



FROM KNOWLEDGE  
GENERATION  
TO SCIENCE-BASED  
INNOVATION



# Data discovery mechanisms and metadata handling in RAIA Coastal Observatory

Artur Rocha (1), Marco Amaro Oliveira (1), Filipe Freire (1), Gabriel David (1), Pedro Monteiro Vilar (2), Begoña Vila Taboada (2), Isabel Iglesias (3), Clara Lázaro (3), Luísa Bastos (3), Ilmer van Golde (4), Jorge Silva (4)

(1) Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC), (2) Instituto Tecnológico para o Control do Medio Mariño de Galicia (INTECMAR), (3) Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), (4) Instituto Hidrográfico

**Speaker: Gabriel David, INESC TEC**

**4as Jornadas de Engenharia Hidrográfica, 21-23 June 2016, Lisbon, Portugal**

Research and Technological Development | Technology Transfer and Valorisation | Advanced Training | Consulting  
Pre-incubation of Technology-based Companies



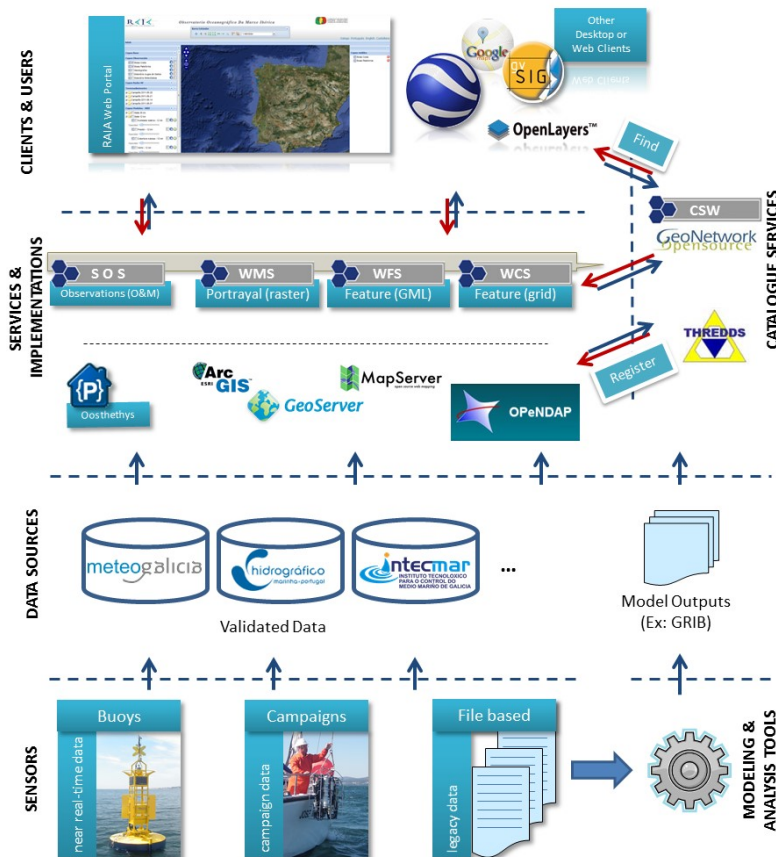
COOPERAÇÃO TRANSFRONTEIRIZA  
ESPAÑA - PORTUGAL  
COOPERAÇÃO TRANSFRONTEIRIZA

# Introduction

- Context: RAIA Coastal Observatory
- Premise #1: Maritime Data and Service Infrastructure, composed of several geographically distributed nodes, which can simultaneously act as data/service providers and consumers.
- Corollary #1: Federated distributed architecture – no central node – each node has the necessary capacity to act as such and remotely invoke services from others to provide value-added services;
- Corollary #2: Adherence to INSPIRE directive, providing compliant remote services;
- Goal: to implement truly interoperable and harmonized services by 2020. How?

# Background

## Information Flow and layers



Several OGC/INSPIRE compliant services in place from different partners in RAIA:

- WMS: map service for portrayal (raster);
- WFS: features of interest, such as beaches, places for barnacle fishing (attributes and geometry in GML/KML);
- WCS: coverages, maps under the form of grid (forecasts, other multi-dimensional spatiotemporal products);
- SOS: time series API for sensor data.

