitude, longitude...): just sizes of matrices.
- ☐ Set of global attributes (title, provider, ...). Global file metadata.
- ☐ Set of variables (temperature, salinity, ...): Actually the dataset.
 - variable type: int, double, char, ...
 - variable dimensions: relation with dimensions mentioned before.
 - variable attributes (units, standard name, description, ...)

NetCDF, ver. 4, many changes expected on file structure.



Standard conventions:

- ❑ NetCDF format defines just the file structure, not its content.
- ❑ Different existing conventions (CF-1.0, GDT-1.3, COARDS, ...)
- ❑ Relation between conventions (e.g. CF extends COARDS).
- ❑ Product dependent (ARGO profilers conventions, OceanSites glider NetCDF convention, ...)
- ❑ Convention establish:
 - Global attributes (who, what, where, when, why, how)
 - Variable attributes (e.g. units)

TMOOS Activities: Data Archive



Metadata integration: NetCDF metadata is extracted from relational DB

```
❑ :title = "SESAME 2008";  
❑ :type = "Glider data file";  
❑ :conventions = "CF-1.0";  
❑ :netcdf_version = "3.6.1";  
❑ :creation_date = "2008-07-01 21:02:18";  
❑ :references = "http://www.imedeia.uib-csic.es/tmoos/gliders/";  
❑ :provider = "CODO (Centro Operacional de Datos Oceanograficos)";  
❑ :history = "Glider data file definition, Version 0.1, MAR 28, 2008";  
❑ :summary = "In the framework of the EU funded SESAME project, the specific objectives of this mission is (...)";  
❑ :cdm_data_type = "Time-Series";  
❑ :data_mode = "R";  
❑ :project_name = "SESAME";  
❑ :pi_name = "Simón Ruiz";  
❑ :platform_code = "maya";  
❑ :ptt = "62605";  
❑ :trans_system = "IRIDIUM";  
❑ :positioning_system = "GPS";  
❑ :platform_model = "WEBB SLOCUM";  
❑ :platform_maker = "Webb Research Corporation";  
❑ :anomaly = "none";  
❑ :launch_date = "2008-07-02 10:00:00";  
❑ :launch_latitude = "36.583";  
❑ :launch_longitude = "0.3166";  
❑ (.....)
```

TMOOS Activities: Discovery



Discovery: Thredds catalog, virtual data organization

- ❑ The THREDDS (Thematic Realtime Environmental Distributed Data Services) project is developing middleware to bridge the gap between data providers and data users.
- ❑ The goal is to simplify the discovery and use of scientific data and to allow scientific publications and educational materials to reference scientific data.

Catalog <http://130.206.32.248:8080/thredds/catalog/catalog.xml>

Dataset



Observational/



Models/

THREDDS Data Server Version 3.16.36 Build Date = 2008-05-05 22:17:07 [Documentation](#)

TMOOS Activities: Discovery



TMOOS Thredds catalog organization:

□ Observational

- Gliders: grouped by year, and one NetCDF file per deployment. File name specifies deployment date and glider name.
- Moorings: grouped by mooring platform, then by year, and finally one file per month. File name specifies year, month and platform name.
- Satellite: grouped by product (SST, Altimetry, Ocean Color). Classified as real time image or analyzed product. Then grouped by year and month. File name specifies image production date.

□ Models

- HOPS
- ROMS
- Input Data:
 - MFS: grouped by year, month and day. File name specifies forecast day and contained variables.
 - HIRLAM: divided in high and medium resolution model outputs. Grouped by year, month and day. File name specifies forecast time.
 - ESEOO: Grouped by year, month and day. File name specifies forecast time and contained variables.
 - Coastlines.
 - Bathymetries: different resolutions and domains.
 - Climatologies.

TMOOS Activities: Downloading, distribution and access



Data Distribution:

Once data discovered in THREDDS catalog, it has to be accessed.

- ☐ HTTP (HyperText Transfer Protocol)
- ☐ FTP (File Transfer Protocol)
- ☐ OPENDAP (Open-source Project for a Network Data Access Protocol) the protocol handles:
 - Transport
 - Translation
 - Subsetting of data
 - Work data remotely as if it was local
 - Access from different programming languages as Matlab, R-Project, IDL, Java, python,...