# Challenges

- Describe data and services with consistent metadata;
- Minimize the effort in including metadata in datasets;
- Increase the efficacy of searches, maximizing the chance of the user finding what he is looking for.

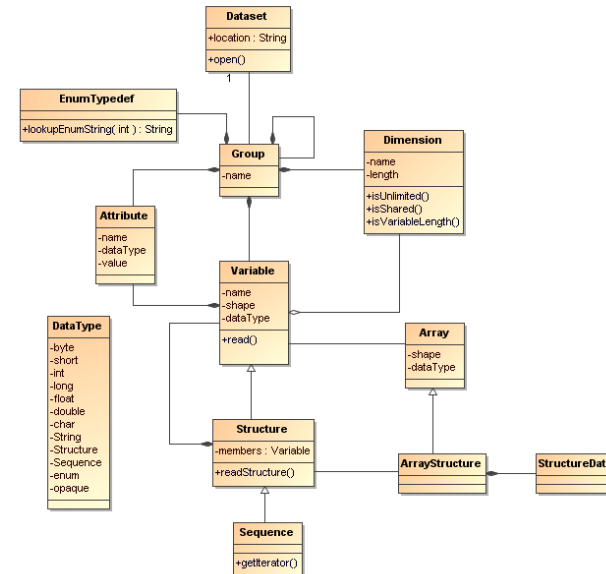
# Methods and Implementation

- Service level:
  - Be more thorough at filling “service” metadata;
  - Make use of INSPIRE Code Lists (controlled vocabularies);
- Dataset level:
  - Choose encodings (and the implementations that support them) that ensure the use of a Common Data Model (CDM). Examples:
    - UNIDATA Common Data Model, used in THREDDS
    - INSPIRE Observations & Measurements, used by OGC SOS
  - Adopt common vocabularies to name variables and unit in datasets, such as the Climate and Forecast (CF). Use other domain-specific controlled vocabularies where CF is not suitable.
  - Annotate at the dataset level only the metadata that is different from the service level one.

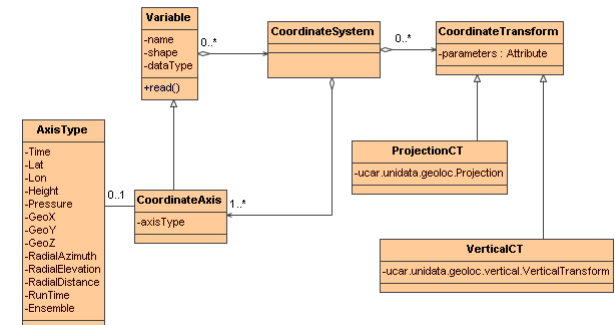
# Implementation Example 1: THREDDS Data Servers

- Used for voluminous file-based data sets (ex: forecasts);
- Uses underlying UNIDATA CDM with intrinsic geospatial traits (CRS);
- Uses CF standard names, which also contain clear rules on how to derive new variable names (1)
- Convention available at: <http://cfconventions.org/standard-names.html>

Data Access Layer Object Model



Coordinate System Object Model



(1) [surface] [component] standard\_name [at surface] [in medium] [due to process] [assuming condition]

# Implementation Example 1: THREDDS Data Servers

- Partners agreed to use at least CF on version 1.4 (some using CF 1.6)
- THREDDS Data Server implements several standard APIs to access data (WMS, WFS, WCS, etc)
- Metadata accessible according to ISO19115-2
- Simple interface to check completeness of the metadata

## OPeNDAP Dataset Access Form

Tested on Netscape 4.61 and Internet Explorer 5.00.

Action:

Data URL:

Global Attributes: **Conventions: CF-1.6, ACDD-1.3**  
 standard\_name\_vocabulary: CF Standard Name Table v29  
 creator\_name: Instituto Hidrografico  
 creator\_url: http://www.hidrografico.pt  
 institution: Instituto Hidrografico, Marinha - Portugal,

Variables: ☐ time: Array of 64 bit Reals [time = 0..24]

time:  
 long\_name: time  
 standard\_name: time  
 units: hours since 2016-06-22T00:00:00  
 calendar: gregorian

Extent Search Score: 6/8

This basic extent information supports spatial/temporal searches that are increasingly important as the number of map based search interfaces increases. Many of the attributes included in this spiral can be calculated from the data if the file is compliant with the NetCDF Climate and Forecast (CF) Metadata Conventions.

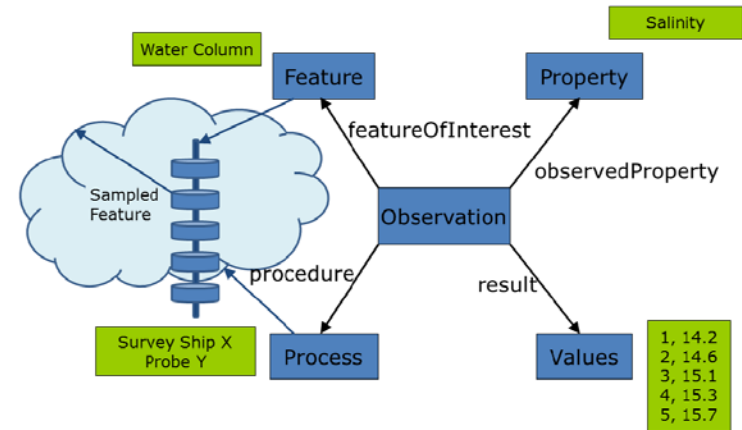
Score	Attribute	Description	THREDDS	ISO 19115-2
1	<a href="#">geospatial_lat_max</a>	Describes a single latitude, longitude, vertical and temporal bounding box. For a more detailed geospatial coverage, see the <a href="#">suggested_ensemble_urls</a> page.	metadata.geospatialCoverage northsouth start	gml:MI_Metadata gml:identification gml:MD_DatasetIdentification gml:extent gml:EX_Extent gml:geographicElement gml:EX_GeographicBoundingBox gml:southBoundLongitude gml:Decimal
1	<a href="#">geospatial_lat_min</a>	Further refinement of the geospatial bounding box can be provided by using these units and resolution attributes.	metadata.geospatialCoverage northsouth size	gml:MI_Metadata gml:identification gml:MD_DatasetIdentification gml:extent gml:EX_Extent gml:geographicElement gml:EX_GeographicBoundingBox gml:southBoundLongitude gml:Decimal
1	<a href="#">geospatial_lon_max</a>	Many of these extent attributes are calculated using the CF Conventions.	metadata.geospatialCoverage westeast start	gml:MI_Metadata gml:identification gml:MD_DatasetIdentification gml:extent gml:EX_Extent gml:geographicElement gml:EX_GeographicBoundingBox gml:westBoundLongitude gml:Decimal
1	<a href="#">geospatial_lon_min</a>		metadata.geospatialCoverage westeast size	gml:MI_Metadata gml:identification gml:MD_DatasetIdentification gml:extent gml:EX_Extent gml:geographicElement gml:EX_GeographicBoundingBox gml:westBoundLongitude gml:Decimal
1	<a href="#">time_coverage_start</a>		metadata.timeCoverage start	gml:MI_Metadata gml:identification gml:MD_DatasetIdentification gml:extent gml:EX_Extent gml:temporalElement gml:EX_TemporalExtent gml:extent gml:TimeSeries gml:beginPosition
1	<a href="#">time_coverage_end</a>		metadata.timeCoverage end	gml:MI_Metadata gml:identification gml:MD_DatasetIdentification gml:extent gml:EX_Extent gml:temporalElement gml:EX_TemporalExtent gml:extent gml:TimeSeries gml:endPosition
1	<a href="#">geospatial_vertical_min</a>		metadata.geospatialCoverage updown start	gml:MI_Metadata gml:identification gml:MD_DatasetIdentification gml:extent gml:EX_Extent gml:verticalElement gml:EX_VerticalExtent gml:minimumValue gml:Real
1	<a href="#">geospatial_vertical_max</a>		metadata.geospatialCoverage updown size	gml:MI_Metadata gml:identification gml:MD_DatasetIdentification gml:extent gml:EX_Extent gml:verticalElement gml:EX_VerticalExtent gml:maximumValue gml:Real

[Identification](#) | [Text Search](#) | [Extent Search](#) | [Other Extent Information](#) | [Create Search](#) | [Contribute Search](#) | [Publisher Search](#) | [Other Attributes](#)



# Implementation Example 2: SOS

- Standard to provide access to time series from sensors and sensor systems;
- Goal is to be able to query data from sensors independently of which sensor is providing it;
- Facilitate data tailoring and fusion for (real-time) observations, enabling cross-domain analysis.