TMOOS Activities: Downloading, distribution and access



Opendap access

- ❑ Web form allows users to check metadata and retrieve data.

OPeNDAP Dataset Access Form

Tested on Netscape 4.61 and Internet Explorer 5.00.

Action: [Get ASCII](#) [Get Binary](#) [Show Help](#)

Data URL:

Global Attributes:

<pre>title: "SESAFE 2008" type: "Glider data file" conventions: "CF-1.0" netcdf_version: "3.6.1" creation_date: "2008-07-01 21:02:18" references: "http://www.imedea.uib-csic.es/tmoos/gliders/"</pre>
--

Variables:

☐ **profile_index:** Array of 32 bit Integers [datum = 0.34392]

<pre>long_name: "profile index" standard_name: "profile_index" units: "ordinal, adimensional"</pre>

☐ **time:** Array of 32 bit Reals [datum = 0.34392]

<pre>long_name: "datum time" standard_name: "time" units: "seconds since 1970-01-01 00:00:00"</pre>

☐ **latitude:** Array of 32 bit Reals [datum = 0.34392]

<pre>long_name: "latitude" standard_name: "latitude" units: "decimal degrees north"</pre>

1 Global attribute

2 Variable access

Made by DrExplan.com

TMOOS Activities: Viewing



- ❑ GeoRSS: Rich Site Summary with geographical information for last data arrival notification.
- ❑ Accessible from different RSS readers.

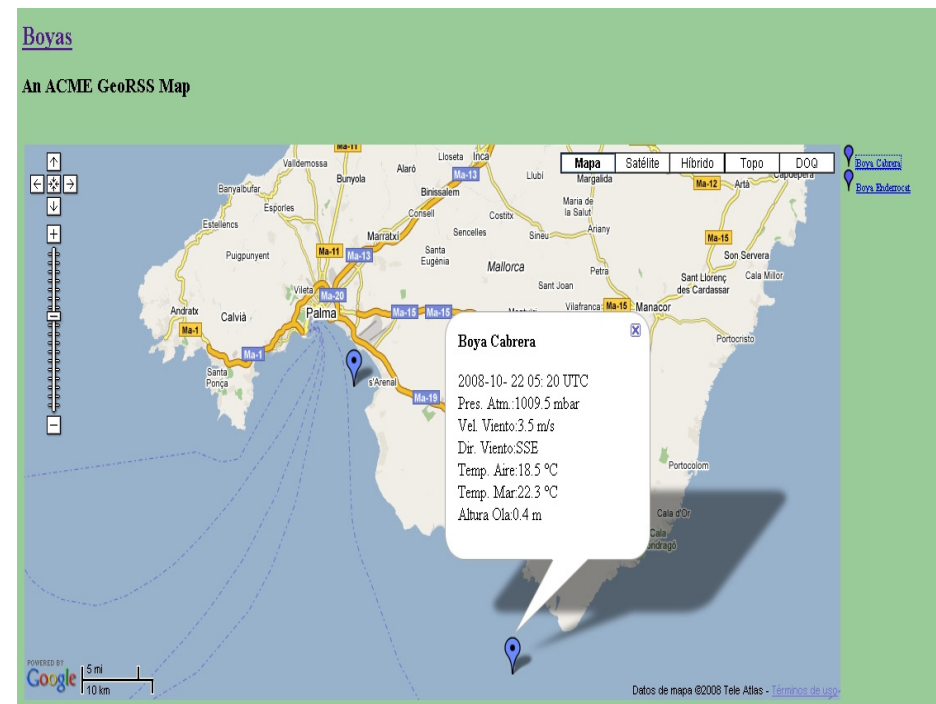
RSS Google reader access



Mobile phone access



GeoRSS Web reader



TMOOS Activities: Viewing



- Observing platforms websites are accessible from main department website:

TMOOS
Tecnología Marina, Oceanografía Operacional y Sostenibilidad

[home](#) || [Mapa Operacional](#) || [proyectos](#) || [publicaciones](#) || [personal](#) || [noticias](#) || [administración](#) ||

Models
HOPS

Observational
Moorings
Gliders
Teledetection

Products
Data Access (Thredds)
Department Wiki
Bug tracking system


Gliders


Numerical models


Teledetection


Moorings

Proyectos
ECOOP-European Coastal Observation Platform

Toma de Datos
Boya Enderrocat
Temp Aire: 0.0oC
Glider MAYA
Temp Aire: 0.0oC
ARGO Float IMEDEA
Temp Aire: 0.0oC
Ver Mapa Oper.

Noticias

Web presentation
Next wenesday will be presented the first versión of department website.

New
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Aliquam diam. Duis vestibulum.

IMEDEA - TMOOS | [Contacto](#)

TMOOS Activities: Viewing



- ❑ Glider website shows real time tracking of glider trajectory and current status estimations:

TMOOS Gliders

Tecnología Marina, Oceanografía Operacional y Sostenibilidad

beta

Panel de control || Misiones || Información || MultiMedia || Enlaces || Documentación

UTC 7:56:06
15/07/2008

Gliders

En Misión
MAYA - seguimiento

En preparación
willy

Info. soporte

Predicción
Modelo de corrientes. IMEDEA
Viento y oleaje. AEMET - PE

Observación
Teledetección. IMEDEA
Red de Puertos del Estado
Imágenes satélite. AEMET

Leyenda

- Última llamada
- Estimación glider
- Objetivo
- Waypoint
- Waypoint conseguido
- Llamada
- Próxima llamada

MAYA en misión SESAME - seguimiento

	Latitud	Longitud	Distancia (km)	Fecha UTC	Tiempos
	36° 59.686'	0° 10.436'	-	2008-07-15 06:47:22	-
	37° 0.4299'	0° 9.7857'	1.6814	2008-07-15 07:56:06	última llamada: 0d 01h 08m 44s
	37° 3.9301'	0° 6.7228'	9.594	2008-07-15 13:17:22	próxima llamada: 0d 05h 21m 15s
	37° 27.94'	-0° 14.4473'	63.952	2008-07-17 02:07:02	próximo WayPoint: 1d 16h 10m 56s