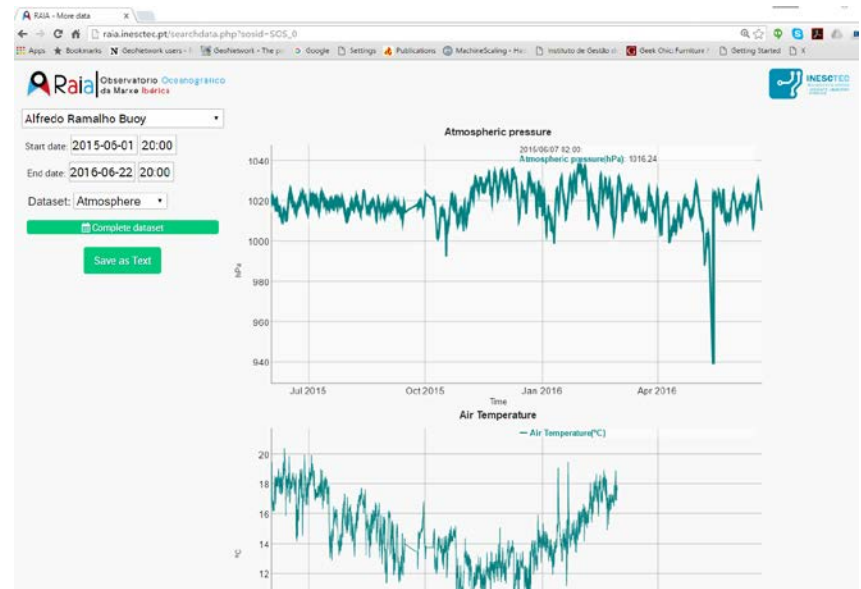
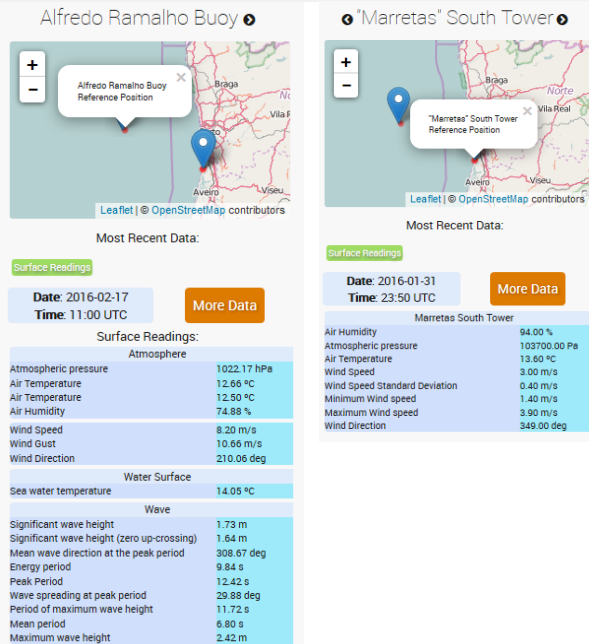
- Advantages:
  - Richer API
  - Includes reference to Feature of Interest and Process (Sensor, Lab Process, ...)
- Disadvantage:
  - Not so efficient for bulkier data (ex: forecasts)



# Implementation Example 2: SOS

- Used in RAIa for Buoys and Wind Towers alike.
- Example containing last observation:

- Example for a more thorough exploration of a sensor time series:



# Example of automatic harvesting of service metadata

RAIA Geocatalogue - RAIA X

geonetix.inesctec.pt/geonetwork/srv/eng/catalog.search

Get started

Search over 45 data sets, services and maps, ...

Search ...