Mapa Relieve

50 mi 100 km

Rango batimetría: 0 -20 -50 -100 -200 -500 -1000

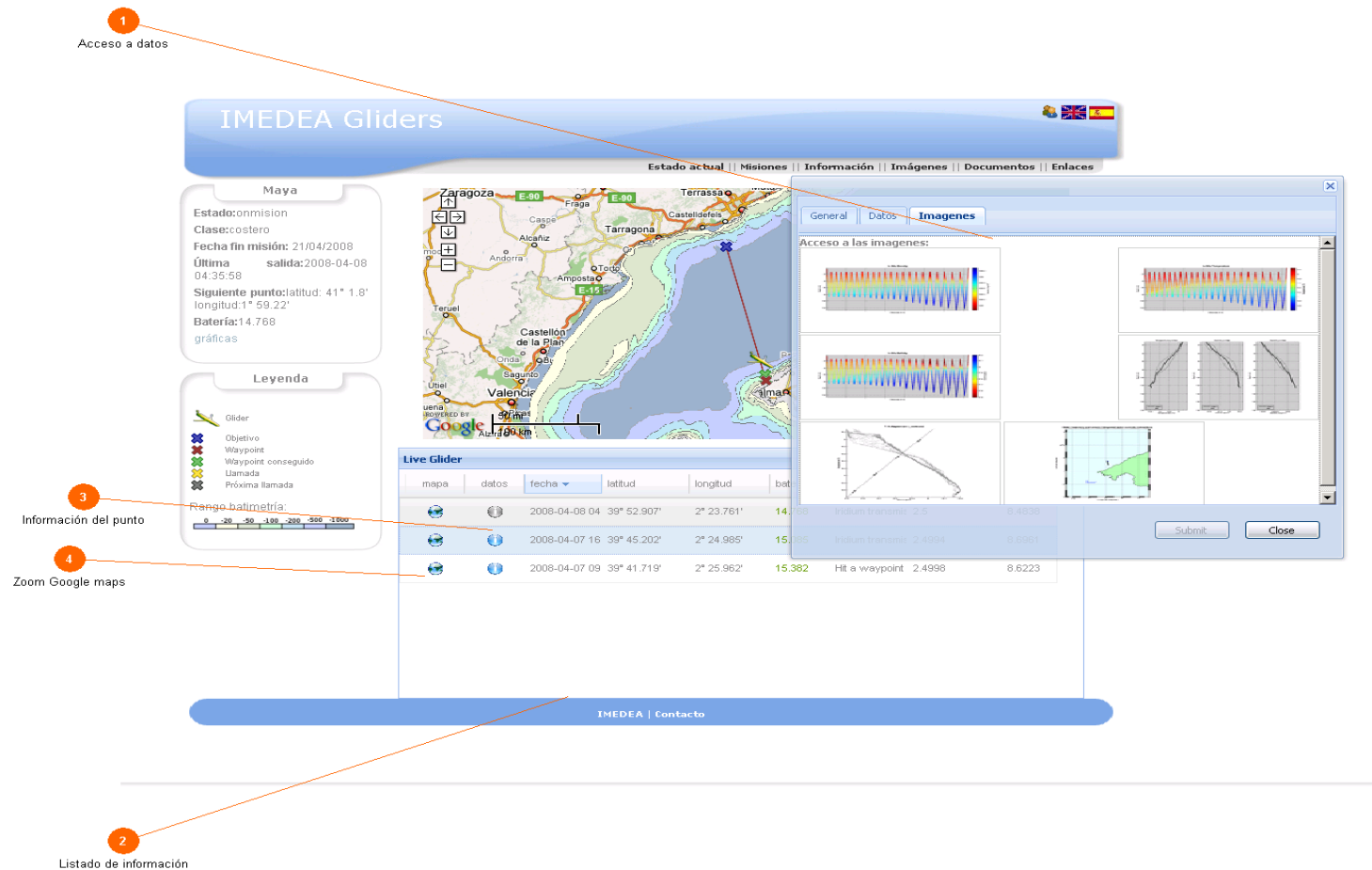
Centrar Mapa :: Batimetría :: Refresco mapa :: Distancias

Posición del mouse
latitud: 37° 4.3626'
longitud: -3° 1.9335'

TMOOS Activities: Viewing



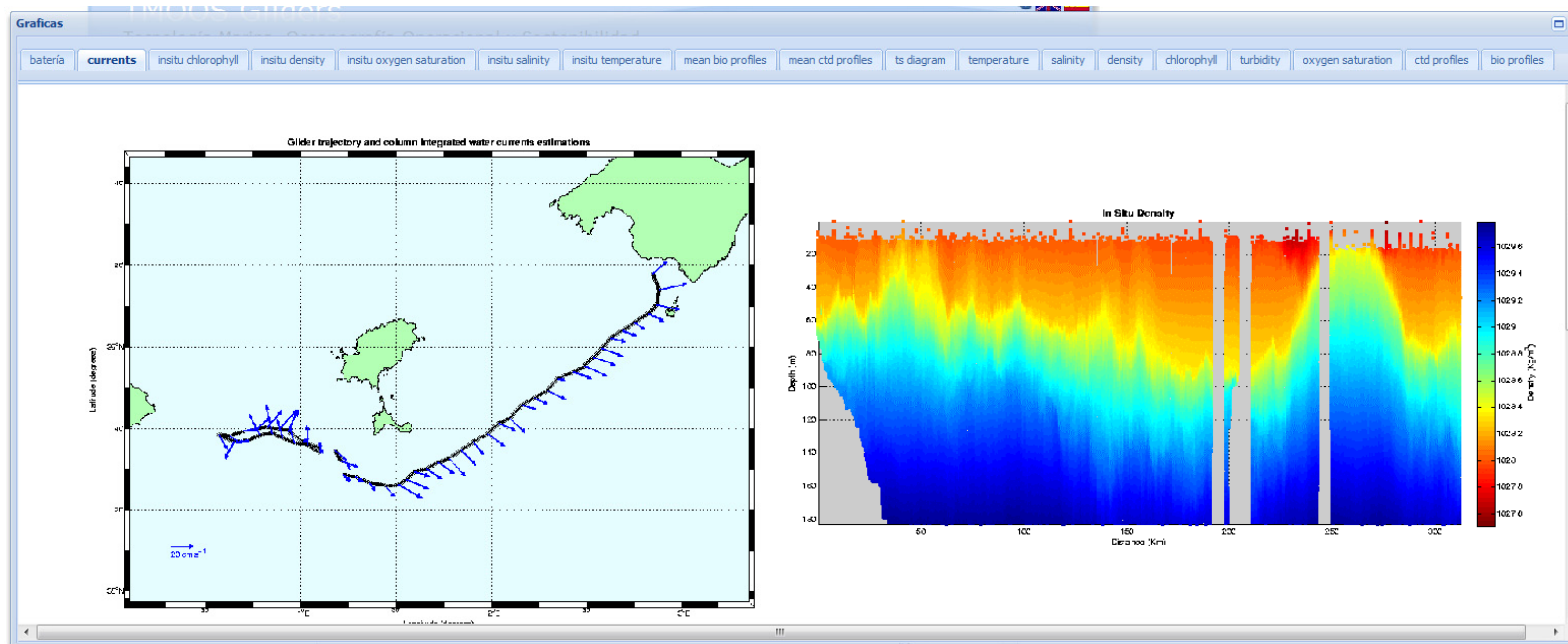
Real time data monitoring: data access and plot viewing



TMOOS Activities: Viewing



- Delayed mode data: historical access and visualization for finished deployments



CODO Development Methodology



Use of standard methodologies for software development and work organization:

- ☐ Development control version (SVN).
- ☐ Software documentation.
- ☐ Bug Tracking system for software improvement and interaction with TMOOS end users (BugZilla).
- ☐ Tasks and deadlines management (Project).
- ☐ Centralized access to documentation (WIKI).



What is a Wiki? How do we use it?

- ❑ A Web based system to centralize and organize internal documentation for TMOOS users.
- ❑ Easy edition of web pages (following some notation) to report users' activity and help other people working on same issues.
- ❑ Incorporates TMOOS Calendar (event notifications inserted by TMOOS users).
- ❑ Sidebar presents different sections attending to their content: modeling, satellite, glider, instrumentation



What is a Wiki? How do we use it?

- ❑ Wiki also contains our own documentation and links to external documental resources on how the team develops their tools and products:
- ❑ Website: miscellaneous information, directory structures, involved tools and technologies, web multi-language support.
- ❑ Processing applications (e.g. Matlab, Ferret, NetCDF, IDV, ODV...): installation and user guide, useful tricks, quick start...
- ❑ Provided services (e.g. thredds, opendap): location, usage...

CODO Development Methodology: Product List



CODO Product List:

- ☐ Web site development: TMOOS department, Glider, Mooring, Models and Satellite.
- ☐ Thredds data catalog.
- ☐ Department Wiki.
- ☐ GeoRSS last data access.
- ☐ Google map visualization.
- ☐ Standard data access for its usage with viewing tools as ODV (ocean data viewer), IDV (integrated data viewer), ncBrowse,...

CODO Development Methodology: Future Work



Future work to be done:

- ☐ Thredds-GIS integration for data analysis (WCS Web Coverage Services, WMS Web Map Services).
- ☐ Centralize TMOOS metadata management.
- ☐ Semantic data searching tools.
- ☐ Adapt viewing tools to our own data.
- ☐ If EGO groups are interested:
 - Prepare a Wiki with glider information to share among glider working groups.



Gliders Data Management Issues

B. Garau, M. Bonet, S. Ruiz, A. Pascual,, B. Casas, G. Vizoso, A. Alvarez, J. Tintoré

IMEDEA(CSIC-UIB), Institut Mediterrani d'Estudis Avançats, Mallorca, SPAIN