RAIA Data Portal

Here you will find data, services and maps and more.

Browse by **INSPIRE themes** topics

Elevation 2 Buildings 1 Environmental monitoring f... 3 Atmospheric conditions 2 Geographical names 2 Administrative units 7

Browse resources

Dataset 40 Service 5

Latest news Most popular

"Marretas" Towers (SOS, INEGI)  
Service

Alfredo Ramalho Buoy (SOS, Hidrográfico)  
Service

RAIA's ncWMS server (INESCTEC)  
Service

Sea surface wave significant height, Porto (Hidrográfico)  
Dataset

Sea surface wave significant height, Aveiro (Hidrográfico)  
Dataset

Sea surface swell wave significant height, Porto (Hidrográfico)  
Dataset

Sea surface wave period at variance spectral density maximum, Porto (Hidrográfico)  
Dataset

Sea surface wave period at variance spectral density maximum, Aveiro (Hidrográfico)  
Dataset

Sea surface wave direction at variance spectral density maximum, Porto (Hidrográfico)  
Dataset

# Example of automatic harvesting of dataset metadata

RAIA Geocatalogue - RAIA X

geonetix.inesctec.pt/geonetwork/srv/eng/catalog.search#/search?facet.q=inspireThemeURI%2Fhttp%253A%252F%252Frdfrdata.eionet.europa.eu%2F

Apps ★ Bookmarks N GeoNetwork users - R GeoNetwork - The po Google Settings Publications MachineScaling - Had Instituto de Gestão de Geek Chic: Furniture F Getting Started X

RAIA Geocatalogue Search Map Sign in English

Search ...

Sorted by relevancy 1 - 2 on 2

**TYPE OF RESOURCES**

- ☐ Dataset (2)

**TOPICS**

- ☐ Climatology,... (2)

**INSPIRE THEMES**

- ☒ Elevation (2)

**KEYWORDS**

- ☐ Galicia (2)
- ☐ Elevaciones (2)
- ☐ MDE (2)
- ☐ Batimetria (2)
- ☐ ClimatologyMeteorolo... (2)

**CONTACT FOR THE RESOURCE**

- ☐ Intecmar (2)

**PROVIDED BY**

- ☐ INTECMAR (2)

**YEARS**

- ☐ 2016 (2)

**STATUS**

- ☐ Completed (2)

**Modelo Digital de Elevaciones de la Fachada Noroeste de la Península Ibérica de...**

El MDE-NW-025 es un modelo digital de elevaciones continuo del espacio marítimo/terrestre de la fachada noroccidental de la Península Ibérica.

Departamento GIS de Intecmar

**Modelo Digital de Elevaciones de la Fachada Noroeste de la Península Ibérica de...**

El MDE-NW-200 es un modelo digital de elevaciones continuo del espacio marítimo/terrestre de la fachada noroccidental de la Península Ibérica.

Departamento GIS de Intecmar

Map showing the location of the dataset in the Northwest of the Iberian Peninsula, highlighting the Galicia region and the Principado de Asturias.

About Github Powered by geonetwork 3.0.4.0

# Conclusions and Future Work

- Implementation of services according to standards has made possible the automatic harvesting of metadata from services and datasets that they contain, thus reducing the effort in annotating datasets;
- Catalogue services can index each other's contents;
- Harmonization requires use of standard code lists and controlled vocabularies;
- A continued effort is needed to promote the implementation of more interoperable services and catalogues, as well as better interoperation between catalogue service and newer services such as SOS;
- More advanced user interfaces need to be developed to facilitate the finding and usage of data by specific user communities.



ДЗЯКУЙ  
ESKERIK ASKO  
KOZOMON  
БЛАГОДАРЯ  
謝謝  
OBRIGADO  
THANK YOU  
DANKIE  
CẢM ƠN BẠN  
GRACIAS  
KIITOS  
MERCI  
DANKE  
GRAZIE  
TAKK  
GRÀCIES  
DIO LCH  
DZIĘKUJĘ  
GRATIAS AGIMUS TIBI  
მადლობა  
TAK  
FALEMINDERIT  
DěKUJI  
شكرا  
ありがとう  
PĀKKA PĒR  
விபரதாரிமே